An areal view of Africa ................................................................. 2
Areal contact in Nilo-Saharan ...................................................... 21
Niger-Congo languages ................................................................ 44
The Kalahari Basin area as a “Sprachbund” before the Bantu expansion ........................................ 68
South Africa and areal linguistics .................................................. 92
Jharkhand as a “linguistic area” ..................................................... 113
Sri Lanka and South India ............................................................ 136
The Transeurasian languages ....................................................... 145
The changing profile of case marking in the northeastern Siberia area ...................................... 181
Languages of China in their East and South-East Asian context ............................................. 203
Language in the Mainland Southeast Asia area ................................ 228
Southeast Asian tone in areal perspective ...................................... 252
The areal linguistics of Australia .................................................... 277
Languages of the New Guinea region ............................................ 300
Languages of Eastern Melanesia .................................................. 351
The Western Micronesian Sprachbund .......................................... 376
Native North American Languages .............................................. 397
The areal linguistics of Amazonia ................................................... 446
Linguistic areas, linguistic convergence, and river systems in South America .................. 473
14 An areal view of Africa

Bernd Heine and Anne-Maria Fehn

1 Introduction

Comparative work on African languages in the past has focused on the genetic relationship patterns among these languages. But there is now a growing interest in areal relationships, and some of the language groupings that were earlier proposed to be genetic units have more recently been redefined as contact-induced, areal groupings (see the contributions in Heine and Nurse 2008 and Hieda et al. 2011).

The present chapter provides an overview of major achievements that have been made in the areal classification of African languages. On the one hand, it will be concerned with smaller groups of areally related languages and the major regions of the continent. On the other hand, it will look at Africa as a possible macro-area. The main interest of the chapter will be with what, following Weinreich ([1953] 1964), is referred to as replication, that is, the contact-induced diffusion of structures and/or meanings rather than that of phonetic/phonological or morphological material.

The questions that are our main concern in this chapter are the following:

(1)  
   a. Are there any geographically defined language groupings in Africa?
   b. Can the African continent be set off from the rest of the world as a linguistic area of some kind?

Question (1a) will be the subject of Section 2, while Sections 3 and 4 are devoted to (1b). In line with Heine and Kuteva (2005: 182-187) and Heine (2011) it will be argued in Section 5 that the study of areas of grammaticalization offers a promising perspective of the diffusion of contact-induced phenomena in both linguistic micro- and macro-settings of language contact. In the final Section 6, some conclusions are drawn.

2 Areal groupings within Africa

Two main types of areal classifications can be distinguished in African linguistics. On the one hand they are based on a single property or domain of language structure that is argued to be diagnostic of areal relationship; we will refer to them as single-feature classifications. On the other hand they are feature-bundle classifications, involving a combination of several properties.

One of the properties that students of African languages recruited to arrive at generalizations on the areal patterning of languages is word order. Based on a survey of the order of meaningful elements in African languages, Heine (1975; 1976) concludes that there are a number of linguistically defined areas cutting across boundaries of language families. One such area consists of a large part of West Africa where Mande, Gur (Voltaic), and western Kwa languages are spoken. In addition to these languages, which are traditionally classified as Niger-Congo, this area also includes Songhai, a
language usually classified as belonging to the Nilo-Saharan phylum (Greenberg 1963). What characterizes this area most of all is the presence of a possessor - possessee word order syntax which is not restricted to the noun phrase but has also affected the structure of the clause (see Claudi 1993). Another area, called the Rift Valley Convergence Area, is defined by the presence of verb-initial (VSO) syntax, very rarely encountered elsewhere in Africa. The languages of this East African area belong to Greenberg’s (1963) Nilo-Saharan (Surma, Kuliak, Eastern Nilotic, and Southern Nilotic) and Khoisan phyla (Hadza).

Word order is also the subject of a more recent attempt by Dryer (2011) to look for areal features shared by African languages across the continent. Based on a world-wide survey of the word order arrangement of nouns and their modifiers in over 1600 languages, Dryer concludes that languages in Africa exhibit a greater tendency to place modifiers after the noun than languages in other parts of the world; we will return to this work in Section 4.

Another property figuring in classifications on areal grouping concerns phonology: Can Africa be divided into significant, geographically defined phonological zones? This question is answered in the affirmative by Clements and Rialland (2008): The authors propose to divide the continent on the basis of phonological distinctions into the following zones:

Table 1. Linguistic zones in Africa

North Africa: Its phonological properties coincide largely with those of Arabic and the Berber languages.

East Africa: It encompasses Ethiopia, Eritrea, Djibouti, and Somalia; nearly all its languages are usually classed in the Afroasiatic phylum.

Sudanic belt: It includes the vast savanna that extends across Sub-Saharan Africa bounded by the Sahel on the north, the Atlantic Ocean on the west and southwest, Lake Albert on the southeast, and the Ethiopian-Eritrean highlands on the east.

Center: It is almost exclusively Bantu-speaking and is characterized by the linguistic features typical of Bantu languages.

South Africa: It includes semi-desert, savanna, and temperate coastal regions. Its phonological characteristics derive from those of the Khoisan and Bantu languages.

---

1 With the term phyla (singular phylum) we refer to the four African “families” proposed by Greenberg (1963), namely Afroasiatic, Nilo-Saharan, Niger-Congo (i.e., his Kongo-Kordofanian), and Khoisan. In doing so, we follow a widespread convention in African linguistics. Not all of these phyla are nowadays generally recognized as valid linguistic units. Nilo-Saharan and Khoisan in particular have so far not been satisfactorily established as genetically defined groupings.

2 The only other clear cases of VSO-languages reported so far are the Berber languages of northwestern Africa, a few Chadic languages, and Krongo, a Kordofan language nowadays considered to belong to the Kadu branch of Nilo-Saharan (Heine 1976).

3 Whether the Khoisan languages of Southern and Eastern Africa do in fact form a linguistic family, as claimed by Greenberg (1963), is nowadays a matter of much dispute. Hence we are using the term “Khoisan” in a loose sense. See the contribution by Güldemann and Fehn, this volume.
spoken within it. This zone contains some of the richest consonant and vowel inventories of the world’s languages.

*Rift Valley:* It includes much of the eastern branch of the Great Rift Valley in northern Tanzania and southwestern Kenya. In this region, languages of all four of Greenberg’s phyla (or super-families) are found.

![Map 1. Phonological zones in Africa according to Clements & Rialland (2008).](image)

These zones are primarily geographic and only secondarily linguistic in nature, they must be viewed as a first approximation to areal phonology, where each zone stands for a prototype, or an ideal type: For example, a number of the languages of the South Africa zone have more phonological features in common with languages of the Center, and many features of the Rift Valley zone can also be observed in other zones.

The declared or implicit goal underlying feature-bundle classifications is to search for sprachbund-like linguistic areas (or convergence areas). The most frequently mentioned and most widely recognized linguistic area is Northeastern Africa, frequently
called the Ethiopian Language Area (or Ethio-Eritrean Sprachbund). Within roughly the last two millennia, the highlands of Ethiopia appear to have favored cultural and linguistic exchange on a massive scale, with the effect that the languages of this region now share a number of linguistic properties (Ferguson 1976; Crass and Meyer 2008). The languages included in this area are mostly genetically interrelated, belonging to the Cushitic, Omotic, and Semitic branches of the Afroasiatic family, but some languages of the Nilo-Saharan phylum are also included. First proposed by Ferguson (1970; 1976), the notion of an Ethiopian Language Area has not gone unchallenged (Tosco 2000), but Crass and Meyer (2008) find further support for it, adding more properties characterizing the area, which is not restricted to the nation state of Ethiopia but also includes languages spoken in Djibouti, Somalia, Kenya, and Sudan.

While Ferguson (1976) proposed 8 phonological and 18 grammatical features to define the Ethiopian area, Crass and Meyer (2008) add another 12 criteria in support of the grouping, most of which are morphosyntactic in nature. Half of these criteria concern processes of grammaticalization shared by the languages of the area and, hence, are suggestive of grammaticalization areas (see Section 4).

Another grouping that qualifies as a linguistic area is the Tanzanian Rift Valley Area (Kießling, Mous and Nurse 2008). It includes languages of all African families (or phyla) that have been identified by Greenberg (1963), namely Southern Cushitic languages of Afroasiatic, Southern Nilotic of Nilo-Saharan, some Bantu languages of Niger-Congo, and Sandawe and Hadza of Greenberg’s Khoisan phylum. The authors adduce a range of altogether 15 features to substantiate their sprachbund hypothesis. Five of the features are phonological, they include presence vs. absence of a lateral fricative /ɬ/, of ejective obstruents, or of a 7-vowel system. The largest number of features are morphological, being either structural (presence vs. absence of a preverbal clitic complex, of categories such as verbal plurality, applicatives, ventives, or the number of past and future tense distinctions) or formal (presence of subjunctive -ee or irrealis laa). Further criteria are syntactic, concerning the linear arrangement of constituents, or conceptual, involving polysemy and semantic transfer strategies (see Kießling et al. 2008, Table 6.5).

In more recent research on linguistic relationships in Africa there appears to be a gradual shift of interest from genetic to areal linguistics. It may therefore not be surprising that some language groupings that were analyzed by earlier authors as genetic units are now reanalyzed in terms of contact-induced change and areal relationships. Roughly a century ago, Westermann (1911) proposed an essentially genetically defined stock, namely the Sudanic languages (Sudansprachen). As Güldemann (2008) argues convincingly, however, many of the Sudansprachen of Westermann are more plausibly re-classified as forming an areal grouping, referred to by him as the Macro-Sudan Belt, which is similar to, but should not be confused with, the Sudanic belt of Clements and Rialland (2008; see above). Defining properties of this grouping are presence of (a) logophoric markers, (b) labial-velar consonants, (c) vowel harmony of the ATR (advanced tongue root position) type, (d) a word order pattern S (Aux) OV-X, where the object (O) precedes the verb, (e) another word order pattern V-O-NEG, where the

---

4 With the adverb ‘essentially’ we remind readers that at that time, no clear distinction between genetic and other kinds of linguistic relationship was made.
negation marker (NEG) is placed clause-finally, and (f) labial flap consonants. Not all the features are found in all languages of the belt, but there is a massive clustering of the features that appears to be distinctive.


The Kalahari basin of southern Africa also forms some kind of linguistic area; it provides an instance of a refuge area where people have been living over centuries and probably millennia without much interference from outside. It is the homeland of traditional hunter-gatherer populations speaking languages belonging to the “Khoisan” phylum of Southern Africa. Güldemann (1997) argues that the Kalahari basin convergence area is not confined to languages conventionally classified as belonging to the Khoisan cluster but also includes a Bantu language, namely Tswana, but more evidence is required on this issue (see Güldemann and Fehn, this volume).

A number of additional groupings, suggestive of contact-induced relationships, have been proposed but have so far not received wider recognition as linguistic areas.

5 Labial-dental flaps begin with the lower lip placed behind the upper teeth. The lower lip is then flipped outward, striking the upper teeth in passing. Its occurrence outside Africa is extremely rare, while in Africa it is found in more than a hundred languages in the Chadic family (Margi, Tera), Ubangian (Ngbaka, Ma’bo, Sera), Central Sudanic (Mangbetu, Kresh), and Bantoid (Ngwe, some Shona dialects).
3 The search for “Africanisms”

In an attempt to isolate areal within Africa and separating Africa from other regions of the world, Greenberg (1959) proposed a number of what he called ‘special’ features of African languages. The properties listed by him include in particular a number of lexical polysemies, such as the use of the same term for ‘meat’ and ‘wild animal’, of the same term for ‘eat’, ‘conquer’, ‘capture a piece in a game’ and ‘have sexual intercourse’, and the use of a noun for ‘child’ as a diminutive, or of ‘child of tree’ to denote ‘fruit of tree’.

Around this time, students of African languages began to search for what – following Meeussen (1975) – tends to be called “Africanisms” (Greenberg 1983: 3; Heine and Leyew 2008), referring to properties that satisfy the following set of criteria:6

(2) a They are common in Africa but clearly less common elsewhere.
   b They are found, at least to some extent, in all major geographical regions of Africa south of the Sahara.
   c They are found in two or more of the four African language phyla.

Larochette (1959) presented a catalog of linguistic properties characteristic of Congolese Bantu (Kikongo, Luba, Mongo), an Ubangian language (Zande), and a Central Sudanic language (Mangbetu), but a number of the properties proposed can also be found in other regions and genetic groupings of Africa. Building on this work, Meeussen (1975) proposed an impressive list of Africanisms, that is, phonological, morphological, syntactic, and lexical properties widely found in African languages across genetic boundaries. Another range of properties characterizing many African languages was proposed by Welmers (1974) and (1977).

In 1983, Greenberg proposed distinguishing between areal properties that are exclusive to Africa though not found everywhere within it, and properties that are especially common in Africa although not confined to the continent (Greenberg 1983: 3; cf. (2)). As an example of the former he mentioned clicks; as instances of the latter he discussed in some detail the following ‘characteristics’ (Greenberg 1983: 4): (i) coarticulated labial-velar (or labiovelar) stops, (ii) labial-dental (or labiodental) flaps, (iii) the use of a verb meaning ‘to surpass’ to express comparison, and (iv) a single term meaning both ‘meat’ and ‘animal’. He demonstrated that these four properties occur across genetic boundaries and, hence, are suggestive of pan-African traits, especially since they are rarely found outside Africa.

Search for areal properties across Africa is associated to some extent with creole linguistics (see e.g. Boretzky 1983). In an attempt to establish whether, or to what extent, the European-based pidgins and creoles on both sides of the Atlantic Ocean have been shaped by African languages, researchers of creoles have pointed out a number of properties, more widely found in Africa, and/or a set of characteristics of African languages. Perhaps the most detailed study is that by Gilman (1986). Gilman proposed a larger catalogue of pan-African areal properties, arguing that a large number of African-like structures in Atlantic and other pidgins and creoles are best explained by the

6 ‘Africanisms’ correspond to what Greenberg (1983: 3) called African areal properties, that is, properties ‘which are either exclusive to Africa, though not found everywhere within it, or those which are especially common in Africa although not confined to that continent’.
influence of those areal properties which are widely distributed among the languages of Africa.

A number of the properties that are clearly more widespread in Africa than elsewhere are not considered here, for the following reasons. First, because they appear to be genetically determined. The presence of gender or noun class systems is a case in point. Most instances of such systems to be found in Africa are presumably genetically inherited. This can be assumed to apply on the one hand to the nature-based noun class systems found in Niger-Congo and the non-Khoe languages of the Khoisan unit, and on the other hand to the sex-based gender systems of Afroasiatic and Khoe (Central Khoisan) languages.7

Perhaps surprisingly, we will also not consider presence or absence of click types as distinct phonemic units to be a relevant property, although it appears to be the only property that is confined exclusively to Africa, and although it satisfies most of the criteria proposed above. The reason for doing so is the following: The main goal of this section is to find out whether African languages resemble one another more than they resemble other languages and what factors can be held responsible for such resemblances. To be sure, clicks occur in three of the four African language phyla, not only in all Khoisan languages. Rather, they are also found in South African Bantu (Niger-Congo) languages such as Zulu, Xhosa, Gciriku (Dciriku), or Yeyi (see Maddieson 2003: 31-7; Bostoen and Sands 2012 for more details), and in the Cushitic (Afroasiatic) language Dahalo (Tosco 1991: 4). Still, their occurrence is geographically restricted to southern Africa and three East African languages, that is, it is not representative of areal relationship in Africa as a whole.

Furthermore, although Khoisan languages are among the phonologically most complex languages in the world, some of them distinguishing more than 110 distinct phonemes (see e.g., Traill 1994; Heine and König forthcoming), this fact is ignored here since it does not appear to be a characteristic of Africa as a linguistic area. Furthermore, such complex sound inventories are restricted to a few Kx’a (Northern Khoisan) and Tuu (Southern Khoisan) languages.

In the following we will discuss a catalogue of properties that have been proposed to be characteristic of Africa as a linguistic area (especially by Greenberg 1959; 1983; Larochette 1959; Meeussen 1975; Gilman 1986). Our selection is to some extent arbitrary in that we will ignore some properties that have been mentioned by other authors but where we are not entirely convinced that they are possible candidates for the status of ‘Africanisms’. Note that we restricted ourselves to properties that are suggestive of areal relationships, that is, that are not confined to genetically defined language groupings.

A general phonological property that has been pointed out by a number of researchers of African languages is the preponderance of open syllables and an avoidance of consonant clusters and diphthongs (Meeussen 1975: 2; Gilman 1986: 41). Furthermore, tone as a distinctive unit is characteristic of the majority of African languages, in many cases on both the lexical and grammatical levels.

Ignoring click consonants, there are a number of consonant types that are widespread in Africa but uncommon elsewhere (see Clements and Rialland 2008 for detailed treatment). Prime examples, among others, are coarticulated labial-velar (or

7 It is possible that the presence of gender systems in the Eastern Nilotic languages (Maa, Teso-Turkana, Lotuxo, Bari) is the result of language contact with Cushitic languages, but the evidence for this is not conclusive.
labiovelar) stops such as \( kp \) and \( gb \) (Meeussen 1975: 2; Greenberg 1983: 4; Gilman 1986: 41). There are also corresponding nasals and/or fricatives, but they do not show the wide distribution of stops, and their occurrence is largely predictable on the basis of stops (Greenberg 1983: 4). The distribution of this property is clearly areally constrained: Labial-velar stops occur in a broad geographical belt from the western Atlantic to the Nile-Congo divide, and they are also occasionally found outside this belt, e.g. in Katla and Giryama (see Welmers 1974: 47-8). Still, they are found in three of the four African phyla; only Khoisan languages have no labial-velar stops. Also, in the Afroasiatic and Nilo-Saharan phyla, their occurrence is restricted essentially to one branch each, namely Chadic and Central Sudanic, respectively (Greenberg 1983: 7; Güldemann 2008).

Perhaps even more characteristic are labial-dental (or labiodental) flaps, where the teeth touch well below the outer section of the lip, which is flapped quickly outwards and downwards. They have been found in all African phyla except Khoisan, e.g. in Chadic of Afroasiatic (Margi, Tera), Niger-Congo (Ngwe, Ngbaka, Ngbaka Mabo, Ndogo-Sere, some Shona dialects), and Nilo-Saharan (Kresh, Mangbetu) (Gregersen 1977: 31; Greenberg 1983: 4, 11; Clements and Rialland 2008; Güldemann 2008). Still, their occurrence is confined to a relatively small number of languages, and even there they show restrictions in their use as phonemic units; not infrequently, these sounds are found only in special vocabulary such as ideophones. In their survey of 250 African and 345 non-African languages, Clements and Rialland (2008) did not find a single non-African language with such flaps, but at least 70 African languages did.

A third type of consonant that is widespread in Africa can be seen in implosives, which — following Clements and Rialland (2008) — we define as non-obstruent stops. To be sure, there exist in some non-African languages, such as the Indonesian language Auye (Mike Cahill, p.c.), but these languages are rare. Furthermore, word-initial prenasalized consonants, for the most part voiced stops, are widely found in Africa (Meeussen 1975: 2; Gilman 1986: 41), although they occur most of all in Niger-Congo languages.

An outstanding property relating to the vowel system can be seen in the presence of cross-height vowel harmony based on distinctions of the tongue root position, commonly known as ATR (advanced tongue root) vowel harmony. It is widespread in Niger-Congo and Nilo-Saharan languages across the continent but appears to be rare outside Africa (see Clements and Rialland 2008, Güldemann 2008 for discussion.

Morphological properties that have been mentioned as areal characteristics of African languages include reduplication of nouns and adjectives, used to express a distributive function (e.g. Swahili \( tano tano \) ‘five each, in fives’; Gilman 1986: 40). Within the verbal word, many African languages are characterized by a wide range of derivational suffixes for functions such as reflexive, reciprocal, causative, passive, stative, andative (itive), and venitive (ventive), and these suffixes can be combined in sequence (Meeussen 1975: 2; Gilman 1986: 43). However, both these properties can also be observed widely in non-African languages.

A conspicuous feature of nominal morphology is the paucity of languages having case inflections. Those African languages that do distinguish grammatical case are mostly marked nominative, that is, in such languages it is the accusative rather than the nominative case that is unmarked (marked nominative languages are crosslinguistically exceptional). And a perhaps unique property of case systems is the presence of cases marked exclusively by tonal inflection, which so far has been found only in African languages.

Morphological properties that have been mentioned as areal characteristics of African languages include reduplication of nouns and adjectives, used to express a distributive function (e.g. Swahili \( tano tano \) ‘five each, in fives’; Gilman 1986: 40). Within the verbal word, many African languages are characterized by a wide range of derivational suffixes for functions such as reflexive, reciprocal, causative, passive, stative, andative (itive), and venitive (ventive), and these suffixes can be combined in sequence (Meeussen 1975: 2; Gilman 1986: 43). However, both these properties can also be observed widely in non-African languages.

A conspicuous feature of nominal morphology is the paucity of languages having case inflections. Those African languages that do distinguish grammatical case are mostly marked nominative, that is, in such languages it is the accusative rather than the nominative case that is unmarked (marked nominative languages are crosslinguistically exceptional). And a perhaps unique property of case systems is the presence of cases marked exclusively by tonal inflection, which so far has been found only in African languages.
marked nominative languages but apparently nowhere else in the world (König 2006; 2008).

With regard to word classes, African languages have been said to be characterized by a paucity of adjectives and in a number of languages adjectives are claimed to be absent altogether. Those contents typically expressed in non-African languages by adjectives are likely to appear as verbs of state in Africa (cf. Gilman 1986: 40). On the other hand there is a word class of ideophones that appears to be remarkably salient in many African languages (Meeussen 1975: 3). While languages in other parts of the world have ideophones as well, African languages have been found to have them in distinctly larger numbers. Furthermore, ideophones expressing color distinctions have so far only been found in Africa (Kilian-Hatz 2001; Voeltz and Kilian-Hatz 2001).

In their arrangement of words, African languages of all four phyla exhibit a number of general characteristics, such as the following: while on a worldwide level languages with verb-final syntax (SOV) appear to be the most numerous, in Africa there is a preponderance of languages with subject – verb – object (SVO) as their basic order: Roughly 71% of all African languages exhibit this order (Heine 1975; 1976: 23; see also Gilman 1986: 37). Furthermore, the placement of nominal modifiers after the head noun appears to be more widespread in Africa than in most other parts of the world. Thus, in Heine’s (1976: 23) sample of 300 African languages, demonstrative attributes are placed after the noun in 85%, adjectives in 88%, and numerals in 91% of all languages. Another characteristic in the arrangement of meaningful elements relates to verbal structures: in most African languages, pronominal subject clitics or affixes precede the tense markers (93%), which again precede the verb (83%), while adverbs follow the verb 93% (Heine 1976: 24).

An arrangement of basic word order that occurs in a number of languages across the continent but which is fairly uncommon outside Africa concerns what nowadays tends to be referred to as “SOVX” order. In languages having this order, the direct object precedes the verb but the indirect object and adjuncts follow the verb. SOVX languages are likely to have postpositions and to place the genitival modifier before its head while other nominal modifiers follow the head noun (cf. the type B of Heine 1976).

With reference to information structure, front-focusing of nouns by means of some kind of cleft-construction has been mentioned, frequently used obligatorily in word questions, for example *who went?* is expressed by *who is it who went?* (Gregersen 1977: 50-51; Gilman 1986: 39). In addition to noun phrase focusing there is also front-focusing by means of verb-copying, where the verb appears first in the focus position and is repeated in the main clause (Gilman 1986:39); the exact distribution of this phenomenon across Africa, however, is unknown. A striking characteristic of African languages can be seen in the widespread change of verb form, or the use of special auxiliaries, to express focus distinctions (Creissels et al. 2008: 138).

In addition, there are construction types that are said to be found in a number of African languages but to be rare outside Africa. One of them is called anastasis by Meeussen (1975: 4), consisting of the swapping of subject and complement participants within the clause, e.g., the possibility to express ‘Worms enter the corpse’ by ‘The corpse enters worms’. It is unknown how widespread anastasis is in Africa, and it would seem that it is not all that uncommon in other parts of the world (Felix Ameka, p.c.).

---

8 Unlike, for example, an English medio-passive get-sentence like *The thieves got arrested by the police*, anastasis is not formally marked in any way.
Logophoric marking constitutes another construction type that has been claimed to be specifically African. Logophoric pronouns indicate coreference of a nominal in the non-direct quote to the speaker encoded in the accompanying quotative construction, as opposed to its non-coreference indicated by an unmarked pronominal device (Güldemann 2008). Thus, whereas (3a) illustrates logophoric marking, (3b) is a plain, non-logophoric structure.

(3) Ewe (Kwa, Niger-Congo; data of authors)

a é gbɔ bɛ ye- dzó.
3.SG say that LOG- leave
‘She said that she left.’

b é- gbɔ bɛ é- dzó.
3.SG- say that 3.SG- leave
‘She said that she left.’

Logophoric structures are with very few exceptions concentrated in a large belt extending from the southeastern corner of Ethiopia to the east up to the Niger River in the west and are found in three of the four language phyla (Güldemann 2008; see Section 2).

Finally, there are a number of conceptualization strategies that might qualify as Africanisms. This applies in particular to what is called the goose-file model of spatial orientation (Heine 1997: 12-14), to be found in at least three of the four African language phyla, described by Meeussen in the following way: 9

Imagine a place from which a house can be seen, and further away a small hill. In such a situation the hill will be referred to in African terms as being ‘in front of the house’, and the house as being ‘behind the hill’, whereas in European languages the reverse expressions will be used. (Meeussen 1975: 3)

The following example from the Kuliak language may illustrate the goose-file model, where an item to be located is conceptualized not as facing the speaker but rather as facing the same direction as the speaker.

(4) So (Kuliak, Nilo-Saharan; own data)

nékɛ yɔG sú- o sóG.
be.at people behind- ABL hill
‘There are people in front of the hill.’

There is another conceptualization strategy that has been proposed as an Africanism (Meeussen 1975), being one manifestation of what is usually called the inclusive or inclusory construction, which is used in reference to a plural that refers to a set of individuals and includes two explicit constituents. The form the construction typically takes in African languages is illustrated in (5).

---

9 According to the “goose-file” model, speakers and hearers are somehow conceived as following one another like in a goose file, rather than facing one another when talking.
(5) Swahili (Bantu, Niger-Congo; own data)

```latex
sisi na wewe
we and you
‘I and you’
```

It is unknown how widespread this construction type is; it is by no means restricted to Africa, being found in various other parts of the world (Singer 1999; Moravcsik 2003: 479).

Another strategy which is not restricted to Africa but is perhaps more widespread in Africa than elsewhere and which can be found in all four language phyla, can be seen in affirmative answers to negative questions where the speaker wants to know if the propositional content of the question is correct or not, e.g., ‘Didn’t you sleep?’ – ‘Yes, I didn’t’ or ‘No, I did’ (Meeussen 1975:4; Gregersen 1977:44; Felix Ameka, p.c.).

Perhaps the most conspicuous domain where one might expect to find Africanisms involves lexical and grammatical polysemies. The following are a few examples that have been pointed out by researchers of African languages.

Within the domain of nominal polysemy, a paradigm case is where the same noun is used for ‘meat’ and ‘animal’ or, alternatively, that there are different but etymologically related nouns for ‘meat’ and ‘animal’ (Greenberg 1959, 1983: 4) – a case described by Lichtenberk (1991) more appropriately as heterosemy. Perhaps remarkably, if one of the two meanings is derived from the other, then it goes from ‘meat’ to ‘animal’ rather than vice versa\(^{10}\). This is suggested at least by the fact that whenever the two are distinguished by means of some derivational, compounding or other mechanism then it is the item for ‘meat’ that is likely to be unmarked and ‘animal’ to be marked; cf. the following examples (for an example from the Bantu language Tonga, see Greenberg 1983: 16):

<table>
<thead>
<tr>
<th>Language</th>
<th>‘meat’</th>
<th>‘animal’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausa (Chadic, Afroasiatic)</td>
<td>nāmà</td>
<td>nāmàn dājì ‘wild animal’ (‘meat of the bush’)</td>
</tr>
<tr>
<td>!Khoisan (North Khoisan)</td>
<td>ǁhã</td>
<td>ǁhã-mà (‘animal-DIM’)</td>
</tr>
</tbody>
</table>

To be sure, such polysemy can also be observed in other parts of the world, but it appears to be much more frequent in Africa than elsewhere.

Another nominal polysemy that has been claimed to be “pan-African” is that of nouns denoting both ‘hand’ and ‘arm’, or nouns denoting both ‘foot’ and ‘leg’ (and ‘wheel’) (Gilman 1986: 43). Note, however, that these polysemies are also widespread outside Africa; a worldwide survey shows that 50 out of the 109 languages analyzed have

\(^{10}\) Greenberg (1983:16), however, says that this is not always so: ‘The most conspicuous exception is the Grasslands languages where the form bep or the like is found in many languages with the meaning ‘meat’ while the nama root survives as ‘animal’. It would seem that this fact does not invalidate the hypothesis of a directionality ‘meat’ > ‘animal’; rather, it might suggest that – for whatever reasons - an earlier meaning ‘meat’ received a new form of expression.
a ‘hand’/‘arm’ polysemy and 42 out of 109 languages a ‘foot’/‘leg’ polysemy (see Heine 1997: 136).

Examples of polysemies involving activities include verbs for ‘eat’, which are said to also denote ‘conquer’, ‘capture a piece in a game’, and ‘have sexual intercourse’ (Greenberg 1959), verbs for ‘die’, which tend to have many non-literal meanings in African languages such as ‘be in a painful condition’, ‘break down’ (cf. Meeussen 1975: 4), verbs for ‘lie (down)’ also meaning ‘sleep’, or verbs for ‘hear’ (to a lesser extent also ‘see’) also denoting other kinds of perception, such as ‘smell’, ‘feel’, ‘taste’, ‘understand’ (Meeussen 1975: 4-5). Meeussen (1975: 4) furthermore notes that the use of words for ‘good’ also tend to express ‘nice’, ‘beautiful’, and ‘fine’ in African languages. The status of some of these polysemies as cases of Africanisms, however, is far from clear. For example, meaning ranges expressed by verbs for ‘die’ in African languages may also be found in Australia or the Americas (Felix Ameka, p.c.), and much the same applies to polysemies involving ‘hear’ (see e.g. Evans and Wilkins 2000 for evidence on Australian languages).

4 Africa as a linguistic area

In much of the work examined in the preceding section there is an implicit assumption to the effect that the African continent can somehow be distinguished as a linguistic unit from other regions in the world. This is the topic of a study by Heine and Leyew (2008). On the one hand, they observe that the following generalization proposed by Greenberg (1983: 27) is in fact correct: “Ideally, if what is meant by an African areal characteristic is one which is found everywhere in Africa but nowhere else, then clearly none exists” (Greenberg 1983: 27). On the other hand, they argue that it is possible to set the African continent off from the rest of the world at least by means of quantitative generalizations.

Using a sample of 99 African languages from all four language phyla and all major geographical regions of the continent plus 50 non-African control languages, Heine and Leyew (2008) analyze eleven properties that were used by previous authors to isolate “Africanisms”. On the basis of quantitative data such as the ones presented in Table 3 they suggest that it is possible to predict with a high degree of probability that if there is a language that possesses more than five of the eleven properties then this must be an African language. Furthermore, arguing that three of the eleven properties, namely labial-velar stops, implosive stops and ATR-based vowel harmony, have a higher diagnostic value than other properties, they conclude that if there is a language anywhere in the world that has two of these three properties then this must be an African language.

Table 3. Quantitative distribution of 11 typological properties according to major world regions (Sample: 99 African and 50 non-African languages; Heine and Leyew 2008: 30).

11 These properties are: (a) labial-velar stops, (b) 2 implosive stops, (c) lexical and/or grammatical tones, (d) ATR-based vowel harmony, (e) verbal derivational suffixes, (f) nominal modifiers follow the noun, (g) semantic polysemy of ‘drink’ or ‘pull’ for ‘smoke’, (h) semantic polysemy ‘hear’ or ‘see’ for ‘understand’, (i) semantic polysemy of ‘animal’ and ‘meat’, (j) comparative constructions based on the schema [X is big defeats/surpasses/passes Y], and (k) noun ‘child’ used productively to express diminutive meaning.
<table>
<thead>
<tr>
<th>Region</th>
<th>Total of languages</th>
<th>Average number of properties per language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>10</td>
<td>1.1</td>
</tr>
<tr>
<td>Asia</td>
<td>8</td>
<td>2.6</td>
</tr>
<tr>
<td>Australia/Oceania</td>
<td>12</td>
<td>3.0</td>
</tr>
<tr>
<td>The Americas</td>
<td>14</td>
<td>3.4</td>
</tr>
<tr>
<td>Africa</td>
<td>99</td>
<td>6.8</td>
</tr>
<tr>
<td>Pidgins and creoles</td>
<td>6</td>
<td>2.3</td>
</tr>
<tr>
<td>All regions</td>
<td>149</td>
<td></td>
</tr>
</tbody>
</table>

That the generalizations proposed by Heine and Leyew (2008) raise some problems is argued by Dryer (2011). In particular, he draws attention to the possibility that quantitative distributions such as the ones presented in Table 1 may be due to coincidence and that future work should aim at a more balanced sample of languages, especially of non-African languages. Dryer (2011: 310) finds only two of the properties included by Heine and Leyew (2008) that argue for sub-Saharan Africa as a linguistic area, namely postnominal modifiers and presence of tone systems.

### 5 Grammaticalization areas

Another domain where Africa provides a wide range of common properties concerns grammaticalization processes, whereby the same conceptual schemas and constructions are employed to develop grammatical categories, and in situations of language contact these processes can be transferred from one language to another.

The notion of linguistic area or sprachbund has not gone unchallenged over the last few decades (e.g., Stolz 2002), and rather than looking for compact linguistic areas, a number of scholars working on language contact prefer to adopt an areal perspective where linguistic areas are only of marginal concern, if at all (cf. Muysken 2008: 4 and to a certain extent Hickey 2010, this volume). As has been argued by Heine and Kuteva (2005: 182-187), most linguistic areas that have been identified so far can be analyzed, at least to some extent, as consisting of grammaticalization areas. With this term, they refer to groups of geographically contiguous languages that have undergone the same process of grammaticalization as a result of language contact (Heine 1994; Kuteva 1998; 2000; Stolz and Stolz 2001: 1549).

The following example may illustrate the process concerned. With few exceptions, Slavic languages are known for their lack of definite and indefinite articles. Now, in

---

12 Three of the six pidgin and creole languages are spoken in Africa and the rest in the Americas and in New Guinea.

13 Grammaticalization is defined as the development from lexical to grammatical forms, and from grammatical to even more grammatical forms. Since the development of grammatical forms is not independent of the constructions to which they belong, the study of grammaticalization is in the same way concerned with constructions, and with even larger discourse segments (see e.g. Traugott and Heine 1991; Heine, Claudi and Hünnemeyer 1991; Hopper and Traugott 2003; Bybee, Perkins and Pagliuca 1994).
situations of language contact where speakers of Slavic minority languages were exposed for centuries to intense contact with dominant languages having articles, these minority languages developed article-like constructions. In some cases, these developments gave rise to full-fledged articles similar to those of the respective model language German or Italian. These contact-induced developments involved canonical processes of grammaticalization, whereby e.g. a minority language such as Upper Sorbian in Eastern Germany grammaticalized a demonstrative attribute to a definite article and their numeral for ‘one’ to an indefinite article on the model of German (Breu 2003; 2004). The result is that the Slavic minority language Upper Sorbian now shares with German the same grammaticalization area of articles.

Contact-induced grammaticalization processes of this kind must have happened in Africa on a massive scale. While we lack appropriate historical records for reconstruction, grammaticalization theory allows to reconstruct at least an outline of some of these processes (Heine 1994; 2011; Leyew and Heine 2003; Heine 2011).

Perhaps the most widely discussed example is that of comparative constructions, more precisely of comparatives of inequality, based on what in Heine (1997) is called the action schema, taking either of the forms ‘X is big defeats/passes Y’ or ‘X defeats/passes Y in size’, i.e., the use of a verb meaning either ‘defeat’, ‘surpass’ or ‘pass’ to express comparison (Meeussen 1975: 4; Greenberg 1983: 4; Gilman 1986: 39), as seen in the following example.

(7) !Xun (W2 dialect, Khoisan; Heine and König forthcoming)14

\[\begin{align*}
\text{!xo} & \quad \text{má} & \quad n\llaa\à & \quad !\text{'aľa} & \quad \text{gùmì.} \\
\text{elephant} & \quad \text{TOP} & \quad \text{be.big} & \quad \text{pass} & \quad \text{cow}
\end{align*}\]

‘An elephant is bigger than a cow.’

The grammaticalization from either of these verbs to a grammatical marker denoting the standard of comparison (cf. English than)15 must have taken place in some form or other in many African languages across genetic boundaries. To be sure, this grammaticalization has also been observed in some other parts of the world. Especially in Mainland Southeast Asian languages (Sinitic languages, Thai, Vietnamese, and Hmong-Khmer languages) verbs for ‘to cross’ have given rise to standard markers of comparison (Ansaldo 2004: 490-493), but outside Africa it is extremely rare, while at least two thirds of the African languages that have been documented in some detail appear to have undergone this process in some form or other (Heine 1997: 126-9; Leyew and Heine 2003; Heine and Leyew 2008).

---

14The graphs ‘!’ and ‘ǁ’ stand for an alveolar and a lateral click type, respectively.

15To be sure, this contact-induced grammaticalization occurs also in other parts of the world, for example in Sinitic languages, Thai, Vietnamese, Hmong and Khmer, where verbs for ‘to cross’ have given rise to standard markers of comparison (Ansaldo 2004: 490-493), but outside Africa it is extremely rare, while roughly 80% of the African languages have it (see Heine and Leyew 2008).
Other possible pan-African grammaticalization areas are provided by processes leading from nominal to spatial concepts. Body part terms used metaphorically for deictic spatial distinctions are found throughout the world; for example, nouns for the body part ‘back’ are the conceptual source for spatial terms for ‘behind’ in many languages. But there are some developments that are likely in Africa but unlikely to happen elsewhere (Meeussen 1975: 3; Gilman 1986: 42). Such developments include, but are not restricted to, the grammaticalization of body parts for ‘stomach/belly’ to spatial concepts for ‘in(side)’, or of ‘buttocks’ or ‘anus’ to ‘below’ and/or ‘behind’ (Heine 1997: 37).

A number of other grammaticalizations that were proposed are less convincing as Africa-specific processes of areally defined diffusion. This applies on the one hand to sex distinctions used for the grammaticalization of the deictic spatial concepts ‘right(side)’ (> ‘male, strong hand’) and ‘left(side)’ (< ‘female, weak hand’) (Gilman 1986: 42). On the other hand, this also applies to the grammaticalization process leading from verbs for ‘say’ to quotative markers, complementizers, purpose clause markers, etc. (Larochette 1959; Meeussen 1975: 3; Gilman 1986: 44). But such processes appear to be also fairly common outside Africa (see Heine and Kuteva 2002) and, hence, do not appear to have a pronounced areal dimension.

6 Conclusions

Discussion in the present chapter was severely limited in scope and raises a number of problems that had to be ignored in this bird’s-eye view of linguistic areality in Africa. In particular, we were not able to do justice to the wide range of studies that have been or are being presently carried out on the dynamics of language contact and its areal implications in various parts of the continent. And we were also not able to do justice to the wide range of topics that researchers in areal linguistics are concerned with in Africa.

One major problem that we were confronted with concerns diachronic reconstruction. There are hardly any historical documents to support the reconstruction of historical events in Africa. Accordingly, the study of areal diffusion processes must rely on other means for reconstruction. To this end, we applied findings of grammaticalization theory to reconstruct areas of grammaticalization (Section 4). The theory draws on observations on regularities made in languages that are historically well documented to study languages for which there are no or hardly any earlier historical records. But it goes without saying that this theory is limited in scope; it works well when applied to morphosyntactic phenomena but is hard pressed when it comes to phonological, lexical, or semantic processes.

The processes discussed in the chapter are all hypothesized to be due to language contact, that is, to have been externally induced. At the same time, many of them can also be described as involving language-internal developments. That external and internal linguistic changes are by no means mutually exclusive is not really new (see e.g. Thomason 2001; 2003; Heine and Kuteva 2005, Hickey 2012), but what that exactly means with reference to our understanding of linguistic change and of language structure is an issue that would seem to need much further attention in future research.

Abbreviations
ABL = ablative; DIM = diminutive; LOG = logophoric marker; SG = singular; TOP = topic marker; 3 = third person.

References


15 Areal contact in Nilo-Saharan

Gerrit Dimmendaal

1 Introduction

The phylum referred to today by the name Nilo-Saharan is spread over a vast area mainly south of the Afroasiatic phylum and north of the Niger-Congo phylum. The core of Nilo-Saharan was established more than fifty years ago by Joseph H. Greenberg. In his earliest genetic classification of African languages, Greenberg (1955) proposed a Macro-Sudanic family (renamed Chari-Nile in subsequent studies), consisting of a Central Sudanic and an Eastern Sudanic branch plus two isolated members, Berta and Kunama. This family formed the core of the Nilo-Saharan phylum as hypothesized in Greenberg (1963), where a number of groups were added which had been treated as isolated units in his 1955 classification, namely Songhay, Saharan, Maban and Mimi, Nyangian, Temainian, Koman and Gumuz. The present author, however, prefers to treat Songhay (spoken mainly in Mali) as well as Koman plus Gumuz (in the border area between Ethiopia and Sudan) as independent families (Dimmendaal 2011). This also applies to the Kadu languages spoken along the southern range of the Nuba Mountains, Sudan, which had been classified as Kordofanian, i.e. as members of the Niger-Congo phylum by Greenberg (1963), but which has been argued by Bender (1996) to be part of the Nilo-Saharan phylum.

Whereas there is little disagreement on the genetic unity of lower level units of Nilo-Saharan, scholars like Bender (1996) and Ehret (2001) disagree on the subgrouping at deeper historical levels. The present author has defended the following subgrouping in a number of publications, e.g. Dimmendaal (2010 a), where it is argued that Northeastern Nilo-Saharan languages innovated a case-marking system as well as other typological properties; Dimmendaal (to appear a) provides further evidence for the subgrouping presented in Figure 1. This classification will also form the basis for a survey of areal phenomena below, since the subgrouping corresponds to important typological differences between its extant members.
Figure 1. The subclassification of Nilo-Saharan

The approximately 60 Central Sudanic and 100 Northeastern Nilo-Saharan languages cover an area stretching from Eritrea and Ethiopia in the East across Northeastern Africa towards Niger with a southward extension into Eastern and Central Africa. The present typological survey is based on a summary of the existing literature on areal contacts in Nilo-Saharan with neighbouring groups, but some new perspectives are added too, for example on areal features shared between Central Sudanic and neighbouring Ubangian languages as well as on the areal distribution of (split) ergativity.
Map 1. Nilo-Saharan

Convergence as a manifestation of areal contact has to do with multilingualism and its linguistic reflexes, shift-induced interference and metatypy. A second purpose of the present contribution is therefore the identification of areal sources for these phenomena. A third and final aim is to show that typological similarities between languages are not necessarily the result of areal contact, since self-organising principles may also result in parallel developments. As argued below, these latter principles rather than areal contact probably help to explain the presence of so-called Marked Nominative case systems in Afroasiatic as well as Nilo-Saharan languages.

2 Central Sudanic

Within the Nilo-Saharan phylum, there is a relatively sharp typological cut between Central Sudanic languages on the one side and Northeastern Nilo-Saharan languages on the other. Whereas most languages involved share widespread prosodic features of African languages south of the Sahara such as Advanced Tongue Root harmony and tone,
these two primary branches of Nilo-Saharan differ in their morphosyntactic and pragmatic structures. Whereas Central Sudanic languages have a basic constituent order SAuxOV or SVO, Northeastern Nilo-Saharan languages tend to be verb-final, with the exception of a group of Eastern Sudanic languages, which are verb-initial or verb second, as shown below. This dichotomy between the two primary branches of Nilo-Saharan coincides with a number of other typological differences, such as the extensive number-marking system, the use of case, converbs, and coverb plus light verb constructions in Northeastern Nilo-Saharan. Whereas converbs refer to dependent (subordinate) verb forms with a reduced inflectional morphology, coverbs are complements of ‘say’/’do’ verbs belonging to different syntactic categories like noun, adjective or adverb. These properties are essentially absent in Central Sudanic languages, where we find properties shared with neighbouring languages belonging to the Ubangian language family, such as a reduced derivational morphology and word structure, and the frequent use of discourse particles in postverbal position. This latter typological zone is discussed first below.

Map 2. The Central Sudanic-Ubangi contact zone

Greenberg (1966: 87) and Houis (1970: 59-66) made reference to a language type which is particularly common in West Africa and which is characterized by genitives preceding rather than following the governing noun and postpositions instead of prepositions.

This type of constituent order is common, for example, in Gur, Kru, Kwa, Mande, and Songhay. Heine (1976: 41-42) uses these properties as definitional features for his
type B languages (as against type A or SVO, C or VSO, and D or SOV languages). Type B languages are further characterized by a constituent order Verb Adverbial Phrase and an SVO order, sometimes alternating with SOV, according to the same author. Heine (1976: 42) further points out that type B languages are also common in Central Sudanic. This well-defined branch of Nilo-Saharan is usually divided into two primary branches: a Western (or Bongo-Bagirmi) branch and an Eastern (or Moru-Mangbetu) branch (Tucker and Bryan 1966, Boyeldieu and Nougayrol 2008). The Western branch consists of (Sara-)Bongo-Bagirmi plus Kresh, whereas the Eastern branch consists of Moru-Madi, Mangbutu-Efe, Mangbetu and Lendu-Ngiti. S(Aux)VO constituent order, prepositions and possessed/possessor order are common in the Western branch. The common constituent order in the Eastern branch is SAuxOV alternating with SVO, with postpositions rather than prepositions, and the possessor preceding the possessed in genitive constructions. As shown by Andersen (1984), the alternation itself depends on aspectual marking in a clause; if the position after the subject is filled by an auxiliary (marking aspect), the main verb occurs after the object. The following examples from Moru (interlinear glossing added by the present author) illustrate the SVO/SAuxOV alternation:

(1) mī-zī ngagà
   2SG-call boys
   ‘you (sg) called the boys

(2) ɲ-á ngagà ū-zī
   2SG-AUX boys NF-call
   ‘you (sg) are calling the boys’

Similar constituent order patterns occur in neighboring Ubangi languages. Greenberg (1963) refers to the Ubangi languages by the name Eastern and argues that it forms a subgroup with Adamawa within Niger-Congo. Boyd (1989: 192) arrives at the following subgrouping.

1. Gbay
2. A. Gbanda
   B. Ngbandi
   C. I. Sere
   II. a. Ngbaka
   b. Mba
3. Zande

Dimmendaal (2011:89-90) prefers to treat Ubangi as an isolated language family instead. It has been argued by Moñino (2012) that Ubangi does not even form a genetic unit in and by itself. However, whether Ubangi constitutes an areal rather than a genetic grouping is irrelevant for the present study, since areal features of Central Sudanic are at stake. And the genetic unity of Eastern and Western Central Sudanic is beyond any doubt (Boyeldieu and Nougayrol 2008: 11).

In the Mba group within Ubangi, the genitive order is possessor-possessed, whereas postpositions (though rare) occur as well (Tucker and Bryan 1966: 123, 126, 131). In at least one member of the Mba cluster, Ndongo SAUXOV constituent order
alternates with SVO, depending on Aspect, as in neighbouring Central Sudanic languages:

(3) á-mlba ngàràgʊ mè
    1SG-AUX   child   beat
    ‘I had beaten the child’

Constituent order in other Ubangi members (the Zande cluster, Gbaya, Gbanda, Ngbandi, Sere, Ngbaka) is similar to that in Western members of Central Sudanic, in that S(Aux)OV is the common constituent order, as are prepositions whereas in genitive constructions the possessor follows the possessed. Hence, they are classified as type A languages by Heine (1976). Gbaya, Gbanda, Ngbandi and Zande constitute expansion zones (in the sense of Nichols 1992), as shown on Map 2. The current distribution of Sere, Ngbaka and Mba at the periphery north, east, and south of these two expansion zones on the other hand suggests that their ancestral communities were either engulfed by expanding Zande communities or their ancestral communities were expelled from areas now occupied by Zande speakers. Typologically, these peripheral Ubangi groups pattern along with the Western branch of Central Sudanic.

Of course, the similarities in terms of constituent order, the position of adpositions and the order in genitive constructions could all be a coincidence. But a number of other common typological properties of languages in this area suggest that contact played a role too even though the direction of areal diffusion is not clear. For example, Central Sudan languages are characterized by a restricted system of derivational morphology (with only a few affixes; Dimmendaal 2014a) compared to the other primary branch, Northeastern Nilo-Saharan; instead, compounding is commonly used as a lexical strategy in the former. Also, nominal and verbal roots tend to be monosyllabic in Central Sudanic, with a preference for open syllables ((V)CV) and up to four tonal registers. This again parallels the system in neighboring Ubangi languages; see Tucker and Bryan (1966: 26-166) for examples.

A further typologically interesting areal feature shared between Central Sudanic and Ubangian languages, first described by Tucker (1967: 247-262, 408-412), is the presence of up to three discourse markers usually occurring sentence-finally or postverbally. These particles, expressing how the speaker views the state of affairs being described in an utterance or how (s)he wishes it to be seen by the hearer, tend to be monosyllabic elements which form a phonological word with verbs when these precede; when more than one such marker occurs, these particles tend to form a phonological word together. Tucker (1967: 166-176, 247-262, 408-412) gives examples from a range of Eastern members of Central Sudanic, e.g. from Logo:

---

1 Tucker (1967) refers to these languages with the name “Eastern Sudanic”, as they constitute the easternmost representatives of the ancient grouping “Sudanic”. His use of the term should not be confused with the term Eastern Sudanic in Greenberg’s classification, which refers to a different branch of Nilo-Saharan.

2 Such markers, which have also been referred to as discourse markers, attitude markers, or social-expressive and emotive elements, have been studied in detail for various Indo-European languages, but far less so in an African context.
The clitic nature of these particles is evident from their variable syntactic position, as shown for Logo (Tucker 1967: 247)

(6) ma-fule’bi dre
    1SG-kill water-buck PART
    ‘I killed a water-buck’

(7) ma-fu-dre le’bi
    1SG-kill-PART water-buck
    ‘I killed a water-buck’

The position of these particles within a sentence presumably depends on their scope, which may involve the verb or a verbal phrase, as in the examples above. Blackings and Fabb (2003) discuss such particles for the Central Sudanic language Madi, which belongs to the same cluster as Logo, under the heading “modals” (pp. 451-466) and “adverbials with a discourse function” (pp. 532-536). The modal ra, for example, expresses epistemic modal force:

(8) a. kò-mū
    3:DIR-go
    ‘she should go’

b. kò-mū rá
    3:DIR PART
    ‘she should definitely go (ensure that she goes)’

Santandrea (1970: 62-63, 75, 123) discusses these discourse particles for Yulu-Kara (which belong to the Western branch of Central Sudanic) and points out (p. 62) that “[s]ome (e.g. Tucker) have called [these] postpositions, but I think the rather vague term, particles, [is] quite suitable for the purpose.” The author further points out (p. 127) that these particles “…are so closely jointed to the preceding word that in pronunciation, they form a single word with it”, as in the following example from Yulu:

(9) ake luu’bo-lee
    3PL.AUX find-PART
    ‘they will find it/him/her’

Similar particles, again usually occurring post-verbally or sentence-finally and also behaving prosodically as clitical elements, have been reported in studies of neighboring Ubangi languages. Within the Zande cluster, this phenomenon is restricted to one language, Barambu-Pambia, according to (Tucker and Bryan 1966: 153), which borders on Central Sudanic languages such as Mangbetu:

(10) nye-nzí tùngúà
    1SG-AUX work PART
    ‘I have done work’
Tucker and Bryan (1966: 83) further observe: “These languages are characterized by a great number of words and/or Particles which are … postpositional to Verbs….Many of these (…) correspond to the Postpositions of MORU-MANGBETU” (Tucker and Bryan 1966: 83).

Cloarec-Heis (1986: 62-67) describes this phenomenon for the Banda-Gbaya-Ngbandi languages, and contrasts these syntactic elements with “obligatory elements (“syntagmes necessaires”). In his analysis of the Ubangian language Yakoma, a member of the Ngbandi cluster, Boyeldieu (1995: 130) discusses this phenomenon of postverbal particles under the heading of “modalité de proposition”, as in the following example, where the particle expresses a kind of counter expectation:

(11) à hwɛ̀wɛ̀ ‘it has already been done’  
    it be.finished PART

3 Northeastern Nilo-Saharan and (former) contact zones

In his typological survey of African languages, Heine (1976) pointed out that features associated with Afroasiatic languages in the “Ethiopian convergence area”3 extend into Nilo-Saharan languages spoken to the west of this area. These common features involve a verb-final constituent order and extensive case marking, which are also features of Nilo-Saharan subgroups such as Maban, Fur, or northern members of the Eastern Sudanic branch as well as Kunama (see Map 3 for the distribution of these language groups).

---

3 This has been proposed by different authors, e.g. Leslau (1945), by Ferguson (1970). The notion has also been partly criticized by Tosco (2000).
There is also no historical-comparative evidence that case marking was lost in Central Sudanic languages (as far as present knowledge goes). Boyeldieu (2013) shows that one language, Sinyar, a language in Chad bordering on Northeastern Nilo-Saharan languages with case systems (all of which have a system of Differential Object Marking) has a Marked Nominative case system, a phenomenon which is also attested in a group of Eastern Sudanic languages, Nilotic and Surmic, discussed in section 4 below.  

---

4 Given the geographical distance between Sinar on the one hand and Nilotic or Surmic on the other hand but also the dramatic typological differences between Sinyar and Nilotic plus Surmic in terms of other morphosyntactic strategies or constituent order, there is no reason to assume that
ngàar-rí  wéññí
stew-NOM:SG  3:be.nice
‘the stew is nice’

sùdàn-nè  ùññàbí  ìngiltáràa
Sudan-NOM:SG  3:be.hot.pass  England:ACC:SG
‘Sudan is hotter than England’

However, as pointed out by Boyeldieu (2013) it is not at all clear that Sinyar is a member of the Central Sudanic family; Dimmendaal (to appear b) treats this language as a linguistic isolate.

Additional typological features supporting the claim by Heine (1976) that the Ethiopian convergence area extends into Sudan and Chad have been listed since by Dimmendaal (e.g. 2007, 2008). These properties include a tripartite number marking system, converbs as a discourse strategy, and coverb plus light verb constructions, which are illustrated briefly below.

Most Northeastern Nilo-Saharan languages have rich number-marking systems of a type which is also attested in Afroasiatic languages in Ethiopia and Semitic languages in the Middle East. Prototypically, such languages have a tripartite number-marking system, whereby either the plural or collective form is morphologically unmarked (with corresponding singulative marking), or the singular (with corresponding plural marking on the non-basic form); alternatively, both the singular and plural are inflected for number. The examples below stem from Dimmendaal (2000):

Syrian Arabic (Semitic, Afroasiatic)
(14) tɔrk-i  ‘a Turk’
  tɔrk  ‘Turks (collective)’
  ʔatrɔk  ‘Turks (plural)’

Masalit (Maban, Nilo-Saharan)
(15) anyiŋ-ɡi  ‘fly’
  anyiŋ  ‘flies’
  mama  ‘maternal uncle’
  mama-ɡe  ‘maternal uncles’
  mal-ɡo  ‘chattel’
  mal-ta  ‘chattel’

Central Sudanic languages on the other hand, have a restricted system of plural marking (mainly on nouns referring to animate or human entities); again, there is no historical evidence that these languages lost the tripartite system of number inflection.

Converbs constitute another common strategy, characteristic in particular of narrative discourse in Afroasiatic languages in Ethiopia and Nilo-Saharan languages spoken in a zone ranging from Ethiopia and Eritrea across northern and central Sudan into Chad, as shown by Amha and Dimmendaal (2006). The authors further point out that these constructions have often been referred to by such labels as ‘gerunds’, ‘particples’,

these systems have a common historical origin. Similar variation between languages with Differential Object Marking as against Marked Nominative case systems are found, for example, in Cushitic (Afroasiatic). Hence, such parallel systems are more likely to have come about independently through autogenetic (language-internal) processes, i.e. through self-organization.
‘consecutives’ and other names in the past. This is one reason why this areal strategy has gone largely unnoticed in the general typological literature on converbs.\(^5\)

One of the most detailed analyses of converbs in a Northeastern Nilo-Saharan language is to be found in Jakobi and Crass (2004: 165-176) on the Saharan language Beria. In this language, spoken in Chad and Western Sudan, both converbs and main verbs take pronominal subject and object marking. With converbs, however, mood is not expressed, whereas plurality marking and aspect marking can be omitted. The first type of converb in Beria expresses mainly successive actions or events (apart from a number of other functions),

\[(16) \ bélé \ tine=e-n-e \ sè-nɔ̃gɔ́ \ goats \ thief-AUX-sSG:SU-CONV \ eat-IMP:NEG \ ‘don’t steal goats and don’t eat them’\]

The second type of converb construction is used in clauses expressing a purpose or a volitional act.

\[(17) \ ékɔ̀l-dɔ̀ \ gèr=t-e \ gù-g-í \ school-ADV \ read-AUX-1SG:SU-CONV \ go-1SG:SU-AFF-PERF \ ‘I went in order to study in/at school’\]

Such dependent verb forms are deeply rooted in the structure of the Omotic branch within Afroasiatic, for example. Based on two alternative morphological strategies for verbs, Rapold (2008) distinguishes between converbs and medial verbs in the Omotic language Bench. Both involve finite verbs in dependent clauses, but medial verbs involve co-subordination rather than subordination, as with converbs. This is reflected in the use of different pronominal anaphors as well as tense-aspect marking on these two types of dependent verbs. Converbs in Bench are less finite than medial verbs, for example, in that they carry no person-sensitive markers (Rapold 2008: 179). The following is an example with a medial verb whose interpretation in terms of tense or mood depends on the clause-final verb (Rapold 2008: 176-177).

\[(18) \ cɛ̃gɛ̃l-ʃɛ̃l \ fàl-ʃɛ̃ \ ‘clap and sing (Pl)!’ \]

Whether there are also Northeastern Nilo-Saharan languages which distinguish between converbs and medial verbs is not known. Omotic languages also manifest switch reference between pronominal subjects in clause chaining. Again, to date no such system has been identified yet for any Northeastern Nilo-Saharan language.\(^6\)

---

\(^5\) Additional instances of convergence may be found at the lexical level. Hayward (1991) presents various examples with respect to Semitic, Cushitic, Omotic. Whether at least some of these extend into Northeastern Nilo-Saharan has not been investigated yet.

\(^6\) Cyffer (2002) argues that there is a typological break within the Saharan branch in that languages like Kanuri match up typologically with Chadic (Afroasiatic) and other language groups in West Africa.
A further typological property shared by Afroasiatic languages in Ethiopia and Northeastern Nilo-Saharan groups is the frequent use of coverb plus light verb (‘do/say’) constructions, as in the following examples from Tama, a member of the Eastern Sudanic branch (data with the author), where they commonly express positions of the body:

\[(19) \quad \text{wut nV-} \quad \text{‘fall down’} \]
\[(19) \quad \text{wii nV-} \quad \text{‘return’} \]
\[(19) \quad \text{salla nV-} \quad \text{‘pray, prostrate’} \]

In many Northeastern Nilo-Saharan languages in the Chad-Sudan area, such light verbs (e.g. nV- in the case of Tama) also tend to be used to accommodate verbs borrowed from Arabic. These languages differ as to the number of coverb plus light verb constructions they have. In Saharan languages like Kanuri, more than 90 percent of the verbal predications are built with the verb ‘say’ (Hutchison 1981). The degree of productivity of this lexical strategy also varies between different Afroasiatic languages (where these light verbs are particularly common in combination with ideophonic words; see Amha 2001 for a detailed analysis of this phenomenon in the Omotic language Wolaitta).

Dixon (2002: 184 ad passim) describes a similar variation in terms of productivity of coverb constructions for languages in Australia, and argues that these languages appear to go through cycles in this respect. Whereas in some Australian languages, verbal predications are formed mainly through coverbs plus light verb constructions, other languages use this strategy next to a set of basic verbs, and in still others coverbs are absent. This corresponds to the situation in Northeastern Nilo-Saharan, where converbs are extremely common in the Saharan branch; in Eastern Sudanic groups like Taman, they are used next to a set of underived verbs, whereas in Nilotic or Surmic they are absent.

Given the roughly west-to-east areal distribution of these typological features in Northeastern Nilo-Saharan and given the former presence of a major riverine system in this area, it has been suggested by Rilly (e.g. 2007) and Dimmendaal (e.g. 2007, 2008) that this former tributary of the Nile, the Yellow Nile or Wadi Howar, most likely functioned as an important diffusion zone between Afroasiatic languages (or languages with typological features similar to Afroasiatic languages in Ethiopia today) and Northeastern Nilo-Saharan. This former river (shown on Map 1) connected the Ennedi Mountains in Chad with the White Nile between the third and fourth cataract roughly between 10,000 BP and 3000 BP. As family-internal evidence in Nilo-Saharan suggests that Northeastern Nilo-Saharan acquired these features, and as these properties appear to be deeply rooted in Afroasiatic languages in Ethiopia (and beyond, in the case of number marking), they probably entered the former through areal contact.

Until recently, this scenario had to remain somewhat speculative. But these claims on population movements and areal influence through language contact have found independent support from archaeological research and genetics (DNA research and osteology), as shown by Becker (2011). Archaeological findings suggest that a “Pre-Leiterband” population of hunters and gatherers entered the Wadi Howar area in what is now Sudan from the east around 10,000 BP when wetter periods set in on the African continent. (Before that, human populations were restricted for thousands of years to zones along major rivers like the White Nile or higher elevations.) These hunter-gatherers spread across the Wadi Howar riverine system within a relatively short period and lived in semi-permanent settlements. Some 2,000 years later, pastoralists representing the
Leiterband culture (after the pottery tradition associated with their cultures) and originating west of the Wadi Howar in Chad moved eastwards and absorbed many of these Pre-Leiterband populations. This scenario is supported by osteological as well as DNA research (Becker 2011). If the Pre-Leiterband populations indeed spoke Afroasiatic languages, they would have transferred features such as extensive case marking, converbs and other typological features onto their new primary languages, which belonged to the Nilo-Saharan phylum.7

4 The typological break within Eastern Sudanic

The desertification of the Wadi Howar area over the past 4,000 years must have forced people to migrate into wetter or higher elevations west, south and east of this zone, as archaeological research in the area indeed has confirmed (see, for example, Becker 2011). This is also reflected in the geographical distribution of Northeastern Nilo-Saharan languages, in particular of the largest subgroup, Eastern Sudanic, consisting of a Northern branch and a Southeastern branch. Whereas the Northern members are scattered over a vast area in Chad, Northern and Central Sudan plus Eritrea, the Southeastern members in the Nuba Mountains (in Sudan) as well as in areas west, east and south of this area with an extension into Central Tanzania (see Map 3).

Figure 2. The subclassification of Eastern Sudanic

---

7 Thomason and Kaufman (1988) have shown that shift-induced interference and heavy borrowing in a language shift situation are common causes for the kind of dramatic typological changes illustrated above.
The Northern members of the Eastern Sudanic branch share the set of typological properties described for other members of Northeastern Nilo-Saharan above. But the Southeastern members differ rather dramatically from the former. Whereas most languages belonging to the Southeastern branch of Eastern Sudanic have retained the extensive number marking system as a lexical phenomenon, their morphosyntactic structures differ in several respects from those found in the Northern subbranch. Instead of a verb-final constituent order, a verb-second or verb-initial order is common in Nilotic and Surmic as well as Gaahmg (the only Jebel language still spoken today). This change in constituent order is also accompanied by a different case system, namely marking of postverbal subjects (A-roles), and absence of case marking on objects in these Southeastern Eastern Sudanic languages; in addition, these groups manifest a reduced case-marking system compared to Northern Eastern Sudanic, with only remnants of peripheral case marking (Dimmendaal 2005). The common system elsewhere in Northeastern Nilo-Saharan is one whereby the subject occurs preverbally and is not marked for case, whereby objects are subject to a differential object marking system (Dimmendaal 2010 a).

Map 4. Languages allowing for postverbal subjects (A-roles)

---

Cases discussed here involve contact phenomena in situations with more or less stable patterns of bilingualism. Alternatively, of course, language loss may occur. Whereas Nubian languages influenced Sudanese Arabic in various respects (something which cannot be discussed here for reasons of space), the latter is now gradually replacing Nubian languages in Sudan.
The dramatic change in constituent order and case-marking strategies in languages belonging to the Southeastern branch of Eastern Sudanic - all spoken east and south of the Nuba Mountains – was probably triggered through areal contact.\(^9\) The Kadu languages, spoken along the southern ranch of the Nuba Mountains in Sudan, which constitute an independent family according to the present author, are all verb-initial. A verb-initial constituent order with case marking on postverbal subjects is also found in the Kuliak languages, which are spoken in Northeastern Uganda (see Map 3). Andersen (1995) characterizes Berta (which is spoken northeast of the Nuba Mountains in Sudan) as a language with Topic Verb Subject Object Adverbial constituent order, and shows that this language has a case system, whereby both S (the subject of intransitive predications) and A (the subject of transitive predications) are inflected for case in postverbal (but not in preverbal) position.

\[(20)\]
\begin{align*}
a. \quad & bùuŋù ìid-óo gài \quad \text{hyena bite-PAST dog:NOM} \\
& \quad \text{‘the dog bit the hyena’} \\
b. \quad & gài ìid-óo bùuŋù \quad \text{dog bite-PAST hyena:NOM} \\
& \quad \text{‘the hyena bit the dog} \\
c. \quad & qól-á-láa gài \quad \text{eat-PERF-NTS dog} \\
& \quad \text{‘the dog has eaten it’}
\end{align*}

Exactly the same system, with case inflection for postverbal subjects (S and A), usually referred to as a Marked Nominative case system (König 2006), is found in a range of Nilotic and Surmic languages. The case system of the Kuliak languages in northeastern Uganda is more complex (see Schrock 2013), but verb-initial predications are also common in this family. Nominative case may be expressed through segmental inflection, tonal inflection or a combination of these, sometimes in one and the same language.

Apart from Nilotic and Surmic languages with Marked Nominative case systems, we find languages belonging to the same subgroups which only allow subjects of transitive predications to occur postverbally. Hence, only (postverbal) A is marked for case, whereas S occurs preverbally and is not marked for case. Such split ergative systems are attested in Western Nilotic languages like Anywa, Päri and Shilluk, in Surmic languages like Majang or Tennet as well as in Gaahmg (Jebel group). The following examples illustrate these phenomena for Gaahmg, which is spoken east of the Nuba Mountains in Sudan (example adapted from Stirtz 2013):

\[(21)\]
\begin{align*}
& \text{nám- ãn-s-i} \quad \text{aggaàr(-) \break CAUS-AP-COMP-PASS.A hunter.GEN-ERG} \\
& \quad \text{‘something was broken by a hunter making someone break it’}
\end{align*}

\(^9\) Two other branches of Southeastern Eastern Sudanic, the Daju and Temeinian group, which are spoken in the Nuba Mountains and west of this area (extending into Chad) again have retained the number-marking system, but there are no traces of case marking (as far as present knowledge goes); moreover, SVO is the common constituent order, although languages like Sila and Njangulgule also allow for a verb-final constituent order, as in other Northeastern Nilo-Saharan languages in the area.
The system of postverbal case marking for Southeastern Eastern Sudanic languages probably constitutes a shared innovation of this subgroup, as the actual (ergative) case morphemes are cognate. Whereas in Gaahgm (Jebel) and in Western Nilotic Anywa plus Päí, the alternation between -ɛ/-e and -ɪ/-i is phonologically conditioned, the Surmic language Tennet uses -ɪ/-i with proper names, whereas -ɛ/-e occurs with other types of postverbal subjects. What is more, Tennet has a Marked Nominative system for main clauses, i.e. both A and S can occur postverbally and consequently are inflected for case, whereas in subordinate clauses only A can occur in this position, given rise to a split ergative system.

\[(22) \, \text{óröng} \, \text{Lowór-i} \, \text{kåkǻt} \, \text{Lohám-i} \, \text{áríz} \]

want Lowor-NOM spear:SUBJ Loham-NOM bull:ABS

‘Lowor wants Oham to spear the bull’

\[(23) \, \text{óröng} \, \text{Lowór-i} \, \text{Lohám} \, \text{kikíya} \]

want Lowor-NOM Loham:ABS come:SUBJ

‘Lowor wants Loham to come’

The postverbal Agent (ergative) marker -ɛ/-e in these Southeastern members of Eastern Sudanic is cognate presumably with an Instrumental case marker with the same form in Northern Eastern Sudanic, whereas the case marker -ɪ/-i probably is cognate with a Genitive case marker in the Northern branch of Eastern Sudanic, as argued in Dimmendaal (2014b).10 According to the same author, the Marked Nominative system, i.e. the case-marking strategy whereby both S and A are inflected for case in Nilotic and Surmic, results from an extension of postverbal subject marking from A to S roles, i.e. from transitive towards intransitive predications; this is still the situation for subordinate as opposed to main clauses in Tennet above.

The fact that Nilotic and Surmic languages are spread over a huge area across South Sudan and extending into the Democratic Republic Congo, Uganda, Kenya, Tanzania and Ethiopia, and the fact that they are relatively closely related genetically, suggests that we are dealing with expansion zones in the sense of Nichols (1992). East of this expansion zone in the border area between Sudan, South Sudan and Ethiopia, and west of this zone in the Nuba Mountains, we find accretion or residual zones, i.e. areas characterized by genetic as well as typological diversity. Many of the languages in these

10 Western Nilotic Shilluk uses an alternative strategy, namely a proclitic marker, in combination with Agent nouns or noun phrases (Miller and Gilley 2007). The same prenominal strategy is found in the Niger-Congo language Tima, spoken in the Nuba Mountains.
areas are poorly studied. But initial evidence is now emerging that properties such as verb second structures, split ergativity and the postverbal marking of subjects are in fact found in a number of languages in these residual zones. Uduk, a member of the Koman family (treated as an independent language family in Dimmendaal 2011) has been reported to have OVS (or OVA) order with formal marking of the postverbal subject (Don Kilian, p. c.). There is at least one Niger-Congo language in the Nuba Mountains, Tima, with split ergativity (Dimmendaal 2010b), but there may be other languages in this linguistically heterogeneous area manifesting the same type of syntactic alignment.

Surmic languages bordering on Omotic (i.e. Afroasiatic) languages in Ethiopia show some initial typological convergence with the latter through the presence of ejective consonants. These of course are common in Afroasiatic languages in Ethiopia, but within the Eastern Sudanic branch of Nilo-Saharan they are confined to Southeastern Surmic languages bordering on Omotic, and in the Kuliak in northeastern Uganda, where they may be due to (former) contacts with Cushitic languages.

Nilotic languages belonging to the Southern Lwoo cluster and bordering on Niger-Congo languages in South Sudan and Uganda also manifest clear evidence of areal contact with their distant relatives Central Sudanic as well as with Ubangi. In her monograph on the noun morphology of Western Nilotic languages, Storch (2005) identifies several layers of chronologically organized stages of historical influence, amongst others resulting in the formal reduction of nominal number marking in Southern Lwoo. As shown by the same author (Storch 2005: 216-379), these languages also have various paired prefixes stemming from derivative morphemes (some of which go back to lexical roots). In Thuri, for example, one finds a deverbal stem ŋàt-dwáár (singular) / jò-dwáár (plural) ‘hunter’ (Storch 2005: 201). This latter system formed the basis for an even more elaborate system of number marking by way of prefixation in the Southern Lwoo language Luo, as shown in the following.

## 5 The Nilotic borderland

The southernmost member of the Lwoo cluster within Western Nilotic, Luo, is spoken in Kenya, Tanzania and Uganda, where its speakers have been in contact with speakers of Bantu languages for more than a century. The socio-historical context in which these contacts between essential pastoral speakers of Nilotic languages and mainly agricultural speakers of Bantu languages occurred, probably over several centuries, is summarized in Dimmendaal (2001). In the case of Luo, language shift among speakers of Bantu languages towards this Nilotic language was accompanied by heavy unadapted borrowing, e.g. by taking over the corresponding noun class alternation from the source language, sometimes combined with number suffixation, as in mIsùmba (sg) / wa-sumb-ní (pl) ‘bachelor(s)’. Additional noun-class prefixes emerged through a reinterpretation of specific lexical roots as prefixes, as with the irregular singular / plural root for ‘guest’ ja-/jo- from *jal- / *jol- (Dimmendaal 2001).

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>jà-lùò</td>
<td>jò-lùò</td>
</tr>
</tbody>
</table>

Luo has several such prefixes, parallel to the noun-class system of neighbouring Bantu languages. The actual prefixes are not cognate with those found in Bantu (except in
borrowed nouns), but their formal as well as semantic structure is similar. Thus, the prefix *ra-* is used in Luo to express instruments (*ra-*ŋɪ̂(y) ‘mirror’ < ŋɪ̂yɔ ‘recognize’) as well as in referring to a person with a handicap (*ra-*ŋɔ ‘lame person’). This parallels the use of the noun-class prefix *ki-* covering exactly the same semantic domains in neighbouring Bantu languages.

A further grammatical domain with a clear Bantu influence is found in the verbal system of Luo, where a four way distinction for past tense is found.\(^{11}\) A similar tense-marking system (with three distinctions for past tense) is also found in Kalenjin, a language cluster whose speakers have also been in close contact with Bantu speakers.\(^{12}\)

<table>
<thead>
<tr>
<th>Luo</th>
<th>Kalenjin</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ne-</em> ‘earlier today, recently’</td>
<td><em>ka-, kə-</em> ‘today in the past’</td>
</tr>
<tr>
<td><em>nyo-</em> ‘yesterday’</td>
<td><em>kɔ-, ko-</em> ‘yesterday’</td>
</tr>
<tr>
<td><em>yand(e)-</em> ‘a few days ago’</td>
<td><em>kɪɪ-, kii-</em> ‘before yesterday’</td>
</tr>
<tr>
<td><em>ne-</em> ‘long ago’ (tonal difference on verb stem from ‘earlier today’)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Past tense markers in Luo and Kalenjin

However, whereas the Nilotic languages Luo (Western Nilotic) and Kalenjin (Southern Nilotic) both show influence from neighbouring Bantu languages in their verb systems, there are also interesting differences between the two. Kalenjin retained the rich, archaic tripartite system of number marking by way of suffixes of Northeastern Nilo-Saharan (discussed in section 3 above), and there is no sign of convergence towards Bantu in this grammatical domain. In other words, convergence towards Bantu at the nominal level in Luo was “pre-conditioned” by the already highly reduced system of number-marking suffixes in Luo. These divergent patterns of replication in Western Nilotic Luo and Southern Nilotic Kalenjin show that the typological distance between languages is another factor affecting the (short term) outcome of language contact.

In the case of the Datooga cluster within Southern Nilotic, there is evidence for areal contact with neighboring Bantu and Cushitic languages. Kießling, Mous and Nurse (2008) have described the complex linguistic situation in the Tanzanian Rift Valley area, where West Rift Cushitic languages are spoken next to Bantu, the linguistic isolate

\(^{11}\) See Nurse (2008) for a description of this widespread system in Bantu languages.

\(^{12}\) Language shift from a Bantu language towards Kalenjin and corresponding shift-induced interference from the primary language lies at the heart of this convergence process, as argued in Dimmendaal (2001). The same process presumably occurred when hunter-gatherers speaking Afroasiatic (or typologically similar) languages in the Wadi Howar area shifted towards Nilo-Saharan, thereby causing replication of grammatical features in the newly acquired primary languages.
Hadza, as well as Sandawe (which is either a linguistic isolate or a language which is genetically related to Central Khoisan), Eastern Nilotic Maasai, and Southern Nilotic Datooga. The authors also point towards “… a progressive trend away from postpositional marking of spatial relations … towards prepositional marking. This kind of reinforcement of prepositions is symptomatic of a more general trend in West Rift towards head-initial order… Syntactic position and conceptual models must be viewed as reflecting a Datooga substrate, probably originating in a large number of Datooga speakers shifting to pre-Iraqw and imposing Datooga syntactic and semantic structures onto pre-Iraqw” (Kießling, Mous, and Nurse 2008: 215). Datooga has incipient preverbal clitic clustering (comprising four structural slots in front of the verb). This is atypical for (Southern) Nilotic, but characteristic of neighbouring West Rift Cushitic languages, where this complex expresses tense, subordination, sequentiality and focus.

\[(25) \text{ák-àjà  gába síisí  gùurs-á  òorjéedàà-ɲi} \]

SEQ:AFF-FUT  every person  call:APPL-3  son-POSS.3SG

‘then everyone will call his son’

The typological convergence described for different Nilo-Saharan groups above contrasts dramatically with the typological diversity observable for languages in the Nuba Mountains in Sudan. In this accretion zone, we find over 40 languages belonging to Nilo-Saharan, Niger-Congo as well as Kadu, which differ rather dramatically from a typological point of view in almost every respect. Nilo-Saharan languages in this area differ amongst each other and from unrelated languages in the area in terms of their prosodic structures, varying between pitch accent systems and tonal systems with two or three registers. Moreover, some languages have a verb-final structure whereas others are verb-initial, using either case or head marking on the verb as grammatical strategies. The question why virtually no convergence or “language union” (as Trubetzkoy called this phenomenon) is found in the Nuba Mountains cannot be addressed here for reasons of space, but is discussed in detail in Dimmendaal (2015: 25-63).

6 Some conclusions

The Nilo-Saharan phylum covers a vast area across Eastern and Central Africa. The extant languages are spoken both in expansion zones and in residual zones, as shown above. Much of what one observes in terms of language contact today probably played a role in the development of this phylum over the past millennia.

The areal source for some of the dramatic changes is not always obvious. Within the Northern branch of Eastern Sudanic, there are two closely related languages, Nyimang plus Afitti, which deviate rather dramatically from close relatives such as the Nubian languages. Hill Nubian languages, for example, are spoken in the same area in the Nuba Mountains, and are characterized by rich number-marking systems of the type discussed above in section 3. This extensive number-marking system has disappeared from Nyimang and Afitti, where number marking on nouns is restricted to plural marking for nouns referring to human beings.

Data from the Nilo-Saharan area further show that replication of grammatical features in contact situations is facilitated by the proximity of grammatical systems. This can be observed when comparing historical processes in Nilotic languages like Luo or the
Kalenjin cluster, whose speakers were, or are, in close contact with speakers of Bantu languages.

Marked Nominative case systems are attested in Nilo-Saharan subgroups such as Nilotic, Surmic and Berta. In all three branches, this property is associated with postverbal subjects (A-roles). As shown by König (2008), such case systems are also attested in Omotic (Afroasiatic) languages in Ethiopia, where subjects inflected for case precede the verb. Given the geographical distance between most of these languages, and the fact that there is no other lexical or grammatical evidence for convergence between these Afroasiatic and Nilo-Saharan groups otherwise, it seems more likely that these systems developed independently of each other. These latter “autogenetic” processes may also form the ultimate explanation for the presence of Marked Nominative case systems in the Afroasiatic branch Omotic on the one hand and in the Eastern Sudanic branch within Nilo-Saharan on the other. Most of their extant members are not spoken in adjacent areas. Moreover, they differ in virtually every other typologically relevant respect: constituent order (verb-final versus verb-initial or verb-second), the presence versus absence of converbs, the extensive versus restricted use of case, to mention but a few features. The alternation with ergative systems, sometimes in the same language (as in Tennet) in Eastern Sudanic, suggests that Marked Nominative systems derive from the former. The frequently observed formal morphological link between Nominative case and Genitive case in Omotic (Afroasiatic) languages suggests a historical link between these two case forms. Genitive case constitutes one of the etymological sources of ergative constructions cross-linguistically (primarily for nominalized constructions). However, to date, no Omotic language has been reported to have a split ergative case system. The question why (split) ergative systems are so rare and presumably unstable remains one of the future challenges for areal linguistics in an African context.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>agent</td>
</tr>
<tr>
<td>ADV</td>
<td>adverbial</td>
</tr>
<tr>
<td>ABS</td>
<td>absolutive</td>
</tr>
<tr>
<td>AFF</td>
<td>affirmative</td>
</tr>
<tr>
<td>AP</td>
<td>antipassive</td>
</tr>
<tr>
<td>APPL</td>
<td>applicative</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>COMP</td>
<td>completive</td>
</tr>
<tr>
<td>CONV</td>
<td>converb</td>
</tr>
<tr>
<td>DIR</td>
<td>directive</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>M</td>
<td>medial verb marker</td>
</tr>
<tr>
<td>NF</td>
<td>nonfinite verb marker</td>
</tr>
<tr>
<td>NTS</td>
<td>followed by a non-topical</td>
</tr>
<tr>
<td>NOM</td>
<td>Nominative</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PERF</td>
<td>perfect</td>
</tr>
<tr>
<td>SEQ</td>
<td>sequential</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SU</td>
<td>subject</td>
</tr>
<tr>
<td>SUB</td>
<td>subjunctive</td>
</tr>
</tbody>
</table>

References


Santandrea, Stefano. 1970. *Brief Grammar Outlines of the Yuluand Kara Languages*. Bologna:


16 Niger-Congo languages

Jeff Good

1 Niger-Congo: A stock spanning many areas

Niger-Congo is one of the largest language families in the world—perhaps even the largest—spanning an enormous area of Africa from the southern edge of the Sahara desert at the north of its extent to the very south of the continent itself. It further spreads across a diverse range of ecological environments, from rain forest to desert, as well as a number of apparent linguistic areas. Its internal genealogical diversity is similarly remarkable, as well as controversial in some key respects. Any discussion of the areal linguistics of the family can only begin to scratch the surface, especially if sociocultural dimensions of its language dynamics are to be taken into account. The goal of this chapter is, therefore, merely to give some general sense of the “flavor” of the areal patterns of the family’s languages and highlight select topics for the study of areal linguistics generally that these patterns raise.

The discussion begins with an overview of the geographic distribution of the family and the state of the art with respect to the genealogical classification of its languages in section 2. A brief summary is then given regarding salient features that most Niger-Congo languages have in common in order to set a “baseline” for discussion of areal patterns within the family, which is the subject of section 3. The paper then shifts in section 4 from a more geographic-centered view of areality to consider the relatively underexplored issue of how an appreciation of the social dynamics of speakers of Niger-Congo languages is likely to play a crucial role in coming to a fuller understanding of the family’s areal patterns. In section 5, the paper concludes with a brief discussion of possible future directions for the study of Niger-Congo areal linguistics.

2 Genealogical and geographic overview

2.1 Genealogical overview

Niger-Congo, in the broadest sense, is the largest referential (if not genealogical) language group in world (Williamson and Blench 2000: 11), dominating Subsaharan Africa geographically. The extent to which the “traditional” group of Niger-Congo languages forms a true genealogical unit is not clear, however, and this obviously complicates any examination of the areal linguistics of the family. Overviews of Niger-Congo can be found in the chapters of Bendor-Samuel (1989) and in Williamson and Blench (2000), with a more up-to-date discussion of the state of scholarship on the comparative linguistics of the family to be found in Dimmendaal (2011: 85–92). Only a brief summary is given here, and Dimmendaal’s (2011) treatment is given the most serious consideration as an especially recent statement on the composition of the family, though it should be stressed that not all specialists would agree with the full range of his conclusions. For purposes of reference, Map 1 presents a map of the major African
language groups, and Map 2 presents a map of the subgroups of Niger-Congo, as assumed in the discussion here.¹

¹ Map 1 represents something closer to a “splitting” approach to African language classification than the more familiar four-family “lumping” approach that has dominated much of the work on African languages since Greenberg’s influential proposals, and the interested reader can consider the evidence for each approach by comparing the overview references given above. As will be clear from the discussion below, Map 2 represents something along the lines of “core” Niger-Congo, and language groups (e.g., Mande, Ijo, Dogon) which are treated as part of Niger-Congo in other sources are not necessarily indicated in Map 2 but, rather, Map 1. These maps essentially follow the treatment of Dimmendaal (2011), though they have been updated somewhat to make them more in line with the current state of the comparative evidence. The maps in Map 1 and Map 2 were produced by Monika Feinen, cartographer at the Institut für Afrikanistik, University of Cologne, and I am grateful to her for making them available to me for inclusion in this chapter and to Gerrit Dimmendaal for helping me get in touch with her.
There is a relatively uncontroversial “core” to Niger-Congo comprising the following subgroups: Benue-Congo, which includes many languages of southern Nigeria, the Bantu languages, and much in between; Kwa, found in southern West Africa from Côte d’Ivoire to the Nigerian border in a contiguous band along the Atlantic coast moving substantially inland; Gur, found in an inland region of West Africa, including the bulk of Burkina Faso, southern parts of Mali, northern parts of Côte d’Ivoire, Ghana, Togo, and Benin, as well as a small portion of western Nigeria, and whose southern and eastern borders are with Kwa and Benue-Congo languages respectively; Adamawa, also an inland group, found in scattered areas of Nigeria, Cameroon, Chad, and Central African Republic; Kru, mostly found in a strip along the West African coast directly to the west of Kwa, including the southwestern part of Côte d’Ivoire and much of Liberia; the so-called Atlantic group of languages found along and near the Atlantic coast, in an area detached
from the other groups given above, in parts of Senegal, the Gambia, Guinea, and Sierra Leone; and some languages commonly referred to as Kordofanian found in the Nuba mountains of Sudan, a region quite distant from the rest of the family.

Even if the membership of the languages of these groups as part of Niger-Congo is considered uncontroversial, there are still numerous questions concerning their subgrouping. To pick just one example, the validity of the languages grouped under the label ‘Atlantic’ as a coherent subgroup of Niger-Congo has never been properly established, with some even suggesting it has not been abandoned merely due to “scholarly inertia” (Childs 2003: 47).

In addition to the above, various other groups have also been placed under the Niger-Congo umbrella, but their genealogical status with respect to the family remains controversial. Perhaps the most significant of these from an areal perspective is the Mande group, which occupies a large portion of western West Africa. While the inclusion of Mande within Niger-Congo has been part of textbook presentations of the family for decades, see, e.g., Williamson and Blench (2000: 18), the evidence for its inclusion in Niger-Congo is comparatively weak, in large part because languages of the group neither exhibit functioning noun class systems nor clear-cut remnants of them (Williamson 1989b:36, Williamson and Blench 2000:19). If Mande is not part of Niger-Congo, then some of its noteworthy features which appear to show up in less robust form in other Niger-Congo languages would be good candidates for treatment as resulting from contact-induced change, this affecting languages found in an area of historical Mande influence (see section 3.4). For example, Mande languages are known for their rigid S-Aux-OV-Other clausal syntax, see Gensler (1994, 1997) for discussion in a comparative context. If they are not part of Niger-Congo, then cases of other Niger-Congo languages showing Mande-like word order patterns, of the sort documented in Güldemann (2007), see also Heine (1975: 35–36, 1976: 41–42) and Creissels (2006), would be candidates for having undergone syntactic change due to Mande influence. On the other hand, if Mande is part of Niger-Congo, then one must also consider the possibility of shared inheritance. Problems like these do not prevent us from describing contemporary areal patterns in the family, though they do suggest that one must avoid over-interpreting the evidence at hand with respect to specific historical scenarios.

Where relevant, further details of Niger-Congo genealogical classification will be discussed below. Since it can be daunting for an outsider to keep track of the details of a family of this size, the key points to bear in mind are that (i) there is an uncontroversial core to the family, (ii) a number of other groups have been proposed to be part of the family but the evidence is less clear, and (iii) even within the “core”, significant aspects of subgrouping remain obscure.

2.2 Geographic overview

There is one aspect of the geographic spread of Niger-Congo whose significance in terms of the areal patterns of the family seems hard to overstate: The family traverses an enormous north-south distance of around 5000 kilometers, ranging from the south of the Sahara down to the Kalahari and the southeast coast of the continent. Across this distance, one finds a wide range of ecological environments, from rainforest to savannah to even Mediterranean climates in parts of South Africa. Moreover, differences in altitude can produce distinctive ecologies even in regions that are nearby to each other, thus highland areas such as the Cameroonian Grassfields can be markedly cooler than nearby
lowland regions, with concomitant impacts on food production, presence of tropical
disease, etc.

The north-south distribution of the family should not be viewed merely as an
interesting curiosity. As discussed by Diamond (1997: 183–189), see also Diamond and
Bellwood (2003: 599), the relative consistency of patterns of daylight and seasons across
the east-west axis facilitates spread of domesticated plants and animals, and, by
extension, people and languages. Clearly, Niger-Congo patterns of language spread,
especially the Bantu spread (see section 3 for further discussion) managed to overcome
this bias, but the dominating pattern of areality in the family, associated with what
Güldemann (2008) has termed the Macro-Sudan Belt, does seem to fit within this pattern,
capturing a number of branches of Niger-Congo across an east-west axis, not a north-
south one. In addition, if we exclude the Bantu languages from consideration due to their
relatively recent expansion and their status as a subgroup of Benue-Congo, see
Schadeberg (2003: 155), rather than a primary branch of the family, Niger-Congo’s
overall pattern of geographic spread is much more strongly east-west than north-south.

A key concern for understanding the areal linguistics of Niger-Congo, then, is
establishing the extent to which its areal patterns may have been shaped by ecological
factors and determining what factors allowed Bantu to exceptionally spread from a point
around the Cameroon-Nigeria borderland to contemporary South Africa. Aspects of these
issues will be taken up below, though the present state of our knowledge in these regards
is largely incomplete.

3 The areal patterns of the family

3.1 Niger-Congo and “African” features

Any consideration of the areal linguistics of Niger-Congo must start with some sense of
the features that are found throughout the family, bearing in mind the complications
surrounding its precise composition discussed in section 2.1. For instance, SVO word
order predominates in the family, as it does in Subsaharan Africa generally (Dryer
2013b), though there are exceptional subgroups such as Mande, which, as mentioned in
section 2.1, exhibits an unusual type of SOV word order where only a single argument is
permitted before the verb and other arguments after. Similarly, languages of the Ijoid
group of the Niger Delta region of Nigeria, which is, at best, questionably Niger-Congo
(Dimmendaal 2011: 92), exhibit a more canonical SOV syntax, where the verb tends to
be final in the clause (Williamson and Blench 2000: 23), as do languages of the Dogon
group, see, e.g., Heath (2008: 17), though Dogon’s status as a Niger-Congo language is
also not proven (Dimmendaal 2011: 90). If we exclude these groups from Niger-Congo
(as is done in Map 1 and Map 2), SVO word order would dominate it much more than
otherwise. If we were to include these groups, then, perhaps, SVO would be considered
more areal in nature, prominent in the western parts and southern parts of the family but
not as much so in the “central” ones.

By contrast, another noteworthy feature of Niger-Congo, that is general to the family,
is the presence of tone. Like SVO word order, this is also a general African feature, see
Maddieson (2013b), but, while there are Niger-Congo languages lacking in tone, such as
Wolof and Swahili, these are relatively restricted in their distribution rather than characterizing an entire major subgroup.

Not surprisingly, given Niger-Congo’s expansive range over Subsaharan Africa and the ancient origins of its spread, it is difficult to find typological characteristics that would uniquely characterize Niger-Congo in opposition to the other major language families of Africa that it is in contact with.\(^2\) Thus, in the summary of “African” morphosyntactic features given in Creissels et al. (2008: 149–150), which lists around twenty different properties that seem especially common in African languages, many (if not all) of the points would apply just as well to Niger-Congo alone, such as, for instance, the relative lack of case, see, e.g., Creissels et al. (2008: 87–91), or the presence of applicatives (Polinsky 2013).\(^3\) This overlap between “Niger-Congo” and “African” features could be due, in principle, to a contact situation where Niger-Congo features were spread to the rest of Africa with the family itself or where Niger-Congo “absorbed” features common to much of Africa during its expansion, with the most likely possibility, of course, being some mix of the two.

From a historical-comparative perspective, the most significant features of Niger-Congo languages are probably their noun class systems (Heine 1980: 99) and verbal extensions, the latter of which constitute a class of suffixes involved in marking valency alternations, certain kinds of aspects, and other derivational or quasi-derivational functions (Hyman 2007a). The presence of each of these features has been reconstructed for Proto–Niger-Congo (see Williamson (1989a: 37–40) for noun classes and Voeltz (1977) for extensions), and noun class systems, in particular, have been a key diagnostic for family membership (see, e.g., Schadeberg (1981: 122–124)). However, while these are important historical features for the family, patterns of change, in many cases connected to areal influences (see section 3.2), have dismantled these systems in even some “core” Niger-Congo languages (Hyman 2004, Good 2012). Thus, these can only be considered to be Niger-Congo features from a diachronic perspective rather than a synchronic one.

3.2 The Macro-Sudan Belt set against the periphery

Almost certainly, the most significant areal patterns of Niger-Congo are found in a region which Güldemann (2008, 2011) has recently termed the “Macro-Sudan Belt”, but which has long been known to be linguistically distinctive, see, e.g., Westermann (1911), Wallis (1978), even if the role of areality as the central factor in explaining this was not recognized (Güldemann 2008: 170–174).\(^4\) This large region runs roughly from the Atlantic Ocean in the west to the Ethiopian Highlands in the east and is bounded in the

---

\(^2\) Heine and Leyew (2008) suggest that Africa, in its entirety, may even form a linguistic area (see also Greenberg 1983 for earlier discussion along these lines). This conclusion would seem to require a relatively weak sense of the notion “linguistic area”, however, which is why the focus here is on possible areas within Africa en route to understanding Niger-Congo patterns.

\(^3\) König (2008) provides a monograph-level overview of reported instances of case marking in African languages.

\(^4\) Like this chapter, Dimmendaal (2001a: 365–387) also considers how features connected to the Macro-Sudan belt relate to areal patterns in Niger-Congo.
north by the Sahara and Sahel and in the south by the Central African rainforest. It contains Afro-Asiatic and Nilo-Saharan languages, as well as languages from essentially every major subgroup of Niger-Congo. The Bantu languages on the whole fall outside the region, while the other members of its parent subgroup, Benue-Congo, lie within it, allowing divergence between Bantu and its closest relatives to serve as a useful indicator of areality, though as we will see other “peripheral” groups, such as Atlantic, are also relevant to establishing certain patterns as areal as well.

Perhaps the most salient linguistic feature of this region is the presence of labial-velar stops, such as kp and gb. Consonants of this kind are uncommon in almost the entire world except for within the Macro-Sudan Belt. And, as Maddieson (2013b) writes, “The pattern is clearly strongly areal, as it is compact and crosses the boundaries of language families.” For present purposes, what is additionally significant is that it does not just cross the boundaries of language families but also subgroups of Niger-Congo. Thus, if we look at Benue-Congo, there is a clear divide between the Bantu languages, and the rest of the subgroup: Bantu languages generally lack labial-velars while other Benue-Congo languages generally have them. Moreover, on the whole, those Bantu languages which do show labial-velar consonants are found at or near the southern edge of the Macro-Sudan Belt (Güldemann 2008: 157).

Other features which can be associated with the Macro-Sudan Belt are: the presence of ATR harmony (Clements and Rialland 2008:50–53, Güldemann 2008:158–159); the presence of labial flaps (Olson and Hajek 2003, Clements and Rialland 2008:41–42, Güldemann 2008:165–166); the presence of nasal vowels and the lack of contrastive nasal consonants (Clements and Rialland 2008:45–49, Hajek 2013); implosive consonants (Clements and Rialland 2008:55–60, Maddieson 2013a); the presence of three or more tone levels (Wedekind 1985:109, Clements and Rialland 2008:70–74, Maddieson 2013c); special “lax” prosody for marking yes-no questions other than via final rising pitch (Clements and Rialland 2008: 75–80); the presence of logophoric markers (Güldemann 2003); S-Aux-OV-Other word order, see section 2.1 and Güldemann (2008: 159–163); VO-Neg word order (Dryer 2009, Güldemann 2008:163–164); and the presence of serial verb constructions (Dimmendaal 2001a: 382–387). On a macro-level, the relative typologically similarity of languages within the belt can be seen in the treatment of African language typology in Cysouw and Comrie (2009: 202) which shows that, when looking at a number of typological features in aggregate, there is a pattern of east-west influence within the Macro-Sudan Belt not seen strongly elsewhere on the continent.

To these general Macro-Sudan patterns, which cross-cut major language families, one can add a more specific set of changes that have been discussed specifically for Niger-Congo languages in the center of the Macro-Sudan Belt. These changes are most strongly associated with languages of the Kwa subgroup and involve what one might broadly characterize as morphological reduction. This occurs both in the noun and noun phrase, where noun classes can show significant attrition, to the point of being lost entirely in synchronic terms (Good 2012), and in the verb and verb phrase, where, for instance, verbal derivational suffixes are lost with serialization strategies often taking their place (Hyman 2004), see also Nurse (2007) for relevant comparative discussion of the Niger-Congo verb complex. There is a phonological aspect to this reduction as well, wherein roots tend to be reduced to CV, possibly with additional relic prefixes in nouns (Williamson 1985). This set of characteristics has been grouped together as comprising key features of the Kwa “type”, which encompasses not only Kwa languages but also
many Benue-Congo languages, especially in parts of Nigeria. At the same time, not all
genealogical Kwa languages are good instances of the Kwa type. For instance, Ghana-
Togo Mountain languages do not show the noun class attrition associated with the Kwa
type (Heine 1968: 112–130 and Schuh 1995). Thus, this seems to be a good example of a
properly areal pattern within Niger-Congo.

The characteristics listed above do not, in fact, uniformly characterize the entire
Macro-Sudan Belt. For instance, the labial flap is isolated to a central area of the region
moving roughly southeast from Nigeria to the border of Uganda. Other features spread
past it, such as ‘surpass’ comparatives, which are also found in eastern Bantu languages
to the south of the belt. And, while Kwa-type languages are found within the Macro-
Sudan Belt, they are restricted to a central area within it, as just mentioned. Moreover, in
some cases, categories like those listed above mask significant variation. For example,
while the presence of more than three tone levels is, indeed, characteristic of the Macro-
Sudan Belt, this blanket classification leaves out the fact that there are pockets within the
region of higher tonal complexity. One finds languages with four or five tone levels in
places such as southern Côte d’Ivoire where this is found for languages of the Kru, Kwa,
and Mande groups, the first two of which are uncontroversially considered Niger-Congo
and the last questionably so (see section 2.1).

At the same time, on the whole, the array of features found more or less within the
belt is striking, covering both phonological and syntactic features with no obvious
coherence, some of which are also otherwise cross-linguistically rare (e.g., labial-velar
stops, ATR vowel harmony, VO-Neg word order), strongly suggesting areality is at play.
The region’s vast geographic spread and the fluidity of the locations of its characteristic
features indicate that, rather than viewing it in terms of long-term convergence due to
multilingualism of the sort associated with prototypical sprachbunds (Thomason and
Kaufman 1988: 95–97), it is better conceptualized as a “zone”, generalizing from the
sense of the word in the terms spread zone and accretion zone of Nichols (1992, 1997).
That is, what makes it an “area” is not the specific dynamics holding among its speaker
communities at any given moment but, rather, the fact that the region is conducive to the
spread of features within its borders but not across them. As discussed in section 2.2, the
fact that the Macro-Sudan Belt is primarily oriented along an east-west axis is probably
not a coincidence here but, rather, is connected to the tendency for subsistence strategies
to generally spread more effectively from east to west than north to south, see also

In this regard, the region resembles a spread zone in the sense of Nichols (1992,
1997), though one where features spread rather than whole languages, which is quite
different from the more canonical case of a spread zone, the Eurasian steppe, where
features have spread more directly with languages themselves (Nichols 1997: 369).
Indeed, a striking feature of the Macro-Sudan Belt is the extent of its linguistic diversity
since it overlaps notably with the Sub-Saharan Fragmentation Belt of Dalby (1970: 163),
characterized by a greater density of distinctive languages and language groups than in
other parts of the continent and stretching from Senegal to the East African highlands
across a width averaging 700 miles to south of the Sahara (Good 2013). This pattern of
feature convergence within a region of linguistic fragmentation is likely connected, at
least partly, to a distinctive language ideology that characterizes the Niger-Congo area,
see section 4 below.

3.3 “ Peripheral” Niger-Congo
The Macro-Sudan Belt, just discussed, encompasses the “center” of the Niger-Congo family, and it is largely by comparing Niger-Congo—or salient sub-parts of it, such as the region occupied by Kwa-type languages—within and outside the Belt that the areal features of Niger-Congo languages found within it can be established at all. One of the “peripheral” groups used for such comparison, in particular, has long had a special place in Niger-Congo linguistics. These are the Bantu languages. In strict geographic terms, they hardly seem peripheral, occupying the greater part of the southern half of the African continent, ranging from Cameroon in the west to Kenya in the east to South Africa in the south. However, their current distribution is the result of a relatively late expansion in Niger-Congo terms, though its beginnings are not necessarily particularly recent, with Nurse and Philippson (2003: 5) suggesting it commenced around 5000 years ago. More important than the precise timing is the consensus that treats Bantu as a relatively minor offshoot of Benue-Congo in genealogical terms (Schadeberg 2003: 155), a conclusion first explicitly reached by Greenberg (1949).

Bantu languages are generally believed to be relatively conservative in a Niger-Congo context, especially in the morphological domain. This is quite clear, for example, when they are set against many of their closest relatives in Benue-Congo and Kwa, which have been affected by the Kwa-type changes described in section 3.2. In all likelihood this is due to Bantu’s status as a geographic “offshoot”, which protected it from waves of change within the Macro-Sudan Belt. There are still many open questions regarding the dynamics of the Bantu expansion (see, for example, Vansina (1995) for a significant, relatively recent proposal and Pakendorf et al. (2011) for an up-to-date synthesis bringing data from genetic investigations into the picture). However, the end result from an areal perspective is clear: A vast region of Sub-Saharan Africa has become much more linguistically homogenous than its neighboring regions, creating a distinctive linguistic area in its own right. Here, however, the overall consistency is primarily due to language spread, rather than convergence, though ongoing contact among Bantu languages has been extensive (Schadeberg 2003: 158), which presumably has served to reinforce spread-induced homogeneity.\footnote{In the present context, it is worth pointing out that, before the arrival of Bantu, there is evidence for the presence of an older (non–Niger-Congo) linguistic area which has been referred to as the Kalahari Basin. See Güldemann (2010: 572–573).}

Nevertheless, one should be wary of overemphasizing the homogeneity of Bantu languages. Bantu languages of the “canonical” type showing, for instance, well-developed noun class systems (Katamba 2003) and significant prefixing and suffixing verb morphology (Hyman 2011: 4) are not found throughout the Bantu spread area. In the northwest region of the Bantu area, in particular, one finds languages with highly reduced noun class systems, including one language, Komo, which is reported to have no noun classes (Harries 1958: 269, Maho 1999: 127–142 provides a general overview). Some of the explanation for such reduced systems may be due to the effects of contact with non-Bantu languages. Komo, for instance, is spoken in a region of the Democratic Republic of the Congo where Ubangian and Central Sudanic languages are found.\footnote{Current consensus would seem to place the Ubangian languages outside of Niger-Congo despite a common previous classification linking them to Adamawa languages (Dimmendaal 2008: 842). Central Sudanic languages have been classified within the Nilo-Saharan phylum.} Moreover, the presence of labial-velars in the language (Harries 1958:265–266, Thomas 2011:11)
suggests an affinity with the Macro-Sudan Belt (see section 3.2) unlike most of Bantu. Similarly, a language like Ewondo, found at the northwest extreme of the Bantu area in Cameroon, shows a comparatively high degree of analyticity with respect to the marking of Tense-Mood-Aspect, as evidenced by the ability of pronominal objects to appear between verbal auxiliaries and the main verb (Redden 1979: 166). This is presumably connected to contact with closely-related Bantoid languages to the north that have been partly affected by the changes associated with the Kwa type discussed in section 3.2.

To pick another, less extreme case, the so-called interlacustrine languages seem to represent a distinctive area within the larger Bantu region (Bastin 2003). These are languages found “between the lakes”, roughly to the west of Lake Victoria around southern Uganda, Rwanda, and Burundi. By contrast, the Bantu languages found in the equatorial forest, another region that can be defined via physical features, are much more heterogenous, suggesting that, despite common ecology, there has not yet been noteworthy areal convergence (Grégoire 2003). As a final example, Güldemann (1999a) discusses some unusual patterns of suffixation on nouns in Bantu languages found in disjoint areas in the eastern and southern portions of the Bantu zone. However, even if Bantu is not as homogenous as often implicitly portrayed, it still is the case that the region is much more homogenous than the rest of the Niger-Congo area, especially in terms of lexicon. The shorter time depth of the subgroup is clearly part of the explanation for this, but sociocultural factors have presumably also played a role, as will be discussed briefly in section 4.4.

Another peripheral language group of Niger-Congo that should be considered here is Atlantic.7 As mentioned in section 2.1, the languages subsumed under Atlantic are not clearly a genealogical grouping, but this does not prevent them from being examined from an areal perspective. Found at the northwest edge of the Niger-Congo area, languages in this group are, like Bantu, not part of the core of the Macro-Sudan area (Güldemann 2008: 152). From a Niger-Congo perspective, what is striking about some languages of this group is their relatively elaborated morphology, for instance in the domain of noun class systems (Corbett 2011) and verb extensions (Hyman 2007a). For instance, some varieties of Fula have exceptionally large numbers of noun classes, numbering twenty or more (Arnott 1970: 67–109, Breedveld 1995: 295–460), making them much more “Bantu”-like than “Kwa”-like, despite the fact that Kwa-type languages intervene between Atlantic and Bantu (Hyman 2004, Good 2012). Indeed, the parallels between the verbal structure attributed to Proto-Bantu (Meeussen 1967: 108–111) and that of the Atlantic language Bijogo, spoken on islands off the coast of mainland Guinea-Bissau (Segerer 2000:369, 2002:369), are particularly striking, especially given the geographic distance separating Bantu from Bijogo.

The historical interpretation for the similarities between Bantu and Atlantic is controversial, with the central question being the extent to which their similarities represent Niger-Congo archaisms as opposed to parallel developments, presumably facilitated by both groups’ marginal status with respect to the Macro-Sudan Belt (Nurse 2007, Güldemann 2011, Hyman 2011). Nevertheless, from a synchronic areal

---

7 Due to its relative lack of study, I leave out detailed discussion of the most “peripheral” Niger-Congo group, Kordofanian, which is markedly detached from the rest of the family. However, based on what we know of the family, its languages show reasonably robust noun classes and verbal extensions (Good 2012: 303, fn.7; Hyman 2007b: 154), bringing them roughly in line with the other peripheral groups.
perspective, this does not prevent us from speaking of a comparatively morphologically elaborated Niger-Congo “periphery” against a more morphologically reduced Niger-Congo “core”, with extensive clinal variation in between as evidenced by, for instance, Gur languages (see the papers in Miehe and Winkelmann (2007)).

3.4 Contact among language groups and families

The discussion to this point has largely focused on high-level areal linguistic patterns within Niger-Congo, but, of course, there is also the issue of the role of more conventional kinds of language contact in shaping Niger-Congo grammatical patterns (see Childs (2010a) for an overview and discussion of language contact in Africa). Probably the most well-known such case involve the presence of clicks in southwestern Bantu languages (Herbert 1990, 1997, Vossen, Bostoen and Sands 2012). The presence of these in Bantu languages is universally attributed to contact with Khoisan languages (Güldemann and Stoneking 2008: 99), though it is not the case that clicks appear simply due to the borrowing of words with these sounds, since clicks can be found in words that are not borrowed from Khoisan languages, as in the form ǀumáte ‘tomato’ in Manyo (Gciriku) (Bostoen and Sands 2012: 125), a clear European loanword. The phonological category of click may have entered Bantu languages via Khoisan, but this has not prevented expansion in their range of use.

While clicks in some Bantu languages are the most salient outcome of contact between Bantu and Khoisan, other influences are reported as well, though these have not seen the same level of attention. Louw (1986: 152–153), for instance, discusses apparent cases of morphological borrowing from Khoe into the southern Bantu language Xhosa, such as a form that derives adjectives from other adjectives and nouns with a meaning comparable to English -ish (see also Louw 2013: 443 and Güldemann 1999b). However, Vossen (1997: 361) indicates that such effects were unusual in Bantu-Khoisan contact situations. Moreover, the presence of clicks has been attributed, at least in some cases, to the presence of avoidance registers in Bantu societies. This may have prompted the borrowing of clicks as a means to create lexical items fulfilling requirements of these registers, wherein words containing the same syllables as a specific set of names had to be avoided. Since clicks would not have originally been found in native Bantu words, any syllable in which they appeared would have automatically met the restrictions of such registers (see Herbert 1990: 304), which may have motivated their borrowing even in the absence of especially intimate language contact. All told, the observed patterns suggest relatively limited language contact between Bantu speakers and Khoisan speakers without, for example, the presence of long-term stable bilingualism (Vossen 1997: 362).

Less generally well-known contact situations are found in the northern parts of the Niger-Congo area, though these have had more significant grammatical effects than the Bantu-Khoisan case (even if the actual grammatical impact is less salient than the borrowing of clicks). One such contact situation is found between Benue-Congo and Chadic languages in central Nigeria, where there is evidence for lexical, phonological, morphological, and syntactic convergence without a clear general pattern of the directionality of the changes as emanating from one group or the other (Wolff and Gerhardt 1977, Gerhardt 1983, see also Storch 2009: 301–303 for more recent discussion). For instance, one finds the relatively uncommon intransitive copy pronoun construction in languages of both groups—without a clear original source—wherein a
copy of a subject pronoun appears after the verb in intransitive clause, see Atindogbé et al. (2011) for an overview.

Another significant region of contact influence in the northern Niger-Congo area involves the impact of Mande languages on various branches of the family. As discussed in section 2.1, there is no consensus at present regarding Mande’s genealogical affiliation, and it is variously classified within Niger-Congo or as an isolate family. Nevertheless, even among those who would place it within Niger-Congo, it is considered to represent an early branch from the rest of the family (Williamson and Blench 2000: 18), and it is grammatically and lexically distinctive enough to make contact between it and the rest of Niger-Congo more comparable to contact between distinct families than, say, contact among varieties in a dialect continuum, which is characteristic of other parts of Niger-Congo, especially the Bantu area.

The most salient apparent contact effect of Mande on other Niger-Congo languages involves effects on word order. While there is not consensus on all historical details, Heine (1976: 57–58) identifies a region of West Africa that he terms the “Mande nucleus” defined by the word order properties of its languages, for instance showing genitive-noun and SOV word order (which can be seen in recent surveys such as Dryer 2013a, 2013b). The special “Mande” SOV pattern, characterizable as S-Aux-OV-Other above (see section 2.1 and section 3.2) is even more widespread as a contact effect than surveys indicate, since it often appears as a common variant word order in languages otherwise characterizable as SVO (Heine 1976: 41, Güldemann 2007). Childs (2003:195–203, 2010a:699–704, 2010b) discusses Mande influence on Atlantic and provides historical background accounting for why the influence appears to have mostly involved imposition of Mande patterns on other languages rather than the reverse.8 However, this is not to say that Mande languages were not affected at all by contact. Bird (1970), for instance, suggests that some degree of simplification affected the Mande language Manding due to widespread second language acquisition.

The discussion here has emphasized cases where Niger-Congo languages were impacted by contact with other language families (or divergent members of their own family) but, of course, Niger-Congo languages have also significantly impacted languages in other families. Dimmendaal (2001b) describes cases where Nilotic languages have been affected by contact with Bantu languages, for instance with respect to the development of a set of tense markers in languages of the Kalenjin cluster which are functionally parallel to tense prefixes in nearby Bantu languages (Dimmendaal 2001b: 89–93, Heine and Kuteva 2005: 144–147). A more striking example is found in the mixed language, Ma’á (Mous 1994, 2003a, 2003b, Thomason 1997). Ma’á grammar is “more or less identical to that of Mbugu” a Bantu language closely related to Pare (Mous 2003a: 209–210). The difference between Ma’á and Mbugu is in the lexicon. Ma’á makes use of a “parallel lexicon” containing morphological forms with generally the same syntax and semantics as corresponding forms in Mbugu, but differing in phonology. This parallel lexicon consists of a number of words of Cushitic origin, assumed to be elements retained from an earlier East Cushitic language spoken by the ancestors of the Ma’á speakers, making it another apparent case of Niger-Congo impact on a language of another family—in this case, almost to the point of complete language shift.

8 See Dombrowsky-Hahn (1999) for a detailed study of a contact situation between a Mande language and a language from the Gur subgroup which shows a similar asymmetry.
4 Sociocultural dynamics and language ideologies

4.1 From areal states to areal processes

The discussion up to this point has focused on what one might call areal “states” rather than areal “processes”, extending ideas of Greenberg (1978, 1995) to areal patterns. However, it is clear that, if we want to fully understand the areal patterns of a family like Niger-Congo, we should be aiming to not only document them as they are now but also to understand the dynamics holding among the languages of the family that were the driving factor behind the development of those areal patterns in the first place.

At least in a Niger-Congo context, this issue has not received extensive investigation in the context of family-wide patterns of areality. However, there is work on the language dynamics of specific communities of clear relevance and which appears to be broadly consistent with observed family-wide patterns. In section 4.2, I will discuss some cases of language contact in urban situations, which are comparatively well studied, and in section 4.3, I will compare these cases with selected work on the linguistic and cultural dynamics of rural regions. In section 4.4, I will then sketch out the implications of the dynamics found in these two domains for our understanding of Niger-Congo areal linguistics.

4.2 The formation of urban varieties in Niger-Congo

Language use in urban Africa has been the subject of comparatively extensive research, especially when set against non-urban areas of the continent, see the collected papers of McLaughlin (2009a) and the detailed sociolinguistic investigation of Accra in Kropp Dakubu (1997). While the languages of urban regions are, of course, not limited to those of the Niger-Congo group, including not only other languages of Africa but also colonial languages, the family’s domination of the continent means that they are an integral part of the urban linguistic environment across the Sub-Saharan region.

One of the better studied cases of a Niger-Congo urban language variety is that of urban Wolof in Senegal (McLaughlin 2008, 2009b), see also Irvine and Gal 2000: 47–59) for the broad historical context. Urban Wolof is a prestige variety, functioning as an informal national language (despite French being the official language of the country) and is most saliently distinct from rural varieties by virtue of extensive borrowing from French. It has further become associated with urbanization in Senegal in general, rather than being the variety of a specific city (McLaughlin 2009b: 84) and with a distinctive urban-oriented identity (McLaughlin 2008: 155–156). The significance of urban Wolof in the present context relates to something that failed to happen in Senegal: Rather than French or “regular” Wolof becoming the language of the new urban populations, a new linguistic variety emerged, based on a Niger-Congo variety that was already present but with significant intermixing of “foreign” elements. This urban variety, moreover, has not replaced rural varieties but, rather, is “additive” in its sociolinguistic effect.

While each case of language contact has its own history, the particular result of contact in the Wolof case, the creation of a new variety, does not appear to be at all unusual in a Sub-Saharan African context. Kiessling and Mous (2004), for instance, discuss a number of recently formed urban youth languages in Sub-Saharan African areas dominated by Niger-Congo languages. Not all of these are primarily based on Niger-
Congo varieties. For instance, Camfranglais, a youth language of Cameroon, is primarily French-derived, but with significant influence from Cameroonian Pidgin and several other Cameroonian languages (Kiessling and Mous 2004: 306). In other cases, such as that of Isicamtho (Childs 1997, 2003:212–216, Kiessling and Mous 2004:310), the new language is primarily based on a Niger-Congo variety (in this case, Zulu). However, less important than the particular linguistic mixture is the fact that urban social contexts of this part of the world seem especially prone to the formation of youth varieties, and, as Kiessling and Mous (2004: 305) point out, a notable feature of them is the fact that they do not merely arise but are also named, enhancing their visibility as lexicogrammatical codes.

Thus, as with Urban Wolof, urban youth languages again point to a pattern characteristic of Sub-Saharan Africa where language contact leads to the development of recognized new linguistic varieties. I am not aware of any work which suggests this is specifically a Niger-Congo pattern, as opposed to encompassing Sub-Saharan Africa more generally, but it certainly is the case that this linguistic culture implicates much, and in all likelihood most, of the Niger-Congo area. Of course, it is difficult in general to ever specifically separate out any “Niger-Congo” pattern from a “Sub-Saharan” one given Niger-Congo’s spread on the continent.

Other examples of such new variety formation are not hard to find (Childs 2003: 214–215 and Kiessling and Mous 2004). A relatively well-studied case is that of Sheng, originating in Nairobi and which is described as involving a mixture of elements of Swahili, English, and other languages of Kenya (Mazrui 1995, Rudd 2008). A less well-studied parallel variety has also arisen in Nairobi known as Engsh, which is more heavily English-based than the Swahili-dominated Sheng (Abdulaziz and Osinde 1997). While the general interpretation of the development of these varieties is that they are connected to the construction of new identities in urban contexts, as will be discussed in section 4.3, there is reason to believe that the processes that underlie their development are not specifically “urban” in nature but, rather, have their roots in patterns of language use also found in more traditional contexts.

Alongside varieties specifically associated with urban environments, the development of various contact varieties of Niger-Congo languages, some of which have been treated as partially pidginized or creolized, also bears mentioning at present as providing additional cases where new language formation has taken place as new social contexts emerge. Mufwene (2003) presents an overview of significant issues with respect to contact languages of the Bantu area, which are comparatively well studied in this regard.

4.3 The continuity of urban and rural language dynamics

It is clear that many urban contexts dominated by Niger-Congo languages have proven to be fertile ground for the formation of new language varieties. A question that arises from this is whether there is something special about urban contexts in Africa that causes them to not merely be multilingual, but to also promote the creation of new languages on top of indigenous and colonial ones, or whether this urban pattern may represent simply a new instantiation of processes that characterized the Niger-Congo area long before modern patterns of urban living.

In fact, there is a fair amount of evidence that what we see in urban environments is not particularly distinctive in historical terms. For instance, urban youth vernaculars, such as those discussed in section 4.2, are typically seen as having developed, at least partly,
out of a desire for secrecy. This is suggested, for example, by Isicamtho’s origins in the milieu of criminals (Childs 1997: 345) and Sheng’s purported origins as a variety for groups of teenagers seeking special in-group communication (Abdulaziz and Osinde 1997: 49). Such varieties show immediate parallels to secret languages used for ritual purposes such as what is described in Storch (2004: 344–345) and Lüpke and Storch (2013: 81–82) for Jukun languages of the Benue-Congo subgroup. As Dimmendaal (2011: 252) describes it, “modern youth languages constitute a continuation or extension of ancient traditions of language manipulation.” He, thus, draws an explicit link between urban and rural language dynamics. More generally, there is rich documentation of what one might call a tendency towards language “inventiveness” found throughout Sub-Saharan Africa, where new varieties emerge to serve specific social functions, often via what appears to be deliberate linguistic manipulation (Storch 2011 and Lüpke and Storch 2013: 77–122).

It is instructive here to return to the case of Wolof given Irvine’s (1978) study of variation in the language in a rural context. In this case, variability in the use of noun classes was associated with a speaker’s social position. Wolof, thus, provides us with an instance where variation is tied to social status in both urban and rural contexts (see section 4.2), but where the locus of variation is different in each environment. Urban Wolof is most saliently distinct by virtue of its French borrowings, while rural varieties construct distinctiveness using the “native” material of noun classes. In both cases, one can observe a creative and active pattern of variety formation to serve social ends.

Di Carlo and Good (2014), on the basis of a case study of a small, linguistically diverse and rural Bantoid-speaking region of Cameroon known as Lower Fungom, have argued for the prominence of a language ideology in Sub-Saharan Africa that is primarily indexical rather than essentialist in orientation. That is, languages are not treated as markers of “essential” and immutable ethnicities, as is typical of Western language ideologies in the form of the so-called Herderian equation of language, culture, and nation (see, e.g., Hymes 1968, 1972, Foley 2005). Rather, their primary function is to index a speaker’s association with a particular socio-political group (in the Lower Fungom case, a specific village), comparable to the notion of a “community of practice” as developed in the sociolinguistic literature (see, e.g. Bucholtz 1999, Eckert 2000).

Di Carlo and Good (2014) further suggest that this ideology can be correlated with patterns of change that do not align well with standard models of language diversification as involving tree-like or wave-like patterns, but, rather, are more likely to involve language mixing as a means to rapidly create new identities. This would be a clear rural parallel to the formation of a variety like Sheng, discussed in section 4.2, with the major difference that, in the rural Lower Fungom case, the apparent mixing is only among languages of the Bantoid subgroup of Niger-Congo, rather than taking place between a Niger-Congo language and an Indo-European one. It seems important to bear in mind, in this regard, that language mixing is most visible when involving distantly related languages. This means that there might be many cases of mixing among closely related Niger-Congo languages that have yet to be detected (if even detectable).

---

9 Avoidance registers fall into this class as well and, as discussed in section 3.4, an avoidance register is likely to have been one of the routes through which clicks entered southern Bantu languages.

10 Good (2013) considers this region form an areal perspective.
It is useful here to consider ethnographic work like that of Kopytoff (1987), which tries to identify general cultural patterns in Sub-Saharan Africa and, by extension, Niger-Congo (see also Zeitlyn and Connell 2003). Kopytoff (1987: 24) specifically characterizes societies in this part of the world as being built around groups of solidarity, ranging from the kin group to the village to the kingdom. A given individual would be attached to several such groups as a means to ensure personal security. Following Di Carlo and Good (2014), one of the most salient means to signal such group affiliation would be through the use of the linguistic varieties associated with these groups. While this is an emerging, rather than established, area of research, it suggests that a distinctive underlying cultural pattern may be responsible for noteworthy aspects of language use and diversification both in contemporary urban and traditional rural Sub-Saharan Africa. In particular, in this part of the world, new social contexts may be especially prone to the development of new linguistic varieties.

When this idea is linked to complex patterns of multilingualism observable in the Niger-Congo area (Lüpke and Storch 2013: 13–48), it suggests one of the defining areal features of Niger-Congo may be the way a particular language ideology produces language dynamics associated with the frequent emergence (and, necessarily, loss) of language varieties to fulfill different social functions. These dynamics can involve creative use of ambient grammatical material, either through manipulation or mixing, to construct new varieties and are led by speaker communities where individual mastery of multilingual repertoires is the normal state of affairs. To the extent that this sketch may be correct, it has consequences for long-term patterns of historical change in Niger-Congo, and this is briefly explored in the next section.

4.4 From local dynamics to areal patterns

If we take seriously the idea that patterns such as those described in section 4.2 and section 4.3 are not merely isolated cases of new variety formation but, rather, represent instantiations of a more general Niger-Congo pattern, then this has clear consequences for our understanding of the areal dynamics of the family. In particular, it seems clear that repeated instances of the formation of new varieties involving creative mixing and manipulation of “ambient” grammatical material would be unlikely to result in the tree-like and wave-like patterns of diversification, that have long been considered to be the normal state of affairs for language change, and which are still frequently applied to Bantu data (see, e.g., Holden and Gray 2006).

This is not to say that Niger-Congo diversification would never have involved such types of change. For instance, the genealogical division between some groups in contact, such as Gur and Kwa has been viewed as completely unambiguous, indicating that tree-like diversification must have taken place at some (distant) point in Niger-Congo’s history. But, at the same time, one is confronted with long-standing problems, such as the fact that clear criteria for distinguishing Bantu languages from their closest relatives have never been devised despite the general impression that there is something coherent about the group (Nurse and Philipsson 2003: 5–7). More strikingly, despite success in reconstructing Proto-Bantu elements, it has not even been possible to arrive at consensus on Bantu-internal subgrouping (Schadeberg 2003).

It seems worth considering that this may not be due to a failure on the part of comparative linguists to find the “right” grammatical and lexical features for such tasks but may rather be due to the historical forces shaping language change in the family.
itself. Perhaps millennia of new variety formation, following the patterns outlined above, have obscured tree-like patterns of divergence, or even prevented them from developing in the first place in many cases. If so, this suggests that the future of Niger-Congo areal studies may need to consider much more closely the special role that the cultural configurations of Niger-Congo societies and, in particular, their language ideologies, may have played in shaping the current areal patterns of the family. This would require moving away from models that simply treat language change as resulting from “natural” processes of diversification and, instead, view it as inherently culturally embedded (Heggarty et al. 2010: 3830).

Adopting this perspective may also play a role in understanding why the Macro-Sudan Belt shows grammatical convergence in a manner reminiscent of what is found in spread zones, despite showing extreme linguistic diversity (see section 3.2). Multilingual communities in this region may have required the active presence outwardly distinct languages for social purposes while frequently exchanging grammatical features among. The end result would be typological convergence without large-scale language shift.

5 Moving forward in Niger-Congo areal studies

As obvious as it may be, it bears repeating that the present state of our knowledge of the areal linguistics of Niger-Congo is severely limited. The vastness of the family, in terms of numbers of languages, speakers, and geographic spread presents daunting challenges for synthesizing the data required to come to a full understanding of its areal patterns. And, we are also confronted with a lack of consensus regarding the composition of the family and its internal subgrouping (see section 2.1).

The usual response to such a situation—that we need more data—of course, applies, to some extent. However, simply adding more data on top of the already considerable amount of unsynthesized data already available will not suddenly reveal a clear picture, where there was, previously, a murky one. Instead, we would seem to need new models for understanding the nature of language contact and language change in the Niger-Congo area which build the observed realities of widespread multilingualism and rapid formation of new varieties into their foundations. Only by arriving at an improved understanding of how Niger-Congo cultures have impacted Niger-Congo languages will we be able to seriously explore salient linguistic questions such as what has allowed labial-velars to spread widely within the Macro-Sudan Belt but not beyond it, why the striking and “marked” Khoisan feature of clicks was able to be integrated into Bantu languages without significant grammatical change, or why it has been so difficult to find clear linguistic dividing lines for many subgroups despite linguists’ intuitions that they should exist—just to pick three of the many important areal questions posed by the family.

References

11 One can, of course, attempt to fit a tree onto any comparative linguistic data, as Holden and Gray (2006) do with Bantu lexical data, but this sort of work begins with the assumption that the basic nature of divergence will result in tree-like patterns and cannot, therefore, test the validity of the model in the first place.


Heine, Bernd. 1975. Language typology and convergence areas in Africa. Linguistics 144. 27–47.


17 The Kalahari Basin area as a “Sprachbund” before the Bantu expansion

Tom Güldemann and Anne-Maria Fehn

1 Introduction

Güldemann (1998 and following publications) not only challenged the “Khoisan” family hypothesis established by Greenberg (1950, 1963) and popular among non-specialists ever since, but also proposed the areal concept “Kalahari Basin” comprising the indigenous non-Bantu languages of southern Africa. If the linguistic isoglosses shared by these languages are compatible with a historical assessment in terms of multiple and partly long-standing contact, the areal approach is a viable explanation for the emergence of the modern linguistic panorama, as opposed to the genealogical hypothesis. Since the areal approach was proposed more than a decade ago research on linguistic isoglosses and contact-induced convergence across the Kalahari Basin has increased considerably. This article summarizes the earlier results, supplements them with new findings, thus giving more substance to the “Kalahari Basin” concept, and embeds it in the general discussion about linguistic areas.

1.1 The three independent “Khoisan” families of southern Africa

The classification of the non-Bantu languages of southern Africa has been controversial since the first linguistic data were scrutinized from a historical perspective. By the second half of the last century linguists and non-linguists alike (but not necessarily language specialists) had largely settled down on Greenberg’s (1950, 1963) lumping proposal of a “Khoisan” language family, which even includes two isolated languages in eastern Africa, Sandawe and Hadza. Before embarking on the following discussion, it is important to recognize that this genealogical classification can no longer be followed, first of all because Greenberg’s evidence was insufficient and no progress has been made in more than 60 years in substantiating his claim by means of standard historical-comparative methodology (cf. Westphal 1971; Sands 1998; Güldemann and Vossen 2000; Güldemann 2008b).

1 This paper greatly benefitted from research undertaken within the EuroBABEL project ‘The Kalahari Basin area: a “Sprachbund” on the verge of extinction’. We are grateful for the funding provided by the Deutsche Forschungsgemeinschaft (DFG). Anne-Maria Fehn would further like to thank the a.r.t.e.s. Graduate School at the University of Cologne and the German Academic Exchange Service (DAAD) for supporting research and fieldwork on Ts’ixa. We are also indebted to Hirosi Nakagawa and Christfried Naumann for valuable comments on an earlier version of this paper, and to Falko Berthold and Linda Gerlach for sharing their data on N!aqriaxe with us.
Lineages and Languages (L) or language complexes (LC) and (sub)branches selected dialects and dialect groups

(1) **Khoe-Kwadi**

<table>
<thead>
<tr>
<th>A Kwadi</th>
<th>single L↑</th>
</tr>
</thead>
</table>

B Khoe

*Kalahari Khoe*

- **East**
  - Shua: Cara, Deti, ǂXaise, Danisi, etc.
  - Tshwa: Kua, Cua, Tsua, etc.
- **West**
  - ?Ts’ixa
  - Khwe: ǁXom, ǁXo, Buga, ǁAni, etc.
  - Gǁana: Gǁana, Glui, etc.
  - Naro: Naro, Ts’ao, etc.

*Khoekhoe* (Cape K.)↑ LC

- !Ora-Xiri) LC
- (Eini)↑ LC
- Nama-Damara LC
- Haii
  - Aakhoe

(2) **Kx’a**

<table>
<thead>
<tr>
<th>A Ju</th>
<th>single LC: North: Angolan !Xuun varieties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>North-central: Ekoka !Xuun, etc.</td>
</tr>
<tr>
<td></td>
<td>Central: Grootfontein !Xuun, etc.</td>
</tr>
<tr>
<td></td>
<td>Southeast: Tsumkwe Jul’hoan, Epukiro Jul’hoan, etc.</td>
</tr>
</tbody>
</table>

B ǂ’Amkoe single LC: ǂHoan, Nlaqiaxe, Sasi

(3) **Tuu**

<table>
<thead>
<tr>
<th>A Taa-Lower Nossob</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Taa</strong> single LC:</td>
<td>West: West !Xoon, (Nlull’en)</td>
</tr>
<tr>
<td></td>
<td>East: East !Xoon, ’Nloha, (Nlamani), (Kakia), etc.</td>
</tr>
<tr>
<td><strong>Lower Nossob</strong></td>
<td></td>
</tr>
<tr>
<td>(!’Auni)↑</td>
<td></td>
</tr>
<tr>
<td>(!’Haasi)↑</td>
<td></td>
</tr>
</tbody>
</table>

B ǃUi

<table>
<thead>
<tr>
<th>Nǃng: Nluu = (ǂKhomanji) = (ǂNhuki), Langeberg, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(!’Xam)↑: Strandberg, Katkop, Achterveld, etc.</td>
</tr>
<tr>
<td>(ǂ!’Ungkue)↑</td>
</tr>
<tr>
<td>(ǂ!’Xegwi)↑</td>
</tr>
</tbody>
</table>

Note: ↑ = extinct, (…) = older data sources

Figure 1. The three lineages commonly subsumed under southern African “Khoisan”

The opinion of specialists is converging on the recognition of three independent lineages in southern Africa, as shown in Figure 1; the evidence for the assumed affiliation of Kwadi and ǂ’Amkoe, on which our knowledge is still incomplete, with Khoe and Ju respectively, still needs to be extended (see Güldemann 2014a for the most recent survey on classification and terminology). While Kwadi is crucial for an understanding of the
area’s history, the following discussion excludes it, because the data on this extinct language are largely insufficient. This situation also holds for the Lower Nossob group of the Tuu family.

The three lineages, Tuu, Kx’a, and Khoe-(Kwadi), are intended here to constitute the Kalahari Basin area and referred to accordingly - this despite the fact that later colonizers like Bantu languages and Afrikaans may have come to share one or other local feature through language contact (see section 3).

1.2 Previous areal and contact-oriented research

While Greenberg (1950, 1963) is responsible for the heretofore popular but spurious “Khoisan” concept, he (1959: 24) is probably also the first who entertained contact-induced convergence in the major relevant area, here called the Kalahari Basin. Since he did not give any data and discussion, one can only speculate about the exact evidence he had in mind. Heine’s (1976: 56) research on word order patterns in Africa yielded more concrete indications of areal convergence. It is clear, however, that any areal concept for the relevant languages must remain highly problematic from a methodological perspective as long as they are simultaneously considered to be genealogically related.

Such a problem ceases to exist under a non-genealogical approach. Inspired by Nichols (1992), this was pursued for the first time by Güldemann (1998) which introduced the present areal concept. Although prefiguring the idea of a linguistic area before the Bantu expansion, the aim of this initial study was not yet to show any homogeneity across the area but rather its overall diversity compared to other parts of the continent and its potential nature as a “residual/accretion zone”, which would cast doubt on the “Khoisan” hypothesis. In this article, the area also had a narrower geographical extension, excluding languages which are viewed here as belonging to its northern and eastern periphery.

The study (ibid.: 152-4) did, however, identify several isoglosses across a subset of Kalahari Basin languages, thus establishing a typological group called since then “Non-Khoe”. In the present terminology, this entity comprises two of the three Kalahari Basin lineages, viz. Tuu and Kx’a. Studies like Güldemann and Vossen (2000) and Güldemann (2000, 2013e) reiterate this finding and elaborate it in observing a basic typological split between this group on the one hand and Khoe-Kwadi on the other. The following list summarizes the Non-Khoe isoglosses proposed in the four aforementioned studies, all written but not published around the same time.

Non-Khoe isoglosses

1. low complexity (in terms of Nichols 1992)
2. little bound morphology, phonological word often the same as lexical root
3. neutral alignment
4. clusivity
5. verb-medial clause order
6. very few or even no ditransitive verbs
7. verb serialization
8. head-initial noun phrase order with exceptional head-final genitive
9. inalienable possession
10. head of juxtapositional genitive conveys derivation and locative flagging
11. multi-purpose oblique (MPO) preposition (possible circumpositional flagging)
12. noun categorization by means of a special type of gender system
13. irregular number marking, including nominal and verbal root suppletion

The shared and often quirky properties of the sound structure of Kalahari Basin languages (cf., e.g., Güldemann (2001) for a more recent treatment of consonant systems) have long since been known and mostly considered to be inherited from an assumed Proto-Khoisan. Under a non-genealogical approach these would be obvious signals of convergence.

Within the somewhat different context of historical research on Bantu languages, the areal homogeneity detected by Heine (1976) regarding the morphosyntax of genitives was discussed in more detail by Güldemann (1999); an additional finding of this study was that historically derived structures of nominal flagging and derivation are also widely shared across the Kalahari Basin.

Around the same time and later, research explicitly focusing on language contact started to identify non-phonological isoglosses that bridged the dichotomy of Non-Khoe vs. Khoe-Kwadi also on a more local scale. As will become clearer below, these are important for modeling the linguistic history of the Kalahari Basin as a whole.

Traill and Nakagawa (2000) treated a contact zone in the central Kalahari between Glui from the Kalahari Khoe group of Khoe-Kwadi and the East !Xoon variety of the Tuu language Taa, and also indicated the involvement of western !'Amkoe of the Kx’a family. Their discussion focused largely on lexical isoglosses. However, looking at Traill’s (1980) phonological survey of the wider area, it becomes clear that the analysis of sound structure yields a similar picture, e.g., the languages’ sharing of exceptionally large consonant inventories and a high frequency of clicks against non-clicks.

Güldemann (2002, 2006) focused on another zone where Khoe languages share a number of features with Non-Khoe neighbors, viz. the Cape area in South Africa where Khoekhoe languages from Khoe-Kwadi and !Ui languages from Tuu were spoken. The studies did not focus on shared lexicon but on structural similarities in phonology and morphosyntax, which are listed below.

**Structural similarities between Khoe and Non-Khoe languages**

1. relatively small consonant inventories
2. high click frequency compared to nonclicks
3. fricativisation of complex egressive stops (in the north)
4. complex pronouns and clusivity
5. lexically fixed gender, unusual association of feminine sex/gender and large size
6. nominal derivation on non-canonical verbal, adjectival, and pronominal hosts
7. inclusory pronouns in “recapitulative” coordination
8. high load of verbal reduplication
9. particular distribution pattern of temporal predicate operators
10. lexically complex predicates
11. clause-second marking for sentence type and information structure
12. low semantic sensitivity of certain participant flagging
13. similar tagging of reported discourse and proposition (aka “complement”) clauses

In section 2 we recapitulate Kalahari Basin isoglosses from earlier research and entertain some new ones, the minimal requirements being that a feature is recurrent and involves at least one language in each family and the relevant languages are not all in contact today. Another important criterion is that a feature is sufficiently marked, locally or even better globally, so that multiple independent origin is less likely. The following isogloss list is not considered as complete, nor is every feature thought to be an established areal trait, because some isoglosses are not distributed evenly geographically and/or within families and are not even rigorously defined yet, so that their status as areal features requires more research.

2 Isoglosses across Kalahari Basin languages

2.1 Phonetics-phonology

Beach’s (1938) ground-breaking work on the phonetics-phonology of Khoekhoe provided the first scientific basis for describing and comparing Kalahari Basin languages regarding their complex sound systems, in particular, typologically quirky clicks, as well as highly skewed root phonotactics. An equally important achievement was made later by Snyman (1975) and Traill (1985) with respect to the treatment of complex consonants, both clicks and non-clicks, and of vocalic phonation types comprising nasalization, pharyngealization, glottalization, breathiness, and stridency.

Phonetic-phonological commonalities across the Kalahari Basin languages have been taken for granted, also because they were normally thought to reflect common descent, even by Greenberg (1963: 67) who generally strived to exclude purely typological features as genealogical evidence. Later studies relying on a much better data base and more fine-grained analyses actually found a considerable amount of diversity between languages (cf. Traill 1980; Güldemann 2001, 2013e).

Nevertheless, the features shared across the three families are numerous enough, largely absent in Bantu outside the area (cf. Maddieson 2003, Hyman 2003, Kisseberth and Odden 2003), and often so quirky, that they represent an unmistakable signal of linguistic convergence. Moreover, the isoglosses affect different domains of sound structure such as consonant types, suprasegmental features, and phonotactics.

Segments shared across the Kalahari Basin are lingual ingressives aka clicks, glottalic egressives aka ejectives, uvular stops, aspirated obstruents, and tautosyllabic obstruent-obstruent clusters.2 Their possible co-occurrence and a strong tendency to series formation are the most important factors for the emergence of some of the largest consonant inventories in the world’s languages.

Of the relevant suprasegmental features, nasalization is universal (though overlooked entirely by Hajek 2005), and pharyngealization is attested in all three families; glottalization and breathy voice are so far restricted to the Non-Khoe lineages Tuu and Kx’a. In terms of pitch prosody, all Kalahari Basin languages share the feature of register tone systems largely relevant for lexical distinctions (pace Clements 2000: 2).

---

2 There is an alternative analysis according to which the relevant complex consonants are not clusters (cf. Miller 2011). This approach does not alter the segments’ status as shared and rare.
157-158); the number of tone levels, still controversial for some languages, ranges between two and four. This picture is different from Bantu, where tone has a strong grammatical import, with predominantly two tone levels, H and L, often with “significant asymmetries between H and L suggesting privative analysis of H vs toneless” (Kisseberth and Odden 2003: 59).

Another important isogloss is characteristic bimoraic patterns of lexical roots, namely basic C(C)VCV, and derivative C(C)VV and C(C)VN, which in addition involve a very skewed phoneme distribution. For details, the reader is referred to Beach (1938), who discovered the phenomenon, and Nakagawa (2010), an innovative recent account. (In view of data provided by Chebanne (2000) and Snyman (2000), it remains to be determined whether all East Kalahari Khoe languages fully comply with these patterns).

2.2 Lexicon

Shared lexicon across the Kalahari Basin has always been assumed, commonly invoked as evidence for “Khoisan”. While Greenberg (1950, 1963) and similar studies by non-specialists lacked comparative rigor, Köhler (1975: 312-3) and Traill (1986) also discussed substantial lexical isoglosses crossing the major family boundaries, being arguably due to inheritance.

However, such isoglosses are far more extensive in bilateral comparisons (cf. Köhler (1973/4: 185-9) for Jul’hoan–Caprivi Khwe, Snyman (1974: 40-2) for Jul’hoan–Namibian Khoekhoe, and Traill and Nakagawa (2000) for East !Xoon–Glui; see also Sands (2001: 201)) - these would have resulted largely from language contact. Sands (2001) and Honken (2006) also recognized the possibility of extensive borrowing across the entire area. This hypothesis was studied more systematically by Güldemann and Loughnane (2012) for one specific lexical domain, viz. body parts and related terms. Starting out from a bottom-up reconstruction within the three secure lineages, many purported “Macro-Khoisan” lexemes can be argued to have emanated from a single family and entered the others by way of language contact. Dense lexical distributions in the Kalahari Basin can thus be explained alternatively by linguistic convergence, whereby at least three different patterns should be distinguished regarding their geographical scope and time depth. They are exemplified in Table 1 by examples proposed by Honken (2006: 77-78, 81).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Meaning</th>
<th>Tuu</th>
<th>Kx’a</th>
<th>Khoe-Kwadi</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I)</td>
<td>stupid, insane</td>
<td>Taa</td>
<td>‘Amkoe</td>
<td>Jul’hoan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>!Xômô</td>
</tr>
<tr>
<td>(II)</td>
<td>tin, box, pot</td>
<td>E. !Xoon</td>
<td>tôô</td>
<td>Jul’hoan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tôô</td>
</tr>
<tr>
<td>(III)</td>
<td>dirt(y)</td>
<td>Proto-Kx’a</td>
<td>*KX’URI</td>
<td>Owambo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>#Hoan</td>
<td>lx’órî</td>
<td>!Xuun</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>glx’órî</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Jul’hoan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lx’órî</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note: here items in bold italics are loans, those in capital italics are the SOURCE ITEM OF BORROWING

Table 1. Examples for three major patterns of lexical borrowing in the Kalahari Basin

As mentioned above, a borrowing first pattern (I) is a multitude of localized contact situations, some of them more recent, illustrated in Table 1 by a borrowing from Namibian Khoekhoe into Jul’hoan. A second pattern (II) is cross-areal lexical transfer from prestigious Khoekhoe varieties spoken by pastoralists into a large number of forager (aka “San”) languages from all three families, viz. virtually all Tuu languages, southern Ju varieties from Kx’a, and Naro and possibly other West Kalahari Khoe languages. This is a phenomenon with a time depth of several centuries in Namibia and even longer in South Africa. The example in Table 1 shows that Namibian Khoekhoe tōō-s/p ‘tin, can’, ultimately from Afrikaans doos and/or German Dose, is the source of borrowings in at least three genealogically unrelated San languages. As proposed by Güldemann (2006, 2008a), one can posit a third pattern (III) in the form of substrate influence in various stages of Khoe-Kwadi, notably Kx’a influence on (Pre)-Khoekhoe and Tuu influence on (Pre)-Khoekhoe. The relevant comparative series in Table 1 suggests a likely reconstruction *lkx’uri ‘dirt’ for Proto-Kx’a which would have been the source of borrowings into Naro and Proto-Khoekhoe. The root regularly changed to l’uri in northern Khoekhoe which expanded from South Africa into Namibia, providing there the possibility for Jul’hoan to borrow the changed root together with a Khoekhoe adjective suffix -xa, leading to a double reflex.3

Word borrowing aside, there are also other shared lexical patterns across the Kalahari Basin. One such feature is restricted numeral systems, contrasting with Bantu in and outside the area for which numerals higher than three/four are normal, and such systems can be partly reconstructed (cf., e.g., Hoffmann 1953). For Tuu and Kx’a, the few items above ‘three’ are descriptive forms or borrowings; even ‘three’ may sometimes not be a cardinal numeral but mean ‘more than two’ (cf. Güldemann (2012) for the Tuu family, Honken (2013: 253) for t’Amkoe, and Heine and König (2013: 310) for Ju). For Proto-Khoe, Vossen (1997) only reconstructs simplex numerals for 1-4; some modern Kalahari Khoe languages do not even attest anymore for reconstructed *haka ‘four’. The only Kalahari Basin languages with higher numerals are pastoral Khoekhoe varieties like Nama and !Ora, as well as Naro (cf. Visser 2001) which is said to have borrowed numerals above ‘three’ from Khoekhoe in connection with a traditional game (Barnard p.c., Visser 2013: 190). Hahn (1881: 10-6) discusses the possibility that the forms exclusive to Khoekhoe have emerged more recently, as some are morphologically complex and/or show suggestive relations to other lexemes.

3 This comparative series is actually more complex than suggested by Honken’s discussion. On the one hand, the existence of forms in Kalahari Khoe like, e.g., Danisi l’ūri (Fehn field notes) indicate early borrowing on the part of Khoe and multiple independent reduction of velar ejectives to glottal stops. On the other hand, !Ui forms like !Xam lk’warri (Bleek 1956: 340) and Nlng lk’warre (Bleek 1956: 608, presumably with mistranscribed click) are likely loans from Khoekhoe, either from southern varieties like !Ora which did not undergo the relevant sound change or from early Khoekhoe before the change.

<table>
<thead>
<tr>
<th>SIGHT</th>
<th>HEARING</th>
<th>TOUCH</th>
<th>TASTE</th>
<th>SMELL</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ekoka !Xun (Ju, Kx’a); Shona, Ndonga (Bantu)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Jul’hoan (Ju, Kx’a)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shua (Khoe)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gllana-Glui (Khoe); East !Xoon (Taa, Tuu); š’Amkoe (Kx’a); Venda, Tswana, Zulu (Bantu)</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Caprivi Khwe, Ts’ixa, Naro (Khoe)</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Namibian Khoekhoe, !Ora (Khoe)</td>
</tr>
</tbody>
</table>

Table 2. Polysemy in “experience” perception verbs across the Kalahari Basin

Following Viberg’s classification into SIGHT, HEARING, TOUCH, TASTE and SMELL, Table 2 shows different polysemy patterns in the “experience” class of perception verbs. All Kalahari Basin languages except Ekoka !Xun, Namibian Khoekhoe and !Ora display polysemies covering at least three sense modalities, and all but Ekoka !Xun and Jul’hoan conflate specifically TOUCH and TASTE. The “maximal” pattern where polysemy covers all non-SIGHT modalities is found today in the Central Kalahari area affecting languages of all three families. The reason for proposing this as a specific Kalahari Basin trend is as follows: while such Bantu languages close to/within the area as Venda, Tswana, and Zulu also show this pattern, the overall trend in the family is different, namely towards less polysemy, and if any, rather conflating HEARING and TOUCH. This situation can be observed even in Bantu languages at the areal periphery like Ndonga and Shona, and seems to have even affected peripheral Kalahari Basin languages like Ekoka !Xun.

A distinct word class of ideophones was long thought to be absent in “Khoisan” (cf. Samarin 1971: 160-161, Childs 1994: 179). Given the salience of ideophones in Bantu, this might be viewed as an areal feature of Pre-Bantu southern Africa. However, ideophones do feature in such languages as !Xam (Bleek 1928-30: 171, 1956), Taa (Traill 1994), Ju (Dickens 1994, Heine and König 2008), and Namibian Khoekhoe (Haacke and Eiseb 2002). This finding and studies dedicated to the topic like Kilian-Hatz (2001) for

---

4Table 2 reflects the maximal meaning range of individual polysemous lexemes; it is possible that a language has a more specific verb in a certain domain in addition to the polysemous item. The list of languages (sources) is as follows: East !Xoon (Traill 1994), Ekoka !Xun (König and Heine 2008), Gllana-Glui (Nakagawa 2012), Jul’hoan (Dickens 1994), Khwe and Ts’ixa (Brenzinger and Fehn 2013), Namibian Khoekhoe (Haacke and Eiseb 2002), Naro (Visser 2001), Ndonga (ELCIN Church Council Special Committees 1996), Shona (Dale 1975), Shua (Fehn field notes), Tswana (Viberg 1984), Venda (Murphy 1997), Zulu (Doke et al. 1990), !Ora (Meinhof 1930), š’Amkoe (Berthold and Gerlach, p.c.).

5This class was chosen from Viberg’s (1984) three-way distinction, activity, experience, copulative, because it provided relatively complete and reliable data across the area.
Caprivi Khwe, Nakagawa (2012, in press) for Glui, and Brenzinger and Fehn (2013) for West Kalahari Khoe in general call for a reevaluation of the earlier claim (cf. also Childs 2003: 120). The apparent lower frequency of ideophones compared to Bantu aside, there might still be areal traits in the Kalahari Basin concerning the semantic profile of this class, offering an interesting field for future research. In particular, the more recent studies report a notable richness and salience of taste and food texture ideophones; this is not typical in Bantu and possibly even remarkable cross-linguistically, because these are low on implicational hierarchies entertained for ideophones (cf. Dingemanse (2012: 663) where they are merely subsumed under a more generic category “other sensory perception”).

Since lexical isoglosses beyond borrowing represent a largely untapped topic, future research, e.g., on additional domains like metaphors, lexical taboo etc. are likely to yield new insights into Kalahari Basin contact history.

2.3 Morphosyntax

As mentioned, Heine (1976) noted the universal presence of head-final genitives in Kalahari Basin languages irrespective of word orders elsewhere, and Güldemann (1999) treated the fact that such juxtaposed genitives are widely employed to convey notions of location, natural sex, and diminutives, which also results in host-final nominal flagging and derivation from the grammaticalization of compounds. This phenomenon even extends to the marking of number and gender, because former nominal heads could have encoded these categories and bequeathed them to grammaticalized structures, as illustrated in (1) and (2).

(1) Common Taa

\[ qáé \]
mother.3

\[ qáé-tù \] (< *tuu ‘people’) > \[ qáru \]
mother-ASS.P.4 mothers.4

(2) Juǀʼhoan

\[ lxó-má \]
elephant-DIM.S

\[ lxó-mhí \]
elephant-DIM.P

little elephant, e. calf little elephants, e. calfs (Dickens 2005: 27)

An exclusive-inclusive distinction in pronouns is also widespread in the Kalahari Basin; an early assessment is contained in Güldemann (1998). The feature has not been reconstructed for Proto-Khoe (Vossen 1997); Güldemann (2002, 2006: 111-2) gives concrete evidence that the opposition in Khoekhoe is due to contact with Tuu languages. This indicates that clusivity entered Khoe languages only after contact with local Non-Khoe languages. Since clusivity has been attested in a few more languages, including the Gǁana-Glui group of Kalahari Khoe (Ono 2010), the East Kalahari Khoe subgroup is now the only one without any language known to have the feature.

After Güldemann and Vossen (2000: 109) identified verb serialization as a universal feature in Non-Khoe, Güldemann (2006: 117-9) observed that “lexically complex predicates” (comprising serial and compound verbs) are also found in Khoekhoe and proposed that this is related to contact interference with Tuu. The short treatment suffered from an imprecise characterization of the relevant multi-verb construction(s)
(henceforth MVC), which partly spurred the studies by Haacke (2014) and Rapold (2014). They rightly point out that MVCs in Kalahari Khoe employing the “verb juncture” take care of functions very similar to those conveyed by MVCs without a segmental linker in Khoekhoe (and Non-Khoe); Rapold provides convincing evidence that juncture verb constructions were even a feature of Proto-Khoekhoe, contributing crucially to the emergence of the link-less constructions.

Before this background, and in contrast with neighboring Bantu languages, MVCs can be regarded as a prominent feature of the Kalahari Basin, straddling all major lineages in the area. The following sources contain relevant data on this topic: Güldemann (2013d) on ǀXam; Güldemann (2013c, f) and Kießling (2014) on Taa; Collins (2002, 2003) on ǂ’Amkoe; Collins (2003), Dickens (2005) and König (2010) on Ju; Haacke (1999, 2014) and Rapold (2014) on Khoekhoe; Visser (2010) on Naro; and Kilian-Hatz (2008) on Caprivi Khwe. In addition, field notes were consulted from Berthold and Gerlach on the N!aqriaxe variety of ǂ’Amkoe, from Nakagawa on Glui, and from Fehn on Ts’ixa and Shua. Note that the language-specific characterization of relevant structures as serial verbs, compounds, etc. does not always follow identical criteria.

Below six different types of MVCs (mostly “asymmetrical” in terms of Aikhenvald 2006), which are attested in virtually all major subgroups of the three families, are presented (a lineage is also considered to possess a MVC type if it is synchronically more grammaticalized). Each example features under a. the N!aqriaxe dialect of the Non-Khoe language ǂ’Amkoe, and under b. the Kalahari Khoe language Ts’ixa employing the juncture. This array does not attempt to be complete but only illustrates the extent of shared types across the area.6

(3) Sequential cause-effect
a. mā ēn ’l’áú ’n’áá
1S TAM fall sit
I fell into a sitting position
b. náxá=mín=tí kò muùn-à ’aàn
snake=M.S DEM.REF=M.S 1S IPFV see-JUNCT know
I recognize this snake

(4) Accompanying manner
a. mā ná llq’áá tsáá
1S TAM sing come
I am coming while singing
b. tí kò pere glláì
1S IPFV flee:JUNCT run
I run like a fugitive

(5) Accompanying posture (‘sit’, ‘stand’, ‘lie’ etc.+V2)

6 It goes without saying that there exist other more localized MVC types which may also involve contact transfer. Thus, Güldemann (2006: 118-9) discusses the non-causative variant of the “switch-function” type (cf. Aikhenvald 2006); this has a western distribution in the Kalahari Basin in occurring in ǀXam and various Khoekhoe varieties (cf. Haacke 2014 for new extensive data) as well as in Ju dialects (König 2010).
In some languages the last MVC type is an essential ingredient for a special pattern of TAM encoding discussed briefly by Güldemann (2006: 116-117): perfect~stative~relevance is marked exceptionally AFTER the verb or the entire clause. This directly relates to the general TAM morphotactics argued to be shared between !Ui and Khoekhoe in the Cape, namely that this postverbal marker is part of a core system that also involves a preverbal imperfective opposed to a perfective zero. Modern data from ″Amkoe (Berthold and Gerlach p.c.), Glui (Nakagawa 2014, Ono 2014) as well as Ts’ixa and possibly even some eastern Kalahari Khoe languages (Fehn field notes) indicate that such a system is in large parts a wider isogloss of the Kalahari Basin.

Another relevant feature initially identified by Güldemann (2006: 119-22) only for the Cape area concerns markers for sentence type and information structure. These bisect the clause into a pragmatically specific prefield and a postfield containing the rest of the clause. While they are mostly particles, in !Xam the relevant element, whose several allomorphs are represented here as an underlying velar nasal =NG, is transcribed as being attached to the subject topic (cf. Güldemann 2013d: 421), as illustrated in (9).
This feature, too, turns out to have a wider western Kalahari Basin distribution straddling all three families. Table 3 lists languages with such elements in declarative sentences; the diverse terminology in the fourth column betrays that their functional analysis is still far from conclusive.

<table>
<thead>
<tr>
<th>Dialect or language</th>
<th>Family, branch</th>
<th>Form</th>
<th>Label</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ekoka !Xun</td>
<td>Kx’a, Ju</td>
<td>má</td>
<td>Topic</td>
<td>König (2006)</td>
</tr>
<tr>
<td>North Ju’hoan</td>
<td>Kx’a, Ju</td>
<td>m</td>
<td>Verb particle</td>
<td>Dickens (1994: 234, 2005: 44)</td>
</tr>
<tr>
<td>N!aqriaxe</td>
<td>Kx’a, !’Amkoe</td>
<td>ki</td>
<td>-</td>
<td>Berthold and Gerlach (field notes)</td>
</tr>
<tr>
<td>East !Xoon Tuu, Taa</td>
<td>!Xam</td>
<td>ní</td>
<td>Indicative</td>
<td>Traill (1994: 193)</td>
</tr>
<tr>
<td>!Xam</td>
<td>!Ui</td>
<td>=NG</td>
<td>Emphatic nominative</td>
<td>Bleek (1928-30: 87-8)</td>
</tr>
<tr>
<td>N</td>
<td>Ing</td>
<td>!Ui</td>
<td>ke</td>
<td>Declarative</td>
</tr>
<tr>
<td>Standard</td>
<td>Khoe-Kwadi, KK</td>
<td>ke</td>
<td>(Indicative) Declarative</td>
<td>Hagman (1977), Haacke (2013: 335)</td>
</tr>
<tr>
<td>!Ora</td>
<td>Khoe-Kwadi, KK</td>
<td>tje</td>
<td>Subjekt-Determinativ</td>
<td>Meinhof (1930: 49-50)</td>
</tr>
</tbody>
</table>

Table 3. Clause-second elements in declarative clauses in Kalahari Basin languages

Güldemann (2010b) discusses these elements in Tuu languages more extensively and presents data to the effect that !Xam and Taa possess yet younger bisected constructions whose clause-second elements encode term focus and entity-central theticity. A similar analysis can be applied to other languages (cf. Güldemann and Witzlack-Makarevich 2013 for (Nama) Khoekhoe and Güldemann and Pratchett 2014 for South Ju’hoan). This provides a new perspective on the possible emergence and early function of clause-second elements. In some languages, the markers in declarative clauses are in complementary distribution with other particles marking questions, such as !Xam ba/xa, N|Ing xa(e), Khoekhoe kha, and the different reflexes of the interrogative particle in Ju reconstructed as *re by Heine and König (2013: 319). Visser’s (2013) data on Naro seem to indicate that syntactic phenomena revolving around a clause-second syntactic pivot do not necessarily depend on a segmental marker of the kind described above but can be rendered by subject-referring PGN (person-gender-number) markers alone. More in-depth morphosyntactic and discourse-oriented research on all relevant languages is required to clarify whether the different phenomena referred to here are indeed related to each other, and if yes, what their underlying common denominator is.

Note that subject PGNs in Khoekhoe also occur in clause-second position immediately before the markers discussed here.
Another intriguing feature, originally identified by Güldemann and Vossen (2000: 110) for Non-Khoe, has also turned out to be of wider relevance: Tuu and Kx’a languages display an extremely versatile preposition, called here multipurpose oblique (MPO) marker,\(^8\) which is virtually void of semantic content; it rather flags any postverbal term beyond the valency of a single transitive verb and thus marks a templatic syntactic clause slot rather than specifying the term with respect to its semantic relation to the predicate (cf. Güldemann 2004). Low semantic sensitivity of basic participant flagging also turns up in Khoekhoe in the form of the nominal suffix -\(\text{a}\) which marks both subjects and objects as long as they occur after the clause-second pivot (cf. Güldemann 2006: 122-123).

The limited import of semantics in certain participant flagging is related to yet another factor, namely the strong tendency that animacy often ranks higher than semantics for the assignment of a more central grammatical clause relation. This is one reason why in the Kalahari Basin, as opposed to neighboring Bantu, transitive patient and ditransitive recipient are frequently marked alike and the ditransitive theme follows the MPO (=“secundative” alignment in terms of Malchukov, Haspelmath and Comrie 2010), as illustrated by (10) from Jul’hoan.

(10) \(\text{dà’áma jàn l’àn ha bá kò màrì}\)
    child good give 3S father MPO money
    The good child gave his father money. (Dickens 2005: 40)

It is noteworthy in this respect that some Kalahari Khoe languages tend to use their postposition \(\text{kà}\) with a wide functional spectrum and that precisely this MPO-like postposition flags the ditransitive theme in \(\text{l}l\text{Ani and Ts’ixa}\) (cf. (7)b. above).

Another typologically remarkable feature in various unrelated Kalahari Basin languages is that relative-(like) constructions render nominal modification that is cross-linguistically conveyed by simple attributive numerals, other quantifiers, interrogatives, and even possessors. Ju varieties, as the most extreme case, show this entire range: the earliest such case recorded by Dickens (1997) is the verbal nature of deictics in Jul’hoan, which Heine and König (2013) and Lionnet (2014) show to hold for the entire group; for the other categories see Dickens (2005) and Heine and König (2013). A similar situation holds in Taa, notably for deictics involving (earlier) motion verbs (cf. \(t\text{V}('\text{VV})-\text{jà kV}\) (proximate) and \(t\text{V}('\text{VV})-\text{sà kV}\) (remote); Traill 1994: 154), quantifiers (Güldemann 2012, 2014b), and attributive ‘which’ (Güldemann 2013c: 411). As shown in (11), the West Kalahari Khoe language Glui uses its clausal attributive construction for a similar functional range.

(11) \([\text{Noun}_x \text{kà MODIFIER PGN}_x]\)
    a. \(\text{glàëkò kà ñà-m ki mùù sì}\)
       woman\(_x\) REL PRO-3M.S HOD.PST see 3F.S\(_x\)
       the woman who he saw today
    b. \(\text{ŋłàuû kà ŋlin mà}\)
       house\(_x\) REL this 3M.S\(_x\)
       this house

\(^8\)Other terms are “linker” (Collins 2004), “transitive particle” (Dickens 2005: 38-9), and “transitive preposition” (Heine and König 2013: 313).
This cross-linguistically remarkable phenomenon, which may involve several underlying factors, deserves more attention regarding its language-specific profile as well as its areal distribution.

Yet another feature attested in all three families is reduplicative causatives. In Non-Khoe it only occurs sporadically as in lXam (Bleek 1928-30: 171) and Jul’hoan (Miller-Ockhuizen 2001). It can be reconstructed, however, for Proto-Khoe (Vossen 1997: 350), so that the feature may have emanated from this family.

As a final example, dedicated/unique markers of associative plural, which typically though not exclusively combine with personal names and kinship terms, are also found in all three families. Both better known Tuu languages display such a marker, namely -tu in Taa (Güldemann 2013a: 238-9, see (1) above) whose status as an associative plural became evident only in later fieldwork, and -gu in lXam (Güldemann 2006: 131, 2013b: 243). In Kx’a, the situation is not fully clear. In ‘Amkoe, no such marker has been found according to Berthold (p.c.). For Ju, Heine and König (2013: 304-305) report potentially relevant distinct plural markers, reconstructed as *sì and *sin, but do not comment on any semantic difference between them. However, Dickens’ (1994: 263) information on -sín in Jul’hoan, a reflex of *sin, clearly suggests that at least in some dialects this marker is an associative plural. A unique associative plural can be identified in Khoekhoe (cf. Hagman’s (1977: 29) description of so-called ‘háá compounds’). Some Kalahari Khoe languages also possess such a marker, like Glui (Nakagawa, p.c.) and Ts’ixa (Fehn field notes), for which see (12).

Compressing this distribution with Daniel and Moravcsik’s (2005) world-wide survey, the feature may qualify as an areal trait of the Kalahari Basin, because all Bantu languages close to the area and recorded by these scholars, viz. Zulu, Sotho, and Luvale, only have an associative plural which is the same as the normal additive plural. For the record, as opposed to Bantu, Afrikaans has innovated such a marker in the areal context of the Cape, although its origin might not (exclusively) be due to contact with Kalahari Basin languages (cf. Nienaber 1994, Besten 1996).

2.4 Summary

Table 4 summarizes the (potential) isoglosses of sections 2.1-2.3; feature values are as follows:

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>frequent or even universal in relevant group</td>
</tr>
<tr>
<td>(X)</td>
<td>common but with linguistic restrictions</td>
</tr>
<tr>
<td>language/dialect</td>
<td>so far only attestation(s)</td>
</tr>
</tbody>
</table>
The last case holds in particular for East Kalahari Khoe which is the least known group among the languages still spoken. A similar, yet different situation applies to Khoekhoe of Khoe-Kwadi and !Ui of Tuu; here most languages spoken previously in South Africa have become extinct before they could be documented sufficiently. Accordingly, X or a language name only refers here to the subset of documented varieties; they are !Xam and !Nng for !Ui, and !Ora, Nama and other Namibian varieties for Khoekhoe.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Tuu</th>
<th>Kx’a</th>
<th>Khoe-Kwadi</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonetics-phonology</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Lingual ingressives = clicks</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2 Glottalic egressives = ejectives</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3 Uvular stops</td>
<td>Nng</td>
<td>X</td>
<td>--</td>
</tr>
<tr>
<td>4 Aspirated obstruents</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5 Obstruent-obstruent clusters</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>6 Nasalization</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>7 Pharyngealization</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8 Register tone system</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>9 Specific lexical root phonotactics</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Lexical structures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Restricted numeral system</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>11 Specific perception verb conflation</td>
<td>?</td>
<td>X</td>
<td>Jul’hoan (X)</td>
</tr>
<tr>
<td><strong>Morphosyntax</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Head-final genitive</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>13 Host-final locative flagging</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>14 Host-final derivation</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>15 Clusivity</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>16 MVC: V1 cause+V2 sequential effect</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>17 MVC: V1 manner+V2</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>18 MVC: V1 posture+V2</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>19 MVC: V1+V2 motion &gt; path</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>20 MVC: V1+’give’ &gt; dative/benefactive</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
21 MVC: V1+‘exist’ > current relevance

<table>
<thead>
<tr>
<th>22 MVC: non-causative switch-function</th>
<th>23 TAM morphotactics</th>
<th>24 Clause-second pivot</th>
<th>25 Non-semantic participant flagging</th>
<th>26 Non-canonical clausal noun modifiers</th>
<th>27 Reduplicative causative</th>
<th>28 Dedicated associative plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>X X X X X X X</td>
<td>X</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td>--</td>
</tr>
<tr>
<td>ǀXam</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>Glui, Ts’ixa</td>
</tr>
<tr>
<td>G</td>
<td>X</td>
<td>--</td>
<td>X</td>
<td>X</td>
<td>--</td>
<td>?</td>
</tr>
<tr>
<td>?Naro</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>(ǁAni, Ts’ixa)</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: EKalK = East Kalahari Khoe, KK = Khoekhoe, WKalK = West Kalahari Khoe

Table 4. Linguistic features shared across the Kalahari Basin

### 3 Discussion

From a continental perspective, the Kalahari Basin is one of several macro-areas in Africa. Given the results by Clements and Rialland (2008) and Güldemann (2010a), it must be distinguished first of all from the neighboring Bantu spread zone. This entails a certain risk of invoking features for the Kalahari Basin that define it only negatively in opposition to its neighbor, which is genealogically homogeneous and has a specific linguistic signature with numerous shared features. Many such features are likely to be absent in any non-Bantu language and may thus give a strong signal of difference on an areal scale, which indeed holds for the Kalahari Basin.

For example, Kalahari Basin languages predominantly possess a gender system but lack the specific Bantu type. However, this fact is not an areal trait, because the Kalahari Basin is diverse internally in possessing two other distinct types along the basic split of Non-Khoe vs. Khoe-Kwadi (cf. Güldemann 2000). Another case in point is participant indexing on the verb. As opposed to Bantu, no Kalahari Basin language has subject cross-reference but Taa, Khoekhoe, ǁAni, and possibly Deti, have verbal object marking in addition to nominal objects - this is cross-linguistically marked. Pace Güldemann (2010a: 573-574), this should not be viewed as an areal feature either: the lack of subject indexing is typologically common and only a negative criterion with respect to the Bantu spread zone, while exclusive object indexing, though rare, is not really common in the entire Kalahari Basin. Clearly, areal isoglosses must not be negatively defined absences but rather substantial positive features; this is indeed the case for all those proposed in section 2.

The homogeneous Bantu spread zone does, however, play an indirect role for the profile of the Kalahari Basin: the former has so-to-speak “sealed off” the latter from other similar areas in Africa, viz. those hosting more diverse non-Bantu languages. It is
important in this respect that the Kalahari Basin displays non-trivial linguistic affinities to eastern Africa in nominal morphosyntax and phonetics-phonology; these are arguably diagnostic for a hypothesis according to which the Kalahari Basin prior to the Bantu expansion was part of an earlier, far larger linguistic area that coincided with what is called in geography “High Africa” and which would have been largely submerged by Bantu (Güldemann 1999, Güldemann 2010a: 578-579).

A history of “decay” also concerns the Kalahari Basin itself. Recall that so far we have spoken of it as a linguistic area before the Bantu expansion. This raises the question of what happened after the advent of Bantu (and yet other colonizing languages) to the Kalahari Basin languages on the one hand and to these colonizing languages on the other.

With respect to the first issue, the answer is straightforward, when looking at the history and current sociolinguistic status of the relevant languages: virtually all languages in South Africa and southern Namibia have become extinct, and most surviving ones in these and other countries are marginalized, endangered, or even moribund. That is, the Kalahari Basin in the present sense has been in a long process of dissolution through wide-spread language change and language loss induced by later population events. In fact, the data in Table 5 no longer reflect a situation in the present but rather a reconstructed approximation to the past, and the Kalahari Basin’s likely prospects are that its former linguistic profile will vanish as a compact areal signal.

Nevertheless, and this relates to the second issue, it is far from disappearing completely, because it had a noticeable impact on later colonizing languages. We did not attempt here to systematically record whether or not a feature is also found in local Bantu languages and Afrikaans but only mentioned such cases to the extent that they inform the establishment of the Kalahari Basin in the narrower sense. A detailed treatment of this wider topic is a project in its own right. Here we only give a first assessment of three important languages of the area with respect to the features in Table 5. They are, followed by sources used, Nguni (Poulos 1998, Doke 1992), Tswana (Cole 1955, Krüger 2006), and Afrikaans (Donaldson 1993). In addition to language-specific material, there is also considerable literature dedicated to language contact between these and Kalahari Basin languages. The sources most relevant here for Nguni and Tswana are Meinhof (1905), Lanham (1962), Louw (1986), Herbert (1987, 1990), Vossen (1997), and Güldemann (1999). The contact-induced formation of Afrikaans, whose regional, non-standard varieties are especially relevant and taken into account here, has also been studied increasingly (cf., e.g., Luijks 2001, Mesthrie and Roberge 2001/2).

<table>
<thead>
<tr>
<th>Language</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16-22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nguni</td>
<td>X</td>
<td>X</td>
<td>--</td>
<td>X</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td>X</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Tswana</td>
<td>(X)</td>
<td>(X)</td>
<td>--</td>
<td>X</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td>X</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Afrikaans</td>
<td>(X)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td>X</td>
<td>--</td>
<td>(X)</td>
<td>--</td>
<td>(X)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Kalahari Basin features shared by other local languages

The results, given in Table 5, justify the conclusion that substrate interference contributed repeatedly to creating linguistic similarities with Kalahari Basin languages (or at least maintaining existing ones) but has not been strong enough to make the newcomers “full” members of the area.
The following can be observed about the internal profile of the Kalahari Basin in the narrow sense. The best evidence for it are features with a homogeneous geographical and genealogical spread, which are indeed numerous: clicks, ejectives, aspirated obstruents, tautosyllabic obstruct-obstruent clusters, nasalization, register tone system, lexical root phonotactics, restricted numeral system (with the exception of Khoekhoe), head-final genitive, host-final flagging and derivation, and several types of multi-verb constructions.

At the same time, a number of features are not evenly distributed across the area and the three families. Only rarely is a feature found throughout the Khoe family while being sporadic in Non-Khoe, e.g., the reduplicative causative. These are candidates for a scenario in which the feature spread from Khoe into various Non-Khoe languages. The predominant situation is that a feature is well entrenched in the Non-Khoe families while Khoe languages partake in it only incompletely. This situation holds for clusivity, TAM morphotactics, clause-second pivot, non-semantic participant flagging, non-canonical clausal noun modifiers, and dedicated associative plural.

If one conceptualizes a linguistic area in terms of center vs. periphery, the distributions of the last type can be seen as a variation on a more general theme. Khoe displays a geographical cline from the north and east towards the south and west whereby the more its languages have encroached onto the Kalahari Basin the more pronounced is their change towards Non-Khoe patterns (cf. Güldemann 2006: 105). One can tentatively establish the following Khoe-internal group hierarchy of increasing Kalahari Basin character (with the caveat that missing data on East Kalahari Khoe might still change this picture): Shua+Tshwa+Khwe > G!|ana+Naro > Khoekhoe. Güldemann (2008a) has proposed a concrete historical scenario for how this situation would have come about, the main idea being that Khoe-Kwadi is also a colonizing lineage associated with the spread of pastoralism into southern Africa.

The Non-Khoe families Tuu and Kx’a represent the structural core of the Kalahari Basin, to whose profile the many isoglosses restricted to them can be added. As long as the two families are treated as genealogically independent, this finding reflects a yet earlier areality before the advent of Khoe-Kwadi – an idea reminiscent of Westphal’s (1980: 77) concept of a “‘Bush” language province’. Non-Khoe can thus be conceived of as having produced a kind of linguistic founder effect. The resulting areal profile has so-to-speak “percolated up” into a later sequence of colonizing linguistic layers (in chronological order Khoe-Kwadi, Bantu, Dutch-Afrikaans) by means of multiple direct and indirect substrate interference. However, Non-Khoe itself has been receding increasingly due to large-scale language shift.

References


Luijks, Carla. 2001. The Khoekhoe and/or the San: gathering the Afrikaans substrate languages, in Bell and Washburn (eds.), 184-199.


Nakagawa, Hirosi. in press. Glui ideophones, Asian and African Languages and Linguistics 8.


Traill, Anthony. 1986. Do the Khoi have a place in the San?: new data on Khoisan linguistic relationships, Sprache und Geschichte in Afrika 7,1: 407-430.
18 South Africa and areal linguistics

Rajend Mesthrie

1 Introduction

As far as language contact is concerned, South Africa has been reasonably well studied, though previous studies have focused on pairwise comparisons, rather than the broad sociohistorical overview with layered multiple comparisons attempted here. Languages of several major families have co-existed in South Africa since colonial times beginning in the mid-seventeenth century: chiefly Khoesan languages (which are not all related), Bantu (the majority languages), and Germanic (Dutch/Afrikaans and English). To these may be added regionally significant languages no longer spoken – e.g. those of Cape slaves (Malay, Malagasy, etc.), or in decline – e.g. those of Natal indentured workers (Hindi, Tamil etc.) Although South Africa is often characterised by the notorious practice of apartheid, which aimed to keep ethnic groups apart, a large amount of contact between groups of people and their languages nevertheless did occur. This chapter will provide an overview of the extent to which there has been language convergence in South Africa across languages. The current language demographics in South Africa are given in Table 1 for home language speakers of the 11 official languages and Sign Language as returned in the 2011 census. While the census question did not aim to uncover speakers’ full repertoires, it does give a first impression of L1 usage. Unfortunately, the census reports do not give figures for second language usage, crucial in any characterisation of areal tendencies.

<table>
<thead>
<tr>
<th>Language</th>
<th>Speaker numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>IsiZulu</td>
<td>22.70%</td>
</tr>
<tr>
<td>IsiXhosa</td>
<td>16.00%</td>
</tr>
<tr>
<td>Afrikaans</td>
<td>13.50%</td>
</tr>
<tr>
<td>English</td>
<td>9.60%</td>
</tr>
<tr>
<td>Sepedi</td>
<td>9.00%</td>
</tr>
<tr>
<td>Setswana</td>
<td>8.00%</td>
</tr>
<tr>
<td>Sesotho</td>
<td>7.60%</td>
</tr>
<tr>
<td>Xitsonga</td>
<td>4.50%</td>
</tr>
<tr>
<td>SiSwati</td>
<td>2.50%</td>
</tr>
<tr>
<td>Tshivenda</td>
<td>2.40%</td>
</tr>
<tr>
<td>IsiNdebele</td>
<td>2.10%</td>
</tr>
<tr>
<td>Other</td>
<td>1.60%</td>
</tr>
<tr>
<td>Sign language</td>
<td>0.50%</td>
</tr>
</tbody>
</table>
Table 1 shows the robustness of nine Bantu languages and English and Afrikaans. What the present day census misses is the presence of endangered Khoesan languages like /Nu, Gri and 'ora (see the overview in Güldemann and Fehn, this volume). Yet two thousand years ago Khoesan languages would have filled the entire southern African landscape. Bantu languages arrived later - most likely from about the first four centuries CE, though the ancestors of present-day groups followed somewhat later (Herbert and Bailey 2002: 59). Khoesan languages gave way eventually first to Bantu languages and later Dutch. In the process they affected Bantu languages like Xhosa, Tswana and Zulu significantly. They also contributed significantly to the social dialect of Dutch or Afrikaans that they themselves eventually shifted to.

2 Phonetic and vocabulary influences of Khoesan on other South African languages

The greatest influence of Khoesan upon Bantu is in the area of phonology. Unlike Bantu languages outside southern Africa, several of their South African counterparts have clicks as an essential part of their phoneme inventories. Thus clicks relate to contact and synthesis between Khoesan and Bantu in precolonial times. The number of clicks in Khoesan languages varies from as few as 20 to as many as 85, though current research suggests an even higher upper limit (Bonny Sands, p.c.). These numbers encompass different places of articulation from bilabial, dental, alveolar, retroflex, lateral and palatal combined with options for voicing, pre-nasalisation, aspiration and so forth. Of the Bantu languages of South Africa, Xhosa has the most phonemic clicks, then in turn Zulu, Swati and South Sotho (Herbert 1990). Khoesan influence in the phonologies of Xhosa and Zulu is extensive: Lanham (1964), cited by Herbert (1990: 122), calculated that between 21 and 25 of the consonants of Xhosa are of Khoesan origin and occur in borrowed vocabulary. He also estimated that between one-seventh to one-sixth of Zulu words contain clicks. Scholars of Bantu languages assume that Xhosa clicks were borrowed from a Khoe language, possibly 'ora, while those of Swati and Sotho have a San language as their source (Herbert 1990, citing Lanham 1964 and sources therein). No definitive account exists of the nature of the bilingualism that promoted this large-scale phonological borrowing. Herbert (2002) suggested that the cultural practice called hlonipha (respect) in which married Xhosa and Zulu women had to consistently avoid the initial syllables of the names of their in-laws led to a search for alternate ways of expressing the tabooed syllable, including borrowed elements like clicks (see Finlayson 1984). Support for the theory comes from (a) the practice of hlonipha correlating strongly with the inventory of clicks in individual Bantu languages, and (b) the inconsistent way in which clicks are borrowed across languages. The weakness in the argument is how

---

1 The discrepancy between the total for languages and for the population stems from 809 117 returning “n/a” for the language question, presumably indicating a first language different from the 12 languages enumerated in the census (11 official languages, plus sign language). It is a pity that the census did not analyse languages other than the official ones.
clicks were generalised to unmarked rather than taboo usage and transmitted to all speakers, not just resourceful young women. Language shift might prove a more transparent explanation, with speakers of both sexes who slowly shifted from a Khoesan to a Bantu language like Xhosa retaining traces of their earlier phonology and gradually influencing other Xhosa speakers. Such an explanation has parallels in ancient India (Emeneau 1956) and medieval England and Ireland (see Hickey 2010, who terms this an imposition scenario). That this is not a matter pertaining only to the distant past can be seen in the way clicks have survived shift in a small way in one contemporary case: Tony Links (1989) reports on the social dialect of Namaqualand, where descendants of Khoesan people retain clicks in a subset of words pertaining to plant and other traditional names in the language they have shifted to – Afrikaans. Current bilingual speech involving a Bantu language and English are also relevant to the theme of language convergence. Bilingual speakers of Xhosa and Zulu retain clicks when speaking English, though essentially in proper names of people, places and cultural concepts. Insistence on this practice has led to many South Africans using at least the word Xhosa ([ǀǀhosa]) a widely spoken language) with a click, though not all L1 English speakers necessarily produce the correct (aspirated lateral) one. In at least one instance, young people of different language backgrounds that do not necessarily include a Bantu language use the now English slang term (m)njca ([[(m)nla]) ‘cool, nice’, getting the dental click [l] right. In my own interviews of the L2 English of Xhosa speakers there are sporadic instances of clicks being transferred to English. Only four instances of these were noted in about 20 hours of speech data: do (twice by the same speaker) and twenty, by another two speakers, all with an initial alveopalatal click [!] in place of [t] or [d]. While these were “one off” instances, they do suggest how bilingualism and later shift could have caused such large scale areal convergence in a less prescriptive and literate past.

Regarding vocabulary, some basic terms for ‘sheep’, ‘cow’, ‘grass’ and ‘milk’ have been borrowed from Khoesan into Xhosa and Zulu (see Harinck 1969). Branford and Claughton (2002: 202) cite verbs of probable Khoesan origin in Xhosa like -cela ‘to ask’, -qala ‘to begin’ and -nceda ‘to help’. And as is to be expected, several South African toponyms are of Khoesan origin, even if in diluted colonial form: Karoo desert, Tsitsikamma forest, Gariep river/dam etc. The lexical give and take between Bantu languages and English and Afrikaans is enormous. Everyday terms arising out of colonial contact in Zulu include words from English like ingadi ‘garden’, imoto ‘car’, ilondolo ‘laundry’, imakede ‘market’, and words from Afrikaans like ijazi ‘overcoat’ (Afrikaans jas), iphulazi ‘farm’ (Afrikaans plaas), ubontshisi ‘beans’ (Afrikaans boontjie), etc. Today as a result of the technological revolution and globalisation, English (rather than Afrikaans) words continue to enter the Zulu lexicon.

Large-scale borrowing has led to the incorporation of new phonemes into African languages. For example, the traditional form of the English loanword ‘rice’ in Zulu is i-layisi; but this form that turns an /l/ into an /r/ is now considered old fashioned and rural, with younger speakers giving the form i-rayisi, with the newly introduced phoneme /r/. In section 7 I discuss changes in the practice of choosing an appropriate noun class into which to place loanwords.

South African English vocabulary is a monument to language contact, not just in the numerous second-language varieties in the country, but in ‘superstrate’ or L1 ‘Anglo’ varieties. South African English has borrowed extensively from Afrikaans, giving new words to international English in the process (trek, veld, apartheid) and many terms that are less known internationally (braai ‘barbecue’, rondavel ‘small circular building’,
'food taken for a journey' (lit 'roadfood'). The terms cited here are used even in parts of the country where Afrikaans is not widely used (especially parts of KwaZulu-Natal). In areas where English-Afrikaans bilingualism is widespread, the number of words used is even greater: hence erf ‘plot of land’, vlei ‘seasonally intermittent lake area’ and water-blommetjie (an edible water plant) in the Western Cape. Conversely, despite attempts to weed out Anglicisms in Afrikaans, the influence of English is unavoidable. Donaldson’s (1988:169-233) discussion of loanwords, neologisms, calques, hybrids, proverbial and similar expressions ranges over 60 pages - this from an era before the multitude of terms brought by the technological explosion of computers and cellphones. The lexical influence of Bantu languages on English is also notable. In the post-apartheid era, some new terms associated with politics may be noted: lekgotla ‘a meeting of village or political leaders’, batho pela ‘people first’ (in relations to politics). On the whole, new words adopted from the African languages into English of the general populace tend to be informal and associated with youth culture: eish (an expression of hardship or dismay), kasi ‘Black township or section of township’ (a clipping from Afrikaans lokasie), loxion ‘Black area of residence’ (a rehabilitation of the apartheid term location), kwaito (high energy youthful musical style) etc. Thus kasi and loxion are etymologically related, both creative adoptions of the Germanic pair lokasie/location. Advertisers are trying to broaden their appeal by using such informal terms, even if they are unknown to non-African users of English: mahala ‘free of charge’, ringa ‘make a call, speak to someone on the phone’, diski ‘football’ etc. When English words are incorporated into an African language they tend to be less informal or interpersonal: dealing rather with technological and educational developments. Thus, although there is large scale mutual lexical influence across languages, it tends to be asymmetrically structured according to domain.

3 Afrikaans and contact

As we shall see, Afrikaans plays a pivotal role in the diffusion of features into different South African languages. Most analysts would accept the view of Afrikaans as at least a partially restructured variety of early modern Dutch, under pressure of contact with indigenous languages of the Cape and – more importantly – languages brought to the Cape by slaves in the period from the sixteenth to the eighteenth century. Features that became part of Standard Afrikaans that are attributable to such sources include the following:

1. loss of all traces of ablaut and the distinction between strong and weak verbs.
2. loss of all regular verb endings of person, number and tense.
3. replacement of the past tense by perfective aspect, marked by prefixed ge-
4. loss of gender in the article system.

Although this appears to be a set of fairly far-reaching structural changes, it must be noted that other prototypically Germanic features persist in Afrikaans (V₂ order in main clauses; verb final clause order if V₂ is filled by an auxiliary; formal distinction between present and past modal auxiliaries). The features outlined above are of an attritional nature, detailing the loss of important structural distinctions in Dutch. They are not part
of a Sprachbund, attributable to large scale and sustained bilingualism. In the early contact ecology of the Cape, Afrikaans survived and expanded its territory and number of speakers, while the indigenous Khoesan languages and languages of slaves (Bengali, Malay, Buganese etc.) did not. On the other hand there are ‘positive’ adaptations in Afrikaans grammar, showing the addition of structural features under the influence of other languages.

3.1 Double negation

Double negation is fairly complex in Afrikaans, distinguishing it from other Germanic languages. Combrink (1978:79-85) provides an accessible discussion. He notes that with simple monoclausal sentences single negation is the rule when the VP is made up of only a verb (unlike French).

1. Sy eet nie
   she eat NEG
   ‘She does not eat/ She is not eating’

If the VP contains an object, then negation is doubly marked:

2. Sy eet nie   pap   nie.
   She eat NEG porridge NEG
   ‘She does not eat porridge’.

This rule would apply to adjectival and adverbial complements of verbs as well. In main clauses with the auxiliary in V₂ position and the main verb in clause final position double negation is mandatory.

3. Sy sal   nie   eet nie
   She will NEG eat NEG
   ‘She won’t eat’

Combrink (1978:82) notes that whereas Afrikaans inherited the first nie from seventeenth century Dutch, the second nie is “the strange bird in the flock”. Double negation is notable for four characteristics:

1. It is never a stylistic variant of single nie negation, apart from poetic use.
2. It is mandatory in many structures but inadmissible in root sentences with single verb and no object or verb complement.
3. The same particle is used twice (unlike, say, French ne... pas).
4. The second nie is always clause-final.

Although some scholars, notably Pauwels (1958, cited by Combrink 1978) had noted sporadic similarities in Flemish dialects in the Belgian province of Brabant, Combrink finds it implausible on demographic grounds that this should be the source of Afrikaans double negation. He suggests that Khoekhoe influence is more likely “from below”, citing the work of Nienaber (1965) who examined the use of negatives in Afrikaans
sources between 1830-1844. Whereas colonists of Dutch descent use double negation 13% of the time, people of Khoekhoe origins who had shifted to Afrikaans by then show higher usage at 40%. Nienaber (1965) noted that Nama has three negative particles used in different moods. The negative particle for imperatives precedes the verb, but the particles for indicative and subjunctive follow it. Combrink (1978) and Den Besten (2012: 243-245) accept that while Nama and presumably other Khoekhoe languages do not have double negation, the second *nie* of Afrikaans shows the influence of post-verbal negation in non-imperatives of Kho. Double negation was in Combrink’s analysis one of the features that was initially stigmatised in Cape Dutch, the forerunner of Afrikaans, but later accepted and “ennobled” (1978: 85) when the promoters of Afrikaans as a new language emphasised its local innovations to mark it off as independent of Dutch.

3.2 Associative plurals

Afrikaans has evolved a construction in which a third person plural pronoun *hulle* is adjoined to an NP that is typically human or at least animate to form an associative plural carrying the sense ‘X and others, X and friends, X and such’.

4. Pa-hulle
   father-3P
   ‘Father and others, father and friends’

In a detailed survey Den Besten (2012: 31) proposes the influence of Khoekhoe and Eastern Indonesian languages here. From Khoekhoe he cites usage in Nama, which has ‘X + *hã* + PGN’, where PGN denotes a person-number-gender suffix (in this construction it would be a [−sg] suffix). He gives the example:

5. saru-hã-n
   cigarette-ASS-3C.P
   ‘Cigarettes and stuff – smoking things’

Den Besten argued that in transferring to Afrikaans the *hã* element would have been lost and an Afrikaans morpheme like *hulle* (third pl pronoun) or *goed* ‘stuff’ taking the function of the PGN marker. Likewise, although standard Malay does not have associative plurals, eastern parts of the Indonesian archipelago (from which a majority of slaves originated) does: e.g. Ambon Malay has an associative plural formed by noun + third Pl pronoun (Den Besten 2012:32 check page). Hence *Anis dong* “Anis and others’ (where *dong* is a third person plural pronoun). Güldemann (2007) concurs with Den Besten that contact with Khoekhoe and Asian slaves was the main source of this construction. However, he points to associative plurals in the San languages of the Tuu group: example (6) is from lXam, the extinct San language once spoken in the areas closest to Cape Town:

6. mama-gu
   mother-ASS.P

2 Güldemann notes that *gu* is cognate with !Ui *gu-(ken)* ‘stuff, things, creatures’. 
'mother and others, mothers'

The |Xam example, in fact, comes close to the form from the present day Orange River dialect of Afrikaans in semantics and phonetic similarity, see below. *Goed* denotes ‘stuff’ in Afrikaans and is cognate with English *goods*, but the similarity of the initial syllable with Tuu *gu* ‘stuff, things, creatures’ might have played a reinforcing role:

7. pa-goed
   father-ASS.P
   ‘father and others, father and company’.

3.3  Reduplication

This is a pervasive feature of Afrikaans, relevant to a number of word-classes. For some verbs reduplication serves to emphasise continuation and duration as in (8) below taken from Botha’s (1998: 24) full length study of Afrikaans reduplication.

   the thunder rumble-rumble in the distance
   ‘The thunder rumbles continuously in the distance’.

Reduplication of other verbs of a temporally bound nature (e.g. *lek* ‘lick’) denote iteration or repetition. Some verbs of a non-punctual nature (e.g. *skop* ‘kick’), however, carry an attenuative semantics when reduplicated. Punctual, experiential and stative verbs do not generally allow reduplication, unless some contextual material allowing an attenuative reading is present. Hence sentence (9) from Combrink (1978: 78):

9. Sy voel-voel met haar voet hoe warm die water is.
   she feel-feel  with her foot how warm the water is
   ‘She hesitantly/ carefully puts her foot into the water to feel how warm it is’

In the other categories reduplication abounds with intensive or distributive meanings:

1. adverbs: *dronk-dronk* ‘drunkenly’
2. adjectives: *diep-diep* ‘very deep’
3. numerals: *drie-drie* ‘three at a time, groups of three’
4. nouns: *sakke-sakke* ‘sacks and sacks’

Auxiliaries and prepositions are less commonly reduplicated: Den Besten (2012) cites an example of the reduplication of *wil* ‘want’ in the sense of ‘about to, on the verge of’. Prepositions like *aan* ‘on’ may be reduplicated to form a noun for the children’s game *on-on* (a game of tag or touch). In respect of origins Den Besten (2012: 219) summarises the position as follows:

Reduplication in Afrikaans is essentially iconic in nature. The pertinent reduplication patterns almost exclusively derive from languages other than Dutch and were selected on the basis of what one might call an iconicity
principle. This explains why Khoekhoe contributed relatively little to reduplication in Afrikaans. The main influences came from Malay and Asian Creole Portuguese, with a small but important contribution from Khoekhoe, while the majority of the Malagasy slaves came too late to contribute something specifically Malagasy.

For details of the likely influences from Malay and Asian Creole Portuguese the reader is referred to Den Besten, Luiks and Roberge (2002: 278-285).

3.4 Diminutives

Although standard Dutch uses diminutives, Afrikaans shows a greater affinity for this structure. Firstly, Afrikaans seems to attach the diminutive suffix –tjie to a greater range of nouns and new meanings than is possible in ordinary Dutch. Valkhoff (1966: 231) cites examples like dogtertjie ‘girl’, in addition to the Dutch literal meaning of ‘little daughter’. Secondly, Afrikaans has occasional double diminutives, which are not possible in Dutch and which hint at contact-induced restructuring: e.g. baadjie-tjie ‘little coat’; boontjie-tjie ‘little bean’ (Den Besten 2012: 281)

3.5 The tag nè

Afrikaans has an invariant negative tag nè, equivalent to ‘isn’t it/ isn’t that so, don’t you know’ etc.

10. Ek sien jou môre, nè?
    I see you tomorrow, TAG
    ‘I’ll see you tomorrow, right?’

Although European sources of nè are sometimes cited (Dutch nicht waar, French non, West German colloquial ne [nə]); it seems to me that an Asian source is also possible (Sri Lankan Tamil has the exact same form (nè), Bengali has na, and so has Kristang (Malacca Malayo-Portuguese Creole). The particular frequency of this tag among Indian and Coloured speakers of Afrikaans would appear to strengthen the case for an Asian origin, or at least a convergent effect from a variety of oriental and occidental languages.

4 Afrikaans contact effects in relation to other languages

Of particular interest is the fact that a large number of these neologistic Afrikaans forms, arising in a high contact situation, have passed into English. I discuss the features covered in section 3 in the same order here.
4.1 Double negation

Of the above contact features of Afrikaans only double negation has not passed into general South African English. Where double negation occurs it is usually among L2 speakers following a common international pattern (*I didn't see no snakes*) rather than a specifically Afrikaans template. Although Afrikaans double negation has not generally influenced other languages, there is evidence that at one stage it had incipient influence on urban Zulu of the Transvaal. In a comprehensive and carefully documented article, detailing the development of Xhosa and Zulu, Louw (1983) gives an example of *nie* from Afrikaans being adopted to form a triple negative as follows:³

11. A-ngi-hamb-i ni
    NEG-I-go-NEG NEG
    ‘I do not go’

The first two negatives in (11) are standard circumfixes of Zulu (double in form but single in the semantics of negation) to which speakers add an extra *ni* for emphasis (what Louw describes as a “greater negative force”). This third negative is clearly the strange bird of the Afrikaans flock, i.e. the second *nie* of Afrikaans discussed above. It is a free form that rather goes against the agglutinating tendencies of standard Zulu syntax. (Urban Zulu draws upon other salient free forms like *why* and *never* from English, as (variably) expressive of a more modern and young way of speaking.) To my knowledge the use of this borrowed *ni* has not stabilized, and we may well conceive of it as sporadic switching that did not catch on. Even so, it is impressive that the contact features of Afrikaans should continue to prove so attractive to speakers of other South African languages. Louw’s discussion makes it clear that he attributes this usage to Urban Zulu, and not its slang register for special effects (Tsotsitaal).⁴

4.2 Associative plurals

Associative plurals are also common in South African English, as a sort of hypercorrect form in SAE *Johnny-and-them*. As Branford (1991:11) notes, these forms are used for a group of people given in the discourse or known to the hearer. The direct calque is also found in South African Indian English and South African Coloured English: *Johnny-them* for ‘Johnny and friends’.

4.3 Reduplication

This phenomenon is now part of general South African English, though not as prominently as in Afrikaans. Occasional examples can be found in several word classes:

---
³ The article is dated 16 April 1978 by the author.
⁴ Pierre Aycard (p.c., December 2013) informs me that this usage does not occur in his present day data-base of urban Zulu and slang registers of African languages in Soweto.
a) Nouns: *Doctor-doctor, school-school* in children’s games;
b) Prepositions: *on-on* which is a direct calque of the Afrikaans children’s game *aan-aan* cited above, resulting in a compound noun;
c) Adverbs: *quick-quick* ‘quickly’, *now-now* ‘shortly’ (which has a more urgent triplicated form *now-now-now*);

There is also semantic but not lexical reduplication in South African English phrases like *a small little book*. Bantu languages show reduplication independently of Afrikaans. Afrikaans reduplication has not influenced African languages, since for the most part it would clash with the attenuative semantics associated with reduplication in languages like Zulu. Hence Zulu *hamba* ‘to go, to travel’, but *hamba-hamba* ‘to saunter’.

### 4.4 Diminutives

The Afrikaans predilection for diminutives has also influenced South African English, which has an informal diminutive *boy-kie* ‘young men, boys’ (based on an English root and Afrikaans suffix –*kie*). The term *book-ie* ‘a little book’ (not the informal betting term, which is also possible) is also pertinent here. However, the role of Afrikaans might be reinforcive rather than primary in this example, since British English has colloquial endearments like *duckie* and *lovey*.

### 4.5 The tag *nè*

This tag (equivalent to ‘not so?, isn’t it’ see 3.5 above) has proved very mobile in South Africa, and has passed into the colloquial English of mainly Coloured and Indian speakers in many parts of the country.


The tag also occurs in colloquial Xhosa, among speakers influenced by Afrikaans. Although this usage is not recorded in conventional textbooks, it can occasionally be found in oral teaching materials for courses on colloquial Xhosa. It also occurs more generally in the informal street varieties of African languages, in which code-switching is prized. Aycard (2014) notes its occurrence in Urban Zulu of Soweto, where *nè* appears to have replaced the very similar standard Zulu particle *na*, which is also a tag, with negative interrogative interactive force equivalent to ‘aren’t you’.

### 5 The story of *busy* as a case of weak convergence

Proving that Afrikaans is the source for the English innovations is usually taken as unnecessary. Yet one should be cautious in attributing all parallelisms to contact: historical linguists know that independent innovations are possible, or that careful historical work can show what is assumed to be a (convergent) innovation could have older roots than expected. The use of aspectual *busy* in South African English as a semi-
grammaticalised marker of ongoing activity has been studied in this vein by Lass and Wright (1986) and further by Mesthrie (2002). Two examples showing how South African English has lifted the restriction of busy to activity verbs are given in Branford and Branford’s Dictionary of South African English (1991:56):

13. I rushed in and found two infants busy having convulsions – as though there were not enough troubles. K. McMagh Dinner of Herbs, 1968.
14. There was a sign on the kitchen door saying ‘Do not disturb me, am busy praying.’ Sunday Times 28 October 1990.

South Africans can say without any trace of humour or irony things like the following attested examples:

15. I’m busy relaxing
16. You’re busy losing weight!
17. His girlfriend is busy having a baby.
18. His jeans were busy being worn.

As these examples show, the construction denotes particularly intense states or actions rather than ‘merely busy’ activities. The existence of a parallel structure in Afrikaans as the usual way of marking ongoing activity (i.e. progressive aspect) makes it relevant to a study of linguistic convergence. Branford and Branford (1991:56) consider South African English busy to be calqued on Afrikaans besig (om te), made up of the Afrikaans for ‘busy’ followed by the infinitive. While they generally argued against Afrikaans influence on South African English, Lass and Wright concluded that Afrikaans may have played a minor role in lifting the semantic restriction on busy to occur with verbs of conscious activity. Mesthrie (2002) notes that stylistic considerations also come into play, since the collocation of busy with ‘NON-BUSY’ verbs does occur in British and American English, but is usually consciously humorous, ironic or poetic. Thus Agatha Christie can have one of her characters say, “This isn’t exactly a statement you’re asking me to make, is it? No, it couldn’t be, because your Sergeant is busy upsetting the domestic staff” (A Pocket Full of Rye, 1973[1953]:29) Mesthrie (2002:354) also gives examples from American cartoons, and folk-rock music including Bob Dylan’s It’s all right I’m only bleeding, a song steeped in irony. At one level then, the South African usage is a continuation of endogenous developments of the kind found in British English and direct influence from Afrikaans is probably not very strong. Further evidence for this comes from the observation that when some Afrikaans speakers use the construction in their L2 English they follow it with an infinitive. Thus a sentence like A theory was busy to develop in the eighties may be found in occasional academic writing of people with Afrikaans L1, rather than ‘A theory was being developed/ was busy developing in the eighties’. However, South Africa might be more advanced in respect of the bleaching of busy than US and UK varieties as suggested by the decline of constructions like so/very/ever busy + V -ing and of busily + V, all of which occur fairly commonly in the British National Corpus. There is also the related South Africanism busy with, (as in He’s

5 Lifting of the stylistic constraint may even have been incipient in the U.K., judging from a sentence in Agatha Christie’s Curtain;Poirot’s Last Case, 1975: 71): I was so busy staring at him that I did not hear a footfall nearer at hand, and turned with a start when Miss Cole spoke to me.
busy with another patient right now). Branford (1991:56) proposes that this is calqued on Afrikaans besig met. Some role for contact in the use of South African English busy therefore cannot be ruled out altogether. The influence of Afrikaans is therefore likely to be quantitative more than qualitative: bilingual speakers may have contributed to the frequency of the construction, turning it from a marked to unmarked construction for intense activities.

This salient lexico-grammatical item has been borrowed into Urban Zulu of Johannesburg, as the following example from Soweto Zulu shows (Aycard 2014: 169):

19. A-wu-bon-i si-bhizi si-ya-dla?
NEG-2SG-see-NEG 1PL.SC-busy 1PL.SC-PRES-eat
‘Don’t you see we are busy eating?’ (see footnote 8 for abbreviations)

6 English influence on Afrikaans syntax

The influence of Afrikaans on English syntax is mostly of an informal nature. By contrast – and not surprisingly given the status of English as H language and language of higher education – Afrikaans syntax has changed in the direction of what counts in English as greater formality. This can be seen generally in Donaldson’s (1988) book-length treatment of the influence of English on Afrikaans at all linguistic levels. Donaldson concedes that not all scholars of Afrikaans linguistics are in agreement with his claims of the syntactic influence of English, the investigation of Anglicismes being an ideologically controversial area in Afrikaans studies. The main features showing likely English influence are summarised below.

(a) The Anglicism wees ‘was’ co-varying with Afrikaans word in passive constructions (p. 215-218).

(b) Use of een ‘one’ as an indefinite pronoun (‘n bloue, covarying with the Anglicism ‘n blou een - indefinite usage for ‘a blue one’ where Dutch uses an adjective as substantive).

(c) Use of een as an anaphoric pronoun ‘n dom seun en ‘n slim een ‘a stupid boy and a smart one’ with the anaphoric een a likely Anglicism (p. 277).

(d) Occasional non-final position of verbs in subordinate clauses as in (19), where Donaldson (1988: 278) argues that this treatment of van haar as a heavy PP postposed after the verb is inappropriate, since “it is too short an utterance for final positioning”. He detects the influence of English SVO order instead.

20. Ek het ge-hoor van haar.
I have PERF-hear of her.
‘I have heard of her’
(e) Use of COMP element *dat* or $\varphi$ for ‘that’, followed by SVO order of the subordinate clause, as in (21).6

21. Dan moet ons sê dat hierdie is net die eerste stap.  
then must we say that this is only the first step  
‘Then we must say this this is only the first step’

Donaldson (1988:279) acknowledges that examples like (21) are not very common in Afrikaans and possibly have Dutch precedents anyway. However, after certain verbs like *beweer* ‘maintain, claim, assert’, *dink* ‘think’, *glo* ‘believe, trust, think’, and *hoop* ‘hope’ the omission of COMP and use of SVO order in the subordinate clause ‘can be said to be the general rule’, even though *dat* + SOV is still an option.

(f) Preposition stranding leading to non SOV in subordinate clauses. Donaldson cites Ponelis (1985) in connection with the likelihood of English influence in preposition stranding in Afrikaans. Leaving a preposition stranded in final position obviously entails a non-final position for a subordinate clause verb as in (22):

22. drie akker wat hulle boer in  
three acre RELATIVE they farm in  
‘three acres on which they farm...’

(g) Donaldson notes that although the inversion of subject and verb is still the rule in Afrikaans after adverbs (the V₂ rule), there are frequent transgressions after *so* (an *ingeburgerde anglicisme* or naturalised Anglicism). Hence (23) shows the order *jy + kan* ‘you can’, rather than the inverted *kan + jy*.

23. So jy kan vergeet wat ek ge-sê het  
so you can forget what I PERF-say have  
‘So, you can forget what I’ve said’

It seems to me that an alternate explanation – given the rarity of the putative feature – is that *so* is treated as a discourse marker here, rather than a true adverb of sequencing. The following subject plus verb is then exempted from the V₂ rule.

(h) Reduced adverbial phrases functioning as prenominal compounds: ‘*n drie maande lank kursus* ‘a three-month-long course’ rather than ‘*n kursus wat drie maande lank is* ‘a course that is three months long’.

7 English and SLA

The examples of convergence in sections 4 and 5 have shown the influence of Afrikaans on L₁ and L₂ varieties of South African English. However, South Africa is well known for the diversity of its social dialects, mainly along ethnic lines. Cross-linguistic influence

---

6 However, Donaldson (1991: 279) notes that *dat* omission was common in Middle Dutch, without a change of SOV order in the subordinate clause.
from other languages can be seen even more strikingly in the characteristic features of these ethnic varieties. For reasons of space I shall focus on the majority dialect of English in the country – that of Black South Africans (henceforth BSAE). The English of Coloured people will be discussed under code-switching in section 7; on South African Indian English, see Mesthrie (1992).

Black English has now bifurcated into two: (a) a traditional L2 variety and (b) crossover varieties showing considerable overlaps with White SAE, see Mesthrie (2010). Given the social and geographical segregation of South Africa’s race groups in the past it is not surprising that traditional BSAE shows considerable influence from Bantu languages in its phonetics and syntax. It may even be uncontroversial to suggest that traditional BSAE is in many ways phonetically closer to a Bantu language than to White SAE. The characteristics of this basilectal variety may be summarised briefly as follows, based on the work of Hundleby (1964), Wissing (2002) and Van Rooy (2004). It is well known that vowel length tends not to be differentiated in BSAE, and that schwa is rare (mainly occurring in words with more than three syllables). Where vowels are lengthened it is often a phonological effect, the penultimate syllable being lengthened by convention, unless the final syllable is super-heavy (Van Rooy 2004:951). However, it appears to me that in many instances in my data lengthening itself is a secondary effect based on high tone realisations. Thus seventy in is often realised as [sévé:nti]. Tone as a phonological effect is thus present in the basilectal BSAE feature pool. Final devoicing of consonants is common, and combined with other rules can produce forms that differ radically from L1 patterns: in the interviews a word like cards ([kʰats]) sounded identical to cuts ([kʰats]) and is only disambiguated by context. There is a potential four-way ambiguity between word list sets like seed, Sid, seat and sit (as [sid] or [sit], with ‘half-long’ vowels). Regarding the basic five vowel system, unlike other parts of L2 English in Africa BSAE tends to split TRAP into two. With monosyllables (trap, have, act, maths, that, thanks) the most common realisation is [e], but in polysyllables [a] is often used as in salary [salari], advanced [advanst], adamant [adamant] etc. It is possible that the [e] realisations within this set are based on front vowel raising from broader varieties of White SAE (on which see Lass and Wright 1985).

The syntactic features of BSAE have been studied at length by De Klerk (2006) and Botha (2012), with shorter accounts by inter alia Gough (1996), Mesthrie (2006) and Van Rooy (2006). These accounts leave little room for doubting the enormous influence of Bantu languages on traditional BSAE syntax, though some effects may well come from more general second language acquisition processes. Thus Mesthrie (2014) speaks of a robust living substratum for traditional BSAE. This variety contains a number of features many of which are attested in other parts of Africa, and in second language varieties of English the world over:

a) Generalised modal form can be able (for ‘can’ in 24 below)
b) A tendency to use be + -ing progressives where standard English usually disallows it in (25), see Van Rooy (2006).
c) Resumptive pronouns in relative clauses in (26), see De Klerk (2006:150); Mesthrie (2004)
d) Appositional pronouns with topicalised NPs in (27), see Mesthrie (1997)
e) A tendency towards tense neutralization in subordinate clauses in (30)
f) A general tendency not to delete structural elements like infinitive to in (28), see Mesthrie (2006)
g) The possibility of inserting elements like *that* which are disallowed in the standard in sentences like (28).

24. I had to go with a bowl in the street and ask for porridge, so that my mother **can** **be** able to cook.

25. The one I’m **having** presently is a temporary post.

26. Students discovered that the type of education they are trying to give **it** to us...

27. Most of my friends that I grew up with **they** are working at garages....

28. Even my friends were asking me, “Why do you let your child **to** speak Zulu?”

29. As you know **that**, times were very hard then.

30. She made it a point that she **fend** for us... (= “always fended for us”)

31. Although he was so young, **but** he did not fear the authorities.

Mesthrie (2006) points to a general ‘anti-deletion’ tendency in mesolectal BSAE which has three subtypes, showing the following general predilections:

a) to “undelete” elements like **to** in sentence (28) above

b) not to delete elements (e.g. very little pro-drop or copula deletion)

c) to add elements not found in standard English like **but** in the second clause after a conjunction like although in the first clause, as in (31).

This cluster of properties covers a very large portion of the miscellaneous lists of features found in previous studies, e.g. Gough (1996). It brings about a broad typological property which might be stated as follows: “have as few empty nodes on the surface as possible.” I believe that this is a valid characterisation of the agglutinating Bantu languages like Xhosa and Zulu and hence suggests a great deal of syntactic convergence, at least making the L2 English like the substrate.

The reverse phenomenon – influence of English over Bantu language syntax - is perhaps too recent to be prominent, though code-switching effects are treated briefly in the next section. One feature of ongoing convergence relates to English influence over the placement of loanwords into a particular word class in, for example, Zulu. In previous times loanwords were put into a Zulu grammatical class on the basis of its semantics or phonetics. Thus isi-Ngisi ‘English’ is assigned to class 7 with prefix isi- on the basis of its semantics (“language”). And u-tisha ‘teacher’ is assigned to class 1a for “human beings”. Class 5 (with prefix i-) is a particularly common class into which loanwords are placed since it denotes inanimates: hence i-shalofu ‘shelf’. In contrast um-shini ‘machine’, is assigned to class 3 on the basis of the initial consonant ‘m’, which is given a vowel to bring it in line with Zulu phonological and morphological patterns. The noun is treated as a regular member of class 3 inanimates and assigned a plural (in class 4) hence imi-shini ‘machines’. Thus the English word has been resegmented with **m** treated as part of the noun class prefix, and “-achine” or “-chine” as part of the stem. Ngcobo (2013:32) shows how such resegmentation of new loanwords is currently avoided (or resisted) by younger educated speakers, who prefer to maintain the phonological structure of the English root. Thus a common new borrowing like *iselulafoni* ‘cell phone’ is put into class 7 by some speakers, who treat the prefix as **is(i)**-.
and keeping se- as part of the stem (as in English). This ideological practice (signifying a need to follow the segmentation of the prestige language) increases the degree of convergence between the borrowing and donor languages.

8 Phase three: Code-switching and mixing

The development of fluent bilingualism in urban centres worldwide often results in code-switching and mixing, and South Africa has proved no exception. Urbanisation resulting from colonisation and rapid industrialisation after the discovery of gold and diamonds in the late nineteenth century has brought together many languages which would have been previously socially and regionally distinct. Two case studies below will show that code-switching is not just a ‘performance’ phenomenon that juxtaposes two or more codes temporarily. Rather it leaves its mark on those codes when used in subsequent monolingual contexts, leading to restructuring and convergence. The first study continues the theme of convergence between the two closely related languages, English and Afrikaans – this time among mainly “Coloured” people in District Six, Cape Town as studied by K. McCormick (2002). Convergence between Afrikaans and English reaches its zenith amongst people placed on the wrong side of South Africa’s apartheid divide, who consequently had little need of heavy investment in the standard varieties of the two languages. At the time of McCormick’s research in the 1980s and 1990s using standard Afrikaans would have been inappropriate as it was associated with the very officialdom that promoted racial segregation, inequality and a disavowal of the Coloured people, despite their part-European heritage. At the same time using too standard a variety of English would have signalled disloyalty to the community and its solidarity ethic. English was stiff and starchy, standing for king, country and empire rather than the give-and-take of ordinary community life. Given this ideology, there was an openness to code-switching and mixing with the mixed code being used not just in the home and street, but in public domains like meetings. McCormick (2002:230) lists two types of convergence between the two codes. The first shows Afrikaans being influenced by English. Although colloquial and standard Afrikaans of White speakers has been shown in section 5 to undergo word order and other changes vis a vis English, McCormick documents a further degree of changes in the variety spoken by Coloured people in District Six, Cape Town.

a) A tendency to use auxiliary and verb together in main clauses with VO order, rather than a clause-final verb.
b) Occasional use of non-final verbs in subordinate clauses introduced by wat ‘which/ who/what.’
c) Use of demonstrative adjective daai ‘that’ (from standard form daardie) as demonstrative pronoun ‘that’.

More often, it is the Afrikaans substrate that influences the English L2 among Coloured speakers:

a) Use of an invariant tag nè for ‘isn’t it, not so, don’t you’ etc.

7 Ngcobo (2013:27) calls this class 9a, a sub-class of 9, which involves loanwords without a nasal in the prefix.
b) Frequent use of zero in place of third person sg. –s.
c) Frequent use of singular auxiliary *is* for *are* with third person plural subjects.
d) Use of unstressed *do* in past tense, making it somewhat parallel to Afrikaans past tense usage with *het* + *ge-*V stem.
e) Deletion of future auxiliary will, e.g. *I take it later*, in line with one option in Afrikaans *Ek vat haar later*.
f) Tendency to delete –ly as an adverbial suffix.
g) Placement of adverbials before object – e.g. *My mommy don’t make everyday cake*.
h) Placement of time adverbials before place adverbials – e.g. *I’m going now home*.
i) A tendency to use sg. demonstrative forms in place of plurals - e.g. *That is other people’s constitutions*.
j) Use of English prepositions with influence from Afrikaans semantics – e.g. *He cry by the doctor*.
k) Calques based on Afrikaans compounding or phrasal verb patterns –e.g. *to scold someone out* (Afrikaans *skel uit* ‘scold out’); *to throw someone wet* ‘to spray, splash water on someone’ (Afrikaans *nat te spat/gooi*) .
l) Use of *for* with objects after verbs like *tell* (e.g. *I did tell for him that....*).

As McCormick acknowledges, it is hard to be sure whether it is substrate influence at work or general simplification in second language acquisition as many of the above syncretisms are widely reported in research into World Englishes. Of the above only (e), (g), (h), (k) and (l) can unambiguously be related to Afrikaans influence. Nevertheless the differences between the set (a) to (l) and standard English do make the local variety of English closer to the local variety of Afrikaans. A more convincing case for substrate influence is that - apart from feature (i) above - almost none of the features cited above occur in the second-language variety, Black South African English, as summarised in section 6. A few other features from this set occur in Black South African English but only if speakers are in frequent contact with Coloured speakers, especially (a) and (d). Here the pathway of influence is Afrikaans > Afrikaans-English of Coloured bilinguals > Xhosa-English of Black bilinguals.

The second illustration of the convergence fostered by code-switching comes from contact between Bantu languages and English. Once again the pervasive use of English in the education system has had an impact upon the everyday speech of young bilinguals. Thipa (1989:108) gives an example of code-switching in an urban Xhosa educational context:

32. Loo *lecturer* i-clear kuba i-ya-read-a.
   DEM lecturer he-be clear because he-PROG-read-FV\(^8\)
   ‘This lecturer is clear-headed because he reads’

One of the outcomes of code-switching is a partial restructuring of the African languages to include not just lexical elements as in (32) above but structural elements as well. In

---

\(^8\) Abbreviations: AUX – auxiliary; DEM – demonstrative; PERS – persistative; PROG – progressive; FV – final vowel; PERS – persistative; PL – plural; INF – infinitive; INSTR – instrumental; SC – subject concord; SG – singular. OC – object concord.
particular, the use of logical connectors and discourse markers like *and, but, if, even if, I mean, so, because, even though* are very salient amongst students speaking Xhosa in Cape Town (see Deumert et al. 2006). Also salient is the use of the free negative form *never* and the question word *why*. Interestingly, even *busy* as an adjective (e.g. 33) or as a semi-auxiliary denoting ‘in the process of’ (e.g. 19, repeated as 34 below) has to be included here as shown by the following two examples from Urban Zulu and its mixed slang register Iscamtho:

33. be-se-ba-ya-ku-hleka nga-le taal ma u-loko u-busy
    PAST-PERS-3PL-PROG-2.OC-laugh INSTR-this language when 2SG.SC-AUX2 SG. SC-busy
    ‘...they would be laughing at you in this language when you’re busy’
    (Slabbert and Finlayson 2000: 125)
34. A-wu-bon-i si-bhizi si-ya-dla?
    NEG-2SG-see-NEG 1PL.SC-busy 1PL.SC-PROG-eat
    ‘Don’t you see we are busy eating?’ (Aycard 2014: 169)

These stand out in contrast to other bound grammatical categories which do not easily admit English grammatical intrusions: noun class prefixes, concord markers on adjectives and demonstratives, auxiliaries and verb endings, copulatives etc.

Finally, intimate bilingualism between African languages in Soweto and neighbouring areas have resulted in a give-and-take between Sotho and Zulu especially. Gunnink (2012) records the interchangeability of the Zulu and Sotho locatives amongst fluent urban bilinguals: Zulu *e* + noun stem + *ini* and Sotho *ko/mol/o* + noun stem + *-eng*. While this interchangeability does not occur outside this highly multilingual area, Aycard (2014) suggests that in informal youth language registers of Soweto this kind of language crossing is exploited to the full. He proposes that this mixture is in the process of becoming conventionalised and even stabilising as a new language outside the special context of male youth informal street speech. Some English structural elements (the logical connectors cited above, and elements like *never, why* and *busy*) play a key role in this new conventionalisation.

### 9 Conclusion: Overall areal convergence

Several generalisations can be made from the kinds of borrowing illustrated in this chapter. All layers of language are amenable to convergence with other languages in a multilingual milieu. Vocabulary diffusion is as old as language contact in South Africa, and continues to be an active force, with a vengeance in the age of code-switching. As far as phonology is concerned, processes of second language acquisition and language shift can result in the transfer of salient sounds from one language to another, the large-scale incorporation of clicks into Bantu languages being a particularly striking case in point. The phonology of Black South African English shows how far vowel systems can be influenced by speakers’ mother tongues. Syntactic features are also shared in the modern South African linguistic area, via robust processes of calquing, i.e. relexification of

---

9 Copulatives are prefixes whose equivalent in English are forms of copular *be*. 
selected salient syntactic patterns of the \( L_1 \) in a speaker’s \( L_2 \). The influence of English on the syntax of Afrikaans and Bantu languages has been a theme running through this chapter. Educated bilinguals seem to be dancing to an English tune in respect of changes in word order in Afrikaans and the use of English logical connectors in languages like Xhosa.

We might ask what areas of language are impermeable to contact. Morphological influence is perhaps less noticeable, and bound morphemes are rarely transferred, the only examples cited in the literature are forms like –\( rha \) (an adjective forming suffix, equivalent to English ‘-ish’ - Branford and Claughton 2002:202), –\( s \), –\( she \) (a feminine suffix in generalised use) and –\( sholo \) ‘bad, coarse’ (Herbert 2002:302) from Khoe-San languages into Bantu languages. In this chapter the use of the Afrikaans diminutive –\( kie \) in colloquial English words was noted.

Herbert (2002:311) noted that ‘the stability of the Bantu noun classes and, especially, the system of concordial agreement in language contact situations and in Bantu languages used as lingue franche, is remarkable’. It appears to me that the VP is particularly robust in resisting the incursions of code-switching (of both lexical and morphological forms).

Finally, the development of an incipient South African sprachbund can be seen from the way in which some features occur in different bilingual layers, and may even survive not one but two language shifts. Thus reduplication from Malay survived the shift to Afrikaans, and in some communities is surviving a subsequent shift from Afrikaans to English.

References


Emenaua, Murray. 1956. India as a Linguistic Area. Language. 33:3-16.


19  Jharkhand as a “linguistic area” Language contact between Indo-Aryan and Munda in eastern-central South Asia*

John Peterson

1  Introduction

This study presents an overview of linguistic convergences between the Munda and Indo-Aryan languages of eastern-central India and Nepal, with special reference to the Indian state of Jharkhand, which can be considered the “centroid” of a rather loosely defined area of language convergence stretching eastward into Bangladesh, northward through the state of Bihar into Nepal, and southward into the state of Orissa. The convergences to be discussed here encompass both large-scale lexical borrowing from Indo-Aryan into Munda but also a number of convergences of a grammatical nature whose origin is not always known.

2  The languages of Jharkhand and beyond

The state of Jharkhand is home to languages from three families – Indo-Aryan (henceforth “IA”, Indo-European), Munda (Austro-Asiatic) and Dravidian. The present study focuses on contact between Indo-Aryan and Munda, as these have been in contact with one another for many centuries, if not millenia, whereas the Dravidian languages of the region, Kurukh and Malto, appear to be relatively new to this region. Hence we will only refer to these latter two languages on occasion. Since many of the languages of this region still await adequate documentation, we focus our attention on the larger, better described IA and Munda languages of the region, drawing upon published works and data collected during several research trips to Jharkhand. Map 1 shows our area of study.

* The present study summarizes and expands upon an earlier study (Peterson, 2010) and is largely based on data obtained during several visits to Jharkhand. I would like to thank the German Research Foundation (Deutsche Forschungsgemeinschaft) for generous grants which made three of these trips possible (PE 872/4-1, PE 872/1-1, PE 872/1-2). I would also like to thank the Department of Tribal and Regional Languages at Ranchi University, and especially Dr. Ganes Murmu, for their unwavering support over the years. Thanks also go to Ramawatar Yadav for his invaluable insights into Maithili grammar and to both him and Alena Witzlack-Makarevich for their comments on an earlier version of this article. Of course, I alone am responsible for any oversights and errors.
The following gives a brief overview of the major IA and Munda languages native to the region. All figures for numbers of speakers are from Lewis et al. (2013) and, where applicable, refer to the total number of speakers or at least to those of India and Nepal combined.

1) Indo-Aryan
   a) Sadri – the *lingua franca* for much of Jharkhand, especially central and western Jharkhand, by speakers of various Munda and Dravidian languages. Spoken by 3,290,490 people. A number of other, smaller languages, such as Kurmali and Panch Parganiya, are also spoken in Jharkhand and are closely related to Sadri, but unfortunately little work has been conducted on these since Grierson (1903), hence they will only play a marginal role in this study.
   b) Maithili, spoken in Bihar and southern and eastern Nepal by an estimated 32,000,000 people, and the closely related Magahi, spoken to the south in Bihar by an estimated 14,000,000 people.
   c) Oriya, the state language of Orissa (32,137,290 speakers).
   d) Bhojpuri, spoken primarily in Uttar Pradesh, western Bihar and northward into Nepal, with ca. 39,510,000 speakers.

---

1 The original map, which has been modified here and in Map 2 somewhat, is copyrighted by sa PlaneMad/Wikipedia and can be downloaded at the following site [last accessed: 6 August, 2013]:
I am grateful to Arun Ghosh for granting permission to use these maps without the terms of the creative commons share alike license.
e) **Bengali**, spoken in eastern Jharkhand but especially in West Bengal and Bangladesh (193,263,700 speakers).

2) **Munda** – The approximately two dozen Munda languages, which form the western branch of the Austro-Asiatic phylum, belong to one of two groups, which are not mutually intelligible:

a) The **northern group** includes the three largest Munda languages, **Santali**, with 6,218,900 speakers, **Mundari**, with 1,120,280 speakers, and **Ho**, with 1,040,000 speakers. A number of other, much smaller languages belonging to this group are also spoken in or near Jharkhand, such as **Asuri**, **Birhor**, **Korwa**, **Mahali** and **Turi**. As the languages of this group are largely mutually intelligible and as these smaller languages await documentation, we concentrate here on the three larger languages, on which there is considerable published data.

b) The only member of the **southern group** spoken in Jharkhand is **Kharia**, with an estimated 241,580 speakers. Further languages from this group include **Juang**, spoken in central Orissa, as well as a number of other languages spoken in southern Orissa and further to the south in Andhra Pradesh, such as **Sora**, **Gadaba**, **Gorum**, **Remo**, etc. As most of these languages await serious documentation and are outside of our area of study, we concentrate on Kharia from this group.

Map 2 provides an overview of the regions where the larger languages mentioned above are traditionally spoken. As these languages are spoken over extensive areas – for example, through migrations mainly in the nineteenth century, Mundari, Santali, Ho, Kharia and Sadri are now also spoken in eastern Nepal and further to the east – Map 2 provides only a very general indication of the core areas of the respective languages.

Given the number of languages in the region and the fact that many of these have not yet received adequate documentation, the present study lays no claim to being exhaustive. Instead, it builds upon three earlier studies (Abbi, 1997; Osada, 1991; Peterson, 2010) and draws attention to a number of convergences which have so far been noted.
3 Convergence between Indo-Aryan and Munda

In the present section we examine evidence for linguistic convergence in the region of study, beginning in 3.1 with phenomena which are typical of much or even most of South Asia, before dealing in 3.2 with instances of convergence which set this region off from its neighbors.

3.1 Eastern-central South Asia in the larger context

Since Emeneau’s (1956) seminal study, which first brought the topic of linguistic convergence in South Asia to the attention of general linguists, a considerable amount of work has been conducted on the topic of South Asia as a “linguistic area” or Sprachbund and a number of defining criteria have been suggested. In fact, many of the criteria

---

2 See footnote 1.
3 The standard work on this topic is Masica (1976). A more recent overview (in German) is provided by Ebert (2001).
which have been suggested as typical of South Asia are also found in our area. These include, among others, the following:

1. predicate-final word order (SOV). This is of special interest as it is generally assumed that Proto-Munda had a basic SVO order, whereas the Munda languages are all SOV;\(^4\)
2. closely related to this last point is a strong head-final tendency in general;
3. predominantly agglutinative structure or grammatical marking expressed through enclitics;
4. the presence of morphologically marked causatives and, at least in some languages (e.g., Sadri, Kharia, etc.) double causatives;
5. the presence of sequential converbs ("conjunctive participles"), which appear to have been "borrowed" into Munda from IA (cf. Abbi 1997: 140-141; Peterson 2010: 61);
6. the so-called "explicator compound verbs", in which a lexeme denoting the event is followed by one of a small number of "explicator verbs" which derive from lexical verbs and which carry all grammatical marking (cf. e.g., Kharia remaʔ dqothoʔ in (9) or Sadri raı̊h gelAk in (11)). These "explicator verbs" generally denote Aktionsart.
7. the "dative-subject" construction, in which the experiencer appears in a non-nominative case (generally the dative, as the name implies) and the verb agrees in person and number with the stimulus, which appears in the nominative.

There are also a number of traits shared by most languages of this region which are typical of larger portions of eastern South Asia:

**Accusative alignment** – In all languages of this region we find a nominative/accusative alignment pattern with respect to case marking, and predicates in all these languages mark for the person/number/honorific status of the subject (S/A). In some languages, the predicate also marks for the object (P) and certain non-arguments as well (e.g., possessors), such as Maithili and North Munda in general. We find no signs of morphological ergativity in this region, unlike in the IA languages to the west, or in Nepali and the Tibeto-Burman languages to the north, which are either ergative or split-ergative languages.

**Lack of grammatical gender as a productive category**, in contrast to the IA languages to the west and southwest, which possess either two (e.g., Hindi) or three genders (e.g., Marathi). Other IA languages, such as Nepali to the north or our area, do possess two genders (which I consider “feminine” and “non-feminine”) but generally only in formal registers, whereas less formal registers tend to have no gender distinction (cf. Genetti, 1999).

**Numeral classifiers** (cf. Abbi 1997: 142; Osada 1991: 106-108) In addition to their use with numerals, these classifiers also attach in some languages to other units, such as

\(^4\) Cf. Donegan and Stampe (2004) for an account of the process which may have led to these changes in Munda.
demonstratives. However, the use of classifiers varies considerably throughout the region: In many of the IA languages of the region, classifiers may be placed after the noun to denote specificity (Neukom and Patnaik 2003: 24-34). Cf. example (1) from Sadri, where lʌʁki=go refers to a girl who has already been mentioned in the preceding story:

Sadri
(1) u=kʌɾ bad dhan=mʌn=ke lʌʁki=go bʌɾka-l-ʌk
    that=GEN after paddy=PL=OBL girl=CLASS boil-PST-3SG
    sijh-ʌl-ʌk
    boil-PST-3SG
    sukha-l-ʌk ...
    dry(TR)-PST-3SG

   ‘After that, the girl boiled the rice paddy, dried [it] …’

This is not possible in many languages of the region such as Maithili (Ramawatar Yadav, p.c.) or Kharia: Although both of these languages do have classifiers, some of which are homophonous with those of Sadri, these have different distributions. In Kharia they attach only to numerals and a few other quantifiers but not to nouns as in (1), whereas in Maithili one classifier, =ʈa, can attach to nouns, but only if the noun is followed by an enclitic focus marker, such as =e in (u) ram=e=ʈa [(that) Ram=FOC=CLASS] ‘that Ram’. Hence the presence of classifiers as a defining trait of a supposed Sprachbund is only valid to the extent that we do not take their distribution into account. Nevertheless, the presence of often homophonous classifiers throughout this region does testify to convergence between these languages, even if there is no one-to-one correspondence.

3.2 Eastern-central South Asia as a convergence area

Of more interest here are those convergences which set this area off from neighboring regions. The following presents a number of such tentative convergences.

3.2.1 The lexicon

To begin with, there has been massive borrowing between the languages of our region, primarily from IA, to which the politically and economically dominant languages belong, into Munda. In fact, as Abbi (1997) notes, the Munda languages have incorporated large numbers of borrowings from IA into their lexica from virtually all areas, including core areas such as numerals, kinship terms, body parts, etc. This has also led to the incorporation of phonemes from IA into the respective inventories. However, as this is virtually always found in situations of prolonged intense contact, we concentrate in the following on other, less typical types of convergence between the two families. Unless otherwise noted, the processes by which these traits have spread cannot be reconstructed.

---

5 Examples given with sources of this type are from my own texts. The plural marker =mʌn in (1) denotes a large amount of paddy.
at present, but their presence in both language families clearly testifies to this prolonged contact.

3.2.2Alienable vs. inalienable possession

Many IA and Munda languages of the region distinguish between inalienable possession (e.g., with body parts and kinship relations), indicated by enclitic marking for the possessor on the possessum itself, and alienable possession, in which the possessor appears in the genitive and precedes the possessum. Cf. (2)-(4).

<table>
<thead>
<tr>
<th>Inalienable</th>
<th>Alienable Kharia (Peterson 2011: 164)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) (aba=ɲ)</td>
<td>(jɲ=aʔ) (khoɾi)</td>
</tr>
<tr>
<td>father=1SG</td>
<td>1SG=GEN village</td>
</tr>
<tr>
<td>‘my father’</td>
<td>‘my village’</td>
</tr>
</tbody>
</table>

\*Santali* (adapted from Neukom 2001: 32)\(^6\)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) (hʒpən=me)</td>
<td>(am=ak'ofak')</td>
</tr>
<tr>
<td>son=2SG</td>
<td>2SG=GEN house</td>
</tr>
<tr>
<td>‘your son’</td>
<td>‘your house’</td>
</tr>
</tbody>
</table>

\*Sadri* (author’s own data)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(4) (bhəji=har=mən)</td>
<td>(u=mən=Ak) (jəmin)</td>
</tr>
<tr>
<td>sister.in.law=3POSS=PL</td>
<td>3=PL=GEN land</td>
</tr>
<tr>
<td>‘his/her sisters-in-law’</td>
<td>‘their land’</td>
</tr>
</tbody>
</table>

As the IA languages of central India do not otherwise seem to have this distinction, which is found in virtually all Munda languages, it is argued in Peterson (2010: 63-64) that the use of \(=\)har as a marker of inalienable possession in Sadri, which has this distinction only with third-person possessors, is the result of the contact-induced re-analysis of the original postnominal use of \(=\)har as a marker of specificity, like \(=\go\) in example (1), to denote inalienable possession. Thus, the category of inalienable possession in Sadri has most likely been “borrowed” from Munda, although it is expressed by native IA morphology.\(^7\)

3.2.3‘start’ vs. ‘keep on’

\(^6\) Glosses and translations from other sources have occasionally been adapted to the conventions of this study. These modifications do not directly affect our discussion and only serve to unify the presentation of the data.

\(^7\) That not only such categories but also the markers themselves can be “borrowed” is shown by the following example, from Peterson (2010: 64): Compare Sadri \(ke\ ‘who?’\) and its alternative form \(ke=har\ [who=3POSS] ‘who?’\) (with no apparent semantic difference) with Kharia \(ber\ ‘who?’\) and its alternative form \(behar < *ber=har\ ‘who?’\) (also with no apparent semantic difference). Whatever the exact semantics of these complex forms may have been, which cannot be reconstructed at present, the Kharia form is clearly motivated by the Sadri form.
In many IA and Munda languages of the region, one and the same morpheme can have both an inceptive interpretation and either a durative interpretation (‘keep on’) or that of general imperfectivity.\(^8\) For example, the morpheme \(la?\) in Kharia, which derives from Sadri \(lag-\) ‘begin’, can denote the inception of an action or event (5), but much more commonly it denotes general imperfectivity (6).

### Kharia

(5) \(\text{hobne}=\text{te}=\text{ga} \quad \text{ubar} \quad \text{kole}\? \quad \text{kun}=\text{tu} \quad \text{jal}=\text{te}\)

\(\text{ajhe}=\text{kon} \quad \text{“tay} \quad \text{tay”} \quad \text{toro}=\text{na} \quad \text{la}\?=\text{ki}=\text{kivar}.\)

‘Meanwhile (= in that much), two baby parrots got caught in the net and began crying “Tay! Tay!!”.’ [BB, 2:33]

(6) \(\ldots \text{kha}=\text{tiya} \quad \text{lebu}=\text{ki} \quad \text{pujapa}=\text{th} \quad \text{karay}=\text{na} \quad \text{la}\?=\text{ki}=\text{may} \ldots\)

Kharia man=PL sacrifice do=INF IPFV=MID.PST=3PL

‘… the Kharia men used to perform sacrifices …’ [AK, 2:6]

Similar comments hold for Mundari, where the morpheme \(-jan\) depicts an event as having been started and/or as ongoing (cf. the discussion in Hoffmann, 1905 [2001]: 183). Preliminary evidence also suggests that this category once existed in Santali but has since been lost (Peterson 2010: 65, fn. 11). Now consider example (7) from Sadri.

### Sadri

(7) \(\text{rait} \quad \text{bhe}=\text{Ak.} \quad \text{Ab} \quad \text{bud}=\text{ha} \quad \text{bud}=\text{niya} \quad \text{bicar}\)

\(\text{kar}=\text{ek} \quad \text{hel}=\text{Ak.} \quad \text{ki} \ldots\)

‘It became night. Now the old man and the old woman began thinking …’

(adapted from Jordan-Horstmann 1969: 129)

Speakers I consulted confirmed that without the first sentence in (7) and without \(\text{Ab}\) ‘now’, (7) could also have the meaning ‘The old man and the old woman kept on thinking’. Similarly, in Bengali the auxiliary verb \(\text{lag-}\) can mean both ‘begin’ and ‘go on / continue to’ in combination with an infinitive (Radice 1994: 240, note 16).

### 3.2.4 ‘from’ and ‘to’

In many IA and Munda languages of this region we find similarities between markers with an ablative and an allative meaning. Where the exact same marker is used for both

---

\(^8\) This ambiguity does not depend on the Aktionsart of the predicate, at least in those languages for which I have been able to check this.
meanings, I refer to this as the “extentional” function, as it refers to the distance extending to or from an event or location. Cf. Bengali <abadhi> [bobdi] ‘since; until; n. limit’ and Sadri le in (8).

**Sadri**

(8) se=khʌn le ḥʌmr=emu=ke cik bʌθaik kʌhʌl
that=TIME EXT 1NSG=PL=OBL Chik Baraik say-PTCP

ja-t=he. aij le.
PASS-IPFV=PRS.3SG today EXT

‘Since that time we are called “Chik Baraik”. Until today.’

In Munda languages of the region, although the ablative and allative are different, they nevertheless often have strong formal similarities, cf. Kharia tay ‘from’ and the allative postposition khoʔay ‘up to’ (< *khoʔ tay [place ABL] ‘from the place’), Santali: ḥɔbic’ ‘up to; until’ and ḥɔbic’ kho ‘from; since’, with the postposition kho, which can also mark the ablative alone, and the cognate Mundari forms ḥɔbiʔ ‘up to; until’ and ḥɔbiʔ kho ‘from; since’ (Ganesh Murmu, p.c.).

3.2.5 Anticipatory categories

Many languages of the region possess an “anticipatory” predicative category denoting that one event is directly followed by another. This is shown for Kharia in (9), where the “explicator verb” (cf. section 3.1) ḏɔth/ḏo délai, which derives from the homophonous form with the meaning ‘take’, denotes that this action is obligatorily followed by another action. As in (9), the following action is typically introduced by ro ‘and’.

**Kharia**

(9) tay raja jhaʔi ... ḏokloʔ remaʔ ḏoʔ=ʔ ro ho=ki=te
then king all meeting call ANTIC=ACT.PST and that=PL=OBL

“mash=te ate jorne=na ayiʔ?j?”
Messiah=OBL where be.born=INF PRS.COP say=ACT.PST and ask=ACT.PST

‘Then the king called all … to a meeting and asked (= said and asked) them “Where is the Messiah to be born?”’


North Munda languages possess a similar category, expressed by non-cognate forms. These are generally referred to as “pluperfect”, “irrealis”, etc. in the respective studies. Consider (10) for Santali, adapted from Neukom (2001: 80-81). Similar categories are also found in Mundari (Osada 2008: 126) and Ho (Deeney 1975: 39-40).

**Santali** (adapted from Neukom 2001: 80f.)

(10) jɛmɔn=e boʔɔ gɔ'-len=a.

tɛmɔn=ge=kin kilop
as=3SG.SUBJ enter TEL-ANTIC:[MID]=IND then=FOC=3DU.SUBJ close

**εςέι**  \( g \hat{\mathcal{T}} \text{-ked}=e=a. \)

close TEL-ACT.PST=3SG.OBJ=IND

‘The moment he got in, they closed and shut him up.’

Many IA languages of the region, such as Sadri (11), have a similar category, generally referred to as the “conditional participle/converb”.

**Sadri**

(11) \( ... \text{jolha } \ u=ke \ p\hat{\text{f}}\hat{\text{r}}\alpha\text{-le } \text{Aur } g\hat{\text{h}}\text{o}\text{f}a \ up\text{Ar} \ b\hat{\text{A}}\hat{\text{f}}\hat{\text{h}}\text{-le} \)  

\( r\hat{\alpha}<i>h \text{ rem}>=\text{LNK} > \text{ge-l-ak} \)  

‘… the Muslim grabbed it and remained seated (= sat down [and] remained) on top of the horse.’  

(Nowrangi 1956: 165)

Cf. similar forms in Bengali (Thompson 2012: 77) and Oriya (Neukom and Patnaik 2003: 251-252). In addition to their use in the protasis of conditional constructions, these forms typically also fulfil a number of other functions, such as habitual situations where one event/situation precedes another; cf. the “temporal when-conditionals” in Bengali (Thompson 2012: 178) and Neukom and Patnaik (2003: 251-252) on Oriya, especially their examples (89), (90) and (95).

Finally, the North Dravidian language Kurukh possesses a similar category, expressed by \( x\hat{a}c\text{c}-\), which as a free morpheme has the meaning ‘rip’, and \( l\hat{\text{i}}\text{ngh}-\) (meaning as an independent morpheme unknown) (Masato Kobayashi, p.c.; Grignard 1924: 147).

As argued in Peterson (2010: 68), this category probably originated in one language (family) in the region and then spread to others, although the details of this process remain obscure. The anticipatory form \(-\text{le}\) found in IA languages originally derives from the past participle marker \(-\text{l}\) followed by the locative marker \(-\text{e}\) (Chatterji 1926: 1004, §736) but is homophonous in most IA languages with the verb stem \( \text{le}-\) ‘take’, which is often used as a “vector verb” to denote Aktionsart. This is important since, as we noted above, the anticipatory marker in Kharia, \( \hat{\text{d}}\hat{\text{o}}\hat{\text{j}}\hat{\text{d}}/\hat{\text{d}}\hat{\text{o}}\hat{\text{th}} \), is homophonous with and derives from the lexeme \( \hat{\text{d}}\hat{\text{o}}\hat{\text{j}}\hat{\text{d}} \) ‘take’. Finally, the fact that in many North Munda languages the form is homophonous with the “conditional converbs” of IA suggests the possibility that the North Munda forms are borrowings from IA.

However, there are problems with this analysis, since there is considerable evidence that \(-\text{le} \) was not borrowed from IA into North Munda, as \(-\text{l}\) is found in many Munda languages to denote the “anterior” or “cislocative” (cf., e.g., Pinnow 1966: 141, §3.2.12.1.2; Anderson 2007, sections 4.1-4.2), suggesting that it derives from Proto-Munda.

In view of the semantic similarities between these categories in the individual languages, the historical development in Kharia, the phonological similarities between the North Munda and IA forms, and the presence of a semantically similar category with
entirely unrelated forms in Kurukh, I suggest that at the very least the languages of this region mutually influenced one another such that forms which may have already been present and which were already semantically similar became semantically even more similar, while at least Kharia and Kurukh developed this category predominantly through contact with neighboring languages.

3.2.6 A new dual category

In Peterson (2010: 68-70) the emergence of a new number distinction in the Sadri pronominal system due to Munda influence is noted: In Sadri we normally find a singular/plural opposition throughout the entire nominal system, whereas the Munda languages of this region have a singular/dual/plural number distinction. Interestingly, a Kharia-L1 speaker of Sadri (c 50 years old) who has been an active speaker of Sadri most of her life cited in interviews Sadri forms for singular, dual and plural pronouns. Table 1 provides by way of example an overview for the first persons of the “traditional”, L1-Sadri forms, the Sadri forms as indicated by this L1-Kharia speaker, and the corresponding Kharia forms. Note that the Kharia L1-system has an obligatory dual category not found in the Sadri-L1 system. This category has in effect been “transferred” by this speaker into Sadri by the use of the numeral *dui(y)* ‘two’ and a classifier, either *=jhʌn* or *=o*, *=jhʌn* being restricted to humans, whereas *=o* is unmarked in this respect. These dual forms do not violate any principles of Sadri and simply mean ‘we both’, but crucially the Kharia-L1 speaker views them as obligatory, whereas Sadri-L1 speakers readily accepted them as correct but not obligatory.

<table>
<thead>
<tr>
<th>Number</th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sadri-L1</strong></td>
<td>mõe</td>
<td>-</td>
<td>hʌmre</td>
</tr>
<tr>
<td><strong>Sadri-L2</strong> (= Kharia-L1)</td>
<td>mõe</td>
<td>hʌmre dai=jhʌn</td>
<td>hʌmre daiy=o</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hʌmre daiy=ʌ</td>
<td></td>
</tr>
<tr>
<td><strong>Kharia</strong></td>
<td>ɨn</td>
<td>anay (INCL)</td>
<td>anin (INCL)</td>
</tr>
<tr>
<td></td>
<td>ɨnjar (EXCL)</td>
<td>ele (EXCL)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Sadri-L1/-L2 and Kharia forms of the first persons (=mʌn ‘PL’)

There is evidence that a similar process is also taking place elsewhere: Ramawatar Yadav (p.c.) informs me of a similar, non-obligatory dual construction in Maithili. As there is at present no further information on this construction – i.e., how and when it entered the language, the speakers who use it, and to what extent various speakers consider it obligatory – this issue requires further study.

3.2.7 The genitive

The genitive provides an especially interesting field for the study of convergences between Munda and IA. To begin with, the genitive markers in most languages of this
area have very similar forms, regardless of genetic affiliation, with the unmarked form \(=(\lambda)k/=(\sigma)k\) in IA or \(=(a)\varphi\) in North Munda and Kharia (for underlying /ak/ or /ag/).\(^9\)

This formal similarity may be coincidental: In most IA languages, not just those of this region, the genitive marker derives from the verb ‘do’, which generally has the form \(k\lambda r\). For example, in Sadri the genitive can take any of the following forms: \(=k\lambda r/=\lambda k/=k\), largely depending on the last segment of the unit they attach to (cf. Jordan-Horstmann 1969: 45 for details) or Maithili, where the form \(=\lambda k/=k\) is found with all nouns.\(^10\) Other forms are found, although these too derive from the same source, e.g., \(=(e)r\) in Bengali (Chatterji 1926: 753-748, §503), which is also found in Sadri and Maithili with certain pronominals (cf. Jordan-Horstmann 1969: 65-66, §4.4.2; Yadav 1996: 90-92; 108-114).

A similar form (allowing for language-specific phonotactic rules) is found in the North-Munda languages and Kharia:\(^11\)

\[(12)\] Santali \(=ak’\) (=[\(a\varphi\)], Mundari \(=a\varphi\), Kharia: \(=a\varphi\)

It is unclear whether this form goes back to Proto-Munda or whether this similarity is the result of contact. Note that Sora and Juang (both South Munda) mark the genitive by \(=/al\) (Anderson and Harrison 2008: 310-311; Patnaik 2008: 515-516), although all other South Munda languages make use of forms beginning with a nasal and followed by a vowel, which differs in the various languages (cf. the data in the contributions in Anderson (ed.), 2008). While the forms in North Munda, Sora, Juang and Kharia suggest that the form \(=/ak/\) derives from Proto-Munda, the similarity to IA cannot be denied and the forms in other South Munda languages call this analysis into question. Further research is necessary.\(^12\)

In the present study we assume that this similarity is coincidental and will not pursue the topic further, but whatever its origin, the genitive marker has assumed further functions in both IA and Munda, to which we now turn.

(i) In many languages of the region, we find a marker which is homophonous with the genitive used to mark person, generally the third person, singular, in various predicative

---

\(^9\) There are a number of alternate forms in North Munda and Kharia, e.g. Santali \(=reak’\) [\(rea\varphi\)], used with an inanimate possessum, \(=\varphi\), \(=na\varphi\), \(=ya\varphi\) and \(=wa\varphi\) in Kharia, and others, but as these are generally based on the underlying form \(=/ak/\), we take this to be the basic form in Kharia and North Munda.

\(^10\) However, Chatterji (1926: 756) questions this etymology for \(=/k/\) in the “Magadhan” languages, including Sadri and Maithili.

\(^11\) In native morphemes in Kharia an underlying /kl/ or /lg/ in the coda is realized as [\(\varphi\)]. A similar situation holds in Mundari (Osada 2008: 101) and Santali (Ghosh 2008: 26; 30).

\(^12\) In view of the fact that Kharia also has a variant of the genitive beginning with \(=/n/\), i.e., \(=/na\varphi\) (cf. fn. 9), found with demonstratives, it is possible that the North Munda, Kharia, Juang and Sora forms are indeed cognate with the remaining South Munda forms and that the \(=/\varphi/\), to be discussed in the following main text, has been added to this unit in Kharia and North Munda through contact with IA, and that the initial \(=/n/\) of this marker has been lost in these languages, at least in most environments. This topic will be taken up in a future study.
categories. For example, in Sadri a form deriving from the genitive marks the third person, singular of the past tense, which in earlier periods was unmarked for person (i.e., “zero-marking”). Table 2 presents the conjugation of kha- ‘eat’ in the past tense by way of example. /k/ also marks the third person singular and plural of the subjunctive in Sadri, which functions as a kind of third person imperative.

<table>
<thead>
<tr>
<th>Person</th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1      | kha-l-o | kha-l-
| 2      | kha-l-e | kha-l-
| 3      | kha-l-
| 1      | kha-l-
| 2      | kha-l-
| 3      | kha-l-

Table 2. The past tense in Sadri: kha- ‘eat’ (Nowrangi 1956: 93)

Other IA languages of the region in which forms deriving from the genitive mark the third person, singular, include Kurmali (Grierson 1903: 149), Sadri Kol (Grierson 1903: 159) and Panch Pargarniya (cf. the texts in Grierson 1903: 168-172). To the north as well, e.g. in Magahi, /k/ forms part of the marking for virtually all forms of transitive verb marking involving the third person, singular, non-honorific in the composite present, past and future (Verma 1991), and a similar situation holds for Maithili (Yadav 1996: 174; 177), although /k/ is often omitted in the spoken language (cf. Bickel al. 1999: 486-487).

In Bengali, /k/ is found in many forms: In addition to the standard language, where it is found in the imperative of the third person, e.g. kor-uk [do-imp.3] ‘let him/her do’, Chatterji (1926: 989-990, §721) notes that this “pleonastic affix” also occurs in the second, middle-grade honorific past and future and non-honorific third person past and future, dialectally in the third person past habitual, and occasionally in the second person imperative, noting that it is considered archaic except in the case of the third person imperative. According to him this element is documented for Bengali at least since the Middle Bengali period, being found in the Śrī-Kṛṣṇa-Kīrttana, whose date he gives as “before 1400” (Chatterji 1926: 132).

Finally, Chatterji (1926: 990-992, §722) notes that this form was also found earlier in Oriya in the third person – but not in the third person, imperative – although it is not found in the modern language, and also in languages further to the east, such as Early Assamese (to the northeast of the region under study here), where it is still found in the third person imperative in the modern language. As no forms are cited for -k in the verbal paradigms of Bhojpuri to the northwest of Jharkhand, this phenomenon appears to have western Bihar and Jharkhand as its western boundary, while its eastern boundary currently extends into West Bengal/Bangladesh (and further into Assam) and southward approximately to Jharkhand’s border with Orissa.

In North Munda languages, forms deriving from the genitive marker =(a)ʔ can also mark the third person, singular on predicative forms. For example, the third person, singular in Mundari is marked on the predicate by the enclitic forms =e/=i, which are the “original” markers for this person, although these can be followed by the glottal stop in some environments, i.e., =eʔ=iʔ (cf. Osada 1992: 64-65). Cf. the data in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>DU</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCL</td>
<td>=n</td>
<td>=lan</td>
<td>=bu</td>
</tr>
<tr>
<td>EXCL</td>
<td>=lɨ</td>
<td>=li</td>
<td>=le</td>
</tr>
<tr>
<td>2</td>
<td>=m</td>
<td>=ben</td>
<td>=pe</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3</td>
<td>=e/</td>
<td>=i/</td>
<td>=ʔ/</td>
</tr>
</tbody>
</table>

Table 3. Mundari enclitic person/number markers (based on Osada 2008: 120)

As argued in Peterson (2010: 75-76) there appears to be a preference to use the enclitic subject marker with the glottal stop in Mundari, while the predicate-internal object marker is marked by the form without the glottal stop. This is illustrated in (13).

Mundari

(13) biŋ coke=ʔ jom-ja-ʔ=i=a.

‘The snake is eating the frog.’

(Osada 2008: 106)

This tendency in Mundari seems to have been generalized in Ho, where the form =eʔ marks third person animate subjects in general while -ʔ, i.e., the form without the glottal stop, marks animate third person objects (Deeney, 1975: 1 and especially 20). In Santali the glottal stop marks the third person, singular, inanimate object in the imperfective and non-past applicative (Neukom 2001: 121-122). With that, the use of the genitive to mark various predicative forms, while recessive in many languages, is still found throughout Jharkhand (with the exception of Kharia), to the east into West Bengal and Bangladesh and beyond into Assam, and as far north as Nepal.

(ii) The genitive also marks focus in various languages of the region, e.g. Sadri in (14), adapted from Jordan-Horstmann (1969; 107), who glosses it as “Ø”.

Sadri

(14) sob bhai kAh-l-ʔ-ẽ: “hã, cA]-a!” aij huã=e=kAr ja-b!”

‘All the brothers said: “Yes, let’s go! Today we will go there (i.e., nowhere else)!”’

Speakers I consulted all confirmed that =kAr in (14) highlights huã ‘there’, although they unanimously considered it non-standard, unlike the focal marker =e. A similar situation is found in Kharia, where the genitive can also mark focus, although here as well speakers consider it non-standard or even incorrect in interviews.

Kharia

13 The transitive marker has two allomorphs, -d and -ʔ (Osada 2008: 119). This -ʔ does not appear to derive from the genitive marker -k.

14 =kAr seems to reinforce =e in (14). It is possible that the two have slightly differing semantics, although this requires further research.
This use of the genitive as a focus marker may be related to the so-called “genitive subjects” in Oriya (Neukom and Patnaik 2003: 59-61). These are complex subjects consisting of two or more coordinated NPs which appear in the genitive. Despite the genitive marking, these units trigger verb agreement just like nominative subjects. Further research is necessary to determine any possible pragmatic differences compared to nominative subjects.

Oriya (adapted from Neukom and Patnaik 2003: 59)

(16) mo=rɔ ta-nkɔ=rɔ e kamɔ=ɪa kɔ-l-u.

1SG=GEN 3SG-OBL=GEN this work-CLASS do-PST-1PL.INCL

‘He and I did this work.’

It is unclear if forms deriving from the genitive can be used to mark focus in other languages, perhaps due to the non-standard status of this construction. There is one possible case is Bengali: Chatterji (1926: 992, §723) writes that the use of /k/ is “very popular” with certain words, such as ekstu(-k) ‘a little’, the plural marker =gula(-k), and dialectally with kintu(-k) ‘but’. Chatterji (1926: 990, §721) also writes that -k may be used with negatives in all persons and tenses in Bengali. Considering the connection in many languages between negation and focus marking, this fits in well with the Sadri and Kharia (and possibly also Oriya) data. Finally, Santali has a focus marker which Neukom (2001: 126-127) refers to as the “intensive” infix -k’-. Although this would seem to be the same phenomenon at first glance, this form is an infix, hence it is not clear if it is related to the other forms and if so, how it became an infix in this language. Further research is necessary.

(i) An especially interesting convergence is found in the copular paradigms, which often contain a /k/~/ʔ/ which is homophonous with the genitive. Table 4 presents by way of example the paradigm of the non-negated existential/locative copula in Santali, whose stem mena is homophonous with the lexeme denoting ‘remain’. Boxes containing forms with [ʔ] are shaded. Similar forms are found in Mundari and Ho, where however the glottal stop has spread to ALL persons, including the first and second persons, singular, so

---

15 “NPs” in Kharia are marked only once for case, hence =aʔ on choʔdd=aʔ in (15) is not case agreement.

16 For example, the neutralization with negation in English of the categories “simple present” I go and “emphatic present” I do go, both of which are negated by the negative form of the “emphatic” construction, i.e., I don’t go.
that /ʔ/ is now part of the stem of the copula in these two languages (cf. Osada 1992: 118; Deeney 1975: 47-48).

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 INCL</td>
<td>mena-ʔ=ɲ=a</td>
<td>mena-ʔ=laŋ=a</td>
<td>mena-ʔ=bon=a</td>
</tr>
<tr>
<td>1 EXCL</td>
<td>mena-ʔ=ɲ=a</td>
<td>mena-ʔ=liɲ=a</td>
<td>mena-ʔ=le=a</td>
</tr>
<tr>
<td>2</td>
<td>mena=m=a</td>
<td>mena-ʔ=ben=a</td>
<td>mena-ʔ=pe=a</td>
</tr>
<tr>
<td>3 ANIMATE</td>
<td>mena=e=a</td>
<td>mena-ʔ=kin=a</td>
<td>mena-ʔ=ko=a</td>
</tr>
<tr>
<td>3 INANIMATE</td>
<td>mena-ʔ= a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The Santali existential/locative copula (non-negated) (adapted from Neukom 2001: 168)

In Peterson (2010) it is suggested that /ʔ/ in all three languages has spread throughout the respective paradigm via analogical leveling from the third person, singular, inanimate to these other positions, whereas in Santali the first, second, and the third person, singular, animate have resisted this change, perhaps due to a higher frequency of these three forms. We assume that this /ʔ/ derives from the genitive and that its original function was to mark the third person, singular, probably as a nominalizer. 17 This can be explained as follows.

Simplifying somewhat, in Santali, as in North Munda in general, the enclitic subject of a transitive predicate (A) attaches either to the last element preceding the predicate or, if the predicate is the only word in the clause, to the predicate itself, and the marking of the object (P) is marked predicate-internally, before the indicative marking, which precedes the marking for A, as in (17).

Santali

(17) `dal=ɪɲ=\textit{a}=\textit{e}
\textit{beat}={\textit{SG.OBJ}}={\textit{IND}}={\textit{3SG.SUBJ}}
\textit{he will hit me}'

The subject of an intransitive predicate (S) is normally marked similarly to that of A, as in (18).

Santali

(18) `hij-oʔ=\textit{a}=\textit{e}
\textit{come-MID}={\textit{IND}}={\textit{3SG.SUBJ}}
\textit{he will come}'

However, there are a small number of intransitive predicates (the “unaccusatives”) which mark S at the “object” position.

17 Note that =ak’ [aʔ] also functions as a nominalizer in Santali, cf. \textit{rɔrən}=ak’ [dry-PST:MID=NML:INAN] ‘what has become dry’ (adapted from Neukom 2001: 58), from which the other functions such as person and focus marking undoubtedly originated. These will be dealt with in detail in a future study.
In light of this, I argue in Peterson (2010: 76-78) that the paradigm for Santali *mena* ‘exist; COP; remain’ was originally that of the class illustrated in (19): Recall from the discussion in the preceding pages that the genitive often marks the third person on verb forms in the languages of the area, so that we would expect it to appear at the object position if used with predicates which have the same marking pattern as *rɛŋɛc* in (19). This would result in the form *mena=ʔ=a* [remain=3SG.OBJ=IND], which is indeed the form we find in the third person, singular, inanimate in Table 4.

This analysis is supported by the suppletive negative paradigm of this copula, shown in Table 5. Here the stem *bən-* is followed by the marker of the middle voice -*uʔ/-ug, but crucially the third person, singular, animate is marked by the nominalizer =*ic*², used exclusively for animate beings, instead of the usual animate marker of the third person, singular, -*e* (Neukom 2001:170, fn. 72). Note that here as well, subject marking is found in the “object” position.

<table>
<thead>
<tr>
<th>Person</th>
<th>Affirmative</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td>Dual</td>
</tr>
<tr>
<td>1 INCL</td>
<td><em>bən-ug</em>=<em>ŋ=a</em></td>
<td><em>bən-ug</em>=<em>ŋ=a</em></td>
</tr>
<tr>
<td>1 EXCL</td>
<td><em>bən-ug</em>=<em>ŋ=a</em></td>
<td><em>bən-ug</em>=<em>ŋ=a</em></td>
</tr>
<tr>
<td>2 ANIMATE</td>
<td><em>bən-ug</em>=<em>ic</em>²=<em>a</em></td>
<td><em>bən-ug</em>=<em>ic</em>²=<em>a</em></td>
</tr>
<tr>
<td>3 ANIMATE</td>
<td><em>bən-ug</em>=<em>ic</em>²=<em>a</em></td>
<td><em>bən-ug</em>=<em>ic</em>²=<em>a</em></td>
</tr>
</tbody>
</table>

Table 5. The Santali existential/locative copula (negated) (adapted from Neukom 2001:170)

This strongly suggests that /ʔ/ in the non-negative paradigm derives from the genitive/nominalizer, originally to mark the third person, singular, inanimate, which then spread via analogical leveling to other forms – to all forms except the first, second and third, animate, singular in Santali, and to all forms in Mundari and Ho, so that it now forms part of the stem in these two languages.

This marking pattern has spread to a number of IA languages of the area, e.g., Sadri. Consider the four suppletive present-tense copulas in Sadri given in Tables 6 and 7. Boxes containing a form with /k/ are shaded.
Table 6. The identificational copula in Sadri

<table>
<thead>
<tr>
<th>Person</th>
<th>Affirmative</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td>Plural</td>
</tr>
<tr>
<td>1</td>
<td>ahō</td>
<td>ahi</td>
</tr>
<tr>
<td>2</td>
<td>ahis</td>
<td>aha</td>
</tr>
<tr>
<td>3</td>
<td>ahe</td>
<td>ahʌẽ</td>
</tr>
</tbody>
</table>

Table 7. The existential/locative copula in Sadri

The copular forms with the internal /k/ have long puzzled researchers, but in light of the patterns found in North Munda, it is argued in Peterson (2010: 80-81) that the /k/ in these forms, which fulfills no grammatical function, is motivated by the North Munda paradigms and derives from multilingual speakers of Sadri and North Munda. As there is no “object position” in the Sadri verb which this /k/ could occupy, unlike North Munda, this is the most likely explanation. Finally, one form of the Sadri copula, heke, has been borrowed into Kharia, where it is entirely productive, so that a form with an internal /k/ has now found its way from North Munda via IA into South Munda.

/k/ is also very common in copular forms of other IA languages of the area: For example, Tiwari (1960: 178) lists the forms ho-/hokh- (affirmative) and naikhe (negative) for Standard Bhojpuri. However, as Shukla (1981), describing northern varieties of Bhojpuri, cites no forms of the copula with an internal /k/, these forms appear to be restricted to southern Bhojpuri varieties. Chatterji (1926: 683, §431) also includes a number of relevant forms such as Bengali (in modern transliteration) hoibe-k ‘it will be’, nahi-k ‘is not’, and thak-, the latter, like the Santali form mena, also meaning ‘remain, stay’, but also Early Assamese nahi-k-ɔntɔ ‘they are not’, Magahi √hĩ-k, and Maithili chi-k18 and √thik. In another section, he also mentions √ha / √ha-k ‘to be’ for Magahi, among others (Chatterji 1926: 991, §722).

4 Summary and outlook

The IA-Munda contact area provides evidence for a number of convergences between the languages of these two families, although due to the lack of historical data on most languages of the region it is often unclear which phenomena have originated in which family and how they have spread to other languages. Some of these traits are typical of most of South Asia, others are typical of larger areas of eastern South Asia, while still others appear to be restricted to eastern-central South Asia, our area of study.

In addition to massive lexical borrowings from IA into Munda, numerous convergences in the morphosyntax were also found, of which many are restricted to this region. These are summarized in Map 3.19

18 The accuracy of this form was questioned by Ramawatar Yadav (p.c.), a native speaker who is also a scholar of Old Maithili. Further research is required.

19 Ho has been included in the group of languages possessing similarities between ‘from’ and ‘to’ due its otherwise close similarities to Mundari and Santali, although I have not been able to
Map 3. Convergence patterns of the investigated traits

The above map does not include all of the phenomena discussed in section 3: For example, the use of the genitive to mark focus is not indicated in Map 3 as we only have positive data for Sadri and Kharia but no unambiguous data for other languages of the region. Similarly, for lack of data, Map 3 does not contain information on the presence of the dual, which is found throughout Munda but for which we only have very preliminary data on the informal varieties of the IA languages of the region. This is important, since the dual is not found in the standard varieties of these languages, so that data from other registers are required to ascertain to what extent it has spread in this group.

Despite these problems, the data discussed in the previous pages show that it is not possible to demarcate a well-defined “linguistic area” or Sprachbund for this region, as no two “isoforms” are identical: Not surprisingly, Sadri, the lingua franca for much of this region, clearly possesses all of the traits discussed here, but the distribution of these traits in the other languages of the region can vary considerably. Furthermore, we saw evidence that other languages further afield once shared some of the traits investigated here but have since lost them: for example, person marking deriving from the genitive has disappeared entirely from Oriya and seems to be gradually disappearing from North Munda and Maithili as well.

The various phenomena are also not uniform in the languages in which they are found. For example, it was mentioned above (see section 3.1) that virtually all languages of the region possess numeral classifiers, although their distribution can differ considerably in the individual languages. Similar comments hold for the other confirm this. Also, Bhojpuri is only half included among the languages with a /k/~/ʔ/ in at least one copular form as this trait is limited to the southern varieties of this language.
phenomena discussed in the preceding pages as well. In other words, although the languages of our region often possess similar grammatical categories, these categories are just that – similar – and no two languages possess the exact same category with the exact same distribution. Nevertheless, the fact that similar categories are found in so many languages makes it clear that speakers of these languages have long been in close contact, even if there is no one-to-one correspondence between the languages.

Thus, instead of a monolithic, uniform Sprachbund, the traits discussed here combine to define a rather diffuse and constantly changing area of language convergence in eastern-central India and Nepal, which itself is part of an even more diffuse South Asian convergence area – i.e., the notion of South Asia as a homogeneous linguistic area or Sprachbund is far too simplistic and is not supported by the data. Instead, we find a number of convergence areas which partially overlap, so that one area gradually fades into those around it. In our case, it would seem most promising at present to turn our search for further convergences to the north, to the Kiranti (Tibeto-Burman) languages of eastern Nepal, as Ebert (1999) notes a number of interesting convergences between these languages and North Munda which require further study, above all with respect to subordination, although she notes the presence of the “pleonastic $k$”, discussed above in §3.2.7, in Kiranti as well (Ebert, 1999: 397, fn. 19).

There is still much work to be done on the linguistic convergences of this region, and as our knowledge of these languages increases, it is almost certain that some of the phenomena discussed in the preceding pages will eventually have to be removed from the list of convergences but also that others will have be added. To cite just one example, there is a type of reduplication in Kharia in which one or more of the vowels of the stem is replaced by /i/ in the reduplicated form, cf. *gupa* ‘tend (cattle)’ vs. *gupa gupi* ‘finish tending’ or *kayom* ‘talk’ vs. *kayom kiyim* ‘talk something out’ (Peterson 2011: 131). Similar forms are found in Bengali (Thompson 2012: 214), Oriya (Neukom and Patnaik 2003: 132-133) and Sadri (my own data), but to what extent this reduplication is typical of the rest of this area, and to what extent its semantics are similar in the different languages, awaits further study. In the light of evidence such as this, it is highly likely that the number of convergences for this area will increase as research on these languages progresses.

References


Emeneau, Murray B. 1956. India as a linguistic area. *Language* 32:3-16.


## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td>Person</td>
</tr>
<tr>
<td>ABL</td>
<td>ablative</td>
</tr>
<tr>
<td>ACT</td>
<td>active</td>
</tr>
<tr>
<td>ADD</td>
<td>additive focus</td>
</tr>
<tr>
<td>ANTiC</td>
<td>anticipatory</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary</td>
</tr>
<tr>
<td>CLASS</td>
<td>classifier</td>
</tr>
<tr>
<td>CMPL</td>
<td>complementizer</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>DU</td>
<td>dual</td>
</tr>
<tr>
<td>EXT</td>
<td>extensional (case)</td>
</tr>
<tr>
<td>FOC</td>
<td>focus</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>IA</td>
<td>Indo-Aryan</td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
</tr>
<tr>
<td>INCL</td>
<td>inclusive</td>
</tr>
<tr>
<td>IND</td>
<td>indicative</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INGR</td>
<td>ingressive</td>
</tr>
<tr>
<td>INST</td>
<td>instrumental</td>
</tr>
<tr>
<td>IPFV</td>
<td>imperfective</td>
</tr>
<tr>
<td>L1</td>
<td>first/native language</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>LNK</td>
<td>linker</td>
</tr>
<tr>
<td>MID</td>
<td>middle voice</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>NMLZR</td>
<td>nominalizer</td>
</tr>
<tr>
<td>NSG</td>
<td>non-singular</td>
</tr>
<tr>
<td>OBJ</td>
<td>object</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique case</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PERF</td>
<td>perfect</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>PTCP</td>
<td>participle</td>
</tr>
<tr>
<td>SEQ</td>
<td>sequential converb</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SUBJ</td>
<td>subject</td>
</tr>
<tr>
<td>TEL</td>
<td>telicizer</td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
</tr>
<tr>
<td>VOC</td>
<td>vocative</td>
</tr>
</tbody>
</table>
20  Sri Lanka and South India

Umberto Ansaldo

1  Background

The Indian subcontinent, including Sri Lanka, shows traces of three different populations: Austroasiatic, Dravidian and Indo-Aryan (Cavalli-Sforza 1994). Austroasiatic is found today in the form of Munda languages, spoken in Central and Eastern India by several million speakers. Dravidian varieties today occupy the South of India as well as areas of Sri Lanka. Historically however Dravidian populations occupied the whole of the Indian subcontinent; it is likely that they arrived after Austroasiatic and before Indo-Aryan populations (Cavalli-Sforza et al. 1994). Linguistic analysis also seems to suggest that Dravidian populations occupied the area before Indo-Aryan people. Substantial presence of features of the Dravidian type can be found in Old Indo-Aryan, a fact that has prompted scholars to suggest that a language shift took place for speakers of Dravidian languages towards Indo-Aryan varieties (Thomason and Kaufman 1988; Erdosy 1995). This could be explained as a shift prompted by cultural assimilation with a more prestigious neighbor, or it could indicate that differential waves of migration took place. We return to this in section 2.

Indo-Aryan languages are not predominant in South India but they constitute the dominant variety of Sri Lanka in the form of Sinhala. Similarly to South India, Sri Lanka shows traces of three different populations. Besides the Indo-Aryan we find Vedda, aboriginal populations of unclear genetic origins by now extinct; and Dravidian, in the form of Lankan Tamil.

Genetically, one cannot distinguish Dravidians from Indo-Aryans, which points to a deep and prolonged admixture between the two populations. According to Reich et al. (2009) the admixture of Northern and Southern Indian genetic traits could go back as far as 3,500 years. Lankan people are likewise extremely mixed. A study by Kshatria (1995) shows that present-day Sinhalese and Tamils of Sri Lanka are closer to Indian Tamils and South Indian Muslims than they are to Gujaratis and Punjabis of northwest India, Bengalis of northeast India, and Veddas. A genetic tree of 28 South Asian populations published by Cavalli-Sforza (1994: 239) groups South Dravidian, Sinhala and Lambada within the same subgroup, testifying to the close relationship between these otherwise unrelated languages.

2  The influence of Dravidian

It is well known that Emeneau defined a linguistic area as one which includes languages of different families with shared traits that set them apart from other members of those same families. He identified India as such an area, a place where Indo-Aryan and Dravidian resemble each other more than Indo-Aryan resembles other Indo-European languages (Emeneau 1956). Emeneau compiled a rich inventory of features that display

---

1 I would like to thank Kofi Yapko for valuable comments on this paper.
‘Indianization’, i.e. traits that are definitely not Proto-Indo-European and mostly Dravidian in origin. These include:

1. Phonology: retroflex and coronal consonants in contrast with a dental series. The earliest Sanskrit records already display phonemes of this class, which are definitely not Proto-Indo-European and generally not found in other Indo-European languages. In Dravidian, retroflexe consonants, which contrast with dentals, are Proto-Dravidian in origin. This led Emeneau, to conclude that pre-Indo-Aryan and pre-Dravidian bilingualism might have provided the conditions under which pre-Indo-Aryan allophones became redistributed as retroflex phonemes.

2. Grammar (based on earlier work by Bloch): (a) loss of dual: starting from Sanskrit, then paralleled by the rest of the Indo-European domain; (b) loss of infixation; (c) disuse of verbal prefixes, starting in Modern Indo-Aryan and “tied up with the general shift of accent to initial syllables”; (d) absence of prepositions; (e) two themes of personal pronouns, i.e. double stems in personal pronouns in Indo-European languages and the same phenomena found in some Dravidian languages; (f) constructions in which verb stems or non-finite verb forms are strung together in series which are closed by a finite verb form (or other predicate ending). This is a prominent feature in Dravidian and Munda languages, but a feature that makes Sanskrit stand out from other Indo-European languages; (g) past participle constructions based on a nominalised form of a verb (or rather of a predication ending in a verb) followed by a postposition. These are common in Dravidian languages, Buddhist Hybrid Sanskrit, Pali and Asian languages outside India, but not available in Munda languages; e.g. [Marathi] tujhī āī vāryā-pāsūn, [your mother died-since] for ‘since your mother died, since the death of your mother’; (h) echo-word construction, in which a basic word formulated as CVX is followed by an echo-word in which CV is replaced by a morpheme gi- or u- or the like (or C is replaced by m- or the like), and X echoes the X (or VX echoes the VX) of the basic word. The meaning of the echo-word is ‘and the like’; (i) classifiers or quantifiers, a shared feature among all the three families; which spread from Indo-Aryan but which are not an Indo-European phenomenon.

It is clear that Dravidian influence in Indo-Aryan is much stronger than Indo-Aryan influence in Dravidian (Sjober 1992). While Dravidian, especially South Dravidian, borrowed heavily in terms of lexicon, it shows little structural influence from Indo-Aryan. On the other hand Dravidian has a deep structural impact on Indo-Aryan (Krishnamurti 1969). An explanation of this situation was first put forward in Krishnamurti (1969) who suggests that large numbers of Dravidians were responsible for a deep restructuring of Sanskrit – their ‘lingua franca’ – and ended up developing middle Indic. This would suggest that at some point a majority of Indo-Aryan speakers would have been of Dravidian origin. Acquiring varieties of Sanskrit as L2 led to the incorporation of substrate features thus resulting in substantial language change. This analysis is corroborated by Emeneau’s observation that Sanskrit really shows a Dravidian substratum. If this is correct, the evolution of North Indian vernaculars is actually through contact-induced change rather than internal processes of grammaticalization.
The Dravidian family today extends predominantly to the states of Karnataka, Tamil Nadu and Kerala (South Dravidian), Andhra Pradesh and the southern regions of Madhya Pradesh and Orissa (South Central Dravidian) in the South. We also find pockets of Dravidian in Bihar and Nepal (Central Dravidian) and in Pakistan (North Dravidian).

3 Sri Lanka

Within the large linguistic area of India, Gair (1994) proposes a further, southern sub-area, the South-South Asia (SSA) area. This area includes the southern states of India characterized by Dravidian varieties as well as Sri Lanka. The latter houses a number of interesting languages. Besides Lankan Tamil (South Dravidian) and the Indo-Aryan languages Sinhala and Maldivian (Dhivehi), we find Sri Lanka Malay (SLM) and Sri Lanka Portuguese (SLP), both contact languages developed during Portuguese, Dutch and British colonial rule.

Sinhala is somewhat unique among the Indo-Aryan languages as it detached itself before a number of major changes affecting its related varieties in the north (Gair 1998). It is reasonable to believe that it reached its current location around the sixth century BCE. A number of factors contribute to its uniqueness. First of all, Sinhala still retains a strong Indo-Aryan feel, despite having undergone multiple contacts (Gair 1998). Second, it shows influence of indigenous Vedda languages in the lexicon. Though still predominantly Indo-Aryan, the lexicon also shows strong Dravidian influences even in very basic terms and in the kinship system (Gair 1998: 9). Sinhala-Tamil compounds are found as are independent innovations in grammar not found in other neighboring languages. In the gender system, for example, Sinhala has lost the neuter, present in both Indo-Aryan and Dravidian. Instead it has developed a system that distinguishes between animate and inanimate, a distinction also found in SLM (see below). Within animate Sinhala distinguishes between human and animal, and within these two a further distinction between general and feminine is made (Gair 1998: 9).

SLM and SLP are of great historical interest as they are contact languages that developed through the same processes that typically characterize linguistic areas. The original speakers of SLP where either Portuguese or Portuguese slaves from African colonies, as testified for example by the Kaffir population still found in parts of Sri Lanka, who are descendants of African slaves arrived during the Portuguese colonial rule of Sri Lanka. When the Dutch took over Sri Lanka from the Portuguese, these groups started becoming more marginal. The original speakers of SLM, who were Indonesians and Malays deported from the Dutch colonies of Indonesia, grew in number as Dutch power increased. Today there are only a few pockets of SLP left, notably on the East Coast around Battichaloa, in Puttalam (the ‘Kaffirs’) and in the northern area of Trincomalee. SLP is severely endangered. More vital is the situation of SLM communities, both in numbers and in linguistic vitality. Major communities are found in the capital Colombo, Kandy (Center), and Kirinda (South).

In both cases these contact languages have evolved away from their original structural type, to display strong typological affinities of SSA type. They are verb-final, morphologically agglutinative/fusional, phonologically ‘Indian’ and only retain traces of their European/Malay heritage in the lexicon. The list below, selected from Gair (2013)
as the more convincing cases of areal features, show common traits between Dravidian, Sinhala and SLM.²

1. Post-verbal sentence-final question marker when unmarked, question-internal when focused. This feature is found in Sinhala, SLM as well as some Dravidian languages, though not in Tamil (Gair 2013: 171). In Sinhala, for example, (1) is neutral while (2) would be more focussed.

(1) ee minihaa iiye gunapaala-ta salli dunn=da? (Sinhala)
That man yesterday G-DAT money gave=INTERR
‘Was it that man that gave Gunapala money to yesterday?’

(2) ee minihaa=da iiye gunapaala-ta salli dunn-e? (Sinhala)
That man+INTERR yesterday G-DAT money gave-FOC
‘Was it that man that gave Gunapala money to yesterday?’

2. Subordinate clause marking in the form of final verbal affixes or conjunctive forms (Gair 2013: 172). This is an areal feature of South Asia as a whole but it is distinct in SSA in that rather than placing the marker at the left margin, it displays it on the right one. In addition in SSA these markers tend to be multifunctional.

(3) siri koeoema kaalaa gedara giyaa (Sinhala)
S food eat-CP home went
‘Siri ate and went home’

3. Preposed relative clauses occur in Sinhala, Tamil and SLM as the main form.

4. Sentence-final quotative markers ‘say’ are used in all SSA varieties to mark in/direct speech.

5. Nominalization without genetivization via verbal affixes, is found in all varieties, e.g. Sinhala (Gair 2013: 178)

(4) Silva mahattaya ma-ta eeka kiww-a-eka oetta (Sinhala)
Silva gentleman I-DAT that say\PAST-ADJ-EKA truth
‘It is true that Mister Silva said that to me’

Here the subject is in the nominative in an independent sentence. The only difference is in the relativizing verbal affix and the pronoun eka.

6. A variety of complex negative forms including verbal, equational, existential, and subordinate negative verbs, are found in Sinhala, Tamil and SLM.

7. Sentence-final hearsay particles function as evidential clitics.

In addition to these, it has also been noted that the nature of the case system shows metatypic effects that render Sinhala, SL Tamil, SLM and SLP comparable (cf. Bakker 1994; Ansaldo 2009, 2011). Most striking is the fact that all varieties make use of ‘Dative Subjects’, i.e. they use Dative marking for Agents who are perceived as not being particularly in control of an action. As the table below illustrates, the Dative case is used

² Scarcity of recent material on SLP only allow us a partial inclusion of SLP in the discussion that follows.
to mark experiential roles as well as goal, beneficiary and possessive in SLM, SL Tamil and Sinhala.

<table>
<thead>
<tr>
<th>Case</th>
<th>Function</th>
<th>SLM</th>
<th>Sinhala</th>
<th>Tamil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dative</td>
<td>Experiential, Goal, Beneficiary, Possessive</td>
<td>Experiential, Goal, Beneficiary, Possessive</td>
<td>Experiential, Goal, Beneficiary, Possessive</td>
<td></td>
</tr>
<tr>
<td>Nominative</td>
<td>Agent</td>
<td>Agent</td>
<td>Agent</td>
<td></td>
</tr>
<tr>
<td>Accusative</td>
<td>Patient</td>
<td>Patient</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Genitive</td>
<td>Possession</td>
<td>Temporary possession, Locative</td>
<td>Temporary possession, Locative</td>
<td></td>
</tr>
<tr>
<td>Instrumental</td>
<td>Instrumental Source</td>
<td>Instr, Source</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>Comitative</td>
<td>Association</td>
<td>Association</td>
<td>Association</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Case in Sri Lankan Malay, Sinhala and Lankan Tamil

It is interesting to note that typological convergence deeply affects even recent contact languages, as summarized in the table below (based on Bakker 2006: 141-44). Most striking is the fact that both SLM and SLP have developed morphological complexity in the nominal and verbal domain, as well as shifted to a verb-final, right-headed typology.

<table>
<thead>
<tr>
<th></th>
<th>SLP</th>
<th>SLM</th>
<th>Lankan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal case marking</td>
<td>Mostly suffixes (as case markers and postpositions)</td>
<td>All suffixes (as cases and postpositions)</td>
<td>All suffixes (all as cases)</td>
</tr>
<tr>
<td>Discourse particles</td>
<td>Often enclitics, sentence final, or attached to the verb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal categories</td>
<td>Suffixes only</td>
<td>Preverbal, suffixes</td>
<td>Preverbal</td>
</tr>
</tbody>
</table>

Table 2. Lankan features of recent contact languages in Sri Lanka

4 Multilingualism and language contact

It is clear that prolonged contact between unrelated languages has promoted a number of significant common developments in the region. The scenario advocated for the emergence of a SA linguistic area, namely restructuring through second language acquisition, can also be applied to the SSA area. In fact, we need to factor in more than one type of shift and assume different phases of shift to reconstruct what happened in India and Sri Lanka.

The earliest type of contact in Sri Lanka, not considering the aboriginal Vedda languages, was that which occurred between South Dravidian and Sinhala. It seems plausible to assume prolonged contact between these two populations as well as a high degree of bilingualism. This explains why Sinhala looks deeply South Dravidian for an

---

3 Based on the variety of SLM spoken in Kirinda, in the South of Sri Lanka. For a more detailed account of case in the Upcountry variety of SLM (Kandy), see Nordhoff (2009: 583).
Indo-Aryan language. There is corroboration in genetic findings. The study of genetic admixture in Sri Lankan populations reveals that the Sinhalese of Sri Lanka have a higher contribution from Tamils of southern India (69.86% +/- 0.61) compared with the Bengalis of northeast India (25.41% +/- 0.51), whereas the Tamils of Sri Lanka have received a higher contribution from the Sinhalese of Sri Lanka (55.20% +/- 9.47) compared with the Tamils of India (16.63% +/- 8.73) (Kshatriya 1995). The difference between Sri Lanka and India is that here we are most likely looking at a more balanced ratio of populations. While for India the suggestion is of an Indo-Aryan elite whose language was restructured by large Dravidian populations acquiring it as L2, in Sri Lanka the Dravidian population would have been equal or even inferior in numbers to the Sinhalese.

Just as the Tamils of Sri Lanka are closer to the Sinhalese because they were always in close proximity to each other historically, linguistically, and culturally, the Malay and Portuguese of Sri Lanka, despite being more recent arrivals, have undergone convergence to Lankan language and culture. In the last 400 years, speakers of European (Portuguese) and Austronesian (Malay) varieties came into contact with Lankan languages. Here two processes of second language acquisition need to be invoked for the strong typological shift that these languages have undergone. As we saw, though lexically still related to its Malay lexifier, SLM today is clearly a Lankan language in its OV constituent order, its noun phrase and large parts of the VP domain (see Ansaldo 2009). In SLP we also note basic OV structure, postpositions, sentence final particles to mark quotatives, as well as clear Lankan influences in the case system (Smith 1979a, b).

In his analysis of SLM as a product of metatypy (see Ross 2006), Ansaldo (2011) offers a scenario in which bilinguals lead the way in the restructuring process, modelling the vernacular grammar on L2 models. In many multilingual communities around the world, and especially where multilingualism is not institutionally supported through school, education etc., multilingual individuals may experience shifts within their multilingual competence. As pointed out in Matras (2010: 66), bilingual individuals can be seen as possessing an enriched linguistic system which they are capable of appropriately adapting to the context in which they function. In this sense, language contact phenomena can be seen as function-driven interferences that speakers are subject to in goal-oriented communicative interaction. Convergence thus emerges as a process that “offers speakers the opportunity to accommodate and generalize and yet still hold on to a mental demarcation between subsets of word forms within their repertoire” (Matras 2010: 76).

This often goes hand in hand with a process of attrition of the heritage language. Following Thomason and Kaufman (1988), a gradual abandonment of ancestral languages (AL) happens in the transition from monolingualism in AL to multilingualism in L2/L3. This can occur typically in (a) minority groups, (b) colonization scenarios, and (c) nation-expansion processes. In this transition process L2/3-dominant individuals lead the change through intense code-mixing, structural and lexical transfer. As noted in Haugen (1978: 37), the speaker “dismantles and reorders the language he already knows”. One common motivation is the AL speakers’ increasing accommodation to L2/3 under strong social and structural pressure through increased multilingual competence (Ansaldo 2006, 2007; Ansaldo and Lim 2006). In rare contexts where AL is vital, a partial maintenance is indeed possible (e.g. Anglo-Romani, Thomason and Kaufman 1988), typically in cases where the change is gradual.
The type of contact-induced change that we see in this region leads to the emergence of complex structures rather than simplification. While this may be related to a substantial role of young learners in the restructuring, it is impossible not to note the role of the typological ecology in which this contact takes place. In typical Creole scenarios we often note simplification effects, as exemplified by the fact that the inflectional morphology of the lexifier is typically absent in the restructured vernacular, the Creole. This, according to Trudgill (2011) can be explained by the fact that adults were mostly responsible for the restructuring, and adults are allegedly prone to simplification. On the other hand, had children been more involved, complexification might have taken place. Following this line, we might be led to suggest that the varieties discussed in this chapter would have been more influenced by young bilingual/multilingual acquisition than adult acquisition, hence the morphological complexification observed.

I think there is a problem here. As has been noted more than once, it is difficult to imagine a scenario where only children, or only adults, affect the process of language change in a population. While adults and children can be separated in the artificial confines of a linguistic study and we can speak of adult vs. child acquisition, in the real world this separation is impossible. I take it that both adults and children must have been communicating with each other while the restructuring was taking place. I therefore find it extremely difficult to imagine that, ‘naturally’, adults would be gifted with simplification tendencies while children would be gifted with the opposite, complexification tendency. An alternative possibility emerges here to explain what we perceive as simplification and complexification processes.

5 The power of typology

By and large in the region under discussion we observe conflation of non-analytic languages. The type of morphological shift we observe in Indo-Aryan and Dravidian languages is from agglutinative to inflectional. This is a robust feature of the ecology in which language change takes place. When foreign linguistic constructs enter the picture, they are drawn towards this type of morphology. Malay, basically analytic, becomes morphologically more complex and turns into Sri Lanka Malay, an agglutinative language with incipient fusional tendencies (Ansaldo and Nordhoff 2008). The same happens in the case of SLP where, most likely, a somewhat creolized Atlantic Portuguese acquires elaborate morphology. This clearly shows the role that typology plays in contact-induced change. In a different part of the world, the Atlantic, a different ecology is in place. Here large numbers of speakers of strongly isolating languages had to absorb Standard Average European (Haspelmath 1998, 2001; van der Auwera 2011) with mildly fusional features. And this they did, by conforming them to their own, robust isolating types. Thus were born the Atlantic Creoles. Let us try to visualize these two ecologies (albeit with an unavoidable degree of simplification):

India

Majority populations: + INFL (rich morphology) ‡ Outcome: + INFL
Minority populations: - INFL (modest morphology)

Caribbean
Majority populations: 0 INFL (close to zero morphology) †  Outcome: 0 INFL  
Minority populations: - INFL

Figure 1. Typological ecology and contact output

We can see what is happening in both cases: the dominant typological profile wins. It is a matter of type and token frequency, as already argued in Ansaldo (2009) that guides the restructuring. There is no doubt about this in the case of the SSA described above. On the other hand while Atlantic Creoles can be seen as socio-historical continuations of their lexifiers, recent typological work (e.g., Lefebvre ed. 2011) have shown that Creole varieties can be structurally aligned with their substrate languages. In other words, we do not need to invoke the classical tenets of Creole Exceptionalism for the emergence of Atlantic Creoles, just like we do not need to invoke other exceptional explanations for how India is a concentric series of linguistic areas (see Peterson, this volume). Nor do we need to pitch adults vs. children in terms of agency on the restructuring process, which does not make much sense in a naturalistic transmission scenario anyway. It is the interaction of typology and the acquisition of L2 in a rich multilingual environment, in which speakers restructure the new grammar based on their own fundamental competence, which is central. Where morphology is abundant morphology remains, where morphology is scarce morphology recedes.

References


21 The Transeurasian languages

Martine Robbeets

1 Introduction

The present contribution is concerned with the areal concentration of a number of linguistic features in the Transeurasian languages and its historical motivation. The label “Transeurasian” was coined by Johanson and Robbeets (2010: 1-2) with reference to a large group of geographically adjacent languages, traditionally known as “Altaic”, that share a significant number of linguistic properties and include up to five different linguistic families: Japonic, Koreanic, Tungusic, Mongolic, and Turkic. The question whether all similarities between the Transeurasian languages should be accounted for by language contact or whether some are the residue of a common ancestor is one of the most debated issues of historical comparative linguistics (see Robbeets 2005 for an overview of the debate). Since the term ‘linguistic area’ implies that the shared properties are the result of borrowing, I will refrain from a priori attaching it to the Transeurasian region and rely on the concept of “areality” instead, i.e. the geographical concentration of linguistic features, independent of how these features developed historically. Only after evaluating 27 structural features shared across the Transeurasian languages, will I consider how the insights from the data are relevant for historical statements about the way the languages may have come to share these features, considering diffusion, genealogical relationship or an interaction of both factors as possible explanations.

In spite of the strong polarization in the Transeurasian field between so-called “retentionists”, who view the similarities as arising from common descent and “diffusionists”, who view them as arising from areal interaction, detailed characterizations of Transeurasian as a linguistic area are surprisingly rare in the linguistic literature. Poppe (1964) analysed Altaic as a “language type” on the basis of a list of structural parallels shared between Korean, Tungusic, Mongolic and Turkic languages and Rickmeyer (1989) elaborated on this research, adding data from Japanese. Even if these contributions provide an impressive list of shared features, they do not strictly identify Transeurasian as a language area because they do not (i) delimit the language type in relation to its neighbors, (ii) list deviations from the prototypical type in the peripheries, (iii) consider the extent to which the features in question are common or rare across the world as a whole or (iv) attempt to distinguish contact-induced from genealogically motivated features.

In this chapter, I attempt a partial answer to these concerns by providing a typological profile of selected Transeurasian languages, along with their oldest linguistically reliable historical varieties and by comparing this profile with the behavior of languages immediately outside the Transeurasian region. In order to examine external boundaries, I have included adjacent languages to the east (Ainu and Nivkh in the northeast and Rukai in the southeast), to the south (Mandarin Chinese) and to the north (Kolyma Yukaghir, Ket and Eastern Khanty). These languages are taken as horizontal comparative points

---

1 The following sources were consulted for retrieving linguistic data underlying the feature values
representative of surrounding areas such as the Siberian area (Nivkh, Kolyma Yukaghir, Ket, Eastern Khanty) or the Mainland Southeast Asia area (Mandarin) and neighboring families such as Austronesian (Mantauran Rukai), Sino-Tibetan (Mandarin), Yukaghiric (Kolyma Yukaghir), Yeniseic (Ket), Uralic/Ob-Ugric (Eastern Khanty) or Ainuic (isolate Ainu). Although Eastern Khanty can be taken as a representative of the Uralic languages, the main boundary to the west, I have paid less attention to additional western boundaries, excluding sample languages from the Caucasus region or from the Indo-European languages because of the limited space available here.

The vertical comparison points in my analysis consist of a list of 27 features, chosen to maximize positive (+) values for Transeurasian as opposed to neighboring languages. Although all features reflect a certain internal coherence, about half of them (i.e. 13) display deviations from the prototypical type in the peripheries. Where possible, I add an estimation of the degree to which the feature under discussion is common or rare across the world’s languages, relying on the counts in the World Atlas of Language Structures or on other typological research to be specified below.

Given the controversy between diffusionists and retentionists, we cannot simply amass a number of shared features among the Transeurasian languages and allow geographical adjacency to imply the probability of diffusion, without requiring any linguistic support for this. Therefore, historical evidence suggesting the diffusion or the retention of traits may be particularly telling in this particular case. For representatives of the contemporary varieties of the five families belonging to the Transeurasian continuum, I have chosen Turkish (Turkic), Khalkha Mongolian (Mongolic), Evenki (Tungusic), Korean (Koreanic) and Japanese (Japonic) as horizontal comparison points. However, in order to allow a diachronic perspective, their profile will be supplemented by values from the oldest linguistically reliable historical varieties of the individual families, i.e. Old Turkic (eighth-fourteenth century), Middle Mongolian (thirteenth-seventeenth century) and/or Written Mongolian, Manchu (seventeenth-nineteenth century), Middle Korean (fifteenth-sixteenth century) and Old Japanese (eighth century).

In case a diachronic variety does not openly or productively reflect a certain feature, but nevertheless preserves a trace of it, indicating that the value was positive in an earlier stage of the language, the historical variety will be marked with a plus. In this way, we can obtain a glimpse of the unrecorded typological past of the language in question.

The organization of this chapter is as follows. In Section 2, I will set up a typological profile of the Transeurasian languages in relation to that of the selected languages immediately outside the continuum. The linguistic levels discussed will include


The following sources were consulted for retrieving linguistic data underlying the feature values in contemporary Transeurasian languages: Göksel and Kerslake 2005 for Turkish; Janhunen 2012 for Khalkha Mongolian; Bulatova and Grenoble 1999 and Nedjalkov 1997 for Evenki; Martin 1992 and Sohn 1994 for Korean and; Martin 1988 and Kaiser et al. 2001 for Japanese.

The following sources were consulted for retrieving linguistic data underlying the feature values in historical Transeurasian languages: Erdal 2004 for Old Turkic; Street 1957, Weiers 1966 and Rybatzki 2003 for Middle Mongolian; Poppe 1954 for Written Mongolian; Gorelova 2002 for Manchu; Martin 1992 and Lee and Ramsey 2011 for Middle Korean and; Vovin 2005, 2009 and Frellesvig 2010 for Old Japanese.
phonology, lexicon and semantics, morphology and syntax. I intend to treat grammaticalization patterns as a distinct level of analysis because, rather than representing a static feature value, they are concerned with a dynamic force, leading in languages to change from a less to a more grammatical status. I will close Section 2 with a tabular overview, summarizing the presence of the 27 examined features in the selected languages by way of plus (+) and minus (-) values. In Section 3, I will consider how the insights from these data are relevant for general statements about areality, paying attention to the delimitation of areality, peripheral deviations from the prototype, changes in areality and the distinction between diffused and inherited features. In Section 4, I will conclude this chapter.

2 Typological profile of the Transeurasian languages

2.1 Phonology

1. Predominantly polysyllabic root structure The Transeurasian languages, together with their historical varieties, display a preponderance of polysyllabic roots, as do most languages in North Asia. Contemporary and Old Japanese possess a relatively great number of monosyllabic roots, many of which are attributed to root-internal consonant loss and subsequent vowel contraction (Whitman 1990). These phonological reductions argue against Janhunen’s (1997) suggestion that Japanese derives from an originally monosyllabic language. Austronesian languages such as Rukai are typically polysyllabic as well. Mandarin is the only language in the tables that is marked with a negative value. Similar to the languages of Mainland Southeast Asia it is predominantly monosyllabic, but, in comparison to Classical Chinese, it has developed a greater number of polysyllabic roots through compounding (Norman 1988: 86). As such, Japanese and Chinese occupy an intermediate position between the languages of North and Southeast Asia.

2. Absence of complex tonal distinctions None of the Transeurasian languages is tonal in the sense that each syllable is characterized by a distinctive pitch pattern. This is also true for Austronesian languages, such as Rukai. With the exception of Ket, which has been attributed a tone system with five oppositions in recent descriptions by Vajda (2004) and Georg (2007), the neighboring languages of North Asia are typically non-tonal as well. However, Nivkh, Japanese and some varieties of Korean have suprasegmental systems which can be seen as transitional between tonal and non-tonal languages. Nivkh makes distinctive use of two types of tones, whereas Middle Japanese and Middle Korean use a system of pitch accent that differentiates words according to the position of one prominent syllable after which the pitch drops. This system survives in Contemporary Japanese, but it has been lost in Contemporary Standard Korean, where it developed into a vowel length distinction. The two-way tone distinction and the pitch accent system are highly restricted in comparison to complex tonal systems such as in Ket and Mandarin, where each syllable is marked with one out of five distinctive tones. Tonal languages are not only extremely widespread throughout Southeast Asia, but also 42% of languages in Maddieson’s (2005: 58-61) sample of 526 languages across the world are tonal.

3. Presence of vowel harmony Vowel harmony can be defined as a phenomenon whereby
vowels within a domain agree with each other in terms of one or more features (Ko 2012: 7). It is a characteristic feature of the Transeurasian languages, except Japanese, but it is also present in most Uralic languages, including Khanty and in many other languages in North Asia such as in Yukaghir, Nivkh and Ainu. Ket lacks vowel harmony and so do Rukai and Mandarin, as such reflecting prototypical Austronesian and Mainland Southeast Asian behavior, respectively. In Old Japanese, however, there is a restriction on the shape of root morphemes, whereby the vowel \( o_2 \) cannot occur in a root together with the vowels \( u, o_1 \) or \( a \). This phenomenon, known as Arisaka’s law, has been taken as a kind of vowel harmony, but it has been rejected from comparisons with other Transeurasian languages because it applies to roots rather than to suffixes and because it does not reflect palatal harmony, the type of harmony which was attributed to the Transeurasian languages until recently (e.g. Frellesvig 2010: 44). However, in lexicalized verb stems incorporating derivational suffixes, as well as in the noun inflectional suffixes such as the plural suffix and the genitive suffix, there are traces of \( a \sim o_2 \) vowel alternation according to the quality of the vowels in the preceding root, e.g. OJ \( \text{-no} - na\text{-petrified in compounds such as OJ } mi_j\text{-na-moto} \) (water-GEN-base) ‘source, the headwaters’ (Rickmeyer 1989: 316; Robbeets forthcoming c).

4. Presence of tongue root vowel harmony Among the various types of vowel harmony, the most frequently attested ones across the languages of the world are palatal harmony, labial harmony, height harmony and tongue-root harmony. Palatal harmony requires all vowels within a domain to be exclusively front or back. It can be found in most Uralic languages such as in Khanty as well as in the Turkic languages (e.g. Tk. \( \text{ip-ler} \) (rope-PL) ‘ropes’ vs. \( \text{pul-lar} \) (stamp-PL) ‘stamps’). Since the western Mongolic languages Oirat and Kalmuck display palatal harmony as well, it has been proposed that the original system of Mongolic harmony was palatal (Poppe 1955, Svantesson 1985). However, Ko (2012) demonstrated that the original vowel harmony in Mongolic was in fact based on the opposition between the advanced vs. retracted position of the tongue root, rather than on a palatal contrast. He argued that the tongue root retraction system in Khalka (e.g. \( od-\text{o:s} \) (feather-ABL) vs. \( \text{o:d-}\text{r:s} \) (star-ABL)) represents retention rather than innovation. Furthermore, he supported the view that Tungusic vowel harmony is Retracted Tongue Root (RTR) based, as it is in Manchu and Evenki, and that the reduced vowel harmony in contemporary Korean derives from a tongue-root based system in Middle Korean. As far as the harmony-like opposition between \( o_2 \) and \( u, o_1 \) or \( a \) in Old Japanese is concerned, the recent reconstitution of a seven-vowel system in proto-Japonic by Frellesvig and Whitman (2008) implies an underlying opposition between pJ \( *i^+ \), \( *a \) and \( *u, *o, *a \), which does not exclude an original RTR based contrast. Whereas Vovin (1993: 50-51) and Bugaeva (2013: 26-28) reconstruct palatal harmony in Ainu, Shibatani (1990: 15) speculates that the Ainu opposition between \( o \) and \( u, a \) might have its origin in tongue root harmony, but here the indications are even weaker than in the Japanese case. According to Maslova (2003: 35), Yukaghir might be more appropriately described as having tongue root harmony than palatal harmony. Chukchi also displays tongue-root harmony. Although Gruzdeva (1998: 10) suggests that Nivkh leaves traces of height harmony, Janhunen (1981) and Ko, Whitman and Joseph (2014) interpret the system in terms of tongue root harmony. Cross-linguistically, tongue root harmony seems to be concentrated in Niger-Congo and Nilo-Saharan languages. Outside Africa and the Northeast Asian region the phenomenon seems to be rather rare: only Native American languages such as Nez Perce and Coeur d’Alene Salishan are known to have the feature.
(Ko 2012: 11-12). A rough estimate would be that less than 10% of the world’s languages have a tongue root vowel harmony system.

5. Absence of initial velar nasal In most Turkic languages as well as in Mongolic languages and Korean the velar nasal $ŋ$- cannot appear in word-initial position. Japanese lacks a velar nasal phoneme. In the Tungusic languages, with exception of Manchu, however, $ŋ$- can appear word-initially, but generally restricted to a specific phonological environment, notably when it is followed by the sonorants $n$, $r$, $l$, $m$, $y$, e.g. Evk. $ŋene$- ‘to go’, Ma. $genu$- ‘to go together’, Evk. $ŋeːle$-, Ma. $gele$- ‘to fear’, etc. According to Poppe (1964: 4) the initial velar nasal in Tungusic is the result of secondary assimilation of pTg $*g$-, which implies that originally $*ŋ$- was absent in Tungusic as well. The assimilation was probably triggered by influence from languages in the Siberian area, such as Nivkh, which allow initial velar nasals (Anderson 2006). It is under the same influence that initial $ŋ$ became allowed in Dolgan (Turkic), e.g. $gassa$ ‘pipe’. In Khanty, Ket, Kolyma Yukaghir, Ainu and Mandarin $ŋ$- does not occur in word-initial position. Rukai allows an initial velar nasal, e.g. $ŋałai$ ‘saliva’. In Anderson’s (2005: 42) sample of 468 languages, 69 % lack an initial velar nasal. Among the languages of the world that have a velar nasal phoneme, as is the case for most Transeurasian languages, only 35 % do not use it in word-initial position.

6. Absence of initial r- Throughout the Transeurasian languages, the consonant $r$- is not allowed to occur word-initially, except in borrowings (e.g. J $rajio$, K $lətiwọ$, Even $radio$, Khal. $radio$, Tk. $radyo$ ‘radio’). This is also true for Kolyma Yukaghir. Ket lacks a phoneme $r$ altogether. Although initial $*r$- is not reconstructed for proto-Uralic, Khanty is atypical in this sense, e.g. $raγta$ ‘to drop, slide’ and $räγ$ ‘garbage’. Nivkh, Ainu, Mandarin and Rukai also have native words in initial $r$-.

7. Absence of initial consonant clusters None of the Transeurasian languages tolerates initial consonant clusters, although medial clusters are tolerated in Turkic, Mongolic, Tungusic and Korean but, on the face of it, not in Japanese. On the basis of morphological, etymological, dialectal and textual evidence, however, it is safe to assume that the Old Japanese obstruents OJ $b$, $d$, $g$, $z$ resulted from the rephonologization of nasal obstruent clusters pJ $*np$, pJ $*nt$, pJ $*nk$, pJ $*ns$ (Robbeets 2008). Reminiscent of how the Transeurasian languages do not allow for consonant clusters in initial position, Old Japanese did not permit word-initial voiced obstruents except in mimetic adverbs. From the the ninth century onwards, as loans from Chinese began to have a major impact, the restriction was relaxed and initial voiced obstruents began to appear in borrowings and in contracted native forms. The avoidance of consonant clusters is further characteristic of Uralic languages, such as Khanty. Similarly, Yukaghir, Ket, Ainu and Rukai tolerate only single consonants in word-initial position. Wordinitial clusters may comprise at most two consonants in Nivkh, e.g. $mra$ ‘fault’ and $ksynz$ ‘witch’. Although Mandarin lacks consonant clusters, there is strong evidence that in Old Chinese (first millennium BCE) a variety of consonant clusters could occur at the beginning of the syllable as well (Norman 1988: 9-10). The simplification and eventual loss of consonant clusters appears to be a tendency affecting most of the Mainland Southeast Asia area. It is possible that early contacts between Chinese and Transeurasian, that has never tolerated initial clusters, have triggered the development along these lines.
8. Presence of voice distinction for stops Turkic, Mongolic and Tungusic languages share a voiced-voiceless opposition for stops. In Contemporary and Middle Korean, stops display an opposition between lax (p), aspirated (ph) and tensed (p'). Even if the lax stops become lightly voiced between voiced sounds, there is no phonemic voicing distinction. The Japanese voicing distinction for stops is a secondary development. As mentioned in the previous paragraph, voiced stops derive from prenasalized voiceless stops, so originally, Japanese lacked a voicing distinction. Khanty lacks a voicing distinction for stops, a feature characteristic of proto-Uralic, although many contemporary Uralic languages have developed an original singleton-geminate contrast into a voicing distinction. For example, although the contrast between /p/ and /pp/ in proto-Uralic *lapa ‘flat surface; leaf’ and *tappa- ‘to stamp with feet; to hit, knock’ is maintained in the Finnish reflexes lapa ‘shoulderblade, leaf surface’ and tappa- ‘to beat to death, kill’, it has usually developed into a distinction between /b/ and /p/ such as in Estonian laba ‘surface’ and tapa- ‘kill, slaughter’. Ket and Yukaghir display a voicing distinction, but languages to the extreme northeast such as Ainu, Nivkh and Chukchi do not. Mandarin, like Nivkh, has a distinction between aspirated and unaspirated stops, but lacks a voiced-voiceless opposition. Characteristic of most Austronesian languages, Rukai also displays voice distinction for stops.

2.2 Lexicon and semantics

9. Preference for a non-verbal strategy with (extra-family) verbal borrowing As far as the mechanisms of loan verb accommodation are concerned, most recipient languages can be categorized into two distinct groups: borrowed verbs either arrive as verbs, needing no formal accommodation, or, they arrive as non-verbs and need formal accommodation. In Wohlgemuth’s (2009) terminology, the first group represents “Direct Insertion”, while the second group represents either “Indirect Insertion”, when the formal accommodation involves a verbalizer or else, “Light Verb Strategy”, when the borrowed verb is integrated into a complex predicate. Turkic, Mongolic, Korean and Japanese can be assigned to the second group because they display a clear preference for the non-verbal strategy (Wohlgemuth (2009: 159, 161)); for instance, Tk. klik-le- and klik et- << English click; Khal. zee-le- << Mandarin zhà ‘borrow, lend’; K coking ha-, J zyogingu suru ‘to jog’ << English jog; J demo-r- << English demonstrate. Whereas the northern Tungusic languages prefer to borrow verbs through direct insertion, e.g. Evk. vypol̄i̇aj- << Russian vypolnja-t- ‘to fulfill, carry out’, the southern Tungusic languages use verbalizers, e.g. Ud. tancewa-la- << Russian tancewa-t- ‘to dance’ and Na. voprosa-la- << Russian voprosa-t- ‘to inquire, question’. In contrast to the Transeurasian languages, Uralic languages such as Khanty, Austronesian languages such as Rukai, Ainu and Mandarin show a strong preference for direct insertion (Wohlgemuth 2009: 158, 161; Tamura 2000: 267). Yukaghir and Nivkh did not integrate any recognizable verbal borrowings from Russian or from other foreign languages into their lexicons (Fubito Endo p.c.; Ekaterina Gruzdeva p.c.). In Wohlgemuth’s (2009: 157) sample, 55% of languages worldwide are found to use direct insertion, while the remainder prefer non-verbal strategies such as indirect insertion and the light verb strategy.

10. Presence of a two-way proximal-distal distinction in demonstrative pronouns Although Old Turkic displays a two-way distinction in its demonstratives, i.e. OT bo /
bun- ‘this’ vs. ol / an- ‘that’, many contemporary Turkic languages such as Turkish make
a three-way distinction, e.g. Tk. bu ‘this’, şu ‘that’, o ‘that (over there)’. Demonstrative
pronouns in earlier and contemporary varieties of Mongolic and Tungusic exhibit a
proximal-distal distinction: MMO. ene ‘this’ vs. tere ‘that’, Khal. e- ‘this’ vs. te- ‘that’,
Ma. ere ‘this’ vs. tere ‘that’ and Evk. er(i) ‘this’ vs. tar(i) ‘that’. Demonstrative pronouns
in Contemporary and Middle Korean, however, show a proximal-mesial-distal op-
position: K i ‘this’, ku ‘that’, ce ‘that over there’ and MK i ‘this’, ku ‘that’, tye ‘that
over there’. This is also true for Contemporary Japanese: J ko- ‘this’, so- ‘that’, a- ‘that
over there’. In contrast to most accounts of Old Japanese demonstratives, which posit a
three-way contrast between OJ ko2 ‘this’, so2 ‘that’ and ka ‘that over there’, Frellesvig
(2010: 139-142) argued that OJ ka was not a productive member of the demonstrative
system and that pre-Old Japanese had a simple proximal-distal distinction. While Khanty
distinguishes between proximal timi ‘this (here)’ and distal tomɨ ‘that (there)’, Yukaghir,
Ket and Ainu have a three-way opposition, each demonstrative pronoun denoting a
different degree of proximity: Yukagir tiŋ ‘this’ (proximal), adɨŋ ~ edɨŋ ‘that’ (mesial),
taŋ ‘that’ (distal); Ket tu- ‘this, that’ (neutral), ki- ‘this, that’ (proximal); qa- ‘this, that’
(distal) and Ainu ta an ‘this’ (distal), ne an ‘that’ (mesial), to an okai ‘that over there’
(distal). Nivkh makes as many as five distinctions: tyd ‘this’ (near and visible), hyd
‘this, that’ (distant), ad ‘that’ (more distant and visible), aixnt ‘that’ (most distant),
kud ‘that’ (absent). Rukai distinguishes four demonstrative pronouns in terms of visibility
and distance: ‘ina ‘this’ (proximal), ana ‘that’ (mesial), ona ‘that over there’ (distal but
visible), dhona ‘that over there’ (distal and invisible). Mandarin has a two-way
distinction between proximal zhè(ge) ‘this’ and distal nà(ge) ‘that’, which developed
from a three way-distinction in Classical Chinese between neutral, proximal and distal. In
Diessel’s (2005: 170-173) sample of 234 languages, 54% exhibit a two-way distance
contrast in demonstratives, while 38% exhibit a three-way contrast.

11. Inclusive-exclusive distinction in first person plural pronouns
 Among the Turkic
languages, there are no unique pronominal forms that distinguish inclusive from
exclusive person forms. Although Old Turkic and most presently spoken varieties of
Turkic distinguish between a first person plural (Tk./OT biz ‘we’) and an augmented
plural form (Tk. / OT biz-ler ‘we (as a group)’), Nevskaya (2010: 124) argues for a
collective interpretation of the augmented plural, denoting “an isolated group of people
who want to oppose themselves to the others”, rather than an inclusive interpretation as
suggested by Grönbech (1936: 81). The Middle Mongolian distinction between exclusive
bu and inclusive bida is formally preserved in the Khalka oblique paradigm in the
variation between formally exclusive man- and formally inclusive bidn-, but the
functional distinction has been lost. In the Tungusic languages, however, the inclusive-
exclusive opposition is generally well preserved, e.g. exclusive Ma. be, Evk. bu vs.
inclusive Ma. muse, Evk. mut ~ mit. Similar to the Turkic languages, Middle and
Contemporary Korean distinguish between a first person plural (K/MK wuli ‘we’) and an
augmented plural form (K wuli-tul, MK wuli-tolh ‘we (as a group)’ in which K tul, MK
tolh is a collective marker. Contemporary Japanese lacks an inclusive-exclusive
distinction and cannot derive an augmented plural from the first person plural watasi-tati

4 Note that this analysis deviates from the feature values given for distance contrasts in
demonstratives by Diessel (2005: 170-173), since he marks Ainu, Nivkh, Yukaghir and Turkish
as having a two-way contrast.
(I-PL) ‘we’. Old Japanese also lacks the distinction, but the stem OJ wa- ‘I, we’ can be used as a first person plural in the possessive case form, but it can also be augmented with a collective marker -ra ~ -re to OJ ware ‘we’, a form which in its turn has been augmented into ware-ra ‘we’ later in Japanese. As is the case for many Uralic languages, Khanty marks a dual distinction, but not an inclusive-exclusive distinction on its person pronouns. While Ket and Yukaghir lack the distinction, Nivkh distinguishes between exclusive n’yŋ and inclusive mer ~ mir. Although Ainu personal affixes on the verb have an inclusive-exclusive distinction, the first personal pronoun aoka(i) only has a single form. The distinction found in the first person plural pronouns between exclusive wōmen and inclusive zānmen ‘we’ of Beijing and certain other northern Chinese dialects may be due Transeurasian influence. Such a distinction was not found in Old Chinese, and it began to appear in North China during the period of Altaic rule. It is significant in this regard that both Middle Mongolian, spoken under the Yuan dynasty and Manchu distinguish exclusive and inclusive forms. Rukai distinguishes exclusive -nai ~ nai- (NOM) from inclusive -mīta ~ ta- (NOM), a feature characteristic of Austronesian languages. In Cysouw’s (2005: 166-167) sample of 200 languages, 31% distinguish between an inclusive and an exclusive with independent pronouns.

12. Property words may be verbally or nominally encoded Cross-linguistically adjectives have no prototypical encoding strategy of their own: they will align themselves either with verbs or with nominals (Stassen 1997: 30). Across the Transeurasian languages, the encoding of property words appears to be mixed because, at least in the earlier stages, both the nominal and the verbal strategy is used (Robbeets forthcoming c). Generally, this mixed encoding is split in the sense that most property words have only a single encoding option, with the exception of some instances of switched encoding mentioned under feature value 13. In Old Turkic, most property words are nominally encoded, but there seems to be a tendency to apply the verbal strategy to words expressing time-unstable properties such as OTk. būdī- ‘to be(come) big, great’, OTk. isi- ‘to be hot’, OTk. kati- ‘to be hard, firm, tough’, OTk. kiz- ‘to be red’, OTk. tūmli- ‘to be cold’, OTk. us- ‘to be thirsty’, OTk. tīgra- ‘to be tough’, etc. Contemporary Turkic languages maintain only few reflexes of these verbal property words, e.g. Tk. būyūi- ‘to be(come) large’ but in the majority of cases, the earlier verbal property word has been derived through a deverbal noun suffix into a nominal adjective (e.g. Tk. būyūk ‘big’). Similarly, most property words are nominally encoded in Mongolic, but there is a tendency to apply the verbal strategy to less permanent properties in Middle Mongolian such as MMO. ayu- ‘to be(come) afraid’, MMO. ēt- ‘to be ripe, be(come) saturated’, MMO. hiče- ‘to be ashamed’, WMo. qala- ‘to be(come) warm’, MMO. sohta- ‘to be drunk’, etc. Contemporary Mongolic languages such as Khalkha maintain only few reflexes of these verbal property words, e.g. Khal. ayu- ‘be afraid’. The same is true for Tungusic, where contemporary languages such as Manchu (e.g. Ma āka- ‘to be sad’, Ma. bere- ‘to be lame’, Ma. ēbi- ‘to be satiated’) and Evenki (e.g. Evk. ukti- ‘to be hungry’, Evk. uwi- ‘to be satiated’, Evk. buli:- ‘to be sad’) may occasionally exhibit verbal encoding. In Korean, there are property words such as K kanan ha- ‘to be poor’ and phikon ha- ‘to be

5 Note that my evaluation differs from Cysouw’s (2005: 166-167) analysis, which marks Ainu as having an inclusive/exclusive distinction with independent pronouns.

6 Note that my analysis deviates from the feature values inserted for predicative adjectives by Stassen (2005: 480-481), in which Evenki and Manchu are marked as exclusively verbal encoding, in line with the traditional view.
tired’ that consist of a nominal root and the auxiliary ha- ‘to be in the state of’ and whose bases are called “adjectival nouns” (Martin 1992: 189, 190; Sohn 1994: 219-220). However, the majority of property words are inflected in essentially the same way as verbs, e.g. K kwut-, MK kwut- ‘to (become) hard’, K noph-, MK nwoph- ‘to be high’, etc. Some Japanese property words, such as J sizuka, OJ siduka ‘quiet’, J/OJ tasika ‘trustworthy’ are encoded exclusively nominally, while others such as J/OJ taka- ‘to be high’, J/OJ kata- ‘to be hard, tough’ are essentially inflected in a similar way as verbs. In line with most Uralic languages, property words in Khanty are exclusively nominally encoded. This is also true for Ket. In Yukaghir, Ainu and Nivkh, however, property words are exclusively verbally encoded. As in the case with most Transeurasian languages, Ainu property verbs express both the property and the process leading to the property, e.g. pirka ‘to be(come) good’. In line with Mainland Southeast Asian and Austronesian languages, Mandarin and Rukai use verbal encodings for property words. In Stassen’s (2005: 478-481) sample of 386 languages, 27% have mixed encoding in predicative adjectives.

13. Some property words exhibit switched encoding Some property words in the Transeurasian languages, especially in the earlier varieties, further exhibit traces of switching, whereby the same property word can have both nominal and verbal encoding: e.g. OT. ač ‘hungry’ / ač- ‘to be hungry’, OT. keč ‘late, slow’ / keč- ‘to be late, slow’; MMO. bulqa ‘hostile; hostility’ / bulqa- ‘to be hostile’; Ma. jalu ‘full’ / jalu- ‘to be full’, Ma. sula ‘loose, free’ / sula- ‘to be loose, be free’; MK toso- vs. MK toso ho- ‘to be warm’; OJ taka ‘high’ / taka- ‘to be high’, OJ opo ‘big’ / OJ opo- ‘to be big’, etc. None of the neighboring languages, except Tundra Yukaghir, exhibits such behavior. There, two property words, i.e. juku ‘small’ and t’ama ‘big’ occur as noun modifiers without overt adnominalizers, e.g. t’ama-d’ohoje (big-sword) ‘sabre’ in addition to having a verbal encoding, for example, in the deverbal inchoative t’ama-mu (be.big-INCH) ‘to grow, become big’ (Maslova 2003b: 14). Logically, the proportion of languages exhibiting mixed and switched encoding will be lower than 27%, i.e. the proportion of languages with mixed encoding in general.

14. Partial emphatic reduplication of nominal property words Partial emphatic reduplication is a phenomenon whereby the first consonant (if present) and vowel of a nominal property word are repeated with the addition of another consonant to indicate the presence of the property to the utmost degree. Whaley (2000) found that is widespread in Turkic, Mongolic and Tungusic, e.g. Tk. bem-beyaz ‘snow white’, OT kap-kara ‘quite black’, Kal. xob-xoldu: ‘frozen through’, WMo. ub-ulayyan ‘completely red’, Evk. ab-aya ‘very good’. I have not been able to find examples in Manchu, but the phenomenon is present in Sibe, a presently spoken variety of Manchu, e.g. fak-farxun ‘extremely dark’. In Tungusic, emphatic reduplication is restricted to Sibe, Kile-Nanai, Solon Evenki and Oroqen, i.e. the languages spoken on Chinese soil, which have been under strong influence of the Mongolic languages Khalka and Dagur. On the basis of this distribution and because the greatest flexibility, in terms of both the number of reduplicated words and the type of concepts they denote, is found in Turkic, Whaley (2000: 358) argued for a diffusion of the feature from Turkic to Mongolic to Tungusic. Japanese, Korean and the neighboring languages under examination do not display partial emphatic reduplication. In Rukai, however, descriptive verbs are partially reduplicated in comparative constructions (see feature 23).
2.3 Morphology

15. **Morphology is agglutinative** Agglutinative languages connect morphemes linearly in a way that there is a one on one relationship between a morpheme and its meaning. The Transeurasian languages belong to a North Asian and European belt of agglutinative languages together with the Uralic languages, including Khanty and other languages of the Siberian area such as Ket, Yukaghir, Chukchi and Nivkh. Ainu has agglutinative morphology and so do the Austronesian languages, including Rukai. Chinese, the only analytic language under examination, has triggered a decrease of agglutinating features in Tungusic as one moves from north to east and further south. Manchu is the most analytical among the Transeurasian languages; it treats case forms, for instance, as particles rather than suffixes.

16. **Inflectional morphology is predominantly suffixing** Across the strongly suffixing Transeurasian languages, prefixation is rare and restricted to derivational morphology, such as the partial emphatic reduplication (see feature 14 above) and some derivational prefixes in Korean (e.g. K ye-l- ‘young, new’ in ye-l-cwungi ‘a chick out of its shell’) and in Japanese (e.g. ma- intensive in ma-siro ‘snow white’). As is the case for most Uralic languages, Khanty is strongly suffixing and so is Yukaghir. Nivkh is considered to be weakly suffixing. In Ket, nominal inflectional morphology is strongly suffixing, whereas verb inflection is predominantly prefixing. In Ainu and Rukai, inflection makes use of both prefixes and suffixes. Probably due to Transeurasian influence, Mandarin is hard to assign unequivocally to either the isolating or weakly suffixing type, but Sinitic varieties in general tend towards the isolating pole. In Dryer’s (2005: 110-113) sample of 894 languages, 43% are strongly suffixing.

17. **Absence of obligatory numeral classifiers** Although in Turkic and Mongolic some nouns of low countability may be accompanied by a unit of measure by means of which they can be counted, e.g. Tk. sekiz bardak su (eight glass water), OT yeti tutum talkan (seven handful parched.grain) ‘seven handfuls of parched grain’, Khal. gourben debter nom (three volume book) ‘three volumes of books’, etc. these languages do not make obligatory use of sortal numeral classifiers. Similar to the use of collective suffixes for counting people in Old Turkic (OT -(A)gU in e.g. ičäği ‘three together’) and Middle Mongolian (MMo. -(U)A ~ ALA in e.g. qoya’ula ‘two together’), the Tungusic languages use a variety of collective suffixes following numerals from ‘two’ to ‘ten’ such as Evk. -kt(e) and -ni for counting people (e.g. d’u-kte ‘two (people together); (we, you, they) two’), Evk. -gdA/ -ngnA for counting objects, Evk. -llA for counting the number of days (e.g. nada-lla ‘seven days, a week’), Evk. -nu / -pu for counting the number of tents (e.g. ilan-nu ‘three tents’) and Evk. -musa denoting the number of places or directions. However, only Manchu has developed about 70 sortal numeral classifiers, which divide the inventory of count nouns into semantic classes, each of which is associated with a different classifier, such as fesin which is used for objects equipped with a handle, e.g. ilan fesin loho (three CLAS sword) ‘three swords’. These words have original lexical meanings, e.g. fesin ‘haft, shaft, handle’, but under Chinese influence they have grammaticalized into classifiers, which are not obligatory in Manchu. Loho ilan (sword three) ‘three swords’, for instance, is equally possible. Whereas the standard pattern in Middle Korean was to modify a noun with a preposed numeral, e.g.
twu kalh (two knife) ‘two knives’, the most common pattern in Contemporary Korean makes use of a classifier, e.g. pus se:k calwu (writing. brush three CLAS) in which calwu denotes long objects with handles. However, the original pattern surfaces in expressions such as K twu nala ‘two countries’ and the use of classifiers remains optional in Korean, e.g. kalh hana-ka issta (knife one-NOM be.present) ‘there is one knife’. With Chinese influence inundating the language from Middle Korean times onwards, the classifiers developed from native words under Chinese influence or were borrowed as such from Chinese. Note that Middle Korean leaves traces of specialized suffixes to count days, e.g. *(o/u)l in saol ‘three days’, naol ‘four days’ etc. and that some Korean dialects use a suffix -i to count persons, e.g. se:-i ‘three people’, ne:-i ‘four people’, which recalls the use of collective suffixes in the other Transeurasian languages. While there is an extensive list of obligatory classifiers in Contemporary Japanese, e.g. enpitu san-bon (pencil three-CLASS) ‘three pencils’, the system of classifiers is much less developed in Old Japanese, where Chinese influence is restricted to a minimum. Numerals could be used with nouns, without intervening classifiers, e.g. OJ nana se (seven rapid) ‘seven rapids’ and the so-called “classifiers” are restricted to roughly six suffixes, i.e. -ka for counting days starting from the numeral ‘two’, -tu/-ti for counting objects, -ri for persons, -motob for grassy plants, -pe1 for layers and -ka for plants. It is not unlikely that these suffixes originate in collective suffixes. Numeral classifiers are absent in Uralic languages such as Khanty as well as in Yukaghir and Ket. Ainu makes use of a small set of obligatory classifiers such as -n/-iw for persons; -pe/-p for animals and things, and rerko for counting days starting from the numeral ‘two’ (with irregular forms tutko ‘two days’ and rerko ‘three days’). Nivkh distinguishes between 26 semantic classes with different numeral forms for each class. The obligatory use of classifiers is a widespread feature shared by Mandarin and the languages of Southeast Asia, but the use of classifiers in Classical Chinese was the exception rather than the rule. In Rukai the use of classifiers is optional in the sense that it uses a set of unaffixed numerals without classifiers as well as a set of bound numerals which combine with five different sortal classifiers to form verbs. In Gil’s (2005: 226-229) sample of 400 languages, 80% lack obligatory numeral classifiers.

18. Presence of mi-Ti opposition in first vs. second singular person pronouns Nichols (2012) observes that m-T pronominal paradigms with first person labial nasal m and second person apical or palatal obstruent t, c, s, etc. are much more common in northern Eurasia than elsewhere in the world. Janhunen (2013: 213) adds that there is a smaller group of mi-Ti languages extending from Uralic in the west, to Turkic, Mongolic and Tungusic in the east to Yukaghir in the north, in which not only the initial consonant but also the root vowel of the singular stems shows a basic similarity, in that it contains a non-low unrounded front vowel i or e. Although m is absent in the nominative first person singular in the Turkic, Mongolic and Tungusic languages, e.g. Tk ben, OT ben, Khal. biii, MMO. bi, Ma. bi, Evk. biː, it has developed in oblique forms through assimilation to the nasal oblique suffix -n, e.g. OT min-, Khal. min-ii (GEN), MMO. minu (GEN), Ma. min-, Evk. min-. The second person singular forms all reflect a voiceless dental T, i.e. Tk. sen, OT sen, Khal. ciː, MMO. ci, Ma. si, Evk. siː. The Korean pronouns are, among others, first singular K/MK na and second singular K/MK ne. In Japanese, J

Note that my evaluation differs from Gil’s (2005: 228-229) interpretation that Korean has obligatory numeral classifiers.
watasi and OJ wa are among others used in the first singular, while a variety of contemporary pronouns and OJ na are used in the second singular. Although the proto-Uralic first and second singular pronouns *mun and *tun reflect an m-T distinction (Janhunen 1982: 35), Khanty is deviant in having first singular mä and second singular nöŋ. In Yukaghir, however, the mi-Ti opposition is present in first singular met vs. second singular tet. In Nivkh, the distinction is absent in the singular pronouns, first person n‘i vs. second person či, but it is present in the opposition between the first plural inclusive mir/mer and the second plural pronoun čiŋ. The opposition is not found in Ket, Ainu, Chinese and Rukai. In Nichols and Peterson’s (2005: 546-551) sample of 230 languages, 13% display an m-T opposition in first vs. second person pronouns. Logically, languages reflecting a mi-Ti opposition will represent an even smaller proportion.

19. Formation of a secondary oblique stem of personal pronouns With the exception of Korean, the Transeurasian languages share a tendency to form a secondary oblique stem of the personal pronouns by means of a suffix which phonologically may be identified as the dental nasal -n-. In most contemporary Turkic languages, the nominative and oblique forms have merged, e.g. Tk. ben for the first singular nominative and oblique, but in Old Turkic the first singular nominative bän is distinguished from the oblique stem min-, which can be derived from an original pTk *bi-n- (1SG-OBL-). Similarly, the Mongolic and Tungusic languages derive oblique pronominal stems from the nominative roots through a nasal suffix, for instance, in the first person plural pronouns MMO. ba (NOM) vs. man- (OBL) and Khal. bid (NOM) vs. bidn- (OBL) and in the first person singular pronouns Ma. bi (NOM) vs. min- (OBL), Evk. bi: (NOM) vs. min- (OBL). There are no oblique pronominal stems in Contempoary Japanese, but Old Japanese leaves traces of an oblique nasal suffix in some case forms, e.g. in the Eastern OJ first person singular dative wa-nu-ni in alternation with Western OJ wa-ni. Vovin (2005: 229-230) further found that an original Japonic pronominal oblique *-n- is well supported by Northern Ryukyuan dialects where the first person pronoun uses waa- as the nominative and genitive base and extended waN- in the oblique cases. The erosion of the pronominal paradigm in Korean and Japanese may be due to the gradual de-pronominalization in the recorded history of these languages, whereby the system of personal pronouns became replaced by various terms of address and self-reference, probably under Chinese influence. The oblique nasal suffix is an important element in the Uralic pronominal paradigm as well, e.g. the Khanty first person pronoun mä (NOM) vs. män- (OBL). Ket, Yukaghir, Ainu and Mandarin, however, do not derive secondary oblique stems. The third person singular pronoun in Nivkh has both regular and suppletive case forms, e.g. if-øn (3SG-NOM) vs. if-toX ~ e-rx (3SG-DAT/ADD), but here the oblique form is not derived from the nominative base. Rukai personal pronouns have different shapes for nominative, topic, genitive and oblique cases, e.g. the first person singular -lrao (NOM), ilrae (TOP), -li (GEN) vs. -iae (OBL), in which the oblique seems to be formally derived from the nominative base by means of the same i- …-e marking as in the topic form.

2.4 Syntax

20. SOV (Subject-Object-Verb) sentence order Syntactically, the Transeurasian languages pattern as typical SOV languages but the sentence order is not rigid. SOV is also among the characteristic features of the Uralic languages, here represented by Khanty.
Languages to the north such as Yukaghir and Ket or to the northeast such as Ainu and Nivkh are almost all SOV languages, while those to the southeast are virtually all SVO languages. Mandarin, and in fact all major varieties of Chinese, corresponds with Southeast Asia with respect to verb-object order. Like most Austronesian languages, Rukai tends to be verb-initial, but the word order is non-rigid, switching freely between VSO and VOS. In Dryer’s (2005: 330-333) sample of 1228 languages, 40% have SOV sentence order.

21. **GAN (Genitive-Noun / Adjective-Noun) phrase order** A modifier-before-headword word order in the sentence (SOV) is expected to correlate with a modifier-head order within the noun phrase (GAN), whereby adjectives, genitives and modifiers in general occur before the nouns to which they refer. This is the case for the Transeurasian languages, the Uralic languages including Khanty and other languages of North Asia such as Yukaghir, Ket, Ainu and Nivkh. Mandarin, however, runs against the implicational expectation since genitives and adjectives occur before the nouns to which they refer in spite of SVO sentence order. This combination of feature values is absent in almost all the other languages of Southeast Asia and has probably arisen under influence of the Transeurasian languages. Rukai combines an adjective-noun order with a noun-genitive order (ANG). In Dryer’s (2005: 350-357) sample of 1105 languages, 55% have Genitive-Noun order, while out of 1213 languages, 28% have Adjective-Noun order.

22. **Extensive use of converbs** Converbs, also known as gerunds or adverbial participles, can be defined as nonfinite verb forms whose main function is to mark adverbial subordination (Haspelmath 1995: 3). Originally coined by the Altaic scholar Ramstedt, the term converb was adopted from Transeurasian linguistics to denote a cross-linguistic category. The Transeurasian languages are converb-prominent languages in the sense that they use converbs rather than adverbial subordinators as found in many European languages; see the examples below (Bisang 1995, 1998; Johanson 1995; Nedjalkov 1995; Alpatov and Podlesskaya 1995; Sohn 2009; Malchukov 2012).

1. **Turkish**
   
   Ali gel-ince  şaşırd-ı
   Ali come-CONV be.surprised-PST-3SG
   ‘When Ali came, he was surprised’ (Johanson 1995: 314)

2. **Khalkha**
   
   Ger-ees-ee  gar-aad
   house-ABL-REFL exit-PFV.CONV
   deuc-en jil-iin daraa ol-d-lao
   forty-ADN year-GEN after find-PASS-FIN
   ‘She went away from home and was found forty years later’ (Janhunen 2012: 280)

3. **Even**
   
   Dagam-mi, kunte-le d’u-v it-ti-n
   approach-CONV clearing-LOC house-ACC see-PST-3SG
   ‘When he came nearer, he saw a house on a clearing’ (Malchukov 2012: 213)

4. **Korean**
Kiho-nun nol-ko ca-ss-eyo.
Kiho-TOP play-CONV sleep-PST-POL
‘Kiho played and then slept’ (Sohn 2009: 300)

(5) Japanese
Taroo-ga bangohan-o tabe-te furo-ni hait-ta.
Taroo-NOM dinner-ACC eat-CONV bath-DAT enter-PST
‘Taroo took a bath after he ate dinner’ (Alpatov and Podlesskaya 1995: 473)

Although the Uralic languages are characterized by an extensive use of converbs, Khanty is rather atypical in this sense because it has only a single converb in -mɨn, which is the least frequent nonfinite verb form. Yukaghir and Nivkh also use a variety of converbs to link clauses. Ainu, however, employs subordinating conjunctions. Ket has no converbs or serial verb constructions of any kind. In Mandarin, verbs or verbal phrases are merely juxtaposed, the relation between the items being largely unmarked. Rukai marks adverbial subordination through a variety of means such as subordinating conjunctions, changes in word order and nominalized verb forms.

23. Use of a locative existential construction to encode predicative possession The Transeurasian languages show a clear preference to express the concept “X has Y” on the basis of an existential sentence, whereby the possessed noun phrase functions as the grammatical subject of the ‘exist’-predicate, while the possessor noun phrase is in a dative-locative case form. Although locative possessive constructions were standard in Old Turkic, Turkish uses genitive existential sentences as well as locative existential sentences. ‘I have a book’, for instance, can be expressed by Ben-de bir kitab var (I-LOC a book exist) or by Ben-im bir kitab-im var (I-GEN a book-1SG.POSS exist). Middle Mongolian and Khalkha make use of either a conjunctional possessive which construes the possessor noun phrase as the grammatical subject of the copula and marks the possessed with the comitative -tai, e.g. Khalkha Bi nom-tai bai-n’ (I book-COM be-DUR) or else, a locative possesive, e.g. Nad-ed nom bai-n’ (I-DAT book be-DUR). As is the case for most Tungusic languages, Manchu and Evenki employ locative existential constructions, e.g. Evk. Min-du: kniga bisi-n (I-DAT book-3SG); Ma. Min-de bithe bi (I-DAT book be). Korean uses a locative existential construction, e.g. K Na-hanthey chayk-i issta (I-LOC book-NOM exist), but the possessor can also be construed as the topic of the noun phrase, e.g. Na-nun chayk-i issta (I-TOP book-NOM exist). This is also true for Japanese, e.g. Watashi-ni hon-ga aru (I-DAT book-NOM exist) and Watashi-wa hon-ga aru (I-TOP book-NOM exist). Topic possessives may have developed under influence of Chinese, since they represent the standard strategy in Mandarin. Among the strategies used to encode predicative possession in the Uralic languages, we find locative possession such as in Finish and Hungarian, genitive possession such as in Nenets and possession encoded by a transitive verb ‘to have’ such as in Khanty. Whereas Yukaghir employs a conjunctional possessive and Ainu a ‘have’-possessive, Ket and Nivkh use locational possessives. Although many Austronesian languages employ topic possessives, Rukai makes use of locative and genitive possessive constructions. In Stassen’s (2005: 474-477) sample of 240 languages, 20% use a locative existential construction to encode predicative possession.

24. Use of the ablative case form to encode predicative comparison The Transeurasian
languages all form comparative constructions in which the standard noun phrase is constructed in the ablative case form, e.g. Tk. *bu araba-dan daha biyâk* (this car-ABL more big) ‘bigger than this car’, OT *barça-da üzâ-râk* (everything-ABL high-COMP) ‘higher than anything else’, Khal. *ene xun-ees iluu* (this person-ABL good) ‘better than this person’, MMO. *gola-sa gola* (far-ABL far) ‘farther than far’, Evk. *oron-duk gudatmar* (deer-ABL tall-COMP), Ma. *ere niyalma ci sain* (this person ABL be.poor-ADN person) ‘people poorer than me’ and J *chikyu:-yori omoi* (globe-ABL be.heavy) ‘heavier than the globe’. In literary Korean the ablative marker *eyse* ‘from’ can be used in comparative constructions, e.g. K *i eyse te khu-n salang* (this ABL more be.big-ADN love) ‘a greater love than this’, but it is more common to use a comparative particle *pota* ‘than’, e.g. K *kicha pota ppaluta* (train PT be.fast) ‘faster than a train’, MK *nyey pwota thak.wel hota* (past PT superior be) ‘superior to the past’. This particle has grammaticalized from the verb MK *pwo-* ‘to see’ and the transferentive --*ta-ka*, which signals the interruption of an event before its completion, i.e. ‘when one looks at’. It replaced earlier particles for comparison in Middle Korean, such as MK *tukwo* ‘than’ and *lawa* ‘than’. The Uralic languages differ from one another with regard to comparative constructions: languages to the west such as Finish and Hungarian use particle comparatives as in European languages, languages to the east, such as Nenets and Udmurt, mark the comparative standard with the ablative case ending, as in the Transeurasian languages. In Khanty, the marker of comparison is a postposition *niŋǝ* ‘since, from’, which has ablative-like semantics but differs from the standard ablative case ending -*oɣ* or the ablative-ellative ending -*i*. Yukaghir and Ket mark the comparative standard with the ablative case ending. In Nivkh, the comparative suffix -*yk* is traditionally considered as a separate case form as there is no evidence to relate it to the formally similar locative-ablative suffix -*uɣe*; -*ux* (Gruzdeva p.c.). Ainu forms comparative constructions by means of the particle *kasuno* ‘than’. In comparative constructions in Mandarin the standard noun phrase is constructed as the direct object of a verb ‘to exceed’. In Rukai, a comparative construction is formed through partial reduplication (CVV) of the descriptive verb stem. In Stassen’s (2005: 490-493) sample of 167 languages, 47% use locational comparatives, but the proportion of languages that specifically use the ablative case form to encode predicative comparison is logically expected to be lower.

2.5 Grammaticalization

25. Direct insubordination One of the driving forces of morphosyntactic change in the Transeurasian languages is a recurrent tendency to grammaticalize non-finite suffixes to finite suffixes (Robbeets 2009, forthcoming a). In line with Evans (2008: 367), I call this development “insubordination”, i.e. the conventionalized main clause use of what appear to be formally subordinate clauses, but it can be further specified as “direct” insubordination because non-finite suffixes are directly reanalyzed as finite ones, without the omission of a specific matrix predicate (Robbeets forthcoming a). Deverbal noun suffixes such as OTK -(A)r in OTK. *tug* ‘to be born, to rise (of sun) (intr.)’ -> *tugar* ‘sunrise, east’; MMO. -*m* in MMO. *quri* ‘to come together (intr.)’ -> *qurim* ‘feast’; Ma. -*rA* in mute- ‘to be able’ -> mutere ‘ability’; MK -(u/o)m in yel- ‘to bear’ -> yelum and OJ -*sa* in naga- ‘to be long’ -> nagasa ‘length’ develop over intermediate stages of
clausal nominalizers and relativizers into finite suffixes, as illustrated in the examples (1) to (5).

(1) Old Turkic

\[Ölüm-tä \ oz-upan \ ögir-ä \ savin-ä \ yori-r.\]

death-ABL escape-CONV rejoice-CONV be.happy-CONV go.on-FIN

‘Having been saved from death it happily goes on with its life.’

(Erdal 2004: 325)

(2) Middle Mongolian

\[udurit-basu \ ber \ ulu \ busire-m.\]

guide-COND PT NEG believe-FIN

‘Even if you guide them, they don’t believe.’

(Weiers 1966: 144)

(3) Manchu

\[si \ nene-me \ isinji-ci \ uthai \ sin-de \ bu-re\]

you be.first-CONV come-CONV at.once you-DAT give-FIN

‘If you come first, I shall give [it] to you straight away.’

(Gorelova 2002: 256)

(4) Korean

\[onul-un \ sweup-i \ eps-um.\]

today-TOP class-NOM not.exist-FIN

‘No class today.’

(5) Old Japanese

\[punapi \ to2-wo \ mi1-ru-ga \ tomo2si-sa\]

boat.people-ACC see-NML-GEN be.enviable-FIN

‘How enviable it is to see the boat-people!’

(Wrona 2008: 206)

The Uralic languages also display a recurrent tendency towards direct insubordination. Deverbal noun suffixes such as proto-Uralic *-k, *-pÅ, *-mə and *-śÄ are thought to have developed into finite markers for present-day (*-k, *-pÅ) and past (*-mə, *-śÄ) tense, either in proto-Uralic or after the separation of the daughter languages (Collinder 1965: 110-115; Janhunen 1982: 36-37). Eastern Khanty preserves only a faint trace of this development since the finite form of the negative verb can be marked with the perfective participle-əm, as illustrated in example (6). However, the phenomenon is well preserved in the Mansi cognate deverbal noun suffix -əm in uul- ‘to sleep’ -> uuləm ‘sleep’, which has developed into the finite past, illustrated in example (7). Nikolaeva (1999b) also observes the development in the Northern Khanty dialects.

(6) Eastern Khanty

\[mota \ wajaγ \ lök \ ant-im\]

some animal track NEG-FIN

‘There is not a single animal track’

(Filchenko 2007: 429)

(7) Mansi

\[am \ joht-um-m\]

I come-FIN-1SG

‘I have come’

(Collinder 1965: 113)
In Nivkh, there is a single instance of direct insubordination, but the phenomenon does not seem to be recurrent. It concerns the deverbal action noun and infinitive suffix -d’ which has developed over participial use into a finite form -d’, as illustrated in example (8).8

(8) Nivkh
if hum-d’ hyjm-d’
he live-NML know-FIN
‘He knows the living one/ (his) life.’ (Malchukov 2013: 200)

The remaining neighboring languages under discussion display strategies other than direct insubordination in grammaticalizing non-finite suffixes to finite suffixes. In Yukaghir and Mandarin, for instance, clausal nominalization in construction with a copula is the main source for developing new finite constructions. Many Sinitic languages use focus constructions consisting of a nominalizer plus a copula verb; dropping the copula then paves the way for developing finite constructions. The Mandarin shi … de focus constructions, for instance, consist of a copula shi and a nominalizer de, whereas the finite stance construction appears without the copula (Yap and Matthews 2008: 20). Similar processes are found in the Siberian area, for instance in Yukaghir (Malchukov 2013: 192-195). In Kolyma Yukaghir, the deverbal action noun suffix -l in pala:- ‘to escape’ -> pala:l ‘(a situation of) escaping’ has developed into a finite form in subject focus constructions, as illustrated in (9). The intransitive subject ‘I’ takes a focus marker -ek, which is also used to mark nominal predicates, thus pointing to its origin as a copula-like form. As such, the example in (9) can be derived from a cleft-like construction ‘It is me sitting’.

(9) Kolyma Yukaghir
Met-ek moda-
I-FOC sit-FIN
‘I sit’ (Malchukov 2013: 194)

Ket displays yet another strategy to develop finite markers, namely to reduce the matrix predicate to an affix on the former dependent verb. In example (10), for instance, the matrix verb bimbata ‘it is audible’ is reduced to a present suffix -bɛta ~ -bata on verbs expressing sound production (Malchukov 2013: 196-197).

(10) Ket
(10a) tam bis’ɛŋ in’ŋɛj bi-mbata
PT what sound be.audible-FIN
‘a certain sound is audible’ (Werner 1997: 170)

(10b) p-kuta’ɛj-га
1SG.POSS-whistle-FIN
‘I whistle’ (Werner 1997: 187)

8 Note that Kortlandt (2004: 4) identified the Nivkh suffix -d’ with the Indo-Uralic participial suffix *-nt, considering it as evidence of a common origin.
In Ainu, deverbal noun suffixes appear to be functioning as both derivational suffixes and syntactic clausal nominalizers, but there is no indication that they have developed into finite endings. Ainu lacks other nonfinite markers such as participial or converb affixes that could be open to develop into finite markers. Similarly, Rukai does not exhibit traces of direct insubordination.

The languages of the world use a variety of mechanisms for developing finite function on formerly non-finite forms such as (1) verbalization of nominal predicates plus finite copula with subsequent copula erosion; (2) reduction of a finite verb to affix; (3) insubordination through ellipsis of a matrix clause and (4) direct insubordination.

26. Grammaticalization from negative verb to verbal negator

The historical development of negation in the Transeurasian languages involves a recurrent development of an independent negative verb into a negative auxiliary verb, which may move from preposed to postposed position and eventually assume suffix status (Robbeets forthcoming b). All Tungusic languages, except Manchu have preserved evidence supporting the reconstruction of a negative verb pTg *e- ‘not to be, not to exist’. There are some instances of independent use of the negative verb, i.e. without a lexical verb, where it means ‘not to exist, not to live’ as in the Evenki example in (11a). In example (11b), the negative verb acts as a finite auxiliary to the lexical verb, which assumes an invariant adnominal form. In spite of SOV word order, the finite negative verb is preposed to the lexical verb. In emotive sentences, such as in example (11c), the negative auxiliary may move to a postposed position. The Nanai example in (12) represents the final stage of the negative cycle, i.e. fusion, whereby the auxiliary negative verb has assumed the status of derivational suffix on the lexical verb and its phonological form is reduced to lengthening of the stem-final vowel. Although its predecessor Jurchen preserves traces of pTg *e- ‘not to be, not to exist’, Manchu does not, but a similar negative cycle can be reconstructed for the verbal negator Ma. *aku:

(11) Evenki
a. esile e-dyeli-m tadu-gla
   now NEG-FUT-1SG there-ENCL
   ‘Now I will not be (live) there.’  (Nedjalkov 1994: 27)
b. nungan nekun-mi e-ce-n suru-v-re.
   he younger.brother-POSS.REFL NEG-PST-3SG go.away-CAUS-ADN
   ‘He did not lead his younger brother away.’  (Nedjalkov 1994: 11)
c. nungan songo-ro e-ce-n
   he cry-ADN NEG-PST-3SG
   ‘He did not cry [– what’s the use of crying?].’  (Nedjalkov 1994: 8)

(12) Nanai
xola:-ci-si
read.NEG-PST-2SG
‘You didn’t read.’

Like Old Turkic, Turkish has a verbal negative suffix -mA- that can be derived from an original negative verb pTk *ma- ‘not to exist’. The verbal origin is supported by the occurrence of a negative postposition mar in Chuvash, which contains a deverbal noun suffix *-r and can take a nominal argument such as the directive case in deitative constructions. The Middle Mongolian negative verb stem ese- ‘not to be, not to exist’
survived in a number of conjugated forms such as with the past marker -be- in example (13a), but gradually the negative auxiliary became used as an invariant form, transferring its entire inflection to the lexical verb, i.e. the past marker -be is attached to ire- ‘to come’ in example (13b).

(13) Written Mongolian

a. ükü-be-iüü ese-be-iüü
die-PST-INTER NEG-PST-INTER
‘Did [he] die or did [he] not?’ (Poppe 1954: 175)

b. manu baɣši ese ire-be
our teacher NEG come-PST
‘Our teacher did not come.’ (Poppe 1954: 175)

The Middle and Contemporary Korean verbal negator MK a·ni, K an(i) can be derived from an original negative verb *an- and the suffix MK -i that derives both nouns and adverbs from verbs. Gradually, the negator ani is being replaced by an analytic construction consisting of ani augmented by the finite auxiliary MK ho-, K ha- ‘to do, be’, which usually contracts to anh- ‘not to do’ in Contemporary Korean. This seems to reflect the start of the next negative cycle whereby the grammaticalized verbal negator is replaced by a new negative construction in which a negative verb is restored in its function as finite auxiliary to the lexical verb, which is nominalized with the suffix -ci.

(14) Middle Korean

’es·tyey a·ni -wo-no-n-ywo
why NEG come-PROC-ADN-INTER
‘Why is [he] not coming?’ (Martin 1992: 420)

(15) Korean

apenim un ka-ci anh-usy-e
father TOP go-NML NEG-HON-FIN
‘Father is not going.’ (Robbeets forthcoming c)

Old and Contemporary Japanese use an independent negative existential adjective na- ‘to be non-existent, not to exist’, illustrated in (16a), which is thought to derive from the same origin as the Old Japanese negative suffix -(a)n-, illustrated in (16b) (Martin 1987: 821). As such, an original negative verb pJ *ana- ‘not to exist’ seems to have developed into a negative suffix. Negative imperative constructions with na- preserve a trace of the originally preposed position of the negative auxiliary.

(16) Old Japanese

a. s-uru sube₁-no₂ na-sa
do-ADN way-GEN NEG-NML
‘Nothing can be done.’ (Vovin 2009: 483)

b. kiʃmi₁-ga k-i₁-mas-an-u
lord-GEN come-CONV-deign-NEG-ADN
‘You did not come, [my] lord.’ (Vovin 2009: 788)
Similar to the Transeurasian languages, one of the characteristics of the Uralic languages is the expression of negation by means of a construction, comprising a fully inflected negative auxiliary and a largely invariant lexical verb (Comrie 1981; Jahnunen 1982: 37; Payne 1985: 215-221; Honti 1997; Suikkonen 2002: 173). The construction may develop in ways which result in a redistribution of inflectional categories between the negative verb and the lexical verb. Eventually, as is the case for the negative particles Khanty ǝnt or Estonian ei, the negative auxiliary may become totally free of inflections and turn into an invariant verbal negator, which recalls the situation in Mongolic in (13b). However, there are no examples in Uralic in which the negative auxiliary ultimately becomes a suffix, as it does in Turkic, Tungusic and Japanese. In Yukaghir, clausal negation is expressed by a proclitic el-, which usually precedes the verb, but in spite of the formal similarity with the proto-Uralic negative auxiliary *e-, there are no language-internal indications that it originated in a negative verb or auxiliary (Fubito Endo p.c.). Clausal negation in Nivkh is expressed in three ways, i.e. (i) a construction with the negative verb k'au- ‘not to be, not to have’ preceded by the lexical verb in an invariant dative case form; (ii) an incorporation of the negative verb kav-r/gavr- ‘not to be, not to have’ incorporated into the verbal form and; (iii) a negative suffix -rla / -tla. Given the formal similarity between the negative verb k'au- and the negative affix -kavr-, it is not unlikely that they go back to the same source (Ekaterina Gruzdeva p.c.). Among the many negatives used in Mandarin, the most general and neutral negation is expressed by bu, whereas the existential negative mei ‘there is not, has not’ is used to negate the completion of an event ‘not yet’. Both mei and bu originate as verbs (van Gelderen 2008: 225). Ainu uses a negative particle that precedes the verb (e.g. somo ku-oman (NEG 1SG-go) ‘I do not go’). The non-verbal status of somo is indicated by the fact that it does not take any personal affix. There are also no indications that the Ket negative particle bǝ:n or the Rukai negative suffix -ka originate in a negative verb.

The expression of negation via negative auxiliaries is worldwide a minor type to begin with, found in only 40 (17%) out of 240 languages in Dahl’s (1979) sample, which is areally biased towards Uralic and Altaic languages, in 45 (4%) out of 1011 languages in Dryer’s (2005) sample, and in 16 (5%) out of the 297 languages in Miestamo’s (2005) sample. As a consequence, the particular development of negative verbs to auxiliaries to particles or suffixes is hence even rarer.

27. A morphologically simplex first person plural pronoun is complemented by the grammaticalization of the first person pronoun augmented with a collective-plural marker When dealing with the inclusive-exclusive distinction in first person plural pronouns in Section 2.2. (Feature 11), it was mentioned that most Turkic languages and Korean complement their first person plural pronoun (Tk./OT biz ‘we’; K/MK wuli ‘we’) with an augmented collective-plural form (Tk. / OT biz-ler ‘we (as a group); K wuli-tul, MK wuli-tolh ‘we (as a group)’). A similar tendency has been found in the history of Japanese, where the first person singular /plural OJ wa- ‘I, we’ coexists with the same form augmented by a collective marker OJ wa-re ‘we’, a form which in its turn was later augmented into ware-ra ‘we’ in the history of Japanese. Etymologically, the Middle Mongolian inclusive bida, reflected in the Khalkha formally inclusive oblique bidn-, derives from the first person singular MMO. bi ‘I’ and a plural suffix -dA, which also occurs in the plural demonstratives pronouns MMO. e-de ‘these’ vs. te-de ‘those’
(Doerfer 1985: 2; Domii 2006; Nevskaya 2010: 119). Domii argues that originally, *ba and *bi-da complemented each other as plural pronouns and that the distinction between exclusive and inclusive meaning was a secondary development. The Tungusic exclusives Evk. bu and Ma. be can be derived from the first person plural pTg *bö and an augmented plural *bö-(x)e, respectively (Doerfer 1978: 81-83, 95-96; Janhunen 2013: 217), whereas the inclusive Evk. mut ~ mit may go back to pTg *bö plus the collective suffix pTg *-t (Benzing 1955: 1020) and the inclusive Ma. mune may be an extension of this root with the collective suffix -sA (Benzing 1955: 1017-1018). This analysis suggests that successive cycles of plural augmentation on morphologically simplex (or simplified) plural pronouns have triggered the secondary development of an inclusive-exclusive distinction in Tungusic. As far as the Uralic languages are concerned, Khanty makes a commonly found distinction between pronouns in the first person singular (mä ‘I’), dual (min ‘both you and me’) and plural (mön ~ mën ‘we’), but it does not reflect any trace of plural augmentation on the first plural pronoun. Similarly, no traces of plural augmentation on first person plural pronouns are found in Ket or Yukaghir. In Rukai, the first person plural inclusive -mita (NOM) is formally underivable from the exclusive -nai ~ nai- (NOM). The personal pronouns in Ainu have all grammaticalized from person affixes followed by any one of several existential verbs meaning ‘exist’. The first person plural pronoun aoka(i), for instance, consists of the first person plural inclusive transitive subject affix a- and the verb oka ‘to exist’. In Nivkh, however, all plural personal pronouns can optionally be augmented with a plural suffix; the first plural exclusive pronoun, for instance, appears either as n’yən-φ or as n’yən-gu (1PL-PL) ‘we’. In Mandarin, two separate roots for the first person singular wǒ and zān pluralized, along with the suffix -men, into the derived exclusive wōmen and inclusive zānmen ‘we’. However since in Classical Chinese wǒ ‘I’ could be used as a first person plural ‘we’ as well, the exclusive wōmen can be regarded as an instance of plural augmentation.

### 2.6 Overview

In the body of this chapter, I have set up a list of 27 feature labels, chosen to maximize positive values for the Transeurasian languages. These features, inserted as vertical comparison points in the tables below, have been examined for selected representatives among the Transeurasian languages and their linguistic neighbors, which are inserted as horizontal comparative points. In the tables, I summarize the observations made above by introducing plus (+) and minus (-) values in the corresponding cells. This then leads to a quantification of the number of plus values in the last row.

As far as the feature values for the Transeurasian languages are concerned, Table 1 shows the following tendencies. First, the typological coherence seems to be greater for historical than for the contemporary stages of the languages investigated. This suggests that Transeurasian areality has decreased over the last millennium. Second, maximal coherence is found in the Mongolic and Tungusic languages, with minor deviations from the prototype in the Turkic languages in the west and somewhat more in the Japonic and Koreanic languages in the eastern periphery. Third, the deviation from the prototype in

---

9 An alternative analysis, deriving the inclusive MMo bida from the first singular pronoun *bi ‘I’ plus the second plural pronoun *ta ‘you (many)’ is proposed by Janhunen (2013: 215), but the voicing of the medial dental stop would represent an irregular development.
the east does not reflect a gradual loss as we proceed from Korean to Japanese, but rather an en-bloc reduction of features or even a slight increase for Old Japanese.

Table 1. Feature values for selected Transeurasian languages along with their historical stages

<table>
<thead>
<tr>
<th>Tk.</th>
<th>(pre-) OT</th>
<th>Khal. (pre-) MMo.</th>
<th>Ewk. (pre-) Ma.</th>
<th>K</th>
<th>(pre-) MK</th>
<th>J</th>
<th>(pre-) OJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>02</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>03</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>05</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>06</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>07</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>08</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>09</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>22</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>23</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>24</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>26</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>27</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

As far as the feature values for representative neighboring languages are concerned, Table 2 shows the following tendencies. First, the neighboring languages show significantly stronger deviations from the prototype than do any of the investigated Transeurasian varieties. This suggests that it is meaningful to apply the concept of “areality” to the Transeurasian languages in the sense that they reflect a geographical concentration of linguistic features that sets them apart from the selected neighboring languages. Second, Kanthy and Yukaghir show more typological similarity with the
Transeurasian prototype than do other neighboring languages. Note that for at least three of the examined features (i.e. 6, 18, 22), Khanty yields a minus value where the Uralic prototype would yield a plus value. This suggests that “areality” may also apply in a wider, but less coherent sense to the belt of Transeurasian-Yukaghiric-Uralic languages. Third, the investigated languages of North Asia have more typological features in common than those in South East Asia, i.e. Mandarin and Rukai. This suggests a third ring of areality that is the least uniform, involving the languages of North Asia.

Table 2. Feature values for representative neighboring languages

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>02</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>03</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>04</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>?</td>
<td>?</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>05</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>06</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>07</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>08</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>09</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>24</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>25</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>17</td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

3 Interpretation of the observations

3.1 Delimitation of areality
The Transeurasian continuum has clear boundaries which delimit the language type in relation to its neighbours both to the north (Yeniseic, Yukaghiric) and east (Nivkh, Ainu) as well as to the south (Sinitic, Austronesian). Although the observations above are in line with Janhunen’s (2009: 61-62) findings about a certain internal uniformity in the larger Ural-Altaic belt, they also suggest including Yukaghir in this larger belt and they indicate additional boundaries in areality between the Uralic and the Transeurasian languages as such. Among the features that enable us to delimit the Transeurasian languages in relation to their Uralic neighbours are 4. tongue root harmony in Transeurasian (and Yukaghir) vs. palatal harmony in Uralic; 8. voicing distinction for stops in Transeurasian (and Yukaghir) vs. original singleton-geminate distinction in Uralic; 9. non-verbal strategy of verbal borrowing in Transeurasian vs. direct insertion in Uralic; 11. inclusive-exclusive distinction in Transeurasian vs. none in Uralic (and Yukaghir); 12./13 mixed and switched encoding of property words in Transeurasian (and perhaps originally in Yukaghir) vs. nominal encoding in Uralic; 14. partial emphatic reduplication in Transeurasian vs. none in Uralic (and Yukaghir); 18. absence of initial m in the nominative first person singular vs. presence in Uralic (and Yukaghir); 25. development of a negative verb into a suffix in Transeurasian vs. none in Uralic (and Yukaghir); and 27. augmented first person plural pronoun in Transeurasian versus none in Uralic (and Yukaghir). For some features such as 23. and 24, Uralic makes use of a larger variety of strategies than the Transeurasian languages, where all languages uniformly use locative possession or ablative comparatives. It is remarkable that Yukaghir aligns with Uralic rather than with Transeurasian in more than half of the delimiting features (i.e. 9, 11, 14, 18, 25 and 27), although it is geographically adjacent to Transeurasian languages such as Yakut (Turkic) and some northern Tungusic languages, but not to the Uralic languages. In my opinion, this observation is probably not coincidental, but it might reflect the alleged genetic relatedness between Uralic and Yukaghir proposed by, among others, Collinder (1965).

3.2 Deviations from the prototype

Along the margins of the Transeurasian continuum, we can observe examples of gradual loss of Transeurasian features in the western and eastern peripheries, as well as gradual adoption of Transeurasian features, as in the case of Mandarin.

Examples of original Transeurasian features changing in the western periphery under Uralic influence are 4. Transeurasian tongue root harmony, which aligns with the Uralic languages as palatal harmony in Turkic; 12. gradual loss of verbal encoding of property words – mirroring Uralic nominal encoding – as one proceeds from older to contemporary varieties and from Tungusic in the east to Turkic in the west and; 18. the secondary development of m-initials yielding an mi-Ti opposition in first vs. second singular person pronouns in Turkic, Mongolic and Tungusic.

Changes in areality in the eastern peripheries may take place under the influence of the languages to the extreme northeast of the Siberian area or under Chinese influence. Examples of original Transeurasian features in Tungusic and Mongolic changing under Siberian influence are 5. the secondary assimilation of pTg *g- into an initial velar nasal in Tungusic, in line with Nivkh and 11/ 27. the secondary development of an inclusive-exclusive distinction on augmented plural pronouns, mirroring the situation in Ainu and
Examples of Korean and Japanese features aligning with the extreme northeast Siberian area are the lack of voicing distinction in Korean and Old Japanese, in line with Ainu, Nivkh and Chukchi and 10. the development of a mesial demonstrative distinction in Japanese and its presence in Korean, similar to the situation in Yukaghir, Ainu and Nivkh.

Chinese features seem to have diffused into Manchu, Korean and Japanese, for instance, in 1. the gradual increase of monosyllabic roots in Japanese; 2. the development of simple tone systems in Japanese and Korean; 3. the alleged erosion of tongue root vowel harmony in Old Japanese; 12. the relatively strong proportion of verbally encoded property words in Japanese and Korean in comparison to the other Transeurasian languages; 15. the increase of analytical features in Manchu in comparison to the other Tungusic languages; 16. the development of a two-way distinction in demonstratives compared to the three way-distinction in Classical Chinese; 11. the development of an inclusive-exclusive distinction in first person plural pronouns in Beijing and certain other northern Chinese dialects, which was not found in Old Chinese; 16. the weak suffixing tendency of Mandarin as opposed to other Sinitic languages and; 21. the rare combination of SVO sentence order and GAN noun phrase order in Mandarin, absent in almost all the other languages of Southeast Asia.

Geographically, Chinese is located between the Transeurasian languages and the languages of Mainland Southeast Asia, an intermediate position, which it also occupies from the point of view of typology.

Finally, some features in the Siberian languages to the extreme northeast seem to have diffused directly from Southeast Asia, without a Transeurasian intermediary, e.g. 2. the occurrence of two distinctive tones in Nivkh in comparison to the relatively simple pitch-accent systems of Japanese and Korean; 12. the exclusively verbal encoding of property words in Yukaghir, Nivkh and Ainu, similar to Mandarin, but different from mixed encoding in Japanese and Korean; 17. the obligatory use of an extensive list of classifiers in the Nivkh lexicon, and a smaller one in Ainu, recalling the widespread and archaic use of classifiers in Southeast Asia, as opposed to their relatively late development in Japanese and Korean. This observation may gain relevance in the light of theories that derive Ainu from the south (e.g. Murayama 1992; Vovin 1993; Bengston and Blažek 2009).
3.3  Diffused vs. inherited features

A simplistic interpretation of the observations would be to assert that the properties of the Transeurasian language type are universally so common that their parallel occurrence in several adjacent language families is coincidental. This is certainly not the case, however, because the Transeurasian continuum has clear boundaries which delimit the language type in relation to its neighbours both to the west (Uralic), north (Yeniseic, Yukaghiric), east (Nivkh, Ainu) and to the south (Sinitic, Austronesian). Moreover, the relatively low frequency of some features indicates that the shared properties are not due to mere universal principles in linguistic structuring. Above I have provided an estimation of the frequency of 19 out of 27 features. Seven features are not very common (i.e. 5, 9, 10, 16, 20, 21, 23) in the sense that they occur in less than half (50%) but more than a third (33%) of the languages worldwide. Nine features are relatively uncommon in the sense that they occur in less than a third (33%) of the languages worldwide (e.g. 4, 11, 12, 13, 18, 19, 23, 25). Phenomena that are relatively infrequent and randomly spread across the world’s languages but frequent and geographically concentrated in a specific group of languages provide evidence of a historical connection — be it areal or genealogical — between the languages concerned (Croft 1990: 206-207). The strength of the argument increases when a number of features correlate in a particular part of the world, but not in the world as a whole.

It is important to note that the typological similarities among the Transeurasian languages are accompanied by a significant number of correspondences in the lexicon (see Robbeets 2005) as well as in verb morphology (see Robbeets 2007a/b, 2010, 2012) in such a way that — in my own judgement — these languages are likely to be genealogically related. The most plausible family tree, representing the overall relationships is given in Figure 1. The affiliation of the Transeurasian languages remains debated, but even critics such as Janhunen (1996: 220) would agree that before the first millennium BCE the homelands of the individual language families concerned were all located in a compact area in southern Manchuria, along with the homelands of Ainuic and Nivkh speakers.

Figure 1.  Family tree of the Transeurasian languages

Although some of the shared features discussed above, such as 11. inclusive-exclusive distinction in first person plural pronouns; 14. partial emphatic reduplication of nominal property words or 18. mi-Ti opposition in first vs. second singular person pronouns are
almost certainly contact-induced, others appear to be the residue of common ancestral features, as suggested by the following six observations.

1. Geography: isolated position of Japanese Although the Sea of Japan and the Tsushima Strait form a strong geographical boundary separating Japanese from the other Transeurasian languages, Japanese is typologically closer to the Transeurasian languages than geographically less isolated languages such as Ket, Yukaghir, Ainu and Nivkh. Even within a prehistorical contact scenario, this suggests that the Transeurasian characteristics in Japanese did not exclusively arise through diffusion because Ainuic and Nivkh were also present in southern Manchuria.

2. History: older varieties are more prototypically Transeurasian A comparison of typological uniformity between historical and contemporary stages of the languages investigated suggests that Transeurasian areality has decreased over the last millennium. While influences diffusing from adjacent areas such as Mainland Southeast Asia, Siberian and Uralic have demonstrably displaced earlier Transeurasian features in certain contact zones, I find no evidence of Transeurasian features having displaced earlier Chinese, Siberian or Uralic features inherent to the continuum from Japanese to Turkic. Among the examples of displacement of features in contact zones, for instance, we find that initial velar nasals have developed in Tungusic under Siberian influence (5), simple tone systems and classifiers have developed in Japanese and Korean under Chinese influence (2 /17), palatal harmony has developed in Turkic under Uralic influence (4). We furthermore note that nominal encoding of property words has increased in Turkic, Mongolic and Tungusic under Uralic influence, while verbal encoding has increased in Japanese and Korean under Siberian and Southeast Asian influence (12). However, we find no evidence of Transeurasian features entering, for instance, from the Turkic languages and diffusing all over the Transeurasian area, while displacing original and prototypical Sinitic features. This suggests that Transeurasian features are inherent to these languages.

3. Distribution: maximal coherence in Mongolic and Tungusic Maximal structural uniformity is found in the Mongolic and Tungusic languages. This distributional pattern conforms to the expectations for the Mongolic languages within a diffusional scenario, since they constitute the center of the linguistic continuum, but it is not what one would expect for the Tungusic languages, extending towards the northeastern periphery. The structural coherence in Mongolic and Tungusic recalls the separation of Mongolo-Tungusic in Figure 1 as a distinct genealogical unit.

4. Distribution: en-bloc reduction of features in Korean and Japanese Similarly, the collective rather than gradual reduction — if not slight increase — of features as we proceed from Korean to (Old) Japanese is not what we would expect within a scenario of gradual diffusion. It is furthermore difficult to explain how some Transeurasian features, such as 10 and 19, that show a gap in Korean, have diffused into Japanese without a Korean intermediary.

5. Cyclicity: recurrent grammaticalization The features involving shared patterns of grammaticalization in Section 2.5. are particularly good candidates for genealogically motivation because they are recurrent in different forms and at various chronological
stages of the same language. Aikhenvald (2013) characterized contact-induced grammaticalization as “change against the grain” or atypical grammaticalization, while she regarded genealogically motivated grammaticalization as “change that reinforces similarities” because it tends to maintain uniformity between related languages. Given that languages tend to renew their formal encodings in cyclic processes of grammaticalization while maintaining their inherited grammatical categories, new forms are thus expected to grammaticalize along shared conceptual pathways to restore old categories (Heath 1998: 729). Consequently, genealogically motivated grammaticalization is expected to recur on different formal encodings at various points in time, while contact-induced grammaticalization is expected to be restricted to a single formal encoding (or to a very limited number of encodings) during a certain period of contact. The repeated waves of grammaticalization and replacement involved in features 25 to 27 imply that the parallel patterns are genealogically motivated.

6. **Isomorphism: shared features combine with formal correspondences** The observation that some structural features shared among the Transeurasian languages combine with a formal correspondence of the marker reflecting the particular feature is also indicative of genealogical retention. This is, for instance, the case for 9. the non-verbal strategy of verbal borrowing employing a deverbal noun suffix of the common shape *-lA- (Tk. -lA-, Khal. -l-, Ud. -lA-, J -r(a)-) to accommodate for verbal borrowings; 19. the formation of a secondary oblique stem of personal pronouns through a common suffix *-n- in all Transeurasian languages, except Korean; 25. direct insubordination involving deverbal noun and finite suffixes of the common shape *-rA-, *-mA, *-n, *-xA ~ *-kA and *-sA (Robbeets 2009, forthcoming a, c) across all Transeurasian languages and 26. the grammaticalization from negative verb to verbal negator, involving common negative verbs of the common shape *-ana-, *e- and *ma- across the Transeurasian languages (Robbeets forthcoming b). In instances like these in which isostructuralism coincides with form-function isomorphism, the structural correspondence is likely to be genealogically motivated, especially when it concerns an instance of shared grammaticalization (Robbeets 2013). Note that the Uralic languages also display oblique personal pronouns in -n-, direct insubordination in *-k, *-mA and *-sA and grammaticalization of negative verbs in *e-, an observation which seems to point to remote genealogical ties between the Uralic and the Transeurasian languages.

4 **Conclusion**

In this chapter, I have tried to show that the Transeurasian languages form an internally homogeneous linguistic continuum. For this purpose, I have examined the areal concentration of 27 features in the Transeurasian languages, providing a typological profile of some contemporary varieties in relation to historical stages of the languages involved and to selected languages immediately outside the continuum. Comparison with neighbors to the north (Yeniseic, Yukaghiric), south (Sinitic, Austronesian), east (Nivkh, Ainu) and west (Uralic) makes it possible to set up boundaries which delimit the Transeurasian proto-type. Along the margins of the Transeurasian continuum, I have found examples of gradual loss of Transeurasian features in the western and eastern peripheries, as well as gradual adoption of Transeurasian features, as in the case of Mandarin. The data further suggest that the Transeurasian continuum in its turn is part of
a larger Uralic-Yukaghiric-Transeurasian belt of languages, which again is part of a larger area of North Asian languages. Although it is meaningful to apply the concept of "areality" to the Transeurasian languages in the sense of a historically motivated geographical concentration of linguistic features, I prefer avoiding the label "area" with reference to these languages because this would imply that all shared properties are the result of diffusion. Observations relating to geography, history, distribution, cyclicity of grammaticalization and combined isomorphism indicate that this is not the case.

A fuller study would need to take more feature values into account and to insert a larger variety of Transeurasian languages as comparative points. Neighboring languages should also be more diversified and adjacent languages in the west such as Indo-European languages or languages of the Caucasus region should be included. One should also pay attention to structural dependencies between the features and to considering whether particular features can be easier accounted for by diffusion or by genealogical retention. For the latter purpose, it would be particularly interesting to take common diachronic mechanisms, such as shared patterns of grammaticalization into fuller account. Although this chapter perhaps raises as many new questions as it answers, I hope to have contributed here to the understanding of areality among the Transeurasian languages.

**Abbreviations**

a) *Linguistic forms*

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL</td>
<td>ablative</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ADD</td>
<td>additive</td>
</tr>
<tr>
<td>ADN</td>
<td>adnominalizer</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CLASS</td>
<td>classifier</td>
</tr>
<tr>
<td>COM</td>
<td>comitative</td>
</tr>
<tr>
<td>COMP</td>
<td>comparative</td>
</tr>
<tr>
<td>CONV</td>
<td>converb</td>
</tr>
<tr>
<td>COND</td>
<td>conditional</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DUR</td>
<td>durative</td>
</tr>
<tr>
<td>ENCL</td>
<td>enclitic</td>
</tr>
<tr>
<td>FIN</td>
<td>finite</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>HON</td>
<td>honorific</td>
</tr>
<tr>
<td>INCH</td>
<td>inchoative</td>
</tr>
<tr>
<td>INTER</td>
<td>interrogative</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>NEG</td>
<td>negative</td>
</tr>
<tr>
<td>NML</td>
<td>nominalizer</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POL</td>
<td>polite</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>PROC</td>
<td>processive</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>PT</td>
<td>particle</td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
</tbody>
</table>

b) *Languages*

<table>
<thead>
<tr>
<th>Code</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ain.</td>
<td>Ainu</td>
</tr>
<tr>
<td>Ch.</td>
<td>Mandarin Chinese</td>
</tr>
<tr>
<td>Evk.</td>
<td>Evenki</td>
</tr>
<tr>
<td>J</td>
<td>Japanese</td>
</tr>
<tr>
<td>K</td>
<td>Korean</td>
</tr>
<tr>
<td>Khal.</td>
<td>Khalkha</td>
</tr>
<tr>
<td>Khan.</td>
<td>Khanty</td>
</tr>
<tr>
<td>Ket</td>
<td>Ket</td>
</tr>
<tr>
<td>Ma.</td>
<td>Manchu</td>
</tr>
<tr>
<td>MK</td>
<td>Middle Korean</td>
</tr>
<tr>
<td>MMO.</td>
<td>Middle Mongolian</td>
</tr>
<tr>
<td>Niv</td>
<td>Nivkh</td>
</tr>
<tr>
<td>WMo.</td>
<td>Written Mongolian</td>
</tr>
<tr>
<td>OJ</td>
<td>Old Japanese</td>
</tr>
<tr>
<td>OT</td>
<td>Old Turkic</td>
</tr>
<tr>
<td>pJ</td>
<td>proto-Japonic</td>
</tr>
</tbody>
</table>
References


Bugaeva, Anna 2013. Causative constructions in Ainu: A typological perspective with remarks on diachrony. [Ms.]


<table>
<thead>
<tr>
<th>pK</th>
<th>proto-Koreanic</th>
<th>pTk</th>
<th>proto-Turkic</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMo</td>
<td>proto-Mongolic</td>
<td>Ruk.</td>
<td>Mantauran Rukai</td>
</tr>
<tr>
<td>pTg</td>
<td>proto-Tungusic</td>
<td>Yuk.</td>
<td>Yukaghir</td>
</tr>
</tbody>
</table>


Maslova, Elena 2003a. A Grammar of Kolyma Yukaghir (Mouton Grammar Library 27.) Berlin: Mouton de Gruyter


Robbeets, Martine forthcoming a. Insubordination and the establishment of genealogical relationship. In: Evans, Nicholas and Watanabe, Honore (eds.) Dynamics of Insubordination. (Typological Studies in Language.) Amsterdam: John Benjamins


Acknowledgement

The research presented here was realized with financial assistance of the DFG (Deutsche Forschungsgemeinschaft), supporting a collaborative research project “Die transeurasiatischen Sprachen: Kontakt in der Familie”, at the University of Mainz from January 2010 to December 2013. I thank Anna Bugaeva, Fubito Endo, Andrey Filchenko, Ekaterina Gruzdeva, Seongyeon Ko, Fukui Rei and Elisabeth Zeitoun as well as the editor, Raymond Hickey, for their helpful feedback.
22  The changing profile of case marking in the northeastern Siberia area

Gregory D. S. Anderson

1 Introduction

A range of phonological and morphosyntactic features can be found across various genetically unrelated Siberian languages that do not cluster in such manner outside of the region, and so it has been argued that Siberia exhibits the core features of a linguistic area as conventionally understood (Anderson 2001, 2003, 2006). A closer inspection of the features discussed reveals that it is primarily the northeastern part of Siberia where the majority of the languages with the highest concentration of these features are found. In this paper I examine several features found widely among the languages of northeastern Siberia relating to case marking and discuss possible developmental histories for certain of these. Like most linguistic areas there is no one source for the full set of features found. In individual instances, features may be characteristic of proto-languages of specific genetic units and not others, and thus diffusion, borrowing, or interference through shift from one to another may explain their appearance, when the comparative data suggest such (a) feature(s) may be non-original. In particular, all of the features discussed in Anderson (2006) can be reconstructed to Proto-Northern Tungusic (and most to Proto-Tungusic) and as such, Northern Tungusic languages appear to be central to the development of the Northeastern Siberia linguistic area. Further, Northern Tungusic languages are by far the most widespread in the region and bilingualism between Evenk- or Even- with other languages that an explanation of Northern Tungusic as the vector of diffusion in Siberia would necessitate is or was common throughout the region (e.g. Evenki-/Even- with -Sakha/Yakut, -Yukaghir, or -Koryak, etc.). However, not all such case oppositions found in northeastern Siberia that are diagnostic of the area per se are of Tungusic origin and like most linguistic areas, the history of direct borrowing, calquing, diffusion and convergence is complex and varied, and in particular the role of Yukaghiric languages in the development of the linguistic area should not be downplayed. In this study I present four such features found in northeastern Siberia languages with respect to case: i) an instrumental vs. comitative contrast, ii) a prolatative case form, iii) a dative vs. allative case contrast, and iv) the use of particular case forms with semi-finite verb forms to create various functional types of subordinate clauses. Further, with respect to the use of case marking to encode subordination, the pre-Russian contact profile of northeastern Siberian languages (and thus of the northeastern Siberia linguistic area) differs from contemporary usage norms due to the fact that all Siberian languages have experienced intense language contact with Russian in the past two to five centuries. Russian influence

1 There is also local bilingualism between Evenki and Sel’kup, Evenki and Khanty, and even in specific instances, Evenki-Ket as well (Edward Vajda, p.c. 2013). The complex history of Tungusic-Turkic contacts and interactions in the history of the development of the Dolgan can be found in Stapert (2013), while Pakendorf (2007) addresses similar genetic and linguistic inter-connections between Turkic and Tungusic elements in the history of Sakha (Yakut).
is now pronounced across most indigenous Siberian languages, such that a new structural convergence type is emerging (Grenoble 2000, 2010, 2012a, 2012b, forthcoming; Anderson 2005, 2015; Oskolskaya and Stoynova 2013) that differs in obvious ways from the previous typological profile, specifically in the morphosyntax of complex sentences. This shift is briefly examined at the conclusion of this study.

1.1 Indigenous languages of Siberia

Indigenous Siberian languages currently number slightly more than three-dozen in total, though we know there were once more as a number have not survived into the twenty-first century. Alas it seems that most indigenous Siberian languages will not likely survive the twenty-first century. While numbering only a handful of languages, there is still a considerable amount of genetic linguistic diversity attested within the ranks of indigenous Siberian languages. These are two of the reasons that (eastern) Siberia is considered to be a language hotspot (Anderson 2010, 2011). Indeed, no fewer than twelve independent genetic units and one genetically unclassifiable mixed language may be reckoned within Siberia, almost half of which only occur in this region. A list of currently extant or recently extinct Siberian languages with their genetic affiliations is given in Table-1 (roughly structured in a west to east configuration). Traditionally, there were local connections and interactions between different groups such that chains of local bilingualism extend(ed) across almost the entire region. Over the past 200 years or so, bilingualism in Russian has extended to be now nearly universal in indigenous Siberian communities.

<table>
<thead>
<tr>
<th>Samoyedic:</th>
<th>Nenets</th>
<th>Enets</th>
<th>Nganasan</th>
<th>Sel'kup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ob-Ugric: 6</td>
<td>Khanty</td>
<td>Mansi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeniseic:</td>
<td>Ket</td>
<td>Quu</td>
<td>Altai Telengit</td>
<td>Teleut</td>
</tr>
<tr>
<td>Turkic:</td>
<td>Qumandy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2 I take Siberia in this work to mean the entire Asian part of Russia east of the Ural mountains, and thus subsuming the Soviet and post-Soviet Federal Russian regions called Siberia and the Far East.

3 Highly mobile and widespread, Evenki of the Tungusic genetic unit disrupts this distribution a bit, as does the enormous territory the Sakha occupy. In addition to these still extant languages, the following now extinct languages have also been considered in this study: Sayan Samoyed (Samoyedic) Arin, Assan, Kott, Pumpokol, Yugh (Yeniseic), Chuvan, Omok (Yukagiric), Sireniki (Eskimoic). There are conflicting reports on whether Kerek of the Chukotko-Kamchatkan genetic unit still has any living speakers or not, but given how recently it must have gone extinct if it has, there are quite likely to be semi-speakers remaining so it is listed among the still extant languages in Table-1.

4 Studies on aspects of these interactions including lexical and structural interference can be found in such works as Tsyndendambaev (1981) or Pakendorf (2007).

5 Tuva in southern Siberia just north of Mongolia stands out as the one area with extensive monolingualism in the indigenous language.

6 Khanty and Mansi each represent several languages or mutually unintelligible dialects. Some Mansi varieties and probably some Khanty ones too are currently extinct. Some eastern Ob-Ugric groups interacted with western Evenki.
Table 1. Languages of Native Siberia

1.2 Case systems of the languages of northeastern Siberia

Northeastern Siberian languages all make use of different nominal case forms, grammatical, locational/directional, instrumental, etc. Many have six to twelve case forms attested. Indeed, with respect to certain characteristic features of grammatical case marking, the languages of northeastern Siberia have little in common, as, for example, northern Chukotko-Kamchatkan languages show ergative alignment and case marking, but Northern Tungusic and Turkic languages show accusative alignment and case marking, and Yukaghiric has a system of case marking based on the relative focal status or discourse-salience of arguments. However, three specific case functions and contrasts do typify the languages of northeastern Siberia, and help to unite these groups in an areal cluster. Like all individual features in any putative linguistic area, any one feature may extend beyond this area, but it is the clustering of them that delineates northeastern Siberian languages from adjacent areas. These three characteristic case phenomena include (i) the functional and formal contrast between instrumental and comitative, (ii) the presence of the prolative or prosecutive case, and (iii) a dative vs. allative contrast.

In general, one can observe a cline in number of contrastive case forms within a given Siberian language from lower to higher when moving from the southwest to the northeast. The most reduced case inventories among indigenous Siberian languages are to be found outside the northeastern region that is the subject of the present study in the South Siberian Turkic languages and various western Mansi and Khanty varieties. The most developed case systems on the other hand are to be found in the northeastern Siberia region, in the Northern Tungusic languages Evenki and Even, and in various Koryak dialects.

---

7 Probable relatively recent new-comers Eskimoic (de Reuse 1994) and Aleut (Bergsland 1997) are exceptions to this.

8 Formally speaking, what have been called case suffixes in Yeniseic (Georg 2007; Anderson 2004) should be analyzed formally as postpositional clitics (Vajda 2008, 2012). Thus, this is more evidence that Yeniseic is anomalous within the Siberian area on numerous levels, and that shared features are often transparently distinct or different in realization, or probably calques on
2 Instrumental vs. comitative case in northeast Siberian languages

Many different types of case functions can be rendered in English using the preposition ‘with’ (or ‘by’), but two salient functions or roles, that of Instrument and Accompaniment are generally kept formally distinct in northeastern Siberia languages through the use of two separate case suffixes. Both an instrumental and a comitative case form are attested in the two Yukaghiric languages. Wadul (2) and Odul (3), have cognate elements, and this suggests that one should reconstruct this case opposition to Proto-Yukaghiric. Functionally the comitative encodes not only accompaniment but also possession, and further represents the default way of conjoining two nouns in Yukaghir.

(2) i. Yukaghir (Wadul) (Krejnovich 1982: 49-50)

\[ \text{sa-lek} \quad \text{pajdruk} \]

stick-INS hit

‘hit with a stick’

ii. Yukaghir (Wadul) (Krejnovich 1982: 46)

\[ \text{ile-jej} \quad \text{la:me} \quad \text{me-qaldej-ŋi} \]

reindeer-COM dog PV-run.off-3PL

‘the dog ran off with the reindeer’, ‘the dog and the reindeer ran off’

(3) i. Yukaghir (Odul)

\[ \text{tʃo:joje-le} \quad \text{nûme-jej} \]

knife-INS dwelling-COM

‘with a knife’ ‘with a dwelling, he has a dwelling’

ii. Yukaghir (Odul) (Krejnovich 1982: 45)

Only the most northeastern languages within the Turkic family, Dolgan and Yakut (Sakha) show such an opposition between instrumental and comitative cases. Two comitative case forms are actually attested in Dolgan (4)-(5), which contrasts with a single instrumental form (6). The second comitative appears to be cognate with elements in other Siberian Turkic languages that mark attributive or possessive adjectives.

(4) Dolgan (Ubrjatova 1985: 122)

\[ \text{o} \text{ʃ} \text{o} \text{lum} \quad \text{beye-lii} \text{n} \quad \text{ooʃpu-ur} \quad \text{kìh-a} \quad \text{kihi-leek} \quad \text{olor-or} \]

child-COM self-COM play-PRS daughter-3 person-COM.II sit-PRS

‘he is playing with the child’ ‘his daughter is sitting with the person’

(5) Dolgan

\[ \text{munu} \quad \text{ikki ilii-tinen} \quad \text{kus-put} \]

this.ACC two hand-3.INS grab-PST.II

‘he grabbed this with both hands’ (Ubrjatova 1985: 121)

areal norms (cf. Fortescue 1998), with Tungusic or most often Sel'kup influence the vector to explain such developments.
Sakha or Yakut (7)-(8) shows a similar opposition between comitative and instrumental case forms, although the form of the possessive comitative in Yakut appears to be cognate with the possessive instrumental in Dolgan. Note that Tenishev et al. (1988) reconstruct the contrast between instrumental and comitative for Proto-Turkic, but the older etymological instrumental -In is retained mainly only in lexicalized adverbials like Tofa *kuhüm ‘during the winter’ or Turkish *günüzün ‘during the day’. The earlier comitative, itself most likely a fused version of the postposition ‘(together) with’, has shifted into the function of an instrumental case in languages such as Xakas (Anderson 1998a), where it is realized as –naŋ, and is also to be found in this function in Yakut as well.

(7) i. Yakut (Sakha) (Ubrjatova 1985: 124) ii. Yakut (Sakha) (Ubrjatova 1985: 123)

\[
\begin{align*}
\text{Alasov} & \text{ ije-}t\text{iniin} \quad \text{kel-le} \quad \text{Alasov} \text{ emeexsin-}t\text{iniin} \quad \text{kel-li-ler} \\
\text{Alasov} & \text{ mother-3.COM} \quad \text{come-PST} \quad \text{Alasov} \quad \text{old.woman-COM} \quad \text{come-PST-3PL}
\end{align*}
\]

‘Alasov came with his mother’ ‘Alasov came with the old woman’

(8) Yakut (Sakha)

\[
\begin{align*}
\text{kini}, \quad \text{araa}h\text{a}, \quad \text{eye-nen} \quad \text{t}o\text{mn-}r \quad \text{ete}
\end{align*}
\]

‘he, by all appearance, would return peacefully’ (Petrov 1984: 49)

A formal and functional opposition between an instrumental and a comitative case was characteristic of Proto-Chukotko-Kamchatkan. As in Dolgan, there are actually two comitative forms, sometimes called the first and second comitative, other times the sociative and the comitative. The instrumental in Chukotko-Kamchatkan is formally a suffix and in some languages functions as an ergative case as well, while the comitative/sociative forms are formally circumfixes. In Chukchi (9), this circumfix consists of a ga- prefix followed by the stem, followed by the instrumental suffix for the so-called first comitative or sociative, or by the suffix –ma for the so-called second or (plain) comitative. In Chukchi, the (second) comitative is formally a phrasal-type circumfix or prefix-cum-enclitic, as seen in (9ii).

(9) i. Chukchi (Skorik 1986: 107) ii. Chukchi (Skorik 1986: 107)

\[
\begin{align*}
\text{ənp+natg-ərgəna-t qorat} & \quad \text{ga-npənatg-ərgəna-qora-ma} \\
\text{old.man-PL reindeer-PL} & \quad \text{COM-old.men-reindeer-COM}
\end{align*}
\]

‘the old men’s reindeers’ ‘with the old men’s reindeers’

When examining the broader comparative picture within Chukotko-Kamchatkan, one finds a variety of different prefixal elements used within the circumfix case marker (10), e.g. (g)a\text{awun-} ~ gA- in the comitative and gA- ~ gAyqə- in the sociative in Koryak.

(10) i. Chavchuven Kamen Palana Chukchi PNCK

\[
\begin{align*}
\text{INS} & \quad -(t)A & \quad -(t)a & \quad -(t)A \sim -(t)a & \quad -(t)A & \quad -(t)A \\
\text{COM-1} & \quad gA- -(t)A & \quad gA- -(t)A & \quad gA- -(t)A & \quad gA- -(t)A & \quad gA- -(t)A \\
\text{COM-2} & \quad \text{gawun-} -(t)A & \quad \text{gawun-} -(t)A & \quad \text{gawun-} -(t)A & \quad \text{gawun-} -(t)A & \quad \text{gawun-} -(t)A
\end{align*}
\]

(Sternitskij 1994: 187)

ii. Koryak Kerek Chukchi Al’utor Itel’men
The sociative in Koryak (Zhukova 1972: 120) is generally used with animate referents, while the comitative rather appears with inanimates, but this restriction is not rigid or absolute (11)-(12). Specifically, with inanimates the sociative is said to encode a greater degree of connectedness (13)-(14). Further, as (13)-(14) also demonstrate, the comitative and sociative belong to different harmonic classes in Koryak, the comitative belonging to the strong (roughly [+ATR]) class and the sociative to the weak harmonic class (roughly [-ATR]).


\[
\begin{align*}
&g^{-}a^{-}kaj^{-}a \sim gayq^{-}a^{-}kaj^{-}a \quad geyq^{-}miml^{-}e \quad ga^{-}javiqqal^{-}u^{-}ta \\
&\text{COM-bear-SOC} \quad \text{COM-bear-SOC} \quad \text{COM-water-SOC} \quad \text{COM-girl-SOC} \\
&\text{‘with the bear’} \quad \text{‘with water’} \quad \text{‘with (his) daughter’}
\end{align*}
\]

(12) i. Koryak ii. Koryak

\[
\begin{align*}
&gaw^{-}n^{-}kaj^{-}i^{-}ma \quad gaw^{-}n^{-}meml^{-}oma \\
&\text{COM-bear-COM} \quad \text{COM-water-COM} \\
&\text{‘with the bear’} \quad \text{‘with water’}
\end{align*}
\]

(Zhukova 1972: 120)

(13) Koryak

\[
\begin{align*}
&e^{f}gi \ \anno \ geyq^{-}migq^\text{-}\text{oj}^{-}e \quad ku\text{-}qeq\text{-}\text{e}^{-}y \quad \text{not}^{-}\text{ja}^{-}y \\
&\text{COM-gun-SOC} \quad \text{went} \quad \text{tundra-ALL} \\
&\text{‘today he went to the tundra with his gun’}
\end{align*}
\]

(Zhukova 1972: 121)

(14) Koryak

\[
\begin{align*}
&\ \ anno \ gaw^{-}n^{-}melgq^\text{-}\text{oj}^{-}y^{-}ma \quad gap\text{-}qalqalin \\
&\text{COM-gun-COM} \quad \text{drowned} \\
&\text{‘he drowned with his gun’}
\end{align*}
\]

(Zhukova 1972: 121)

The instrumental appears to be the oldest of the forms since it has shifted into marking a grammatical function— as an ergative marker of an A argument—thus it may be possible that the comitative forms were innovated secondarily. Note in this regard that in Koryak no comitative case forms of personal pronouns are used, but rather the postposition omakaŋ appears together with locative case marking.

(15) i. Koryak ii. Koryak

\[
\begin{align*}
&gom^{-}\text{-}sk \text{ omakaŋ} \quad muy^{-}\text{-}sk \text{ omakaŋ} \\
&\text{I-LOC} \quad \text{we-LOC} \\
&\text{together.with} \quad \text{together.with} \\
&\text{‘with me’} \quad \text{‘with us’}
\end{align*}
\]

(Zhukova 1972: 45)
Northern Tungusic languages likewise exhibit an instrumental vs. comitative opposition.9 This contrast is likely reconstructible to the Proto-Tungusic stage. Note that as in Yukaghir, the comitative is the default means of conjoining NPs.

(16) i. Evenki

\[ siː  tara  bər-i-t-piː  \quad \text{garpa-kal} \]
\[ \text{you that gun-INS-REFL shoot-FUT.IMPV.2SG} \]
\[ '\text{shoot that one with your gun}' \]

(Bulatova and Grenoble 1999: 9)

ii. Evenki

\[ asiː  kɨnə-ː-l-dзi  \quad \text{amiːn-dulaː-βiː  \quad is-ɨlaː-n} \]
\[ \text{woman ski-PL-INS father-LOC-POSS go-PST-3SG} \]
\[ '\text{the woman went up to her father on skis}' \]

(Bulatova and Grenoble 1999: 9)

(17) i. Evenki

\[ fтαmiː  \quad ənɨkən-nuːn  \quad \text{bira-βa  \quad daβ-dзa-ra-Ø} \]
\[ \text{female.deer fawn-COM river-ACC cross-IMPV-AOR-3SG} \]
\[ '\text{the female deer crosses the river with the fawn}' \]

(Bulatova and Grenoble 1999: 12)

ii. Evenki

\[ biː  əkin-nuːn-miː  \quad təβlaː-m \]
\[ \text{I sister-COM-REFL.SG collect.berries-1SG} \]
\[ '\text{I went with my sister to gather berries}' \]

(Bulatova and Grenoble 1999: 12)

It is difficult to ascribe the opposition of an instrumental case with a comitative one in any of the northeastern Siberian language families to (Northern) Tungusic influence – either directly or indirectly – as there is evidence for such an opposition in Turkic, Yukaghiric and Chukotko-Kamchatkan proto-languages as well, but it is certainly likely that the presence of the opposition in Tungusic may have helped maintain or expand it in the other groups in the region. The Proto-Yukaghiric form is of such shallow time depth as to reveal very little. The variation found in Chukotko-Kamchatkan suggests the feature is not a stable one in this genetic unit. Contact with Tungusic or indeed Yukaghiric might have played a role in one or another of the individual developments. Finally, it is only the northeasternmost Turkic language Yakut and Dolgan with heavy Tungusic adstratum or substratum (Stapert 2013, Pakendorf 2007) where the older ancestral Turkic instrumental

---

9 Note that some researchers have questioned whether the comitative in Evenki formally belongs to the case system and is not rather some postpositional clitic or relator/relational noun, as it can also attach to an accusative form with conjoined definite direct object NPs (see Kilby 1980 for more on the general status of cases in Evenki). It is not clear that this really matters synchronically or diachronically, as other case forms in Evenki likely have their origin in clitics even if now integrated into the case system, cf. the discussion of the prolative below, as it appears to have historically included the dative case form within it.
vs. comitative opposition has been maintained with separate, individual case forms, albeit in a reformed or non-etymological manner, and Tungusic influence in these instances has almost certainly played some role. So, in summary the feature is shared by languages of different genetic units throughout the northeast, but only with respect to the Turkic languages is direct influence from Tungusic likely to have had a role in maintaining or reconstituting an old featural opposition. However, given the age of the functional contrast in Turkic, even there Tungusic influence is not an absolutely certain explanation for the observed phenomena.

3 Prolative or prosecutive case in northeastern Siberian languages

The prolative or prosecutive case form (motion ‘along’ or ‘through’) is characteristic of most northeastern Siberian language families. The prolative marked noun often indexes the semantic role of Route. A prolative case form can be safely reconstructed back to Proto-Yukagiric and Proto-Tungusic. Correspondences are straightforward between Tundra (Wadul) and Kolyma (Odul) Yukaghir:

(18) i. Tundra (Wadul) Yukaghir
   *ugurt'e-do-han waːj ek-uo-j*
   legs-3-PROL too hole-STV-ITR(3)
   ‘there was also a hole along his feet’ (Maslova 2003a: 59)

   ii. Wadul
   iii. Wadul
   iv. Wadul
   *enu-pul-ɣan jɑlɥi-pe-ɣan gorot-qan*
   river-PL-PROL lake-PL-PROL city-PROL
   ‘along rivers’ ‘along lakes’ ‘about town’ (Krejnovich 1958: 57)

(19) i. (Kolyma (Odul) Yukaghir
   *tʃuge-de-ɣen qon-ŋi*
   trace-POSS-PROL go-3PL:ITR
   ‘they went along his trace’ (Maslova 2003b: 113)

   ii. Odul
   *Omolon-gen*
   Omolon-PROL
   ‘along the Omolon (river)’ (Krejnovich 1958: 57)

10 Among other roles; for more on the function of prolative and other case forms in Evenki, see Grenoble (forthcoming). In literary Evenki there are two prolative type case forms. However, according to Grenoble (2000: 110), some of the more marginal case forms in the Evenki dialect she described, like the allative-prolative in -kliː, are recognized but not used by the speakers of that particular Evenki variety.
The forms in Yukaghir appear to be built off the locative case form (Maslova 2003a: 56), and thus prolative, which Nikolaeva (2006: 80) reconstructs as *-ŋkən, may well be a secondary development within the history of Yukaghir, albeit one that occurred prior to the Proto-Yukaghir level.

Evenki generally uses the full or long form of the prolative, while in Even, a clipped form in –li is used alongside the fuller form found also in Evenki. Note that the prolative case appears to be built off of the dative case marker in these Northern Tungusic languages.11

(20) **Evenki** (Bulatova and Grenoble 1999: 11)  
(21) **Even** (Malchukov 1995: 9)

<table>
<thead>
<tr>
<th>Evenki</th>
<th>Even</th>
</tr>
</thead>
<tbody>
<tr>
<td>oro-r hoktoron-duki huki-dʒə-tʃə-tin</td>
<td>ʃu-u-li ʃu-u-l-duki</td>
</tr>
<tr>
<td>deer-PL path-PROL run-IMPF-PST-3PL</td>
<td>house-PROL house-PL-PROL</td>
</tr>
<tr>
<td>‘deer were running along the path’</td>
<td>‘along the house(s)’</td>
</tr>
</tbody>
</table>

Koryak of the Chukotko-Kamchatkan family (22) likewise follows the northeastern Siberian areal pattern of having a prolative case marker. Note the different allomorphs of the prolative case marker found in the following Koryak forms. It is not possible at present to determine whether Tungusic or Yukaghiric is the likely source of diffusion of this case into Koryak (or both), but the prolative case feature is most likely not an old one in the history of Chukotko-Kamchatkan languages.

(22) i. **Koryak**  
ii. **Koryak**  
iii. **Koryak**

<table>
<thead>
<tr>
<th>Koryak</th>
<th>Koryak</th>
<th>Koryak</th>
</tr>
</thead>
<tbody>
<tr>
<td>wajam-gapəŋ</td>
<td>ɣawatʃŋ-epəŋ</td>
<td>jaja-jpəŋ</td>
</tr>
<tr>
<td>river-PROL</td>
<td>woman-PROL</td>
<td>house-PROL</td>
</tr>
<tr>
<td>‘along the river’</td>
<td>‘past the woman’</td>
<td>‘past the house’</td>
</tr>
</tbody>
</table>

(Skorik 1986: 95)

Note that Eskimoic language Siberian Yupik (23) has three prolative case forms, encoding singular, dual and plural (viz., -kun/-gnəkun/-təkun (SG/DL/PL)). Such portmanteau number+case marking is not found in other northeastern Siberian languages generally speaking.

(23) **Siberian Yupik**

<table>
<thead>
<tr>
<th>Yupik</th>
</tr>
</thead>
<tbody>
<tr>
<td>juk ɪgyta-quq sna-kun</td>
</tr>
<tr>
<td>man go.along-3 bank/coast/edge-PROL.SG</td>
</tr>
<tr>
<td>‘the man is going along the edge, bank, coast’</td>
</tr>
</tbody>
</table>

(Menovshchikov and Vakhtin 1983: 92).

---

---

11 Regardless of whether it originated in a postpositional or relational noun clitic in pre-Proto-Tungusic that governed and then subsequently included the dative case within it, as the data might suggest, the prolative does appear to be an old case form in Tungusic, and we can project it back to Proto-Tungusic most likely (Aleksander Vovin, p.c. 2014), though Tsintsius (1948) does not reconstruct this explicitly.
Outside of the northeastern areal complex, prolative case can be found in a small number of other Siberian languages as well. Indeed, prolative case can be even reconstructed to Proto-(Northern-)Samoyedic.12

(24) Nenets
   to-wna
   lake-PROL
   ‘along the lake’
   (Prokof’ev 1937a: 26)

(25) Nganasan
   turku-manu
   lake-PROL
   ‘along the lake’
   (Prokof’ev 1937b: 62)

(26) Enets
   Tau-mone
   Nganasan-PROL
   ‘along the Nganasan’
   (Castrén 1854: 177)

A prolative case form is common among the case inventories of Siberian languages, and can even be reconstructed to various proto-languages, viz., Proto- Tungusic or Proto-Yukaghiric (and Proto-Northern Samoyedic). In Proto-Yukaghiric the form appears to be secondarily created off of the locative form. Southern Tungusic languages also use the prolative, but often the case marker itself is not cognate with the Northern Tungusic one, e.g. -kii in Ul’ch (Sunik 1997). With respect to other Siberian languages with prolative markers, Evenki influence is at least plausible for its development in Proto-Ket-Yugh, since the prolative does not appear to be old within Yeniseic per se. In Koryak and Al’utor on the other hand, where Tungusic influence is manifest in multiple ways, diffusion from Tungusic is quite likely, especially given the lack of the case in Chukchi and Itelmen, but, as mentioned above, Yukaghiric influence cannot be excluded in the development of the prolative in Koryak and Al’utor. On the other hand, Tungusic contact is however extremely unlikely for the development of new prolative case semantics with a formally old case marker in southern Siberian Turkic Xakas (Anderson 1998a).

4 Dative vs. allative in northeast Siberian languages

Northeastern Siberian languages frequently exhibit a formal contrast between dative case forms and allative case forms. Allative case canonically encodes the semantic role of Goal and thus generally indexes physical motion towards a location or an entity. Dative case on the other hand typically marks Recipient arguments in ditransitive constructions, or with other roles typically associated with the syntactic relation of ‘indirect object’.

Both a dative case form and an allative case form show cognate elements across the Tungusic languages of Siberia. It is thus relatively straightforward to reconstruct the contrast and the formal markers of this to Proto-Tungusic based on these correspondence. So one finds the contrast in such Siberian Tungusic languages as Nanai (27)-(28), Udihe (29)-(30), Oroch (31)-(32), Even (33)-(34), Negidal (35)-(36) and Evenki (37)-(38). Note that the opposition between dative and allative cases is lacking in certain Tungusic languages spoken outside of the Siberian area, e.g. in Jurchen or Manchu, where Chinese influence is pronounced, and Mandarin Chinese of course lacks such an opposition of case inflections. Also, unlike some other Siberian languages, in the Siberian Tungusic languages that do have the dative and allative contrast, addressees of verbs of speaking

12 However, while non-adjacent at present, there is evidence for earlier Yukaghiric-Samoyedic contacts, and one can not exclude the possibility that the prolative case in Northern Samoyedic reflects this history. For more on the complex periodization of Uralic and Yukaghiric contacts, see Häkkinen (2012).
are typically encoded by the allative, not the dative, while recipients of 'give' (e.g., Udihe *bu*-, Evenki *ani*-) are encoded by the dative. Dative is typically used for both functions or roles in many other languages of the area, including Russian (which of course lacks a formal allative).13

(27) **Nanai**

\[\text{ogda-du} \quad \text{ogda-tʃi}\]

boat-DAT boat-ALL

‘to the boat’ ‘toward the boat’ (Sem 1997: 184)

(28) **Nanai**

\[\text{ogda-du} \quad \text{ogda-tʃi}\]

boat-DAT boat-ALL

‘to the boat’ ‘toward the boat’ (Sem 1997: 184)

(29) **Oroch**

\[\text{ʊgda-du} \quad \text{ʊgda-ti}\]

boat-DAT boat-ALL

‘to the boat’ ‘towards the boat’ (Lebedeva 1997: 225)

(30) **Oroch**

\[\text{ʊgda-du} \quad \text{ʊgda-ti}\]

boat-DAT boat-ALL

‘to the boat’ ‘towards the boat’ (Lebedeva 1997: 225)

(31) **Udihe**

\[\text{bi sin-du xeleb-wa bu-oː-mi}\]

I you-DAT bread-ACC give-PST-1

‘I gave you (some) bread’ (Nikolaeva and Tolskaja 2001: 524)

(32) **Udihe**

\[\text{zaŋə-ziga Moskwa-tigi ēne-zeye-ti}\]

boss-PL Moscow-ALL go-FUT-3PL

‘the bosses will go to Moscow’ (Nikolaeva and Tolskaja 2001: 517)

(33) **Even**

\[\text{d'uu-du} \quad \text{d'uu-tki}\]

house-DAT house-ALL

‘to the house’ ‘towards the house’ (Malchukov 1995: 9)

(34) **Even**

\[\text{d'uu-du} \quad \text{d'uu-tki}\]

house-DAT house-ALL

‘to the house’ ‘towards the house’ (Malchukov 1995: 9)

(35) **Negidal**

\[\text{min-duː} \quad \text{min-tixiː}\]

I-DAT I-ALL

‘to me’ ‘towards me’ (Tsintsius 1997: 197)

(36) **Negidal**

\[\text{min-duː} \quad \text{min-tixiː}\]

I-DAT I-ALL

‘to me’ ‘towards me’ (Tsintsius 1997: 197)

(37) **Evenki**

\[\text{nunartin bajɑːŋn-dʊi oron-mo anɨː-ra}\]

they boy-DAT deer-ACC give-AOR

‘they gave the boy a deer’ (Bulatova and Grenoble 1999: 9)

(38) **Evenki**

\[\text{tirgakaːkin bira-tkiː oloː-moː-sina-β}\]
noon river-ALL fish-go-INCPL-1PL-EX
‘at noon we went to the river to fish’ (Bulatova and Grenoble 1999: 10)

The presence of a fully developed opposition of allative vs. dative cases that can freely attach to both pronouns and nouns in Proto-Northern Chukotko-Kamchatkan is not supported by the data. In Chukchi, the contrast is restricted to pronouns (39) and this likely reflects the earlier distribution found in the proto-language.

(39) Chukchi
\[
gəmək-ə \quad \text{I-DAT} \quad \text{I-ALL}
gəmək-agtə
\]
‘to me’ ‘towards me’

(Kämpfe and Volodin 1995: 86)

Certain Koryak varieties, as well as Kerek, show a more fully developed opposition between dative and allative case forms (40). This more fully developed system may have been extended to all nouns in Koryak, which, like the use of a prolative case, might possibly reflect Northern Tungusic influence in this language. Note that both the dative and the allative in Koryak are dominant or strong suffixes and therefore trigger assimilation to dominant/strong vowel series in the stem, as seen in the examples below and in such forms as mil’ut ‘hare’ > mel’otay [hare-DAT] as well.\(^{14}\)

(40) i. Koryak
\[
wajam-əŋ \quad \text{[river-DAT]}
\]
‘to the river’

ii. Koryak
\[
aŋpetfe-nanay \quad \text{[father-DAT]}
\]
‘to (his) father’

iii. Koryak
\[
aŋpetfe-najtəŋ \quad \text{[father-ALL]}
\]
‘toward (his) father’

< wejem ‘river’ < eŋpitʃ ‘father’

(Stebnitskij 1994: 144; Skorik 1986: 93)

In Al’utor, the allative is lacking altogether, with functions of the Koryak allative marked by the dative.

(41) i. ‘Palana Koryak’/Al’utor
\[
qlegi ?ənpqglawol-əŋ \\
\text{drop by old.man-DAT}
\]
‘drop by and visit at the old man’s’

ii. ‘Palana Koryak’/Al’utor
\[
rara-ŋ bewweevlin \\
\text{house-DAT set.off.for.PST.3}
\]
‘he set off for home’

(Zhukova 1980: 48-49)

---

\(^{14}\) This strong vs. weak harmony system characterizes not only Chukotko-Kamchatkan languages but also Tungusic ones as well, with related phenomena found in other northern Siberian languages as well such as velar height harmony in Sakha (Yakut); see Anderson (1998b). The phonetic or acoustical characteristics that distinguish the two series have been variably described for Even (and Tungusic more generally) as [±ATR] or [RTR] (Ko 2012), or [±pharyngealized]. Aralova, Grawunder and Winter (2012) find that different Even dialects have different systems acoustically speaking, and that the systems mix properties associated with ATR harmony and pharyngealization.
The spread of a full dative vs. allative contrast in Koryak may thus reflect Tungusic influence. Note that Itel’men lacks both allative and dative case forms.

In Yukaghiric the formal contrast is lacking *per se*, but there is an allative relational noun in Kolyma Yukaghir that has come to be grammaticalized in a referential or possessed form that functions as a type of third person singular allative pronominal form. In this manner, the development of a formal opposition between dative and allative may be being constituted in Yukaghir in third singular pronominals at least. Formally speaking, this new allative pronominal form is a relational noun marked in the dative case and this type of development may reflect the same type of historical process that operated to give rise to case forms based on other case forms which is seen throughout the Siberian region (and elsewhere).

(42) **Kolyma (Odul) Yukaghir**

\[
\text{qodo al’-d-in met jaqa-te-m}
\]

how ALL-POSS-DAT/DIR I arrive-FUT-ITR:1SG

‘how shall I approach him?’

(Maslova 2003b: 268)

(43) **Kolyma (Odul) Yukaghir**

\[
\text{tud-in er-ʃuon el+aː-ŋi-le-k qamie-d’e-ŋi-k}
\]


‘do not do anything bad to him, help him’

(Maslova 2003b: 177)

The spread of a dative vs. allative case to Koryak may be the clearest example of the reflection of the influence of Northern Tungusic on the case system of a neighboring language. The incipient opposition in third singular pronominals in Yukaghir likely reflects the same distributional restriction as is found in Chukchi, and may represent an instance of Chukchi influence in Yukaghir, possibly reinforced by the extensive contact of Kolyma Yukaghir with Tungusic Even as well.

5 **Loss of case-marked clausal subordination: The role of Russian contact**

Syntactically, the basic constituent order SOV characterizes all the languages of the Siberian macro-area, though some languages exhibit stronger tendencies to verb-final order than others, and unsurprisingly other constituent orders are available to encode various types of information structure such as term focus. Further, many of the languages of the northeastern Siberia linguistic area are highly endangered languages, and as such show pronounced influence from Russian morphosyntactically and syntactically. This has in part disrupted the relatively strict SOV structure in most discourse or informational structural configurations. Further, case usage norms are converging with those of Russian in many Siberian languages, so that according to Grenoble (2000: 109-110), agents of passives are now marked by instrumental typically in Evenki, not the original dative. Given the dismal state many Native Siberian languages find themselves in, this Russian influence is becoming quite widespread areally such that a newly-emerging structural type can be suggested. At least one such new feature deals with the system of case marking to encode various functional subtypes of complex sentences that once
characterized the languages of the region. I exemplify this with some evidence from complex sentence structure where the the languages of the area are showing shift from the original Siberian type with non-finite subordinate clauses marked with various case forms that encode different functional sub-types of clausal subordination using non-finite predicates to a new Russianized type or syntactic pattern involving adverbial subordinators, complementizers, etc., with finite verbs.

Pre-Russian contact varieties of Siberian languages signal(ed) clausal subordination or dependency in complex sentences through a range of non-finite or semi-finite structures, which non-exhaustively includes various converbs and nominalized verb formations. Macro-areally, Siberian languages tend(ed) to use case morphology to encode specific functional types of subordinate clauses. Several formal sub-types of this type of case-made clausal subordination are attested in Siberian languages; see Anderson (2004) for details on languages of central Siberia, and Anderson (2005) for restructured and original formations in Xakas. In the most basic widespread and commonly attested type, the cases attach to some overtly nominalized form of the verb, variously called a participle, infinitive and so on.

One particularly common construction of this type is the use of a locative case element on a participial form of the verb to create temporarily subordinate clauses. The person/number of the subject is often encoded by possessive morphology in these subordinate clauses. Individual languages vary as to the relative order that the case and possessive marker on these nominalized verbs occur in, e.g. person+case in Kolyma (Odul) Yukaghir (44), or case+person in Tungusic (45-47). In all instances these case and person markers follow a participial suffix.

(44) **Kolyma (Odul) Yukaghir**

\[ qa:\text{-pe-}gi ajli-\text{-de-}ge \quad \text{“el+qon-}y\text{-i-lek”} \quad \text{mon-}de-ge \]

grandfather-PL-POSS forbid-3.NF-LOC NEG-go-PL-PROHIB say-3.NF-LOC

\[ \text{tamun-}gele u\text{o}rpe-\text{p-ki} \quad \text{el+med-o}l-\text{-}ji \]

that-ACC child-PL-POSS NEG-listen-DES-3PL.ITR

‘their grandfather forbids (it), saying “don’t go” but the children do not obey’

(Maslova 2003b: 372)

Many different case forms can be found functioning to encode case-marked clausal subordination of various types throughout the languages of Siberia. Formally of the second type mentioned above, e.g., with the order participle+case+person, one finds both the dative (45) and ablative (46) cases used to mark different kinds of temporally subordinate clauses in Tungusic Evenki, while the prolicative can be used to form dependent clauses with causal semantics (47).

(45) **Evenki**

\[ bira daga\text{dun }o\text{;}ri-du-\text{v} \quad \text{so}t \quad ed\text{umni-}l-le-n \]

river near become-PRTCPL-DAT-1 very blow.wind-INCH-NFUT-3

‘when I found myself near the river, a strong wind began to blow’

(Nedjalkov 1997: 51)

(46) **Evenki**
In Chukchi (48) at an earlier historical period pre-dating Russian penetration into the region, a formally different structure is found from that which typifies Yukaghiric or Tungusic (or Turkic), albeit one that appears to be calqued on these areal norms. Person marking is absent on the verb in same-subject subordinate clauses in Chukchi, as is an overt marker of nominalization (cf. the Enets and Yeniseic data in note 27), and thus the case marker that encodes the specific functional type of subordinate clause attaches directly to an otherwise unmarked verb stem. Like Evenki, different cases encode different functional subtypes of subordination, e.g., allative to encode temporally preceding action, comitative to encode simultaneous action, etc. In each instance, the Chukchi case form attaches directly to the verb stem, and not a nominalized form as in Tungusic, Turkic or Yukaghiric.

(48) i. Chukchi

\[
\begin{align*}
&\gamma\omega\mu\epsilon-\gamma\theta & \text{nel}/\gamma\omega-n & \gamma\omega m-nan \\
&\text{hang.up-ALL} & \text{pelt-ABS} & \text{I-ERG} \\
&\text{t} \omega-\text{ti} \gamma\theta-e-n & \omega\text{weyt}/\gamma\omega n & \\
&1\text{SUBJ-knock.over-PRF-3OBJ} & \text{vessel.ABS} & \\
&\text{‘when I hung up the pelt I knocked over the vessel’} & & (Kämpfe and Volodin 1995: 106)
\end{align*}
\]

ii. Chukchi

\[
\begin{align*}
&\eta\text{ewosqet} & \text{ga-gantow-ma} & \text{kulil}/\text{er} \text{?u-g?i} \\
&\text{woman} & \text{COM-run.away-COM} & \text{cry.out-PRF} \\
&\text{‘the woman cried out while running away’} & & (Kämpfe and Volodin 1995: 54)
\end{align*}
\]

Thus was the state of complex sentence formation in Siberian languages in the pre-colonial or pre-contact period and is still true of the less restructured speech of certain (often the most elderly) individuals within contemporary communities. However, as alluded to above, Russia has extended its territorial, cultural, and linguistic hegemony to cover almost every part of Siberia, and profound ethno-linguistic consequences have been the result. In general it is true that, except possibly Sakha, all indigenous northeastern Siberian languages and communities are in catastrophic decline. This can be seen when looking at census data over the past half-century (1959-2010), where even in these highly charged and politically sensitive accounts we can see the cumulative effects of ethnic
shame and language shift that almost all Native Siberian language communities have been experiencing over this period (Anderson 2015). In Table 2 are offered data on the self-reported ethnic group affiliation for demographically small Native Siberian communities during the last four Soviet censuses (1959, 1970, 1979 and 1989) and the first two Russian Federation censuses (2002 and 2010) for communities found in northeastern Siberia.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Evenki</td>
<td>38396</td>
<td>35527</td>
<td>30163</td>
<td>27294</td>
<td>25471</td>
<td>24151</td>
</tr>
<tr>
<td>Khanty</td>
<td>30943</td>
<td>28678</td>
<td>22521</td>
<td>20934</td>
<td>21138</td>
<td>19410</td>
</tr>
<tr>
<td>Even</td>
<td>21830</td>
<td>19071</td>
<td>17199</td>
<td>12523</td>
<td>12029</td>
<td>9121</td>
</tr>
<tr>
<td>Chukchi</td>
<td>15098</td>
<td>15767</td>
<td>15184</td>
<td>14000</td>
<td>13597</td>
<td>11727</td>
</tr>
<tr>
<td>Koryak</td>
<td>7953</td>
<td>8743</td>
<td>9242</td>
<td>7879</td>
<td>7487</td>
<td>6287</td>
</tr>
<tr>
<td>Dolgan</td>
<td>7885</td>
<td>7261</td>
<td>6945</td>
<td>5053</td>
<td>4877</td>
<td>3932</td>
</tr>
<tr>
<td>Itelmen</td>
<td>3193</td>
<td>3180</td>
<td>2481</td>
<td>1370</td>
<td>1301</td>
<td>1109</td>
</tr>
<tr>
<td>Yupik</td>
<td>1738</td>
<td>1750</td>
<td>1719</td>
<td>1510</td>
<td>1308</td>
<td>1118</td>
</tr>
<tr>
<td>Yukaghir</td>
<td>1603</td>
<td>1509</td>
<td>1142</td>
<td>835</td>
<td>615</td>
<td>442</td>
</tr>
<tr>
<td>Aleut</td>
<td>482</td>
<td>540</td>
<td>702</td>
<td>546</td>
<td>450</td>
<td>421</td>
</tr>
<tr>
<td>Chuvan</td>
<td>1002</td>
<td>1087</td>
<td>1511*</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 2. Self-reported ethnic identity of Native Siberians 1959-2010

One issue should be mentioned first in order to correctly interpret these numbers, and that is, all demographic factors being equal, population groups should be experiencing gradual increases; if the birth rate is lower than the death rate, the group is in demographic decline. Only four of the northernmost groups Evenki, Even, Dolgan and Yukaghir have shown increasing populations throughout the period between 1959-2010. All of these communities occupy vast and sparsely populated regions, and one is inclined to believe that they reflect real demographic trends of gradual population increase. Ethnic shame plays a role in the decline of self-reported ethnic group membership, rather than actual demographic collapse, for example, in the reported decline of Koryak and Chukchi in 2010. Sometimes trends of this sort are reversed as happened in the case of the Khanty when comparing the numbers of the 1970 and 1979 censuses, with that of 1989. The rise and fall in the reported numbers thus reflect sociological rather than demographic realities for the most part.

Self-reported language use also shows decline between 1959 and 2002 as seen in Table-3. Here the cumulative effects of ethnic shame and actual language shift can be seen dramatically in the massive decline in percentages of self-reported language use across the tables right to left.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dolgan</td>
<td>67.0%</td>
<td>81.7%</td>
<td>90.0%</td>
<td>89.8%</td>
<td>93.9%</td>
</tr>
<tr>
<td>Chukchi</td>
<td>49.1%</td>
<td>70.3%</td>
<td>78.2%</td>
<td>82.6%</td>
<td>93.9%</td>
</tr>
<tr>
<td>Koryak</td>
<td>34.5%</td>
<td>52.4%</td>
<td>69.0%</td>
<td>81.1%</td>
<td>90.5%</td>
</tr>
<tr>
<td>Yupik</td>
<td>23.4%</td>
<td>51.6%</td>
<td>60.7%</td>
<td>60.0%</td>
<td>84.0%</td>
</tr>
<tr>
<td>Even</td>
<td>37.6%</td>
<td>43.9%</td>
<td>56.9%</td>
<td>56.0%</td>
<td>81.4%</td>
</tr>
</tbody>
</table>

15 Sufficient 2010 census data have not yet been available to me.
Yukaghir 40.0% 32.8% 37.5% 46.8% 52.5%
Evenki 21.3% 30.4% 42.8% 51.3% 54.9%
Aleut 32.4% 28.3% 17.8% 21.8% 22.3%
Itelmen 12.1% 19.6% 24.4% 35.7% 36.0%
Yakut 102.8%!! 94.0% 95.3% 96.3% 97.6%

Table 3. Self-reported percentage of language use in censuses 1959-2002

Sources for Table-3:
(http://www.perepis2002.ru/), http://www.hum.uit.no/a/trond/s70.html,
http://www.hum.uit.no/a/trond/s79.html, http://www.hum.uit.no/a/trond/s89.html,

The cumulative effect of these trends is that a newly emerging type of area is being constituted, and there is Russian influence on virtually all levels of morphosyntax and syntax to be observed in everyday use of most if not all Native Siberian languages, at least in the speech of younger speakers. As alluded to above, the diffusion of features across the languages from different genetic units within the northeastern Siberian linguistic area occurred over numerous centuries or even millennia. However, in more recent times, a significant homogenizing force has operated over the indigenous languages of Siberia over a much shallower time depth, but has resulted in a new type of structural convergence. This is the establishment of Russian linguistic hegemony over the vast expanse of Native Siberia. Almost all Native Siberian people use Russian in the majority of communicative contexts, and norms of Russian that differ from those in the indigenous Siberian languages have begun to filter into the structures themselves of the Siberian languages. With respect to the structure of complex sentences one can begin to speak of a new areal type. In particular, Russian prefers finite subordinate and relative clauses introduced by complementizers and relativizers, instead of the non-finite, nominalized or case-marked clausal subordination patterns just examined above.

Many speakers of Evenki show a partially calqued structure based on the Russian model along with frequent code-switching. In the following sentence (49)-(50) we find the borrowed subordinator pogda and a finite verb, but no negative scope operator, as is found in Russian and in some heavily Russianized varieties of other Siberian languages. Thus we find the insertion of calqued or borrowed subordinators and complementizers with finite verbs mimicking the structures found in Russian.

(49) Evenki

goroː o kim is-tsanaː β tar pogda do poselka is-tsanaː β
far when reach-FUT-1SG there until reach-FUT-1SG

17 Such as the Xakas spoken by younger Xakas speakers in Abakan (Anderson 2005), or in certain Ket varieties as well (Kostjakov 1976, Grishina 1977). Nivkh likewise reflects a host of morphosyntactic interference and syntactic calques from Russian (Gruzdeva 2000).
**poka Ljuda-βa baka-džina-ββ**
until Ljuda-ACC find-FUT-1SG

‘It was far until I would get there, until I would reach the village, until I would find Ljuda’

(Grenoble 2012b: 104)

(50) **Evenki**

*hunatra:n-mi: ele doždalas* otkim huru-bru:
girl-REFL hardly (Russ).waited when go-PRF

‘His girlfriend could hardly wait until they would go’

(Grenoble 2012b: 104)

Such influence from Russian has been recognized in Evenki syntax since at least the mid-1960s (Kolesnikova 1966: 19; Grenoble 2010, 2012a, 2012b, forthcoming).

### 6 Summary

The languages of northeastern Siberia share a range of features in their case systems that distinguish them in this clustering from neighboring languages. The one common language in each contact situation that defines the different genetic units found in northeastern Siberia are the Northern Tungusic languages Evenki and Even which are structurally very similar in their case systems. The case features discussed above include the opposition of an instrumental vs. a comitative case, the presence of a prolatave or prosecutive case form, and an opposition between a dative and an allative case form, and finally the use of case-marking on nominalized verbs to create many functional subtypes of subordinate clauses. All features can be reconstructed to Proto-Tungusic, but not every one can be for the other genetic units of the area. In particular, for the first three features mentioned, the languages with a higher degree of Tungusic contact, e.g. Koryak vs. Chukchi within Chukotko-Kamchatkan or Dolgan and Yakut vs. other Turkic languages, one finds a more Tungusic looking profile with respect to these features, and thus diffusion from Tungusic seems plausible as at least a partial explanation for the appearance of the features involved. With respect to case marked clausal subordination, regardless of whether the original system had cases attached to finite(-looking) verbs or nominalized ones, all indigenous languages of northeastern Siberia have been undergoing profound influence from contact with Russian, and one may now begin to speak of the emergence of a new areal type that has replaced or is replacing the system of case-marked clausal subordination with a system of subordinate clauses introduced by adverbial subordinators and complementizers, with finite verbs, etc. Thus, linguistic areas are not static entities only unfolding and developing over millennia, but may be subject to newly introduced yet commonly shared homogenizing forces that can alter the characteristic profile of the area with respect to individual specific features involved within a relatively short time frame, given proper sociolinguistic contexts, as in the case of the rapid expansion of Russian linguistic and cultural hegemony over the vast expanses of Siberia during the past few centuries.

Abbreviations used
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Accusative</td>
</tr>
<tr>
<td>ABS</td>
<td>Absolute</td>
</tr>
<tr>
<td>ALL</td>
<td>Allative</td>
</tr>
<tr>
<td>CAUS</td>
<td>Causative</td>
</tr>
<tr>
<td>COM</td>
<td>Comitative</td>
</tr>
<tr>
<td>DES</td>
<td>Desiderative</td>
</tr>
<tr>
<td>EX</td>
<td>Exclusive</td>
</tr>
<tr>
<td>DETR</td>
<td>Detransitive</td>
</tr>
<tr>
<td>F[EM]</td>
<td>Feminine</td>
</tr>
<tr>
<td>GEN</td>
<td>Genitive</td>
</tr>
<tr>
<td>GER</td>
<td>Gerund</td>
</tr>
<tr>
<td>HAB</td>
<td>Habitual</td>
</tr>
<tr>
<td>.II</td>
<td>Class-II</td>
</tr>
<tr>
<td>INCMPL[TV]</td>
<td>Incompletive</td>
</tr>
<tr>
<td>INAN</td>
<td>Inanimate</td>
</tr>
<tr>
<td>INF</td>
<td>Infinitive</td>
</tr>
<tr>
<td>INS</td>
<td>Instrumental</td>
</tr>
<tr>
<td>LOC</td>
<td>Locative</td>
</tr>
<tr>
<td>NFUT</td>
<td>Non-Future</td>
</tr>
<tr>
<td>.N</td>
<td>Noun</td>
</tr>
<tr>
<td>OBJ</td>
<td>Object</td>
</tr>
<tr>
<td>OBV</td>
<td>Obviative</td>
</tr>
<tr>
<td>PRS</td>
<td>Present</td>
</tr>
<tr>
<td>PRTCPL</td>
<td>Participle</td>
</tr>
<tr>
<td>PX</td>
<td>Possessive</td>
</tr>
<tr>
<td>P/E</td>
<td>Prolative/Equative</td>
</tr>
<tr>
<td>P/F</td>
<td>Present/Future</td>
</tr>
<tr>
<td>POSS</td>
<td>Possessive</td>
</tr>
<tr>
<td>REFL</td>
<td>Reflexive</td>
</tr>
<tr>
<td>SUBORD</td>
<td>Subordinator</td>
</tr>
<tr>
<td>STV</td>
<td>Stative</td>
</tr>
<tr>
<td>SUBJ</td>
<td>Subject</td>
</tr>
<tr>
<td>TERM</td>
<td>Terminative</td>
</tr>
<tr>
<td>1</td>
<td>first person</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
</tr>
</tbody>
</table>

References


Anderson, Gregory D. S. 2004. The languages of central Siberia: Introduction and


200


Stapert, Eugénie. 2013. Contact-Induced Change in Dolgan: An investigation into the role of linguistic data for the reconstruction of a people's (pre-)history. Leiden University, PhD Dissertation.


23 Languages of China in their East and South-East Asian context

Hilary Chappell

After describing the nature of the correspondence between demography and the principal language families or groups found in China, this chapter next outlines the main sets of features used to support the proposal of continental East and South-East Asia as forming a single linguistic area.

Second, in light of these areal features, three case studies will be treated as the main part of this chapter, two located in frontier areas of China and one in its interior:

(i) Gansu-Qinghai in northwestern China where a variety of different Mandarin languages meets Mongolic, Tibetan and Turkic languages
(ii) Guangxi in the south of China where Zhuang (Tai) languages intermingle with western dialects of Cantonese Yue and the little explored Pinghua Chinese
(iii) Hunan in central China where Southwestern Mandarin and Xianghua, an unaffiliated Chinese language, are spoken alongside Hmong (Hmong-Mien) and Tujia (Tibeto-Burman).

The purpose will be to describe the clustering of shared features in these micro-areas, if not the degree of linguistic impact, given the protracted periods of contact between the communities in question. It is only in recent decades that studies on contact-induced linguistic change have begun to be examined in China from the point of view of transferral from non-Chinese languages into Sinitic. The first two case studies report on this recent research. The third is a case study of language contact between related Sinitic languages, a contact situation which is only at an early stage.1

1. China and its languages: a linguistic demography

The principal language families in continental East and South-East Asia are the following six: Sino-Tibetan, Altaic, Tai-Kadai, Hmong-Mien, Austroasiatic and Austronesian, each of which is briefly discussed in turn below, with respect to China, the focus of this chapter.

While 91.5% of China’s population are of Han (汉) stock, that is, belong ethnically to the Chinese nationality according to the most recent 2010 census, the remaining 8.5%
comprise the so-called ‘national minorities’ including Zhuang, Uygur, Tibetan and Korean of which there are 55 recognised ethnic groups.\(^2\)

1.1 Sino-Tibetan (Sinitic, Tibeto-Burman)

As the family name suggests, Sino-Tibetan family comprises the two main branches of (i) Sinitic, the technical term for Chinese languages, and (ii) Tibeto-Burman, a family tree configuration which continues to be contested (for differing viewpoints, see, Sagart (2005) on Sino-Austronesian and van Driem on the Trans-Himalayan linguistic phylum (2011)).

The heartland of Sinitic languages is found in China where the ten subgroups are spread across the territory from Manchuria in the northeast, through central China to Guangdong in the south and Sichuan in the southwest. They have also made incursions into northwestern China (§1.7). The Tibeto-Burman languages, approximately 250 in number, complement this distribution by virtue of their principal locations in Tibet, the western province of Qinghai and in the southwestern corner of China in Sichuan and Yunnan provinces. From the larger Asian perspective, this branch of Sino-Tibetan is essentially centred in the Himalayas and Tibetan plateau, straddling several neighbouring polities including Myanmar (Burma), Bhutan, Nepal and Northeastern India.

The main subgroups of Tibeto-Burman found within the borders of China are the Yi (or Lolo-Burmese), mainly located in southwestern Yunnan, but also scattered across Guizhou and Sichuan provinces (Naxi, Lisu, Hani, Lahu etc), the Jingpho in Yunnan, the Qiangic and rGyralrongic in northern Sichuan, and Tibetan languages in Tibet, Sichuan, Qinghai and parts of Gansu. Its easternmost branch of Tujia is located in Xiangxi, northwestern Hunan (§3.3).

Excepting Karen (Thailand, Burma) and Bai (Yunnan), Tibeto-Burman languages are, in the main, SOV with postpositions and adjuncts preceding the verb, but, unlike Sinitic, not all the subgroups are tonal, while some make use of register distinctions. The branches located to the southwest of China in Nepal and the Himalayan region, including Kiranti, may possess inflectional features such as person agreement.

1.2 Altaic

The Altaic phylum comprises three language families, some of whose members are located in the north and northwest of China and include principally Tungus-Manchu, Mongolic and Turkic. Note, however, that the existence and make-up of such an Altaic phylum or macro-family is the subject of much disagreement. The isolates of Korean and Japanese are also frequently included in this group by some scholars.

\(^2\) According to the Sixth National Census carried out on 1\(^{st}\) November 2010, the population count yielded the figure of 1,339,724,852 inhabitants. This figure has probably increased to more than 1.4 billion at the time of writing. Facts and figures in this section are drawn from several sources, including the sixteenth edition of Ethnologue (Lewis 2009) the online eighteenth edition of Ethnologue (Lewis et al, 2015), Shearer and Sun (2002) for the non-Sinitic languages of China, Xiong and Zhang (2008), Wurm and Li (1987) and Zhang (2012) for Sinitic.
The Manchu language could be described as practically extinct within Manchuria (Dongbei), the northeastern provinces of China, if it were not for the related variety of Sibe (Sibo), spoken in Xinjiang. The fate of Manchu reveals an interesting case of language shift from a conquering nation to that of the conquered: the Manchus became gradually sinicized during the Qing dynasty (1644 – 1911) that they had founded, shifting to Mandarin by the end of the seventeenth century. There are nonetheless reports that some elderly speakers can still be found in two small communities in the northeastern province of Heilongjiang: Heihai 黑海 and Sanjiazi 三家子 (Beffa and Even 2011).

Tungus is represented by two small communities of Oroqen and Evenki speakers in Inner Mongolia.

Mongolic languages are spoken widely not only in the Mongolian People’s Republic to the north of China but also in the neighbouring, autonomous region of Inner Mongolia (part of China), and in the western provinces of Gansu, Qinghai and Xinjiang. This family includes Bao’an, Dongxiang (Santa), Monguor and Eastern Yughur in Gansu and Qinghai.

The Turkic family is spread across Central Asia all the way to its western-most branch in Turkey. In China, it is represented by the Uygur, Kazakh, Uzbek, Tatar and Kirghiz languages, spoken in the far northwestern Uygur Autonomous Region of Xinjiang (former Chinese or Eastern Turkestan), Western Yughur spoken in Gansu and Salar, spoken in Qinghai and Gansu. These historically peripheral regions of China have inevitably produced unusual scenarios of language contact between unrelated languages, including those with the more recent immigrant Mandarin varieties that belong to the Northern, Northwestern Lanyin and Central Plains subgroups (§3.1).

Altaic languages are well-known for their strictly SOV order, postpositions and agglutinative features, including the use of suffixed case-markers (see Robbeets, this volume). Some show vowel harmony.

1.3 Tai-Kadai

The Tai-Kadai language stock, which has also become known as ‘Kra-Dai’ in western literature, has the different name in China of Zhuang-Dong 壮侗. While the heartland for the numerically dominant Zhuang languages is in the Guangxi autonomous region of far south China, this subgroup extends south into Vietnam, and north into Guizhou, where members of the Kam-Sui branch are also found, likewise under a different name in China, that of Dong-Shui 侗水. This group includes Mulao (Mulam) and Maonan. Southwestern Tai languages, known as Dai 傣 in China, are located further to the west in Xishuangbanna in southern Yunnan, and straddle the borders of Burma and Laos. This happens to be the subdivision to which Standard Bangkok Thai belongs. Another important branch within China is constituted by the Li or Hlai languages of Hainan Island; see also Diller et al (2008).

In terms of word order correlations, Tai-Kadai languages represent almost the ideal for SVO languages (Dryer 2003) conforming to the predictions of modified-modifier in most of their syntactic features, unlike Sinitic, which is head-final for its nominal syntax and for certain aspects of its predicate syntax (see Chappell, Li and Peyraube 2007). The ancestors of the Tai-Kadai, known from historical texts as the Bai Yue 百越 or ‘Hundred Yue tribes’ are believed to have had their homeland in a large area south of the Yangzi River, well before the arrival of the first Han military colonizers in the late third

1.4 Hmong-Mien

This language family is known under the different name in China of Miao-Yao苗瑶. The Hmong languages are found as far north as the province of Hunan in central China and extend southwards through Guizhou, Sichuan and Yunnan provinces into Vietnam, Laos and thence, further into Thailand. Mien languages are found in scattered pockets across the highland areas of southern China. Historically, speakers of these languages migrated slowly southwards under pressure from the Han over many centuries and may have pre-dated the ancestors of the Austroasiatics and the Tai in southern China (Yue-Hashimoto 1991).

The SVO Hmong languages are well-known for their high number of lexical tones and large array of consonant initials, including prenasalised and preglottalised series (Niederer 2011, Sposato 2014).

1.5 Austronesian

Apart from Huihui 回辉语, an isolated Chamic language of Hainan, the island province located to the south of Guangdong, a large group of Austronesian languages is located in Taiwan which are known as Formosan, the most ancient members of this large family. While Formosan languages number approximately twenty, many are endangered, if not already extinct. Forming 2% of the Taiwanese population, the aboriginal communities who speak languages such as Atayal, Bunun and Rukai, are located in what have been, until comparatively recently, the more inaccessible mountainous areas of central and eastern Taiwan.

Formosan languages do not form a single branch within Austronesian. Nonetheless, they share many features such as VSO or VOS word order, focus systems, case marking and a well-developed verb morphology.

1.6 Austroasiatic

Austroasiatic languages are generally divided into two main branches: Munda languages spoken in central and eastern India, and Mon-Khmer spoken in Vietnam, Cambodia and parts of Laos and Thailand. It is claimed that the ancestral languages to the Mon-Khmer branch of the Austroasiatic language family were originally spoken in parts of eastern and southeastern China and that substratum effects identified in Southern Min dialects support this hypothesis (see Norman and Mei 1976, Diffloth 2011).

Today, only a few languages from this group are present on China’s soil, including small communities of Va (Wa), Khmu and Blang (Bulang) in Xishuangbanna, Yunnan province, and also Vietnamese in Guangxi, known as Gin, Kin or Jing in China.

In addition to these six large families of greater South-East and East Asia, also present within the borders of China are:
(i) Sizeable Korean communities in the northeast of China (in the three provinces formerly known collectively as Manchuria: Liaoning, Jilin, Heilongjiang)

(ii) Indo-European: Tajik (Iranian) and Russian (Slavic) in Xinjiang

1.7 Sinitic languages

The ten main divisions for Sinitic languages (or Chinese dialect groups) currently recognized are Northern Chinese (Mandarin), Xiang, Gan, Wu, Min, Kejia (or Hakka), Yue, Jin, Pinghua and Hui dialects (as, for example, in the Language Atlas of China, second edition, 2013). Some disagreement exists with respect to the status of Pinghua and Hui, as being independent from the Yue and the Wu dialect groups, respectively. The linguistic diversity of the Sinitic languages is immense and, as yet, is neither fully explored nor acknowledged. Lack of mutual intelligibility is regularly the case even within each of these large, often heterogeneous dialect groups. For example, Southwestern Mandarin is not immediately comprehensible to speakers of the various Northern Mandarin dialects; nor is Hokkien to Teochiu speakers, even though both the latter are classified as Southern Min varieties. Sinitic or Chinese languages could be thus considered to be as diverse as the Indo-European family.

The somewhat artificially-created language of Standard Mandarin was adopted as the official language of China in 1958 and is called pǔtōnghuà普通話 or literally ‘the common language’. The use of pǔtōnghuà has since been very effectively implemented across China through the domains of education, government and media as the national lingua franca. It is modelled on the pronunciation of the Beijing dialect, the lexicon of Northern Mandarin dialects and the grammar of modern vernacular literature (Chen 1999: 124). However, every provincial city has tended to develop its own variety or registers of pǔtōnghuà, defined by the extent to which local elements and pronunciation are mixed in (cf. plastic pǔtōnghuà of Changsha, Wu Yunji 2005).

Table 1 below gives the most up-to-date figures on the ten main groups of Sinitic languages.3

Table 1. The Sinitic branch of Sino-Tibetan

<table>
<thead>
<tr>
<th>Language branch</th>
<th>Region of China</th>
<th>Population (millions)</th>
<th>Representative Varieties</th>
</tr>
</thead>
</table>

3 These figures are based on those given by Xiong Zhenghui 熊证辉 and Zhang Zhenxing 张振兴 (2008:97) and the second edition of the Language Atlas of China (中国语言地图集第二版 Zhongguo Yuyan Dituji Di’er ban, 2012) and have been rounded up. Xiong and Zhang explain that they have used the 2004 China Administrative Regions Yearbook (中国行政区划简册 Zhongguo Xingzhengqu Huajiance) for the population figures. This compares with the sixteenth edition of Ethnologue (Lewis, 2009) and also the online eighteenth edition (Lewis et al, 2015) which are both based on an extrapolation of the 2000 China census, the last census to pose questions on language and dialect use. Lewis (2009) and Lewis et al (2015) estimate 840 millions of speakers for the different Mandarin dialects within China, rather than 799 million as above.
I. Mandarin
Northern, Northeast, Southwest of China 799m Beijing, Tianjin, Nanjing

II. Jin
Shanxi, Inner Mongolia 63m Taiyuan, Huoija

III. Xiang
Hunan 36m Changsha, Chengbu

IV. Gan
Jiangxi 48m Nanchang

V. Hui
Anhui 3.3m Jixi

VI. Wu
Zhejiang, Southern Jiangsu 74m Shanghainese, Suzhou, Ningbo

VII. Min
Fujian, NE Guangdong, Taiwan 75m Hokkien, Teochew

VIII. Kejia
SW Fujian, NE Guangdong 42m Meixian Hakka

IX. Yue
Guangdong & Guangxi 59m Cantonese

X. Pinghua & Tuhua
Guangxi, Hunan & Guangdong 7.8m Nanning, Guilin

TOTAL 1,207,100,000 (1.2 billion)

The enormous Mandarin branch is itself further classified into eight subgroups. The greatest linguistic diversity for Sinitic is concentrated in the southeast of China where eight of the non-Mandarin dialect groups are located (III – X). Only the northern Jin group is co-territorial with Mandarin.

In the next section, we review some of the principal works on typological characteristics of East and South-East Asia, concluding with recent research on linguistic areas in China.

2 Typological features of Sinitic languages in their South-East Asian context

The received wisdom is that in particular, Sinitic, Tai-Kadai, Hmong-Mien and Mon-Khmer show isolating or analytic tendencies while Tibeto-Burman does so to a somewhat lesser extent, particularly its western branches. In general, East and South-East Asian languages do not possess inflectional morphology, that is, they do not use special affixes on their verbs to code tense, nor to indicate person agreement. Nouns do not usually take markers for number (singular and plural), gender (masculine, feminine, neutral), nor for case (nominative, accusative, dative etc.). At word- or morpheme-level, there is claimed to be a preference for monosyllabicity.

However, typology is a relative matter. Sinitic languages do possess small inventories of affixes, predominantly making use of suffixes and clitics. Adpositions, mainly prepositions, are used to code case roles such as direct object, agent in the passive construction, or the comitative. These may be found to form portmanteau morphemes in many dialects when they fuse with following pronouns, one high frequency counterexample to the isolating stereotype that lingers on with respect to Chinese, based largely on the typological profile of Standard Mandarin. In contrast to this, Central Plains and Jilu Mandarin dialects, also Jin dialects, may use changed syllable finals for the verb
to indicate aspe ctual functions such as the perfective (Lamarre 2015). Furthermore, the
use of tone sandhi (tone change), vowel lengthening and reduplication with grammatical
functions is widespread in Sinitic for closed classes such as pronouns and demonstratives
to indicate, for example, the plural forms of pronouns or semantic distinctions in
demonstrative paradigms (Chen 2015).

In addition to these important features, Marybeth Clark (1974, 1985, 1989) who
worked extensively on Hmong, Vietnamese and other South-East Asian languages,
pinpointed the following features which prove to be equally applicable to Sinitic
languages:

1. Widespread use of tone (Tai, Hmong, Sinitic, some Tibeto-Burman languages.)
2. Use of numeral classifiers as part of the nominal structure with quantified nouns
3. SVO or medial word order in the east and southeast (Note: verb-final word order in
the west for Tibeto-Burman)
4. Stative verbs to express qualities as opposed to adjectives
5. Polar “Yes-No” questions formed with verbs and negative adverbs: either VERB-
NEG or VERB-NEG-VERB
6. Serial verb constructions (SVCs)
7. Adversative passive constructions
8. Identity of existential and possessive verbs

This has been elaborated upon by Enfield (2005) *inter alia* (see this volume) who
discusses further features such as:

9. Use of ellipsis and noun phrase movement
10. Polyfunctionality of nouns and verbs
11. Lack of explicit marking on complex clauses
12. Use of rhyming expressions and reduplication

Bisang has also contributed many studies on mainland South-East Asia, with his major
areal study of different kinds of SVCs (1992), a typology of extended uses of classifiers
(1996), and the notion of hidden complexity (2009) that accompanies the lack of explicit
morphological coding.

2.1 Studies on areal diffusion in East and South-east Asia

Some of the earliest pioneering work on the areal typology of Sinitic languages was
carried out by Mantaro Hashimoto 桥本万太郎 who argued that due to centuries-long
contact, Northern Sinitic languages have become substantially altaïcized while Southern
Sinitic languages have similarly become taïcized. His evidence comes in the form of
phonological, lexical and syntactic features (see Hashimoto 1974, 1976a, 1976b, 1986,
1987). Some of these tendencies are listed in Table 2 below for the Chinese languages
found in the two zones.

---

4 Hashimoto could well have usefully included Austroasiatic languages in this continuum which
similarly to the Tai (or Kra-Dai) languages, lie in and to the south of China (see §1.6; also Mei
and Norman 1976).
His treatment of the north-south opposition as a continuum can be clearly perceived in features such as the increasing number of classifiers, tones and consonantal endings to syllables, not to mention the greater preference for monosyllabic morphemes as one moves southwards.

Table 2. Tendencies in Sinitic languages according to Hashimoto’s North-South division

<table>
<thead>
<tr>
<th>Altaicization (North)</th>
<th>Taïcization (South)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress-based and fewer tones</td>
<td>More tones</td>
</tr>
<tr>
<td>Higher proportion of polysyllabic words</td>
<td>Higher proportion of monosyllabic words</td>
</tr>
<tr>
<td>Smaller inventory of classifiers</td>
<td>Larger inventory of classifiers</td>
</tr>
<tr>
<td>Modifier-modified for NPs</td>
<td>Modified-modifier for NPs</td>
</tr>
<tr>
<td>Gender prefixes on certain animal terms</td>
<td>Gender suffixes on certain animal terms</td>
</tr>
<tr>
<td>Indirect Object – Direct Object word order for prepositionless ditransitives</td>
<td>Direct Object – Indirect Object word order for prepositionless ditransitives</td>
</tr>
<tr>
<td>Less complex syllable structure</td>
<td>More complex syllable structure</td>
</tr>
<tr>
<td>Preverbal adverbs</td>
<td>Possibility of clause-final adverbs</td>
</tr>
<tr>
<td>Passive markers &lt; causative verbs</td>
<td>Passive markers &lt; verbs of giving5</td>
</tr>
<tr>
<td>Comparative of inequality: marker(＜COMPARE 比)-standard-adjective</td>
<td>Comparative of inequality: adjective-marker(＜SURPASS)-standard</td>
</tr>
</tbody>
</table>

Hashimoto’s dichotomy cannot be faulted in terms of general trends. However, are they all due to areal convergence? Indeed, for some of the features, further analysis unearths certain complications. For example, passive markers arising from causative verbs is not a purely Northern Sinitic feature in China, as he would claim: Chappell and Peyraube (2006) have argued that Southern Sinitic give verbs must pass through a causative verb stage before they develop into markers of the passive and have also pointed out that the causative > passive grammaticalization has been attested long before any intense contact with Altaic.

For prepositionless ditransitive constructions, known as ‘double object’ constructions in Chinese linguistics, counterexamples to this generalization are the fact that the Northern recipient-theme or Indirect object – Direct object word order is also found in Pinghua Chinese (de Sousa, 2015) and in Southern Min dialects, both Southern Sinitic. With respect to comparatives of inequality, Chappell and Peyraube (2015) have argued that the head-marking type of adjective-marker-standard found in Cantonese, Hakka and many other Southern Sinitic languages, is in fact a native construction, attested in historical documents from as early as the Late Archaic period (5th – 3rd BC). Without further historical evidence from Tai languages, it would be imprudent to uphold any hypothesis of taïcization in this case.

---

5 This feature is only pertinent to Southern and Central Sinitic languages and not to mainland South-East Asia.
In another approach to South-East Asian areal linguistics, Matisoff (1991a, 1991b) has proposed that mainland South-East Asia may be divided into the Sinospheric and the non-Sinospheric or Indospheric zones. The Sinospheric area includes Southern Sinitic languages located south of the Yangzi. In this study, he adduces that several common grammaticalization pathways basically involve the syntactic reanalysis of one of the main verbs in asymmetrical serial verb constructions (SVCs), as in the cases from (2) – (7):

1) modal verbs > desiderative markers, ‘be likely to’
2) verbs meaning ‘to dwell’ > progressive aspect markers
3) verbs meaning ‘to finish’ > perfective aspect markers
4) verbs meaning ‘to get, obtain’ > ‘manage to do’, ‘able to’, ‘have to’
5) verbs of giving > causative and benefactive markers
6) verbs of saying > complementizers, topic and conditional markers
7) formation of resultative and directional compound verbs through verb concatenation

As pointed out in Chappell (2001), nearly all of these pathways of grammaticalization apply to Northern Sinitic as well, apart from the limited use of ‘give’ with a causative meaning. Hence, they may be profitably added to the inventory of shared areal features for East and South-East Asia.

Dryer (2003) has developed Hashimoto’s areal classification further by providing a careful analysis of word order typology in mainland South-East Asia, setting up a continuum from west to east, or SOV to SVO and their principal correlations. He observes that GEN-N and REL-N are two common areal features, regardless of word order. For Sinitic and Tibeto-Burman languages, he concludes that they are influenced by their closest neighbours.

2.2 Five linguistic areas of China

On the foundation of these earlier studies, particularly those of Hashimoto (cited above), Norman (1982, 1988) and Yue (2003), Chappell (2015b) argues the case for five linguistic areas in China for Sinitic languages on the basis of the distribution and intersection of five main types of grammatical constructions: differential object-marking constructions, adversative passives and their related causative constructions, comparative constructions of inequality and ditransitive constructions. It is shown that the parameters used in establishing linguistic areas for unrelated languages can also be clearly applied to related languages involving intralinguistic contact, such as in the case of the different divisions of Sinitic.

The five areas in question involve a refinement of the north-south divide proposed by Hashimoto, based on the clustering of major syntactic features, whereby the northern area continues to be opposed to a southern area, but this is composed itself of three smaller areas. A central transitional area manifesting great linguistic turbulence in terms of linguistic features is sandwiched in between the north and the south:

---

6 Only Karen is the ‘odd man out’ here: it makes use of N-REL.
(i) the Northern Mandarin dialect groups (Central Plains, Jilu, Jiaoliao, Northeastern Mandarin), Jin dialect group
(ii) the Southwestern Mandarin dialects spoken in Hubei, Sichuan, Yunnan, Guizhou, NW Hunan
(iii) the Far Southern Yue (Cantonese) and Hakka dialects in Guangdong and Guangxi
(iv) the Southeastern Min dialect group in Fujian province and Taiwan

and a buffer zone located in between the northern and southern linguistic areas:

(v) the Central Transitional zone Wu, Hui, Gan, Xiang and Jianghuai Mandarin

Significant for the areal linguistics of South-East Asia is the fact that the two Southern Sinitic linguistic areas share their use of SURPASS comparatives with Thai, Lao, Hmong, Vietnamese and Khmer, as well as SUFFER-passives based on contact verbs, but this is in the case of Southwestern only. Note that the Southwestern province of Yunnan shares borders with Burma (Myanmar), Laos and Vietnam. The Far Southern zone shares its southern border just with Vietnam (Guangxi). The latter’s preference for non-grammaticalized serial verb constructions where the first verb is TAKE is also found in several South-East Asian languages according to Bisang (1996). These have not yet become fully-fledged differential object markers (DMOs) as in the Northern area (see Chappell 2013).

3 Case studies of language contact and its outcomes in China

3.1 Northwestern China – the Gansu-Qinghai linguistic area and contact between Altaic and Chinese

The Northwest of China is historically a frontier area where Han Chinese have carried out trade along the Silk Route with many different ethnicities, or have engaged in warfare to occupy and re-occupy the same border territories, if not to be occupied in their turn. This part of China meets with and forms part of Central Asia, where languages of the Mongolic and Turkic families are found (§1.2), such as in the Xinjiang 新疆 Autonomous Uygur Region, former Chinese Turkestan.

To the southeast of Xinjiang, in Qinghai 青海 province, speakers of Tibetan and Mongolic languages intermingle with smaller communities speaking Turkic languages, such as Salar. Lanyin (Northwestern Mandarin), Central Plains and Northern Mandarin are relative latecomers to this area bordering on a vast expanse of deserts such as the Qaidam and Taklimakan and also the Yellow River plateau.

7 The Lanyin or Northwestern Mandarin dialects are rather aberrant in many respects of their syntactic behaviour, due to contact influence from Altaic and so are not included in this list of core Northern Mandarin dialects; see §3.1 for more information on their special features.

8 The Pinghua dialects of Guangxi Autonomous region prove to possess many Northern features, or if not, to be transitional between the Southwestern area and the Southern area (see de Sousa,2015, §3.2 below and Chappell 2015b, for more details).
Across the eastern border of Qinghai, the linguistic situation is similarly mixed in Gansu 甘肃 with large communities composed of a variety of different Turkic and Mongolic languages. Two of the main Turkic languages in this area are Salar and Western Yughur, while the main Mongolic languages are Baonan, Monguor (Tu), Eastern Yughur and Dongxiang (Santa).

In this section, we will concentrate on a small Sprachbund which straddles the border between Qinghai and Gansu and has been proposed on the basis of language contact between varieties of Chinese, Tibetan and Mongolian. Furthermore, the claim has been made that several of these languages are mixed, including Wutun 五屯话, Linxia 临夏话 (Hezhou) and Tangwang 唐汪话 in particular.

What is linguistically striking about the northwestern area is first of all the widespread development of case-marking systems in the Chinese languages located in this zone, undoubtedly the result of contact with the surrounding Altaic languages. Secondly, the Chinese languages in this area have undergone a general word order change to verb-final, SOV, that is, the same word order as in Altaic. Thirdly, the plural suffix, normally restricted to personal pronouns in Sinitic (and sometimes a handful of human, animate nouns), has been extended in use to the pluralization of non-human animate and inanimate common nouns, once more replicating the model of plural suffixes found in many Altaic languages. For example, it is common to find the application of the Mandarin plural suffix –men 门 in the Chinese languages in this Sprachbund to any kind of common noun, producing forms such as HUAR MEN 花儿们 ‘flowers’, HONGQI MEN ‘red flags’, YEZI MEN ‘leaves’ and even to abstract nouns, yielding ‘kindnesses’ – ENQING MEN (see Li Keyou and Li Meiling 1998: 25-26), all unacceptable in standard Mandarin. Further examples of these intriguing contact features are presented immediately in the following.

Gangou 甘沟话, Tangwang 唐汪话, Linxia 临夏话 and Wutun 五屯话 are all fundamentally Chinese or have a Chinese language-base. Nonetheless, quite atypically for Sinitic, they use a mixture of suffixes and postpositions to mark a minimum of four or five different case roles, including a form that syncretizes the direct object with the dative and the benefactive, as well as an ablative/comparative marker, an allative marker and a syncretized form for comitative/instrumental/manner. The use of the accusative and dative postposition –ha or –xa can be seen in examples (1) and (2), which also clearly show the typical SOV word order in this micro-area.

(1) Gangou Chinese, Qinghai 甘沟话 (Zhu et al 1997 : 438-439)
Aijie-liar mian-ha wa-shang, you-ha dao-shang-zhi
she-two flour-ACC take out-PERF, oil-ACC pour-out-CHAIN

---

9 I use the pīnyīn Romanization for standard Mandarin with small capitals wherever the IPA transcription is not available in the original source.
10 Tangwang and Wutun are claimed to be mixed languages by some, including Ethnologue (2009, 2015) and Wurm and Li (1987). For a different conclusion, see Peyraube (2008, 2015) and Djamouri (2015) on Tangwang.
11 Peyraube (2008, 2015) observes that the label ‘case’ for adpositions does not strictly speaking conform to the traditional notion of case. This is a controversial issue which we leave open for further discussion, while noting that the adpositions in these Northwestern Chinese languages fulfil a similar function to case inflections.
youbingzi zha-liao.
friedcake fry-PERF
‘The two took out the flour, poured the oil into a cauldron and then they fried fried-cakes.’\(^{12}\)

(2) Tangwang Chinese, Gansu (Djamouri 2015)

\begin{align*}
&nə nə ɕɛ kaxə-mu-xa-nə wawa-mu-xa-nə \\
&3SG DEM some apricot-PL-ACC-LOG child-PL-DAT-LOG \\
\end{align*}

ʈʂʰ-ki-xa-xa-li̟ɕ
eat-CAUS-RES-POT-PERF
‘He was able to get his children to eat those few apricots he had.’

The ablative marker may double up as a comparative marker, as in Gangou Chinese and in the neighbouring Mongolic language of Monguor, not to mention in Tangwang Chinese as well. It is a source frequently found crosslinguistically in comparatives of inequality (the ‘separative’ strategy of Stassen 1985) but is not attested elsewhere in Sinitic for synchronic data (Chappell 2015b).

(3) Gangou Chinese, Qinghai (Zhu et al 1997: 445)

\begin{align*}
&Zhi-ge huar hao-mei, zhi-ge-ha-sha \\
&this-CL flower good-NEG this-CL-ACC-COMP \\
\end{align*}

han da-zhi zhe-gei more big-GEN pick-CAUS
‘This flower isn’t good, I’ll pick a bigger one than this.’

Finally, the instrumental/comitative case is a regional calque from Altaic languages, based on the morpheme for ‘two’ LIANG ~LIA in the Chinese languages of Linxia (Dwyer 1992: 165168), Wutun (Janhunen et al 2008: 60), Tangwang and Gangou (Peyraube, 2015). While its syncretism with instrumental and manner roles is typical in European languages (Stolz 1999, Heine and Kuteva 2002), it is rare in Sinitic languages, in which the comitative regularly develops into a NP coordinative conjunction, but never into an instrumental marker. In certain languages, the comitative grammaticalizes into a differential object marker as well, via a benefactive/dative use (Chappell 2013, 2015b).

\(^{12}\) This example (1) faithfully reproduces the text in Zhu et al (1997). The grammatical abbreviations used in the glossing are as follows: ACC Accusative Case; CHAIN Clause Chaining Marker; PERF Perfective; and in example (2): CAUS causative; DAT dative; DEM demonstrative; LOG logophoric; PERF perfective; PL plural; POT potential; RES resulative.
The grammatical abbreviations for the Wutun examples are: SOC sociative (i.e., instrumental/comitative case); PRF perfective; EXEC executive; OBJ objective; 1P:SG first person singular; CONT continuative.

14 Zhuang is further classified into Northern and Southern branches in China. According to the traditional division into Northern, Central and Southwestern Tai made by F.K. Li (1977), Northern Zhuang belongs to Northern Tai while Southern Zhuang belongs to Central Tai.
similarly identified the same clause-final position for intensifiers and also the use of a particular type of reduplication for verbs, both directly calqued from Zhuang.

In a comparison of Nanning Pinghua with Nanning Cantonese and Northern Zhuang, de Sousa (2015) has found that Nanning Pinghua is in fact far less influenced by Zhuang than is the local variety of Nanning Cantonese. At first blush, this may appear somewhat counter-intuitive since Pinghua speakers, as de Sousa explains, have been resident in the Guangxi area since the eleventh century, when military troops from the Northern province of Shandong were sent to quell the Zhuang during the Northern Song dynasty (960-1127). In contrast to this, the Cantonese speakers have only arrived relatively recently in the past 150 years, migrating westwards along the river routes from Guangdong province, from the time of the first Opium wars onwards.

De Sousa shows that while Nanning Southern Pinghua follows expectations in revealing more Zhuang loans in its lexicon than does Nanning Cantonese, by way of contrast, in its morphology and grammar, Pinghua unexpectedly exhibits many important Northern Sinitic features including IO-DO word order (or \textit{Verb-Recipient-Theme}) as the unmarked word order in prepositionless ditransitives (also known as ‘double object’ constructions), gender prefixes on terms for domestic animals, the typical use of preverbal adverbs, and a strong dispreference for monosyllabic words (§2.1). The latter is evident in the higher incidence of suffixes, acting as markers of the noun class, which results in the creation of disyllabic words. In contrast to this, the Yue dialect of Nanning Cantonese displays the predictable DO-IO word order (or \textit{Verb-Theme-Recipient}) common to the Yue group of Sinitic with verbs of giving\(^{15}\), gender suffixes, the presence of clause-final adverbs, as remarked upon above, and no dispreference for monosyllabic words at word-level. Two of these features are selected below for exemplification.

(6) Nanning Pinghua \textit{Verb-Recipient-Theme} order in ditransitive constructions

\begin{center}
\begin{tabular}{l}
系 佢 錢, 佢 就 抓 去 賭。

give 3SG money, 3SG then take go gamble
\end{tabular}
\end{center}

\textit{If you give him/her money, s/he will take it to gamble.}

Table 3. Gender affixes in three languages spoken in Nanning, Guangxi (de Sousa 2015)

<table>
<thead>
<tr>
<th>Language</th>
<th>‘hen’</th>
<th>‘rooster’</th>
<th>Affix</th>
</tr>
</thead>
</table>

\(^{15}\) As de Sousa (2015) observes, the \textit{theme-recipient} word order with verbs of giving in Cantonese may well be an internal development. Furthermore, verbs of deprivation allow the \textit{recipient-theme} order. Both word orders for ditransitives are possible in many Tai languages which seems to vitiate the argument in favour of the hypothesis that borrowing took place in the direction from Zhuang, or Tai in general, into Southern Sinitic languages. See also on this topic, Xu and Peyraube (1997).
Another set of features which link Yue dialects more closely with Northern Zhuang or Tai languages can be seen in the category of the nominal classifier. In a similar pattern to Tai languages, Nanning Cantonese displays a far greater polyfunctionality in its extended uses of the classifier, than does Nanning Pinghua. The most striking difference is seen in the use of the appropriate classifier to code possession in preference to the general marker of subordination in Nanning Cantonese, 嘅 kɛ³³, which may also be employed. This preference is shared with Northern Zhuang (Milliken 1998), while the use of classifiers to mark the relation of possession is not possible in Nanning Pinghua. Nanning Pinghua, like Northern Sinitic languages in general, must use the general marker of modification (MOD), which is kə⁵⁵. Compare (7) with Northern Zhuang in (8) and Nanning Pinghua in (9) (data from de Sousa, 2015).

(7) Nanning Cantonese: N\textsubscript{POSSESSOR}–CL–N\textsubscript{POSSUM}

佢隻崽 kʰy¹³ tʃɛk tʃɛ⁵⁵

3SG CL son ‘His/her son’

(8) Northern Zhuang: CL–N\textsubscript{POSSUM}–N\textsubscript{POSSESSOR}

æn vanj mwngz

CL bowl 2SG ‘Your bowl’

(9) Nanning Pinghua: N\textsubscript{POSSESSOR}–MOD–N\textsubscript{POSSUM}

細蘇個 (*隻)狗児

tʃi⁵⁵ tɔ³³ kə⁵⁵ (*tʃət³) kən³³-ɲi¹¹

Little Sū MOD (*CL) dog-DIM ‘Little Sū’s puppy/puppies’
This is, however, a general feature that is also shared by Zhuang with Standard Cantonese. Nanning Cantonese, however, goes one step further in allowing attributive adjectives to be linked by the appropriate classifier to the head noun, just as in Northern Zhuang (for details, see de Sousa, 2015).

De Sousa (2013) argues that the less extensive impact of Zhuang on Nanning Southern Pinghua can above all be attributed to the very low rate of intermarriage that traditionally existed between the two groups before the mid-twentieth century and the ‘social distance’ maintained by the Pinghua communities towards the Zhuang, a situation which has not been applicable to the same degree at all for the Nanning Cantonese communities. In fact, since the arrival of the Cantonese in Nanning in the mid-nineteenth century, there has been large-scale language shift from Zhuang to Cantonese (Kwok 2010) which has tended, therefore, to reinforce the already existing Tai substratum in Cantonese Yue dialects.

3.3 Hunan, Central China – Southwestern Mandarin and Xianghua

Hunan province lies just to the south of the Yangzi River in central China. The main Sinitic languages spoken in Hunan belong to the central transitional Sinitic language group known as the Xiang 湘 group. There are also pockets of Gan speakers in the southwest, scattered Hakka communities and a group of largely undescribed patois or tūhuà 土话 in the south of the province, an area bilingual with Southwestern Mandarin (for locations, see Wurm and Li, 1987, Map B-11). The northwest of Hunan is the main target of this third case study in a linguistically diverse area known as Xiangxi 湘西, from which an autonomous Tujia and Hmong prefecture has been created: Southwestern Mandarin is the lingua franca in this area, rather than Standard Mandarin, reserved for formal, and official domains, if not for communicating with people from outside this area. This being said, Standard Mandarin is very well-understood due to its use in all the different types of media. The main languages are thus:

(i) Southwestern Mandarin
(ii) Hmong
(iii) Tujia
(iv) Xianghua (also known as Waxiang)

According to Xiong and Zhang (2008), there are 260 million speakers of Southwestern Mandarin dialects, a subgroup of Mandarin which is not readily comprehensible with the standard language. It spreads from Hubei through northwestern Hunan to Sichuan, Guizhou and Yunnan provinces, and is itself made up of a large number of dialects. As the speakers of Southwestern Mandarin arrived fairly late in the Xiangxi region, during the Qing dynasty (1644-1911), their language has come to be known as kèhuà 客话 ‘guest language’ by the local population. This is in fact the dominant language, particularly in the towns and cities.

16 Note that the use of classifiers to mark possession is also found in other Sinitic languages, such as the Wu dialects, for which there is no connection with Zhuang. It is not found in Northern Sinitic, however.
The Hmong languages of Xiangxi belong to the northernmost group of Western Hunan Hmong which is itself further divided into Western and Eastern dialects, and is spoken, in the main nowadays, in rural villages. The population is estimated at 770,000 of a total Hmong population of over five million (Wurm and Li 1987: Map C-9) and 900,000 in Ethnologue (Lewis et al 2015). Urban speakers have already long ago undergone massive language shift to Southwestern Mandarin. The same situation applies for speakers of the easternmost Tibeto-Burman language of Tujia whose localities overlap to some extent with the Hmong speakers in Xiangxi, although not entirely. According to Wurm and Li (1987: Map C-10), only approximately 170,000 speakers are left of an estimated 2.8 million (or even less – 71,500 – according to eighteenth edition of Ethnologue).

Xianghua, by way of contrast, remains an unclassified, Sinitic language, whose speakers are estimated to number approximately 300 – 400,000 (Bao and Yan 1986). Xianghua can be divided up into at least four main dialects: those of Guzhang 古丈, Luxi 泸溪, Yuanling 沅陵 and Nanshan 南山. Apart from Nanshan, located near Chengbu in southwestern Hunan, the other three are located in Xiangxi. According to their family genealogies, they claim to have migrated into northwestern Hunan several hundred years ago from Jiangxi, the neighbouring province to the east. A subsequent, further migration south to Chengbu, in southern Hunan, saw the Xianghua speakers settling in and around the mountainous border areas on the frontier with Guangxi. The family histories of two of the communities – Guzhang and farflung Chengbu, over 500 kilometres to the south – both independently confirm each other’s stated origins, for which we at present do not have any documentary evidence.

The Xianghua dialects are similarly under threat of extinction, possibly within the next generation. It is interesting to compare their plight with the Hmong and Tujia population: A sociolinguistic survey carried out by Yunji Wu in 1998 showed overwhelming sinicization of language use for the Tujia and the Hmong. And this, from an early stage, as far as records can tell us. Wu observes that even by the early twentieth century, Hmong and Tujia speakers had already undergone a massive swing to the monolingual use of Southwestern Mandarin, according to the gazetteer of Dong Hongxun (1907). In 1907, from the data that Dong collected in 282 villages in Xiangxi, 72% of Tujia and Hmong speakers, were already L2 speakers of Mandarin. By way of contrast, for the Xianghua community, a mere 9% were L2 speakers while 13% were monolingual having undergone complete shift to Mandarin. The remaining majority of 78% were monolingual Xianghua speakers. Nearly 100 years later, Yunji Wu carried out her own survey in Guzhang county and was able to identify an increase in the shift to Southwestern Mandarin use to over 90% for the Xianghua speakers (reported in Wu and Shen 2010).17

Before the mid-twentieth century, intermarriage was rare between different ethnic groups in Xiangxi, and openly discouraged. In the rural Xianghua-speaking villages, the disappearance of this social ban has not really affected the linguistic situation greatly: intermarriage is not common, simply because the district village populations are fairly uniformly made up of Xianghua families, apart from the case of men who have moved

---

17 According to our own fieldwork, this tendency is much more in evidence in the county towns of Xiangxi than in the rural villages. In Lijiadong village of Gaofeng district, where my fieldwork was conducted, more than 90% of the village of over 1000 inhabitants speak Xianghua in the home and non-official domains, in addition to Southwestern Mandarin.
outside of the village at some stage to work and have brought back speakers of other Chinese languages as their wives.

Living in separate communities could partially explain why it is difficult to detect any linguistic influence on Xianghua from Hmong or Tujia. Instead, we find an increasing influence on Xianghua from Southwestern Mandarin, the local lingua franca. We also find that, unsurprisingly, this influence is more easy to observe with city speakers than with rural speakers, for example, Xianghua speakers in Guyang 古阳市, the town which is the seat of government for Guzhang county 古丈县, versus Xianghua speakers in Lijiadong 李家洞 village, in the remote mountainous district of Gaofeng 高峰乡. 18 This third case study thus outlines the outcome of interdialectal contact in a Southwestern Mandarin Sprachbund which includes this northwestern area of Hunan and neighbouring Sichuan and Guizhou provinces whose borders it shares.

Xianghua shows many archaic features in its lexicon, syntax and morphology, and so may represent a very early branch of Sinitic (Baxter and Sagart 2014). Nonetheless, borrowing of lexical items connected with domains such as politics, government and technical terms from Southwestern Mandarin is the norm for Xianghua.

In some respects, the syntax of Xianghua is also very conservative, when compared with Mandarin dialects. For example, the comparative of inequality in Xianghua makes use of the polarity cognitive schema (Heine 1997) with a complex coordinate structure that combines antonyms. The stative verbs in the predicates of each clause represent antonyms of one another. The comparee occurs in the first clause, while the standard NP occurs in the second. Apart from structural considerations, the only overt marker is the use of the plural classifier sa\textsuperscript{55} some which acts as an adverbial quantifier in this construction. The construction is at an early stage of grammaticalization.

(10) Xianghua polarity comparative:

$$\text{NP}_{\text{COMPAREE}} \text{Verb}_x Q(\text{CL}_{\text{PL}}), \text{NP}_{\text{STANDARD}} \text{Verb}_x Q(\text{CL}_{\text{PL}})$$

$$\text{你} \text{高} \text{些,} \text{我} \text{矮} \text{些。}$$

$$\text{ng}^{25} \text{kau}^{55} \text{sa}^{55}, \text{wu}^{25} \text{ya}^{13} \text{sa}^{55}.$$ 

$$2\text{SG} \text{COMPAREE} \text{tall} \text{some}_{\text{DEG}} \text{1SG} \text{STANDARD} \text{short} \text{some}_{\text{DEG}}$$

Literally: ‘You’re a little tall; I’m a little short.’ = ‘You’re taller than me.’

This structure has not been widely reported to date in the literature on comparatives in South-East and East Asian languages, including the research on comparatives in Chinese linguistics. Nonetheless, in most Sinitic languages, we can also find the predicative use of a gradable stative verb followed by a quantifier which can be regarded as the coding of a non-conventionalized comparative that lacks its standard NP, cf. Hong Kong Cantonese: gāmyaht yiht dī today hot a.bit 今日熱啲 ‘It’s hotter today’. At present, it is not clear to what extent the complex type of conjoined comparative found in Xianghua with its

---

18 Data has been collected during 8 field trips carried out by the author since 2006 in the town of Guzhang and in the village of Lijiadong, Xiangxi.
polarity semantics can be identified in other Chinese languages in the form of a conventionalized comparative.

What is interesting for the present study is a kind of hybrid comparative that has arisen with the borrowing of the SW Mandarin ‘Compare’ comparative into Xianghua. The hybridization leads to double marking, since the Mandarin comparative marker, \(p_<^\text{比} \) ‘compare’, is accompanied by the quantifier \(s_a^{55} \) ‘some’, the only morphological vestige of the polarity structure which remains, apart from the gradable predicative adjective. And this native marker is obligatory in the new comparative. The original complex sentence form with two clauses that coded the Xianghua comparative has been conflated into a monoclausal structure. This can be regarded as one piece of evidence attesting to intense contact (Hickey 2008: 77), since the resultant structure largely reflects Mandarin syntax and morphology for this type of comparative (see Chappell and Peyraube, 2015):

\[
(11) \text{NP}_{\text{COMPAREE}} \quad p_<^{25} \quad \text{NP}_{\text{STANDARD}} \quad \text{Verb} \quad Q
\]

\[
\text{你} \quad \text{比} \quad \text{我} \quad \text{肥} \quad \text{些} .
\]

\[
\text{ȵ}_i^{25} \quad p_i^{25} \quad w_u^{25} \quad f_i^{213} \quad s_a^{55}.
\]

\[
2SG \quad \text{CM} \quad 1SG \quad \text{fat} \quad \text{some}_{\text{DEG}}
\]

‘You’re fatter than me.’

A second example of borrowing concerns the differential object-marking construction whose marker, \(k_a^{55} \) ‘跟’, has evolved from the unusual source of a comitative in Xianghua (Chappell, Peyraube and Wu 2011), exemplified by (12). It is being slowly replaced in the speech of more educated speakers by the Southwestern Mandarin marker \(p_a^{41} \) ‘把’ (13). The tonal value of 41 for \(p_a^{41} \) ‘把’ belies the fact that it is borrowed: The corresponding tone for the Xianghua native morpheme is a 25 contour. 19 Example (14) records the use of SW Mandarin \(p_a^{41} \) ‘把’ in a Xianghua narrative:

\[
(12) \text{Xianghua differential object-marking construction: (NP}_{\text{AGENT}} \quad [k_a^{55} \quad \text{– NP}_{\text{DO}}] \quad \text{– VP}
\]

\[
\text{跟} \quad \text{履} \quad \text{端} \quad \text{来} ! \quad \text{跟} \quad \text{门} \quad \text{闭} \quad \text{下} .
\]

\[
k_a^{55} \quad l_i^{25} \quad t_o^{55} \quad z_\varepsilon^{13} . \quad k_a^{55} \quad m_a^{55} \quad p_i^{33} \quad k_a^{33}.
\]

\[
\text{GEN}_{\text{om}} \quad \text{shoe} \quad \text{carry} \quad \text{come} \quad \text{GEN}_{\text{om}} \quad \text{door} \quad \text{close} \quad \text{PRT}
\]

‘Bring the shoes over here!’ ‘Close the door!’

\[
(13) \text{Southwestern Mandarin differential object-marking construction: (NP}_{\text{AGENT}} \quad [p_a^{41} \quad \text{– NP}_{\text{DO}}] \quad \text{– VP}
\]

\[
\text{酸} \quad \text{肉} \quad \text{它} \quad \text{就} \quad \text{是} \quad \text{把} \quad \text{它} \quad \text{称} \quad \text{起} \quad \text{来} ,
\]

\[
s_u_a^{55} \quad z_u^{13} \quad t_h_a^{55} \quad t_c_i_{\text{om}}^{25} \quad s_i^{25} \quad p_a^{41} \quad t_h_a^{55} \quad t_h_{o}^{55} \quad t_{c_{\text{om}}}^{41} \quad a_{i}^{13} ,
\]

\[
\text{sour.meat} \quad 3SG \quad \text{then} \quad \text{be} \quad \text{BA}_{\text{om}} \quad 3SG \quad \text{weigh} \quad \text{DIR}
\]

‘Buy it, buy meat!’

19 This morpheme is natively used in Xianghua as a classifier for handfuls of substances.
‘As for sour pork, when (you’ve) purchased the meat, …’ (literally: ‘sour pork, when it’s been weighed and bought, the meat’s been bought.’)

(14) Xianghua narrative extract with SW Mandarin DMO: pa41

need     very many  dish only.then can DMO wedding

Xianghua speakers mix the two DMOS in their speech. The frequency is determined to a large extent by the age and level of education of the speaker and the city-rural divide, as Wu and Cao (2011) have observed in a text study. Other features that have been modified are the aspectual and modal system with the borrowing from SW Mandarin of aspectual suffixes to the verb such as liau41 了 < ‘finish’, the borrowing of a negative marker that brings with it an innovative polar question form VERB-NEG-VERB that may eventually replace the native one of CLAUSE-NEG., These additions create an asymmetry in Xianghua vis-à-vis its native paradigms.

The important role of geographical isolation and little intermarriage outside their own traditionally isolated communities suggests an explanation for the scant evidence of Hmong or Tujia linguistic influence, not to mention the high degree of sinicization for the two latter groups, already evident in the early twentieth century.

4    Conclusion

This chapter has considered the main language families present in China and the issues of areal linguistics from both the macro-perspective of East and South-East Asia, with respect to Sinitic languages, and the micro-perspective of three small Sprachbünde within China. First, the influence of Altaic languages on Chinese languages in the Qinghai-Gansu area of Northwest China has been briefly examined to reveal major structural and morphological changes to the grammatical system. Second, features of a Tai substratum were described for Sinitic languages located in Guanxi, in the far south of China, namely Southern Pinghua and Nanning Cantonese, in the form of both ancient and more
recent borrowings. Third, in the case of Xianghua, a conservative Sinitic language of Hunan in central China, we argued that Southwestern Mandarin has not ceased to impact on its linguistic structure from the mid-twentieth century onwards.

References


Chappell, Hilary. 2015b. Linguistic areas in China for differential object marking, passive and comparative constructions. In: Hilary Chappell (ed.).


Chen, Yujie. The semantic differentiation of demonstratives in Sinitic languages. In Hilary Chappell (ed.).


de Sousa, Hilario. 2015. ‘Language contact in Nanning – from the point of view of Nanning Pinghua and Nanning Cantonese’, in Hilary Chappell (ed.).


Dong Hongxun 董鸿勋 1907 《古丈坪厅志》Guzhangping Tingzhi (Guzhang Gazetteer). o details for the publisher are given). For more information, see parts reprinted in Wu and Shen (2010).


24 Language in the Mainland Southeast Asia area

N. J. Enfield

1 Mainland Southeast Asia and its people

Mainland Southeast Asia (hereafter: MSEA) can be broadly defined as the area occupied by present day Cambodia, Laos, Peninsular Malaysia, Thailand, Myanmar, and Vietnam, along with areas of China south of the Yangtze River. Also sometimes included are the seven states of Northeast India, and—although here the term ‘mainland’ no longer applies—the islands from Indonesia and Malaysia running southeast to Australia and West Papua (see Map 1).


MSEA is a tropical and sub-tropical area with rugged and well-forested hills and river systems running from higher altitudes in the northwest to the plains and deltas of the south. Among the biggest rivers are the Mekong, the Brahmaputra, the Red River in North Vietnam, the Salween and Irrawaddy rivers in Myanmar, the Pearl and Yangtze rivers in China, and the Chaophraya in central Thailand. The lower reaches of these river systems are well-fertilized plains, which have attracted people partly because of the mobility the environment affords, but also because of the suitability for paddy rice farming. Paddy farming, in which rice plants are kept continually flooded as they grow, requires management of water via systems of dikes and channels (Hartmann 1998). This method is significantly more productive than upland dry-field methods, and can support larger populations (Bellwood 1992: 90). It also reduces biodiversity.
Map 1. Greater mainland Southeast Asia: present day Cambodia, Laos, Peninsular Malaysia, Thailand, Myanmar, and Vietnam, along with China south of the Yangtze River, Northeast India, and Insular Southeast Asia

MSEA has seen a long and complex history of human movement, contact, and diversification. Evidence from genetics and archaeology suggests that there has been human activity in the area since some 40,000 years ago, when conditions were very different from today. At around 20,000 years ago, global sea levels were 120m lower than now (Chappell and Shackleton 1986; Tooley and Shennan 1987), implying different possibilities for human movement and livelihoods. Then, one could walk on dry land in a straight line from the site of present-day Ho Chi Minh City to Kuala Lumpur, and then in another straight line to Bali and again up to Brunei (Voris 2000; Oppenheimer 2011; White 2011). While a fair amount is known from bioarchaeological evidence about more recent human activity in the pre-agricultural period (Oxenham and Tayles 2006), the time horizon of comparative linguistics is limited to the last few thousand years (for recent reviews, see Enfield 2011a). Just behind that horizon are the beginnings of agriculture in MSEA some 4000 or so years ago.
A widely accepted view is that the people of MSEA once spoke Austroasiatic languages in a ‘continuous distribution’, and that this distribution was ‘broken up by the historical expansions of the Chinese, Tai, Vietnamese, Burman and Austronesian (Malay and Cham) peoples’ (Bellwood 1992: 109; cf. Sidwell and Blench 2011: 338 and passim; Post 2011). By what mechanism did this take place? Some have argued that modern ethnolinguistic diversification in MSEA was associated with demic diffusion (Bellwood 1992; Blust 1994; Higham 2002; Edmondson and Gregerson 2007). This implies the incoming migration of groups of people who rely on agriculture, and who can thereby support large populations. The incomers replace less populous and less powerful existing forager populations (Ammerman and Cavalli-Sforza 1971; Cavalli-Sforza, Menozzi, and Piazza 1993;
Nichols 1992). An alternative to demic diffusion is cultural diffusion, whereby resident populations remain in place, but adopt new practices and ways of speaking. According to O’Connor (1995: 987), ‘there is no direct evidence that an actual influx of immigrants ever displaced earlier peoples’ in MSEA. He argues instead that an ‘agricultural paradigm’ is what diffused, bringing with it a ‘society-shaping complex’ (see Jonsson 2011, 2014 for discussion). For other critiques of the application of a demic diffusion model in MSEA see White (2011) on the view that hunter-gatherer communities have played a central role in shaping modern MSEA ethnographic diversity, and Fix (2011) on the genetics of ethnolinguistic diversification, in which he presents an alternative to the standard account of demic diffusion in the Malay Peninsula, with a model he calls trickle-effect colonization.

Regardless of whether one thinks the historical process of peopling and ethnolinguistic diversification in MSEA was driven primarily by the spread of people or by the spread of ideas—here, more work is needed—the modern distribution of ethnolinguistic groups is clear. In lowland areas, populations are denser, more culturally and linguistically homogeneous, and more closely affiliated with state political power. In upland areas, populations are sparser, more culturally and linguistically diverse, and have limited if any access to infrastructure, education, or power. The dominant lowland populations are clearly distinct from each other in terms of political identity (‘the Thai’ vs. ‘the Lao’ vs. ‘the Khmer’, etc.), but the upland minority populations that straddle these nations have something in common: they are politically and geographically marginalized.

The upland areas in which many MSEA minorities live are conjoined in a single, elongated area, crossing political borders and encompassing ‘virtually all the lands at altitudes above roughly three hundred meters all the way from the Central Highlands of Vietnam to northeastern India’ (Scott 2009: ix). This area has been referred to as Zomia, a term coined by Van Schendel (2002) in making the point that arbitrary research areas can be constructed and reified by ‘academic politics’, as he puts it (cf. Michaud 2010). Van Schendel’s proposal of a Zomia area is a conceptual exercise, useful because it counteracts the politically sanctioned alternatives. The term has gained some recognition (though ironically not without danger of creating the reification it was warning against; Jonsson 2011, 2014), particularly due to Scott (2009). According to Scott, it is not that the inhabitants of Zomia simply share the fate of having been marginalized by states. Instead, he argues, they share a cultural distaste for being governed: they have chosen to remain isolated from central government control.

We do not have space in this introduction for more on the detailed history of human activity—peopling and migration, social contact and cultural shift, state formation and avoidance, war and peace—in MSEA. For further information, see Tarling (1993), Scott (2009), and Enfield (2011a).

---

1 Demic diffusion is the spread of genes. It is usually associated with the outcomes of migration. In world history, this has often involved the movement of groups who have adopted agriculture, and who are therefore more populous and viable than those (e.g., hunter-gatherers) who are resident in the area being entered. Demic diffusion may be associated with population displacement or replacement, but this need not necessarily be the case. There may be genetic admixture between an incoming population and a resident population, such that some fraction of the genes of the resident population survives. Thanks to Mark Stoneking and Dan Dediu for clarification of these points.
2 Mainland Southeast Asian languages

The degree of linguistic diversity in MSEA (i.e., the number of languages per square km) is high (Enfield 2011b), and it is highest in upland areas. Lower language density in lowland areas is likely related in part to geographical factors and their implications for the nature of social networks (see Nettle 1999). In historical demographic processes of the kinds noted above, formerly diverse lowland communities in MSEA have become homogenized by a combination of two processes. One process was ethnolinguistic shift. Some groups stayed where they were but stopped passing on their languages and identities to their children, instead adopting the languages and identities of new dominant groups. This process can be observed all over MSEA today. Another process was out-migration, typically to more isolated hill areas (Scott 2009). Geographical isolation is a force that still promotes language diversity in the region, where former diversity of lowland areas is on its last legs. Many of the lowland languages are heavily endangered or extinct (Enfield 2006, Bradley 2007, Suwilai 2007). This is quickened by effects of the concentration of political power of modern nation states in the lowlands. In recent decades, processes of language standardization in MSEA nations (Simpson 2007) have helped to heavily reduce language diversity.

The languages of MSEA are from five major language families: Sino-Tibetan, Tai-Kadai, Hmong-Mien, Austroasiatic, and Austronesian. There are nearly 600 distinct languages spoken in greater MSEA. If we exclude the China and India data, thus representing the core MSEA area, the number of languages is about half this amount; see Table 1.

Table 1. A breakdown of numbers of languages in MSEA, separated into language families

<table>
<thead>
<tr>
<th></th>
<th>Core MSEA</th>
<th>Greater MSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austroasiatic</td>
<td>122 (44%)</td>
<td>138 (24%)</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>74 (26%)</td>
<td>288 (49%)</td>
</tr>
<tr>
<td>Tai-Kadai</td>
<td>51 (18%)</td>
<td>93 (16%)</td>
</tr>
<tr>
<td>Austronesian</td>
<td>25 (9%)</td>
<td>26 (4%)</td>
</tr>
<tr>
<td>Hmong-Mien</td>
<td>8 (3%)</td>
<td>38 (7%)</td>
</tr>
<tr>
<td>Total</td>
<td>280</td>
<td>583</td>
</tr>
</tbody>
</table>

2 The Andamanese languages are located just outside MSEA as defined here; though we note with interest new work on these lesser-known languages: see Abbi’s recent reference grammar (2012) and dictionary (2013) of Great Andamanese.

3 Data are from glottolog.org, accessed in May 2014. Many thanks to Harald Hammarström for his input and assistance. Core MSEA was defined for this count as Cambodia, Laos, Myanmar, Thailand, and Vietnam; Greater MSEA included this, along with Peninsular Malaysia, areas of India east of 90 degrees (i.e., the states of Arunachal Pradesh, Nagaland, Manipur, Mizoram, Assam, Meghalaya, and Tripura) and China south of the Yangtze river (specifically, the provinces of Zhejiang, Jiangxi, Hunan, Guizhou, Yunnan, Guangxi, Guangdong, Fujian, and Hainan).
The very high linguistic diversity (i.e., the number of languages) in northeast India
and southern/southwestern China adds dramatically to the number of languages
included in the area. It also reverses the relative proportion of Sino-Tibetan and
Austroasiatic languages.

The MSEA area is unusual in global terms in that there is good agreement
among scholars as to the basic language family affiliation of known languages.
There are unresolved issues about lower level subgroupings and there are
unresolved hypotheses about possible macro-groupings. But for every known
language, scholars agree as to which of the five main language families it fits into.
This is unusual firstly because it means that each language’s basic affiliation is
apparently uncontroversial, and secondly because it suggests that there are no
language isolates (Blench 2011: 125-126). For a survey of the historical linguistic
background, see Sidwell (2013).

Following is a list of some of the typological features that characterize MSEA
languages (drawing mostly from Enfield 2005: 186-190, 2011b: 69-70; see further
references there):

*Sound system features*

1. Large vowel systems (it is sometimes difficult to determine how many vowels a
   system has; there are alternative analyses of features such as diphthongs and
   phonation splits).
2. Common underlying structure of vowel phoneme system (often 9-place,
   symmetrical; hi-mid-low by front-central-back).
3. Long versus short vowel distinctions.
4. Many more consonants are possible in initial position than in final position;
   syllables have an initial-and-rhyme structure.
5. Preference for one (major) syllable per word, with many languages featuring
   minor syllables or pre-syllables in an iambic pattern (see Pittayaporn 2015,
6. Lexical contrast is marked by laryngeal features including pitch and phonation
   type, often in combination; tone systems are complex (number of tones ranges
   from 4 to 15 in number, with counts for a language differing depending on the
   analysis chosen); phonation type systems usually distinguish two registers, e.g.,
   ‘clear’ versus ‘breathy’; lexically contrastive pitch and phonation type are
   strongly correlated in functional and historical terms.
7. Gap in voiced stop series at velar place of articulation (no voiced ‘g’).

---

4 Not considered in this chapter are sign languages. The sign language used in Ban Khor,
Thailand (Nonaka 2004) appears to be an isolate, and there are surely more of its kind.
Among spoken languages in MSEA there is Kenaboi, now extinct, and known only from
two early twentieth century word lists. Hajek (1998) refers to Kenaboi as ‘unclassified’
but does not call it an isolate. Benjamin (2006) summarizes and analyses the available
data as far as is possible. His view is that Kenaboi is ‘a specially-invented form of
speech’, a ‘taboo-jargon’ associated with forest collecting trade. Kenaboi had large
proportions of both Austroasiatic and Austronesian vocabulary, along with some
unexplained forms. The data are too tenuous to establish whether it was an isolate or not.
Morphosyntax-semantics system features

1. No inflectional morphology (no case, gender, number, or definiteness marked on noun phrases, no agreement or tense-marking on verbs); note that derivational morphology is widespread and sometimes highly productive in Austroasiatic languages of MSEA (see Alves 2015).

2. Open class items – mostly nouns and verbs – serve functions that are expressed by dedicated functional morphemes (including bound morphology) in other languages, e.g., nominals as prepositions, verbs as aspect markers, comparative markers, adversative passive markers, and valence-changing devices (Clark and Prasithrathsint 1985; Kölver 1991; Ansaldo 1999).

3. Widespread use of verb serialisation (meaning a range of different kinds of predicative structures that use combinations of verbs), with a rich array of types and functions in each language (Bisang 1991).

4. Order of major constituents of the clause tends to be relatively flexible within languages, sensitive to pragmatic factors (though verb-object constituent order is dominant); noun phrases tend to be left-headed, and may have discontinuous constituents, especially when classifiers are involved.

5. Zero anaphora: noun phrases may be ellipsed when their referents are contextually retrievable (this combined with flexibility in constituent order results in quite variable surface options; for a case study see Enfield 2007: 271-284).


7. Large set of labile or ambitransitive verbs, especially of the causative/inchoative or unaccusative type (e.g., Lao hak2 can mean transitive ‘snap’ or intransitive ‘is/has been snapped’).

8. Rich inventories of sentence-final particles that make subtle distinctions in sentence type, stance, evidentiality, and combinations thereof.

9. Rich inventories of ideophones (or ‘expressives’) and other expressive forms, including rhyming four-syllable expressions, and productive elaborative rhyming devices.

10. Numeral classifiers and related systems of nominal classification (see Blench 2015).

11. Complex pronominal systems, with multi-level social-deictic meanings.

Some of the most noteworthy commonalities among MSEA languages concern their lack of marking of certain semantico-grammatical categories. Most notably, as remarked upon in the list above, the languages almost entirely lack inflectional morphology in the usual sense of that term (i.e., including agreement, case, gender/number/definiteness on noun phrases, tense-marking on verbs). For an overview of selected national languages, see Comrie (1990), while Goddard (2005) presents a more topic-oriented approach; see also Vittrant and Watkins (2015).

3 Linguistics of MSEA: Resources and developments

3.1 Conferences and publications
The community of scholars working on MSEA linguistics is steadily growing. The South East Asia Linguistic Society (SEALS)—founded by Martha Ratliff and Eric Schiller at Wayne State University, Detroit, in 1990—will hold its 25th annual meeting in 2015. Prior to 2009, proceedings of SEALS meetings were published in edited volumes. Since then they have appeared in the open-access *Journal of the Southeast Asian Linguistics Society* (for which, see [http://www.jseals.org/](http://www.jseals.org/)). The SEAlang Projects website ([http://www.sealang.net/](http://www.sealang.net/)) is an invaluable resource that makes accessible a range of primary and secondary sources on MSEA languages. Other regular publishing venues for research on MSEA languages include the journals *Mon-Khmer Studies* (an open-access journal, see [http://www.mksjournal.org/](http://www.mksjournal.org/)) and *Linguistics of the Tibeto-Burman Area* (see [http://sealang.net/sala/lbta/htm/index.htm](http://sealang.net/sala/lbta/htm/index.htm)). Some recent interdisciplinary explorations of ethnolinguistic diversification have focused on languages of MSEA and neighbouring places (e.g., Sagart, Blench, and Sanchez-Maras 2005; Enfield 2011a). The last 10 years have seen the publication of multiple landmark overviews of MSEA language families, including Tai-Kadai (Diller, Edmondson, and Luo 2008), Sino-Tibetan (Thurgood and LaPolla 2003; cf. Matisoff 2003a), Austroasiatic (Jenny and Sidwell 2015; cf. Shorto 2006), and the Austronesian languages of MSEA (Thurgood 1999; Grant and Sidwell 2005; Larish 2005; Blust 2013b: 70-75).

### 3.2 New descriptive work

A key measure of progress in an area is the production of reference materials based on new empirical research. Full-sized descriptions of MSEA languages published since the turn of the century include grammars of Semelai (Kruspe 2004), Jahai (Burenhult 2005), Garo (Burling 2004), Deuri (Jacquesson 2004), Mongsen Ao (Coupe 2007), Lao (Enfield 2007), Anong (Sun and Liu 2009), Hainan Cham (Thurgood, Thurgood, and Li 2014), Turung (Morey 2010), the Tai languages of Assam (Morey 2005), Lisu (Yu 2007), Thai (Higbie and Thinsan 2003; Iwasaki and Ingkaphirom Horie 2005) and Cambodian (Haiman 2011). Numerous grammars have been completed as PhD dissertations; just in the area of northeast India, for example, see grammars of Galo (Post 2007), Atong (van Breugel 2014), and Karbi (Konnerth 2014). Sketches or partial descriptions have appeared on languages including Pacoh (Alves 2006), Kri (Enfield and Diffloth 2009), and Arem (Ferlus 2014), and detailed descriptions have appeared of specific domains of grammar such as phonetics/phonology; see for example Watkins (2002) on Wa and Coupe (2003) on Ao. Major dictionaries of minority languages are less abundant; two notable examples are Watkins (2013) on Wa and Svantesson et al. (2013) on Kammu Yùan. An important preoccupation of descriptive linguistics globally is the

---

5 We mention here only a selection of those recent materials that have been published in English, though we note that a substantial descriptive literature on MSEA languages is being published in other languages, including Chinese, French, Indonesian, Thai, and Vietnamese (for some examples, see: Bo 2002; Bon 2014; Buakaw 2012; Chen 2005; Gai 2002; Giaphong 2004; Kosaka 2000; D. Li 2003, 2004; Y. Li 2003; Lidz 2010; Mao, Zongwu, and Yunbing Li 2002, 2007; Mayuree 2006; Ploykaew 2001; Samarina 2011; Seng Mai 2012; Shee 2008; Shintani 2008; Srisakorn 2008; Wayshepa 2010).
documentation of endangered languages; for excellent examples of new empirical work with this orientation in the MSEA context, see Morey (2005, 2010; see also Suwilai 1998, 2008).

A significant amount of new data and analysis from MSEA languages has become available on most if not all domains of interest to linguists, and on most if not all language families and sub-areas of MSEA. As just one example, here we mention the Aslian languages of Peninsular Malaysia. In the last decade or so, we have seen the publication of typological overviews of the Aslian languages as a group (Matisoff 2003b), new reference grammars (Kruspe 2004; Burenhult 2005), other descriptive materials (Burenhult and Wegener 2009; Wnuk forthcoming), new interdisciplinary research on the history and diversification of ethnolinguistic subgroups (Burenhult, Kruspe, and Dunn 2011; Dunn et al. 2011; Bulbeck 2011; Fix 2011; Oppenheimer 2011; Dunn, Kruspe, and Burenhult 2013), and field research on the psychological implications of semantic systems that are indigenous to Aslian languages and world views (Burenhult and Majid 2011; Majid and Burenhult 2014; Wnuk and Majid 2014). Not only is this breadth and depth of new work improving our basic understanding of MSEA languages and their socio-historical contexts, it is also helping to balance our perspective on the MSEA area, with effects on our image of what a Southeast Asian language is typically like (see below). The availability of new descriptive materials means that we can progress in the field by testing existing proposals and by continuously expanding the scope of our work (see Pittayaporn 2009 for a good illustration of this point).

3.3 New methods

As new methods in linguistic research are being developed and applied in linguistics globally, so they are being developed and applied in mainland Southeast Asia. In phonetics and phonology, for example, new instrumental and computational technologies are rapidly transforming the realms of possibility in data collection and analysis, both by making new kinds of measurement possible, and by making the equipment smaller and more portable for fieldwork; see Edmondson and Esling (2006: 172-175) for the use of laryngoscopy to study the phonetics of breathy vocal register in Jianchuan Bai (spoken in Yunnan), and Brunelle (2009) for the use of electroglottography to study register in Cham dialects in Vietnam (see also Brunelle, Nguyễn, and Nguyễn 2010 on Northern Vietnamese). Newly-developed statistical techniques are being applied with interesting results: in historical linguistics, probability-based bioinformatic techniques are being used for exploring cladistic representations of language relatedness (see for example Burenhult, Kruspe, and Dunn 2011); and in areal typology, statistical modelling is being used to test dependencies among phonological features, language history, and language contact (Brunelle and Kirby this volume). In lexical and grammatical work, new field methods are being applied in the exploration of semantic fields, in a range of functional and conceptual domains (see, for example Burenhult 2006; Wnuk and Majid 2014; Enfield 2015). There is an increasing interest in combining methods in order to further our knowledge of the area’s languages, for example in the interdisciplinary collaborations of historical work (Sagart et al. 2005; van Driem 2007; Enfield
And computational power is being exploited in building larger and better databases of, or including, MSEA languages (Dryer and Haspelmath 2013; Donohue et al. 2013).

3.4 Historical-comparative linguistics

Research in historical-comparative linguistics continues apace in MSEA. At the level of sub-grouping, advances are being made in all the major language families. Old hypotheses are being tested with new data and techniques, and new hypotheses are being put forward. The appearance of new data, in particular, has made an important difference, enabling, for example, Pittayaporn (2009) to propose a new reconstruction of Proto-Southwestern-Tai phonology, Sidwell (2009) to offer an improved account of vowels in Proto-Mon-Khmer, and Matisoff (2015; cf. Matisoff 2003a) to re-examine the place of the Jingpho language within Tibeto-Burman. In research on historical Hmong-Mien, Ratliff (2010) has recently provided an assessment of previous work and offers substantial new reconstructions, with consideration of their implications. Historical Austroasiatic has seen substantial developments, including a suspension of the assumption of a highest-level split between Munda and Mon-Khmer. It is no longer widely assumed that ‘Mon-Khmer languages’ represent descendants of a single ancestor language below Proto-Austroasiatic (although the term is still useful with the meaning ‘non-Munda Austroasiatic languages’; for a range of perspectives on this, see discussion in Sidwell and Blench 2011; Diffloth 2011; Sagart 2011; and Van Driem 2011). Similarly, in Sino-Tibetan linguistics, assumptions are being questioned. For example, recent reconsiderations of the position of Chinese in the family have assigned it to a lower-level subgroup rather than the standard placement as a high major branch; more subgroups of Sino-Tibetan are identified, and the time-depth of reconstructed proto-Sinitic is pushed back to well before Old Chinese (Blench and Post 2013; Van Driem 2013).

3.5 Language in social life

Numerous lines of work in linguistics deal with the role of language in social life. An important theme in recent work in MSEA is the sociolinguistics of language endangerment, and associated issues including language protection and revitalization; for an example, see Phthatharathanit (2012) on identity maintenance in Lanna (cf. Bradley 2007, Suwilai 2007). Research on linguistic politeness continues, mostly in relation to national languages, and with reference to the languages’ elaborated systems of social deixis, for example in their systems of personal pronouns, and the pragmatic alternatives that effectively create open class systems for person reference (Cooke 1968; Haas 1969; Luong 1990; Enfield 2015: Ch. 5). The more complex documented systems of person reference are those belonging to the major literate languages of the area, including Thai, Cambodian, Vietnamese, and Burmese (Cooke 1968). There has been recent work in this domain on languages including Lao (Enfield 2007: Ch. 5, 2015: Ch. 5). On Vietnamese, see Sophana (2008) on politeness strategies, and Sidnell and Shohet
(2013) on avoidance strategies (see also Luong 1988). Linking social life to central concerns of historical linguistics and typology, there has been recent work on sociolinguistic conditions for borrowing (Alves 2009); for similar work see Thurgood (2010) comparing two varieties of Cham with the Tibeto-Burman language Anong. A new line of work in MSEA is conversation analysis; Enfield (2013) presents several case studies of Lao language in conversation; Hà (2010, 2013) presents studies of Vietnamese conversation with a focus on the role of prosody, for example in repair and backchannelling (see also Umaporn 2007 on backchannelling in Mon).

3.6 Changing perceptions

Like in any area, linguistics in MSEA is subject to preconceptions. As soon as an idea becomes something of an orthodoxy it is right to revisit and question it. We are pleased that several chapters in this book raise and sometimes challenge certain assumptions about the linguistics of this area.

3.6.1 The idea of a typical MSEA language

Comrie (2007: 45) finds that, on measures taken using data from the *World Atlas of Language Structures* (Hasselmath et al. 2005), ‘Thai turns out to be the most typical of the three major national languages of Mainland Southeast Asia considered here.’ This conclusion is shared by Dahl (2008). This of course does not mean that Thai is the most typical of all MSEA languages, although this is often assumed to be the case. The national languages of the area are the better-described and better-known languages, and they happen to share many typological features that characterize Thai, such as a tendency for monosyllabicity, a lack of productive affixation, and an elaborate numeral classifier system. But there are many MSEA languages whose properties differ from these and many other properties found in Thai and other national languages like Vietnamese. In fact, many languages of the area lack these features. Within MSEA linguistics one’s view of what is typical may depend on one’s academic background, and, especially, on which language one worked on first, or has worked on most, in one’s research career. If, for example, one’s earliest and most in-depth work on MSEA languages was on Lao (as is the case with the present author), then languages like Lao and Thai would seem typical. They are typologically very similar to other major languages like Vietnamese. Another researcher’s background would suggest otherwise. The viewpoint professed by our colleague Gérard Diffloth is that a typical MSEA language lacks lexical tone, has complex phonotactics including syllable-initial consonant clusters, and has productive derivational morphology, quite a contrast from the oft-cited set of features of MSEA languages (see Henderson 1965; Capell 1979; Suwilai 1987; Kruspe 2004; cf. Alves 2001, 2015). The problem with treating the area’s major national languages as reference points is not only that they

---

6 The idea that a language may be ‘typical’ of an area seems to be an intuitive one, but the relevant sense in which a language can be said to be typical is seldom defined.
are a tiny sample but that they are known to be not like the rest, due to factors including (1) they are spoken by very large, often urbanized populations, (2) they are spoken as second languages by large sections of the population, (3) they are official languages, used in major education systems, media and broadcasting, and legal documents.

3.6.2 Nominal classification

MSEA is often cited in typologies of nominal classification as an area that has numeral classifiers (cf. Grinevald 2000; Aikhenvald 2000). Recent research shows that systems of nominal classification in MSEA can be more complex than this. They not only contain the classic numeral classifier type, consisting of a large set of classificatory nominals that are used whenever something is being numerated, but also systems that resemble the noun class systems found widely in Africa and the Amazon, and ancillary systems that resemble numeral classifiers but which are involved in the use of more simple modifiers such as demonstratives and specifiers. Enfield (2007: 119-156) shows that in Lao there are in fact four distinct grammatical systems of nominal classification, of which numeral classifiers are one (see Blench 2015).

3.6.3 Sesquisyllables

Researchers of the sound structure of words in MSEA languages often refer to the idea of ‘sesquisyllables’ and even the property of ‘sesquisyllabicity’. This term was introduced by Matisoff (1973) to refer to the ‘one-and-a-half syllable’ form of words found in many MSEA languages (see Henderson 1952; Shorto 1960; Butler 2015, Pittayaporn 2015). The term has not always been applied in an exact or consistent way. In a narrow sense, it can refer specifically to a syllable with schwa epenthesis between elements of an initial consonant cluster; that is, a syllable whose onset is phonologically /CC/ but phonetically [CəC]. In a broad sense, it can refer to any word that has an iambic structure, with the main stressed syllable coming at the end. Consider the following three words in Kri (Enfield and Diffloth 2009): /cakaanŋ/ [cakaːŋ] ‘to measure something by handspans’, /ckaanŋ/ [c⁹kaːŋ] ‘a hand span’, and /caanŋ/ [caːŋ] ‘buttress of a tree’ (or /kaanŋʔ/ [kaːŋʔ] ‘chin/jaw’). In the broad sense, both /cakaanŋ/ [cakaːŋ] and /ckaanŋ/ [c⁹kaːŋ] are sesquisyllabic, while in the narrow sense, only /ckaanŋ/ [c⁹kaːŋ] is. Recent work—by Butler (2015) and Pittayaporn (2015)—has made a significant advance not only by insisting that we be consistent and precise in the use of such terms, but by turning to empirical and theoretical accounts in order to offer motivated solutions, making the intuitive idea of sesquisyllabicity accountable to current theory and data in theoretical phonology and articulatory phonetics. Butler (2015) calls for more thoughtful consideration of the terms, and seeks to make progress by holding certain phonological ideas of syllable structure accountable to phonetic behaviour that can be experimentally tested. Pittayaporn (2015) takes a broader comparative approach
to the problem, offering a typology of sesquisyllabic languages, defining the distinct meanings that this term can have.

3.6.4 Tone phonetics and phonology

An oft-cited feature of MSEA languages is that many of them are tone languages. When asked what this means, most linguists would agree with Yip (2002:1): ‘A language is a “tone language” if the pitch of the word can change the meaning of the word.’ But as linguists of MSEA languages since Henderson (1952, 1965, 1967) have insisted, it is wrong to think that pitch is the sole or defining feature of a tone system in MSEA (see Brunelle and Kirby 2015 and Sidwell 2015; see also Abramson and L-Thongkum 2009): ‘It is important to recognize that pitch is frequently only one of the phonetic components of “tone” as a phonological category. … A phonological tone is in our area very frequently a complex of other features besides pitch—such as intensity, duration, voice quality, final glottal constriction and so on.’ (Henderson 1967: 171). From this perspective, while tone and phonation type are sometimes considered to be distinct phonological organizations, they should instead be treated as instances of a single sound system property insofar as they each involve the use of laryngeal features for lexical contrast. Pitch contours, distinctions in phonation, and other glottalic effects are all produced in the larynx, by the vocal folds, and are all articulatorily independent of segmental speech sounds produced with the lips, teeth, and tongue (i.e., consonants). Tone and phonation are intimately bound, and not essentially distinct. For this reason we recognize that the sound system of an MSEA ‘tone’ language, such as Vietnamese, is not of a different species from that of a classical MSEA ‘register’ or ‘phonation type’ language such as Kri (Enfield and Diffloth 2009). Most systems that are identified as one or the other (in phonological terms) actually display properties of both (in phonetic terms).

3.6.5 MSEA as a linguistic area

In research on areal linguistics, a great deal of new empirical and conceptual work from around the world has improved our understanding of historical processes of ethnolinguistic diversification, contact, and convergence, while at the same time some of the basic tenets of areal linguistics have come under question (Stolz 2002; Muysken 2008). MSEA has been widely regarded as a classic linguistic area with close parallelism in structure between neighbouring languages that have no demonstrable common ancestor (see Henderson 1965; Capell 1979; Clark 1989; Matisoff 1991; Bisang 1991; Enfield 2005; Comrie 2007; Dahl 2008; Vittrant and Watkins forthcoming). The cause of this parallelism is widely assumed to be language contact. While much work examines typological parallels across language families and interprets these as evidence of effects from language contact, recent work by Sidwell (2015), Ratliff (2015), and Brunelle and Kirby (2015) calls for caution in jumping to that conclusion. If neighbouring but unrelated languages share typological features this can also be a result of parallel language-internal development (cf. Thurgood 1998; Enfield 2005). That possibility is equally
deserving of consideration, and so the idea that convergence is due to contact should not be assumed without question.

4 Summary

Languages in the mainland Southeast Asia area are well known for their seemingly high degree of convergence in many if not all aspects of their grammar. Recent empirical and theoretical advances in the study of mainland Southeast Asian languages are driving important developments in our understanding of the area. While much new work is confirming and enriching our understanding of known patterns of structure in the area, other contributions are challenging orthodoxy, including the idea of what a ‘typical’ mainland Southeast Asian language is like, and even the very notion that the high degree of convergence in the area is a direct result of social contact among historical speech communities.

Acknowledgements

This chapter draws directly on sections of a chapter co-authored with Bernard Comrie that appeared as the introduction to our 2015 edited book Languages of Mainland Southeast Asia: The State of the Art (Berlin: de Gruyter Mouton, 2015): I thank Bernard for allowing me to revise and submit this chapter, based on our earlier co-authored work. For financial support, I am grateful to the Max Planck Institute for Evolutionary Anthropology (Department of Linguistics), the Max Planck Institute for Psycholinguistics (Language and Cognition Department), and the European Research Council (through grant 240853 ‘Human Sociality and Systems of Language Use’). I would like to thank Maarten van den Heuvel for technical assistance in preparing the manuscript for publication, and Angela Terrill at Punctilious Editing (http://www.punctilious.net/) for copy-editing. For comments and suggestions on a draft, I thank Roger Blench, Jeremy Collins, Bernard Comrie, Mark Donohue, David Gil, Pittayawat Pittayaporn, Mark Post, Martha Ratliff, and Paul Sidwell.

References


Majid, Asifa & Niclas Burenhult. 2014. Odors are expressible in language, as long as you speak the right language. Cognition 130(2). 266–270.


25 Southeast Asian tone in areal perspective

James Kirby and Marc Brunelle

1 Introduction

Tone is often presented as one of the quintessential features identifying mainland Southeast Asia (MSEA)\(^1\) as a linguistic area (Henderson 1965; Matisoff 2001; Enfield 2011). For Matisoff, the proliferation of tone languages is ‘[p]erhaps the most striking phonological feature of the South-East Asian linguistic area’ (2001: 291), while the ubiquity of tone tops Henderson’s list of ‘features…typologically characteristic of a South East Asian linguistic area’ (1965: 401). The presence of tone in a large number of genetically unrelated languages has commonly been attributed to areal diffusion, with Chinese most commonly hypothesized as the ultimate source (Matisoff 1973; Pulleyblank 1986; Benedict 1996).

The reference to tone as a ‘feature’ in the preceding citations might suggest that there exists a simple typological dichotomy between languages with and without tone. Upon closer inspection, however, the phonetic, phonological, and typological characteristics of MSEA tone systems differ in important ways. To the extent that this diversity reflects substantive differences between languages, it raises the question of precisely what role contact has played in the evolution of tone in MSEA. In this chapter, we address this question through an examination of the phonetic and phonological properties of MSEA tone systems as well as of proposals regarding their evolution. After briefly discussing tone systems in the broader typological perspective, we present an overview of the phonetic, phonological, and genetic characteristics of MSEA tone systems, emphasizing the rich variability of tonal realization found in the region. Next, we discuss the ways in which languages can become tonal, reviewing evidence for the spread of tone through contact as well as for the idea that much of the observed tonality on the ground in modern MSEA might be traced to a small number of ‘tonogenetic events’ rather than a large number of borrowings. In light of this discussion, we consider whether a re-evaluation of the notion of tone as a canonical indicator of ‘linguistic area’ more generally is warranted. While our treatment is focused on a particular geographic region, we hope that this areal perspective on tone can also be of use to scholars working in other linguistic areas where large numbers of genetically unrelated tone languages are present.

2 Synchronic typology of tone in MSEA

2.1 On the definition of ‘tone language’

---

\(^{1}\) In this chapter, we use the term ‘mainland Southeast Asia’ (MSEA) to refer to the Indochinese peninsula comprising the modern states of Vietnam, Laos, Cambodia, Thailand, Myanmar (Burma), Singapore and the mainland territory of Malaysia.
Before discussing the tonal properties of MSEA languages, it may be useful to briefly review some standard approaches to classifying and typologizing tone systems. As suggested above, the notion of a fundamental dichotomy between ‘tone languages’ on the one hand and ‘non-tone languages’ on the other has tempted researchers since the dawn of prosodic analysis (Lehiste 1970). Unlike (segmental) phonological features such as presence versus absence of voiceless nasals or final consonant clusters, the essential qualities of tone have proven harder to pin down, but definitions of ‘tone language’ almost always involves some reference to the paradigmatic use of pitch. For example, Pike (1948: 3) offers the following: “A tone language may be defined as a language having lexically significant, contrastive, but relative pitch on each syllable”. Hyman (2009:229) gives a similar definition: “A language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes”.

Setting aside the complications such definitions raise for the classification of languages such as Japanese or Swedish (so-called ‘pitch-accent’ languages), the primary role of pitch is also assumed by researchers who subdivide the set of tone languages into ‘simple’ and ‘complex’ based on the number and nature of the pitch movements. Pike himself distinguished between ‘register-tone’ and ‘contour-tone’ languages, with the former encompassing languages with a ‘small, restricted number of tone contrasts between level tonemes’ (1948: 5; emphasis ours), while the latter type comprise languages ‘in which glides are basic to the system, with no level tonemes whatsoever’ (ibid., 8). Similarly, Maddieson (2011) makes a distinction between simple and complex tone languages where simple languages are ‘essentially those with only a two-way basic contrast, usually between high and low levels’ and complex languages are everything else.

Such classifications likely stem from the observation that African- and New World-type tone systems often have phonetically less complex tone systems compared to those of East and Southeast Asia. However, the differences between these systems may rest less with aspects of phonetic realization and more with the fact that the tone systems of African and New World languages tend to exhibit phonological properties such as decompositionality or spreading, as well as a tendency for tone to be employed for grammatical functions (inflection, derivation, etc.). Conversely, tone may have less of a lexically contrastive function in Africa and the Americas, whereas purely tonal minimal pairs are extremely common in many MSEA languages. In any case, there is significant tonal variability in each of these areas, as we show for MSEA in the next section.

2.2 Variation in MSEA tone systems

In principle, a division between simple and complex tone languages could be useful for typologizing a given geographic region. At least in MSEA, however, the closer one stays to the phonetic ground, the less satisfying the simple/complex dichotomy becomes: not

---

2 Note that languages with both of these properties exist in MSEA as well: the contour tones in several Naish languages have been argued to be properly analyzed as sequences of level tones (Michaud and Xueguang 2007; cf. Clements et al. 2011), and tone is frequently deployed to grammatical ends in Hmongic (Ratliff 1992b) and Tibeto-Burman (Henderson 1967) languages, especially those of the Chin group (Hartmann-So 1989; Hyman 2010; Watkins 2013; see also Ratliff 1992a).
only is the boundary between tone and non-tone languages difficult to ascertain, but there is considerable variation in the properties of tone systems even between ‘clearly’ tonal languages (Abramson and L-Thongkum 2009). Here we consider several types of tone-related phenomena commonly encountered in MSEA: variation in pitch realization, phonation type, register, and the domain of tonal contrasts.

2.2.1 Variation in pitch realization

One reason that application of a simple/complex dichotomy to MSEA languages can be misleading is that there exists a large number of languages with just a two-tone system (‘register-tone’ languages in Pike’s sense) but where one of the tones is phonetically realized as a contour, at least some of the time. One such example is Western Khmu (Kammu) (of the ‘tone 1’ variety given in Table 1), which contrasts a low and a falling tone – pitch apparently being the primary or sole phonetic parameter of relevance here. This contrasts with other Western dialects in which phonation type also plays a role.

Table 1. Tones in Khmu dialects (after Premsrirat 2001). The initial voicing contrast maintained in (toneless) Eastern Khmu is transphonologized into a low versus falling contrast in one Western dialect (‘tone 1’), a low versus high contrast in another (‘tone 2’), and a breathy versus falling contrast in a third (‘register’).

<table>
<thead>
<tr>
<th>E. Khmu</th>
<th>W. Khmu (tone 1)</th>
<th>W. Khmu (tone 2)</th>
<th>W. Khmu (register)</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>buːc</td>
<td>pːc</td>
<td>pʰuːc</td>
<td>pːc</td>
<td>‘rice wine’</td>
</tr>
<tr>
<td>puːc</td>
<td>pːc</td>
<td>pʰuːc</td>
<td>pːc</td>
<td>‘to take off clothes’</td>
</tr>
<tr>
<td>glaːŋ</td>
<td>klâːŋ</td>
<td>kʰləːŋ</td>
<td>klâːŋ</td>
<td>‘stone’</td>
</tr>
<tr>
<td>klaːŋ</td>
<td>klâːŋ</td>
<td>kʰləːŋ</td>
<td>klâːŋ</td>
<td>‘eagle’</td>
</tr>
</tbody>
</table>

It is not clear exactly how many MSEA languages have ‘simple’ (i.e., two- or three-tone) systems where one, or possibly both, of the tones are realized with rising, falling, or more complex pitch movements, in part because tone systems are often described solely for their phonological (i.e., contrastive) rather than phonetic properties. At the same time, care must also be taken when designating surface pitch movements to be inherent tonal properties, as non-lexical factors may also influence surface tonal realization. For example, the Tibeto-Burman language Naxi has four surface tones (high, mid, rising, and falling), but the falling tone is an intonational allotone of the high and mid (level) tones, while the rising tone is often the result of a (phonological) process of tonal reassociation (Michaud 2006, 2013).

2.2.2 Phonation types and tone: tone is not (only) pitch

A second factor complicating the analysis of tone in MSEA is that voice quality, or
phonation type, is a crucial aspect of tone in many languages of the region. By this we mean that a voice quality setting (modal, breathy, creaky, tense, etc.) is canonically present as an (obligatory) phonetic cue to the tone category along with pitch (and potentially other features as well). Far from being unusual, tone systems involving different phonation types are a common feature of many widely-spoken MSEA languages such as Vietnamese (Maspero 1912; Nguyễn and Edmondson 1997), Burmese (Watkins 2001; Gruber 2011) and Green Hmong (Andruski and Ratliff 2000); indeed, in an early survey Henderson remarked that ‘… “tone” is seldom, if ever, a matter of pitch alone’ (1965: 404). Such systems are sometimes termed ‘phonation-prominent’ (Matisoff 1973) or ‘mixed pitch/phonation type’ (Andruski and Ratliff 2000); for brevity, we will use the term ‘mixed’ to refer to this type of system throughout this article.

The canonical example of a mixed tone system is Northern Vietnamese, where at least two and possibly three of the language’s six tones (low-falling hôi, broken-rising ngã and falling-glottalized nàng) involve glottalization or laryngealization along with a distinctive pitch movement (Nguyễn and Edmondson 1997; Pham 2003; Michaud 2004), and where the voice quality setting serves as a crucial perceptual cue for listeners (Brunelle 2009; Kirby 2010). While most mixed tone systems involve either breathy or creaky/laryngealized voice quality alongside modally voiced tones, many Hmongic languages have tone systems involving modal, breathy, and creaky phonation types, and tone in dialects of the Tai-Kadai language Nùng is reported to involve both modal and glottalized phonation types along with either creaky or breathy voice (Nicolson 2000).

The widespread tendency of MSEA languages to involve phonation type as an integral part of tonal specification makes for a fuzzy boundary between mixed tone systems and so-called register systems, to which we now turn.

2.2.3 Register systems

In many MSEA languages, especially those of Austroasiatic or Austronesian stock, lexical contrasts are signaled by a ‘bundle’ of (broadly suprasegmental) features, such as phonation type, pitch, vowel quality, intensity, and vowel duration. Such languages have been termed (voice-)register languages in the Southeast Asian linguistic literature (Henderson 1952; Gregerson 1976; Ferlus 1979; Diffloth 1982). Here, register normally refers to a type of phonological contrast arising from the neutralization of voicing in onsets and subsequent phonologization of phonetic properties originally associated with voicing, but in rarer cases, the loss of final laryngeals conspires with voicing neutralization to lead to the development of complex register systems, as in some Pearic and Vietic languages (Di Canio 2009; Ferlus 2004; Enfield and Diffloth 2009). Depending on the language in question, register may be a property of the onset, the rime.

Following Abercrombie (1967), Laver (1980), and others, we use the term ‘voice quality’ as a general term referring to voice settings for a variety of purposes, and ‘phonation type’ to refer to the phonetic realization of voice quality when employed for phonological purposes.

The low-falling (huyền) tone in Northern Vietnamese is also optionally breathy (Nguyễn and Edmondson 1997).

Note that this use of the term must be distinguished both from Pike’s use of ‘pitch registers’ (referring to pitch levels) as well as from its use when referring to the sociolinguistically distinct speech styles associated with many Austronesian languages (Uhlenbeck 1964).
or the entire syllable; phonetically, however, register systems involve a common set of acoustic correlates, listed in Table 2 (the low register derives from former voiced stops and the high register derives from former voiceless stops).

It is important to note that not all of the phonetic properties in Table 2 are found in all register languages. For example, despite evidence that many of these properties are present to some extent in Eastern Cham at a fine-grained phonetic level (Brunelle 2005, 2006), this language primarily contrasts registers through pitch and voice quality. In modern Standard Khmer, register is expressed exclusively through vowel quality, although historical evidence and the acoustic analysis of more conservative dialects suggests that it may have passed through a stage in which phonation type was a prominent cue (Jenner 1974; Wayland and Jongman 2001, 2002). Huffman (1976) and Ferlus (1979) provide good overviews of the range of variation in register systems among Mon-Khmer languages.

Table 2. Possible phonetic correlates of register. High register typically develops from proto-voiceless stops, low register from proto-voiced/aspirated stops.

<table>
<thead>
<tr>
<th>High register (voiceless stops, [*pa])</th>
<th>Low register (voiced stops, [*ba])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher pitch</td>
<td>Lower pitch</td>
</tr>
<tr>
<td>Tense/modal voice</td>
<td>Lax/breathy voice</td>
</tr>
<tr>
<td>Monophthongs/shorter vowels</td>
<td>Diphthongs/longer vowels</td>
</tr>
<tr>
<td>Raised F1 / lower vowels / [+ATR]</td>
<td>Lowered F1 / higher vowels / [-ATR]</td>
</tr>
<tr>
<td>Plain stops/shorter VOT</td>
<td>Aspirated stops/longer VOT</td>
</tr>
</tbody>
</table>

Since tone systems can also make use of phonation types, as we have just seen, this raises the question of how one decides whether one is dealing with a register system or a mixed tone system. Some researchers have suggested that register systems constitute a typological profile distinct from tone languages, including those of the ‘mixed’ variety. Di Canio (2009) provides a careful phonetic analysis of the Takhian Thong variety of Chong, which involves modal, breathy, tense, and breathy-tense phonation types, accompanied by (marginal) pitch differences. This suggests to him a fundamental distinction between register languages and tone languages (Di Canio 2009: ??):

a register language is distinct from a tone language because contrastive phonation type typifies the former, while contrastive pitch typifies the latter. Phonation type is to a register language what tones are to a tone language.

The problem with this definition is that it is not always (or even usually) clear which feature is dominant, acoustically, perceptually, or phonologically. A classic example here is that of Burmese, which has been described both as a register system (e.g. Bradley 1982; Jones 1986) and as a (mixed) tone language (e.g. Watkins 2001; Gruber 2011). While the precise details are somewhat complicated, Burmese syllables can bear one of four registers/tones, shown in Table 3. However, Gruber (2011) has shown that glottalisation, creakiness and the presence of a high pitch target are all important perceptual cues, thus demonstrating that Burmese cannot be analyzed only in terms of pitch or voice quality.
Table 3. ‘Tone’ in Burmese. Examples and notation from Watkins (2001).

<table>
<thead>
<tr>
<th>Tone Type</th>
<th>Notation (Watkins 2001)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>/ma/ [maː˩]</td>
<td>'hard'</td>
</tr>
<tr>
<td>High</td>
<td>/má/ [maː˦]</td>
<td>'towering'</td>
</tr>
<tr>
<td>Creaky</td>
<td>/mə̰/ [mə̰˥˩]</td>
<td>'female'</td>
</tr>
<tr>
<td>Killed</td>
<td>/maʔ/ [maʔ˥˩]</td>
<td>'March'</td>
</tr>
</tbody>
</table>

Abramson and L-Thongkum (2009) also suggest a distinction between tone and register based on the prominence of the primary cue, with tone languages being primarily cued by pitch and voice-register languages primarily cued by phonation type. However, they acknowledge this boundary can be fuzzy, and suggest that tradition and researcher degrees of freedom are likely to play a significant role in the assignment of languages to one category or the other. A similar sentiment is echoed by Watson (1996: 202) in an essay on Pacoh vowel phonology:

thren tends to be a dichotomy in voice quality ranging from breathy to clear to creaky, in pitch ranging from high to mid to low, in voicing of initial consonants, in vowel height between close and open, in vowel gliding between onglided, plain or offglided, and tension from tense to lax. In some cases there has been a general movement from a distinction between voicing in initial consonants to a distinction in vowel quality and/or pitch.

Enfield (2011: 69) suggests treating tone and phonation

...as instances of a single sound system property because they each involve the use of laryngeal features for lexical contrast. While tone and phonation type are often considered to be fundamentally distinct phenomena, in fact most systems that are identified as one versus the other (in phonological terms) actually display properties of both (in phonetic terms; Henderson 1967: 171). Pitch contours, distinctions in phonation type, and glottalic effects are all produced in the larynx (specifically, by the vocal folds), and are all articulatorily independent of segmental speech sounds produced with the lips, teeth, and tongue (i.e., typical ‘consonants’). Tone and phonation are intimately bound, not essentially distinct, and for this reason I do not regard the sound system of a classical MSEA tone language such as Vietnamese to be of a different species from that of a classical MSEA register language such as Kri (Enfield and Diffloth 2009).

While we are broadly in agreement with this position, it is worth noting that laryngeal and supralaryngeal features are at least weakly related via the connection of the tongue root to the larynx via the hyoid bone.

Additional evidence for the fluidity of laryngeal features in MSEA can be found in the comparison of related languages (or dialect continua) where the same set of segmental and suprasegmental properties seem to take on different degrees of prominence in different dialects. In other words, prominence is often unstable. The Khmu dialects described in Table 1 above constitute one such example; Lamet, a Palaungic Mon-Khmer language, is another. Narumol (1982) and Svantesson (1988) describe
versions of the language with two contrastive voice registers, but no pitch (tone) distinction, while Conver (1999) describes a phonemic 2-tone system realized as high and low pitch. Conversely, Lindell et al. (1978) indicate that none of the dialects they studied made use either of tone or of contrastive phonation type distinctions. The nascent tone contrast in some dialects of Khmer shows similar variation, ranging from dialects that maintain differences in voice quality, f0 and vowel quality (Wayland and Jongman 2001) to dialects that retain vowel quality distinctions only, such as modern Standard Khmer.

There are at least two ways one might begin to address the issue of phonetic prominence in tone/register systems. One is to look at the weighting of cues in production by conducting e.g. linear discriminant or factor analyses of acoustic data (e.g. Andruski and Ratliff 2000 on Green Hmong; Abramson et al. 2004 on Suai, Abramson et al. 2007 on Khmu). Such studies may also reveal diachronic changes in progress. In their study of Khmu Rawk, for example, Abramson et al. (2007) found that male speakers no longer produce a measurable phonation type difference between voice registers, suggesting that F0 is becoming a more prominent cue for at least some speakers of this language.

The issue of establishing acoustic separability is in principle a separate undertaking from determining whether a given cue is used by native listeners in perception (see e.g. the previously cited work by Abramson and colleagues; Hombert 1977; Mazaudon and Michaud 2008; Brunelle 2009; Brunelle and Finkeldey 2011; Gruber 2011; Kirby 2014). This type of work involves experimental manipulation of acoustic properties of natural or resynthesized stimuli, using either an alternative forced-choice identification paradigm (and subsequent analysis of error rates and classification trees) or discrimination tasks (which facilitate analysis of reaction time data: see e.g. Gandour 1983; Kirby 2010). The interpretation of perceptual responses may be complicated by the existence of learned and/or inherent perceptual dependencies between cue dimensions (Brunelle 2012). Nevertheless, the available perceptual studies suggest that the relation between acoustic separability and perceptual weighting in MSEA tone and register systems is far from straightforward, and this remains an area of active research.

Even if a prominence hierarchy can be established, it is not clear that languages in which pitch is the most prominent aspect of the system are fundamentally different from languages in which voice quality, or vowel height, or even voicing are the most prominent aspects, or that any of these are different from languages in which no laryngeal cue is prominent. Indeed, given that redundancy is the default state of phonetic contrast (Lisker 1986), one might argue that clear instances of ‘pure tone’ or ‘pure voice quality’ languages are actually rather unusual from a functional perspective, and that register systems or systems in flux would be expected if perceptual robustness were somehow privileged. Thus, while there may be a descriptive utility to terms such as ‘register language’, ‘mixed pitch/phonation type tone language’, etc. it is important to bear in mind that the borders between them are likely to be extremely porous.

As a final note, it is worth pointing out that while pitch and phonation type often cross-cut one another in the prosodic systems of MSEA languages, there are also languages that have segmentally anchored phonation type distinctions in addition to fully developed tone systems. One such example is the endangered Tibeto-Burman language Mpi, spoken in northern Thailand (Ladefoed and Maddieson 1996). This language has a system of six tones, each of which may co-occur with a plain or laryngealized (tense or stiff-voiced) vowel, as shown in Table 4. While this type of system seems to be relatively rare among MSEA languages, further examples may be found among the Sino-Tibetan
languages spoken in China such as Yi and Bai (Edmondson et al. 2001), as well as in the Oto-Manguean languages of Central America such as Itsunyoso Trique (Di Canio 2012).

Table 4. Contrasting tones and phonation types in Mpi (after Ladefoged and Maddieson 1996: 316).

<table>
<thead>
<tr>
<th>Tone</th>
<th>Modal voice</th>
<th>Stiff voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>si</td>
<td>low rising</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>si</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>si</td>
<td>mid rising</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>si</td>
<td>Mid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>si</td>
<td>high rising</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>si</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2.4 The domain of tonal contrast

Another important aspect, implicit or explicit, in many definitions of ‘tone language’ is that the syllable is cast as the relevant domain over which relative differences in pitch are defined. The resulting problem of how to classify ‘marginally’ tonal languages such as Swedish or Japanese has led some prosodic typologists to propose a tripartite classification of ‘tone’, ‘stress’, and ‘pitch-accent’ languages (Ding 2006; van der Hulst 2011), although Hyman (2006, 2009) points out that the unique properties of pitch-accent can be difficult to separate from those of tone and stress.

Languages where lexical tones are associated to units larger than the syllable are hard to come by in MSEA, but one does not have to go too far to find such languages. Many Tibeto-Burman languages of the Bodish and Qiangic subgroups such as Tamang (Mazaudon and Michaud 2008), Naxi (Michaud 2007), and Prinmi (Ding 2001), spoken in nearby China, Tibet, and Nepal, are characterized by ‘cumulative’ tone systems, where distinctive pitch patterns are defined over units determined both prosodically and morphologically (Mazaudon 1973; Evans 2001a; Michaud and Mazaudon 2006; Hildebrandt 2007). In the Tibeto-Burman language Lizu (Chirkova and Chen 2013), for example, monosyllabic words contrast low-(rising) and high-(falling) tones, e.g. /Rŋu/ ‘silver’ vs. /Fŋu/ ‘cow’, while three pitch patterns are observed in disyllabic words: two mid-level pitch contours of equal prominence (/RPmidzɹ/ ‘hare’), a left-prominent falling contour (/LPmidzɹ/ ‘pepper’), and a right-prominent rising contour (/RPmutsɹ/ ‘cat’).

Another way in which the ‘domain’ of tone can differ is in the extent to which it has diffused through the lexicon. Most tone languages of East and Southeast Asia have a restricted tonal inventory in syllables closed by an obstruent (often called ‘checked’ or ‘dead’ syllables). This impoverished inventory is usually attributed to the fact that these syllables preserved their segmental coda during the three-way split stage of tonogenesis (see section 3.1) and therefore did not develop a contrastive tone. Tai-Kadai languages often have similar (though less systematic) restrictions between tones and onsets. In Central Thai, for instance, a high tone may not appear after a voiced onset. In some languages, these tone-consonant interactions are even more radical: tone is phonemically marginal, with pitch-based contrast restricted to certain words or phonological
environments. For example, in several varieties of Khmer, an incipient tone contrast has developed following the loss of /r/ in onset position, leading to a small number of minimal pairs distinguished solely by pitch (Wayland and Guion 2005; Kirby 2014; see also section 3.2), while in the Tibeto-Burman language Kurtöp, tone is only contrastive following sonorants and the palatal fricative /ç/ (Hyslop 2009). It is also not uncommon to find tones that are restricted to certain lexical strata, such as loanwords. For example, although Mal (T‘in), a Mon-Khmer language of Thailand, contrasts both a falling and a rising tone, the rising tone is largely (though not exclusively) used with Thai loanwords (L-Thongkum and Chommanad 2008).

2.3 Some tonal characteristics of the individual language families

To give a different sense of the range of tonal diversity in MSEA, we include here a brief overview of the tonal properties of languages included in the database described in Brunelle and Kirby (2015). At the time of writing this database includes 186 Southeast Asian languages from five families. As noted by Matisoff (2001), migration patterns in Southeast Asia have traditionally been rather different from those in Europe; the result is that the branching-tree model of genetic relationships, already a simplification, is perhaps even less insightful in the MSEA case. For this and other reasons, sub-groupings and -branches are often contested; as such, we restrict our classification to major language families only.

As the preceding discussion makes clear, it is difficult to place MSEA languages into a single category on the basis of their lexical treatment of prosodic properties. Thus, instead of insisting on labels, we describe the languages in terms of the number of contrastive prosodic units, the number of distinct pitch units, and the number of voice quality dimensions they distinguish. In our database, we furthermore make note of properties such as consonant-tone restrictions; maximal canonical word shape (mono-, sesqui-6, or polysyllabic); and the complexity of codas. As we are relying largely on published sources, it is not always clear if descriptions should be interpreted phonetically or phonologically, but this approach has the advantage of allowing a more nuanced overview of the tonal properties of MSEA languages than would be gained by yet another arbitrary classification into a small number of sub-types.

2.3.1 Austroasiatic

Our sample contains 78 Austroasiatic languages (41.9%), all of them Mon-Khmer.7 The vast majority of the AA languages in our sample are sesquisyllabic. Around a third of these (24) are non-tonal, while another third (27) have systems of two tones or registers.

6 A sesquisyllable (Matisoff 1973) is a disyllabic word composed of a main, stressed, final syllable that may contain the full range of phonological contrasts of a language, preceded by an unstressed reduced ‘presyllable’ that is subject to radical contrast neutralization.

7 While often used interchangeably with Mon-Khmer, the Austroasiatic group is also thought to include the Munda languages, spoken in India and Bangladesh. The internal classification of Mon-Khmer languages is complicated; see Diffloth (2005) and Sidwell (2013) for overviews of the issues involved.
This includes languages such as Riang (Luce undated), Conver’s Lamet (Conver 1999), and T’in (Lua’) (L-Thongkum and Chomnanad 2008) which are described as having two pitches but no phonation type differences; languages like Western Bru (L-Thongkum 1979) or Narumol’s Lamet (Narumol 1982), described as register systems distinguished by phonation type differences only; and (most commonly) languages like Mon (L-Thongkum 1988) or Suai (Abramson et al. 2004), for which both phonation type and pitch distinctions are described.

The remaining Austroasiatic languages in our sample have systems of three or more tones. These languages almost always employ a combination of pitch and phonation type distinctions to signal tone categories. This set includes Vietic languages such as Rức (Nguyễn Văn Lợi 1993), Chứt and Thavung (Ferlus 1998), and Northern Vietnamese (Vũ Thăng Phương 1981; Nguyễn and Edmondson 1997), as well as Pearic languages such as Takhir Thong Chong (Di Canio 2009), where each of the four phonemic phonation types (modal, tense, breathy, and breathy-tense) are consistently realized with a unique pitch. The notable exceptions here are Kháng (Edmondson 2010), a language with six tones but no reported phonation type distinctions, and Southern Vietnamese. Kháng has demonstrably had considerable recent contact with Tai-speaking groups, although the same cannot be said for Southern Vietnamese.

It is interesting to note that the tone sandhi phenomena that are so pervasive in language families such as Sino-Tibetan and Hmong-Mien are unknown in Austroasiatic languages, although complex tone spreading processes are attested in Kammu (Svantesson 1983). This could be due to a lack of thorough descriptions, but it is at least worth noting that tone sandhis are so far unattested in otherwise well-described Vietnamese dialects.

2.3.2 Austronesian

While the Austronesian family contains many thousands of languages spoken by hundreds of millions of people, they are relatively thin on the ground in MSEA. There are just 20 Austronesian varieties in our sample (10.8%); of these, 11 are Chamic languages/dialects spoken in Vietnam and Cambodia, with the rest spoken in the Malay Peninsula and its vicinity. Austronesian languages of the Malay Peninsula tend to be disyllabic, while Chamic languages are mainly sesquisyllabic. Cham dialects proper show a tendency to monosyllabicity, the most extreme case being colloquial Eastern Cham, which has become almost entirely monosyllabic (Brunelle 2009a).

The majority of these varieties are atonal, but three dialects of Cham (Eastern Cham, Vietnamese Western Cham and Cambodian Western Cham) and two dialects of Raglai (Cac Gia Raglai and Southern Raglai) have developed register systems combining pitch and voice quality to various degrees (Lee 1966, 1998; Brunelle 2009b). Haroī, another coastal Chamic language, formerly had a register system that was restructured into a complex vowel system, in a manner reminiscent of Standard Khmer (Mundhenk and Goschnick 1977). Moken Dung, a Malayic language of the Andaman sea, is reported to have a two-tone system, although information about its source is scant at best (Naw Say Bay 1995). In short, some mainland Austronesian languages have undergone minor tonal developments, but this seems mostly limited to register.

Some other Austronesian languages spoken outside of mainland Southeast Asia, strictly construed, have also developed forms of tonality. Javanese has a register system
normally described as a tense-lax stop contrast, but which is in practice almost identical to MSEA register systems (Fagan 1988; Adisasmito-Smith 2004; Thurgood 2004; Brunelle 2010); other Malayo-Polynesian languages of Indonesia, such as Sundanese and Madurese, have similar systems. Tsat, a Chamic language spoken in Hainan, has developed a five-tone contrast from laryngeal codas and onset voicing, just like Vietnamese or Chinese (Maddieson and Pang 1993).

2.3.3 Sino-Tibetan

Sino-Tibetan languages, including Chinese dialects, make up 19.9% of our sample (37 languages). All of these languages are tonal to some degree, ranging from the two-tone systems of Bwe Karen (Henderson 1979) or Daai Chin (Hartmann-So 1989) to five- and six-toned Loloish languages such as Akha (Lewis 1973) or Lisu (La Maung Htay 2011). Roughly half of the Sino-Tibetan languages are purely pitch-based (e.g. many Chin languages, Pa’o Karen), while the other half are mixed pitch/phonation-type systems (e.g. Lisu, Sgaw Karen).

Most of the Sino-Tibetan languages spoken in MSEA are polysyllabic, largely because of a more or less opaque concatenation of monosyllabic roots and affixes. There are however many exceptions (sesquisyllabic Burmese, [largely] monosyllabic Yue Chinese) and these languages rank among the most widely spoken.8

In addition to tone sandhi processes (where the surface tone realization is affected by tonal environment), which are especially common in Chinese dialects, the tone systems of many Sino-Tibetan languages display tonal alternations such as spreading, re-association, contour simplification and OCP effects more commonly associated with African tone systems (Evans 2008; Michaud 2008; Hyman 2010, Watkins 2013). Besides these strictly phonological processes, Tibeto-Burman languages often have morphophonological tonal process and grammatical tones (Henderson 1967; Hyman and Haokip 2004).

2.3.4 Tai-Kadai

Tai-Kadai (also Kra-Daï) languages make up another 23.1% of our sample (43 languages). Tai-Kadai languages are mostly monosyllabic, although some languages have acquired sesqui- and polysyllables through borrowing from Khmer, Pali and Sanskrit and through occasional semantic bleaching of compounds. All Tai-Kadai languages in our database are tonal, with systems of 4 to 7 tones. Seventeen of these languages (39.5%) make no recorded use of voice quality, while the rest have mixed tone systems in which complex contour tones are combined with creakiness and/or glottalization. Interestingly, despite the complexity of Tai-Kadai tone inventories, no instances of tone sandhi have to our knowledge been reported in this language family.

2.3.5 Hmong-Mien

8 Note that Burmese also contains a large number of polysyllabic loanwords, and that Chinese languages, especially Mandarin, contain many disyllabic compounds.
There are eight Hmong-Mien varieties in our sample (4.3%), which reflects the fact that most of these languages are spoken in China rather than MSEA per se. All Hmong-Mien languages are highly tonal, having between six and eight tones. In nearly all documented cases, phonation type contrasts (modal, breathy and/or tense/creaky) are an integral part of the tone system. For example, two of the seven tones in Hmong Leng (the low checked -m tone and the mid-falling -g tone) are creaky and breathy-voiced, respectively (Andruski and Ratliff 2000). The exception here appears to be Iu Mien (L-Thongkum 1997; Bruhn 2007), the only Mienic language in our sample to lack a phonation type contrast.

Tone sandhi processes are found in most Hmong-Mien languages, although they are reported to be highly variable and lexicalized (Ratliff 1987). However, tonal alternation is commonly employed for morphological purposes (word classes and compounding) as well as in ideophonic expressives such as poob [pɔ́ːŋ] ‘to fall’ vs. poog [pɔ̤ːŋ] ‘the sound of falling’ (Ratliff 1987, 1992b).

3 Tonogenesis

3.1 The usual path to tone

For those tone languages that can be reconstructed as having a prior toneless state, it is broadly accepted that the origins of lexical tones (dubbed tonogenesis by Matisoff 1973) normally lie in earlier laryngeal contrasts. The now-standard tonogenetic scenario was proposed by Haudricourt (1954) in a convincing demonstration that Vietnamese is indeed an Austroasiatic language despite having a well-developed tone system (although since that time, tones have been found in a number of Austroasiatic languages). Haudricourt showed that Vietnamese underwent a two-step process of tonogenesis in which an initial three-way tone split (conditioned by laryngeal properties of the coda) was followed by a subsequent two-way split (conditioned by laryngeal properties of the onset), resulting in a system of six tones (Table 5). The precise phonetic outcome of the second split varies with language, but in general, voiceless onsets condition higher variants and voiced onsets lower variants; while in the initial phase, final glottal stops gave rise to rising tones while final aspirates have a pitch-lowering effect.

Table 5. Haudricourt’s schematic view of Vietnamese tonogenesis. Following an initial three-way split into level, rising, and falling, a subsequent two-way split, conditioned by the voicing of the initial obstruent, produced the modern six-tone system. (The fourth column shows the diacritics used in Vietnamese orthography.)
Haudricourt then applied this scenario to other languages (Haudricourt 1961) and today there is little doubt that similar two-way and/or three-way splits were involved in tonogenesis in Chinese, Tai-Kadai, Karen, Tibeto-Burman languages of Nepal (Mazaudon 2012), Tsat (Maddieson and Pang 1993) and possibly Hmong-Mien (Ratliff 2013). Furthermore, the two-way split proposed by Haudricourt seems to be the driving force in the development of register in a large number of Austroasiatic (Ferlus 1979) and Chamic (Lee 1966) languages.

Shortly after publishing his canonical scenario, Haudricourt suggested the possibility of an initial two-way split involving a crucial stage of voice quality contrast, where voiced initials induce breathiness on the following vowel, which then causes pitch lowering (Haudricourt 1965). This idea has regularly been revisited since (Egerod 1971; Pulleyblank 1978; Diffloth 1989; Thurgood 2002; Ferlus 2009; Mazaudon 2012), and while it is still unclear if breathiness is a compulsory stage in the two-way tone split, the available data certainly suggest that this is a common evolutionary trajectory. Since then, based on observations by Gage (1985), the possibility that creakiness or tenseness plays a role in the three-way split alongside the glottal stop has been proposed by Ferlus (1998) and Diffloth (1986); while the phonetic explanation to support these accounts has yet to be fully worked out, it is worth noting that a similar account has been proposed for the origin of tones in Athabaskan languages (Leer 1999; Kingston 2005).

3.2 Less common paths to tone

Despite the prevalence of the canonical tonogenetic scenario, other types of tonogenetic mechanisms have been described for Southeast Asian languages. In the Palaungic languages U and Hu, tone is claimed to have developed from differences in vowel height and vowel length, respectively (Svantesson 1989, 1991). Table 6 gives examples of some Hu forms compared to those of related Lamet, demonstrating that the Hu tones are not related to the voicing of the initial consonant.

Segments other than stops and aspirates may also induce tonogenesis. In some dialects of Khmer, loss of /r/ in syllable onsets has led to an incipient tone contrast between words like /kruː/ > [kʰuː] ‘teacher’ and /kʰuː/ > [kʰuːː] ‘venerable’ or /ricing/ > [hien] ‘to learn’ and /hien/ > [hien] ‘to dare’ (Thạch 1999; Wayland and Guion 2005; Kirby 2014). While Thạch (1999) proposes a contact-based explanation and Wayland and Guion (2005) suggest that phonologization of f0 was conditioned by a combination of the high degree of airflow necessary for trilling and subsequent devoicing of the trill, Kirby (2014) argues that this sound change may have arisen via the perceptual reanalysis of changes in spectral balance, coupled with the coarticulatory influence of the dorsal
gesture accompanying /t/>. 

Table 6. Examples of Hu tonogenesis from vowel length (after Svantesson 1989).

<table>
<thead>
<tr>
<th></th>
<th>Hu</th>
<th>Lamet</th>
</tr>
</thead>
<tbody>
<tr>
<td>*short</td>
<td>jäm</td>
<td>jäm</td>
</tr>
<tr>
<td></td>
<td>phịɲ</td>
<td>pịɲ</td>
</tr>
<tr>
<td></td>
<td>ũk</td>
<td>khũk</td>
</tr>
<tr>
<td>*long</td>
<td>jäm</td>
<td>jâam</td>
</tr>
<tr>
<td></td>
<td>ðañ</td>
<td>tâañ</td>
</tr>
<tr>
<td></td>
<td>nasök</td>
<td>jōok</td>
</tr>
</tbody>
</table>

Finally, suprasegmental contrasts may also serve as a source of lexical tone systems. Evans (2001ab) argues that certain Southern Qiang dialects developed tone systems after pitch accent, developed from an earlier lexical stress system, was re-analyzed in the wake of phonological reduction and heavy borrowing from Mandarin. While evidently uncommon in MSEA languages, this type of evolutionary trajectory is reminiscent of the probable path to tone in Germanic languages (e.g. Gussenhoven 2000).

4 Areality and contact

4.1 Contact-induced tonogenesis

The view that tone spread from Chinese to other languages of MSEA (Matisoff 1973, Pulleyblank 1986) is now so well-established that it is often considered as received knowledge (cf. the tonogenetic ‘waves’ posited by Ratliff 2002 and Mazaudon 2012). Unfortunately, in the absence of solid historical and sociophonetic data, claims about contact-induced tonogenesis, however likely they may seem, remain unproven. Even for the most likely cases of contact-induced tonogenesis in MSEA, it is difficult to decide if we are looking at accidentally parallel internal developments or at contact.

Thomason (2001: 59-60) suggests a number of conditions that should ideally be met when making a case for a contact-induced structural change. In addition to clear recipient and donor languages, a strong case will seek to establish that the feature(s) in question were present in the donor language but not present in the recipient language prior to contact. While these conditions may be straightforward to establish when looking at morphemes or even syntactic structures, they are surprisingly difficult to meet in the case of a phonological feature like tone. For instance, Vietnamese is often claimed (quite reasonably) to have acquired tone while under Chinese influence, but it is unclear if this development occurred as a result of this contact or simultaneously with it. The first Sino-Vietic contact probably took place in what is now northern Vietnam around the time of the Qin dynasty (second century BCE), with Chinese administrative control solidifying under the Han empire around 100 BCE (Gernet 1996; Phan 2013). Given estimates that (toneless) Old Chinese was spoken until the early centuries CE (Pulleyblank 1962, 1978; Baxter 1992; cf. Ferlus 2009), and evidence from conservative Vietic languages that maintain four-way laryngeal distinctions or tonal contour systems (Ferlus 1996), it is not
implausible to assume that initial emergence of tone in Vietnamese was an internal development.

In fact, since most cases of tonogenesis in MSEA involve the same regular internal factors (i.e., loss of laryngeal codas and neutralization of onset voicing), it is in general difficult to determine if contact is really playing a role or if we are just looking at independent parallel processes (although as Thomason 2001 correctly points out, the fact that a change can occur through internal factors in another situation is not necessarily a valid argument for rejecting contact). In the end, as discussed by Ratliff (2002), there are only two possible scenarios for contact-induced tonogenesis (besides the adoption of loanwords with their tones): 1) two atonal languages can become tonal simultaneously as bilinguals transphonologize the same laryngeal contrasts in both languages; or 2) an atonal language can become tonal because its speakers, who are bilingual in a tonal language and thus ‘tone-prone’, phonologize previously allophonic pitch variation. Proving either of these scenarios after the fact is probably impossible, at least in the absence of detailed acoustic and perceptual data gathered over several generations, but this does suggest that substantive proposals of contact-induced tonogenesis would include evidence of a high level of bilingualism in at least part of the language community in question.

4.2 Tone as an areal feature

Given all this variation, then, why is tone upheld as an areal feature of MSEA languages? Put differently, what is the evidence for tone as an indicator of areality? Aikhenvald and Dixon (2001) describe a Sprachbund as ‘a geographically delimited area including languages from two or more language families, sharing significant traits (which are not found in languages from these families spoken outside the area)’ (11; emphasis ours). This description is not meant to be used for a single feature, but a quick look at the five MSEA language families suffices to note that tone does not bear the hallmarks of an especially ‘areal’ feature. While tonal Austroasiatic languages seem limited to MSEA, tonality in Austronesian languages is not limited to MSEA. Register is attested in Javanese, for example, and Tsat, a Chamic language of Hainan, is highly tonal. These languages could of course be included in a larger Southeast Asia area, but even then a number of Oceanic (Austronesian) languages of New Caledonia have also developed tone (Haudricourt 1971, Rivierre 1993). Hmong-Mien, Tai-Kadai and Sino-Tibetan are all spoken outside MSEA proper, and all of these families are highly tonal. All the Hmong-Mien languages of China have tones (and some of them are spoken so far north, that it would be hard to regard them as Southeast Asian proper), as do all Tai-Kadai languages, be they spoken in Southern China or in India, such as Aiton and Khamti (Morey 2005). As for Tibeto-Burman languages, many of their representatives in China, but also in the Himalayas, are tonal as well.

Of course, the criterion given by Aikhenvald and Dixon may be too restrictive for tone in MSEA. There could be areal convergence even if some of the language families were tonal before arriving in MSEA. However, in a recent study based on the same database of 186 languages as this chapter, we were unable to establish geographical proximity as a factor of tonal convergence independent of language family and word type (Brunelle and Kirby 2015). In the end, the real question that needs to be addressed is not why MSEA languages are so frequently tonal, but if the number of tonogenetic events in MSEA Austroasiatic and Austronesian in the past two millennia were higher than would
have been the case if it had not been for contact.

That tone cannot unambiguously be shown to be a contact phenomenon (at least as measured by proximity) does not mean that MSEA is not a linguistic area full stop (Haig 2001; Aikhenvald 2006). Linguistic areas are not defined by any single feature, but by a cluster of features (Campbell, this volume), with different features having different weights, and much of the evidence for a MSEA linguistic area is based on convergence of grammatical and lexicosemantic features. In these respects, the Austroasiatic and Austronesian languages of MSEA are rather different from their non-MSEA counterparts. Matisoff (2001) provides a lengthy, though by no means exhaustive, list of some of these shared features, as well as of phonological features other than tone relevant for the establishment of a linguistic area (see also Enfield, 2005; Donohue and Whiting 2011).

4.3 The role of word shape in tonogenesis

If the role of contact in tonogenesis is downplayed, however, how are we to account for the relatively high frequency of tone in MSEA? The chief alternative is to assume that multiple tonogenetic events took place more or less independently in different languages of the region. Mechanistically, this is not problematic; the universality of microprosody (intrinsic F0 or ‘pitch skip’), thought to be the ‘seeds’ of tonogenesis, is well established (House and Fairbanks 1953; Ohala 1973; Gandour 1974; Hombert 1978; Hombert et al. 1979; Hanson 2009). This, however, raises the questions of (a) what types of pressures/conditions might cause languages to rely upon and ultimately enhance these microprosodic differences into tone systems and (b) why the relevant conditions should have appeared in several languages at more or less the same time (like in the two-way split that seems to have occurred in Chinese, Vietnamese and Tai-Kadai around the tenth century).

It is here that contact may play a role, albeit an indirect one, in the spread of tone throughout the region. Matisoff (1973, 2001) has repeatedly emphasized the relationship between monosyllabicity and tone, a correlation that we also found in the statistical study of our database (Brunelle and Kirby 2015; cf. Donohue 2012). Is it possible than instead of a direct relationship between contact and tonality, there is a more complex causal chain in which contact with monosyllabic languages favors monosyllabization, which in turn favors tonality? In our database we find that there is a strong correlation between monosyllabicity and numbers of tones; as you move to languages that retain presyllables, or morphology, one finds fewer or no tones, or word-level tones (as in many Tibeto-Burman languages). This is at least consistent with the idea that loss of segmental material makes a language increasingly ‘tone-prone’.

While it seems unlikely that languages borrow the concept of monosyllabicity per se (much as it would be unusual to borrow a fully-formed tone system), they surely borrow large numbers of lexical items. If language A is primarily sesquisyllabic and language B primarily monosyllabic, and if there exists an asymmetric prestige relationship between A and B such that A borrows more from B than vice versa, this would increase the number of monosyllabic forms in the lexicon of language A. This could have the effect of encouraging further loss of segmental material, driving the language towards a canonically monosyllabic template and increasing the likelihood of phonologizing prosodic properties such as pitch, length, or voice quality. Such a trend would be even more likely if pre-existing structural factors favor monosyllabization (Brunelle and Pittayaporn 2012).
However, despite the empirical correlation between monosyllabization and tone, its mechanistic underpinnings remain unclear. One possible motivation could be driven by functional considerations. Consider the case of a sesquisyllabic language with a laryngeal contrast in final stops. If the presyllables were to reduce and eventually disappear, the burden of lexical contrast would now be borne entirely by the final laryngeal contrast, known to be perceptually fragile (Steriade 2001/2008). This fragility could increase the likelihood that previously redundant phonetic properties, such as differences in pitch or voice quality, could become enhanced (Kirby 2013), eventually transphonologizing if the cues to the segmental identity of the coda are subsequently lost (Hyman 1976). While perhaps intuitively plausible, the specifics of such an account remain to be worked out in sufficient detail to be tested experimentally.

5 Conclusions

In this chapter, we have reviewed the properties of tone and register systems of Mainland Southeast Asia (MSEA), describing both their synchronic diversity as well as a range of theories to account for their development. While MSEA may still earn the title of ‘the ultimate Sprachbund’ (Dahl 2008), the presence of ‘tone’ may not in and of itself constitute a particularly strong indicator of convergence. As we have seen, the tone systems of this region are extremely diverse and it is difficult to establish unambiguous cases where tone (or register) has been spread via contact. More detailed acoustic and perceptual research on tone systems, together with longitudinal studies of speech communities, promise to enhance our understanding of the mechanisms underlying tonogenesis in MSEA and elsewhere.

References


Ferlus, Michel. 2009. What were the four divisions of Middle Chinese? *Diachronica* 26(2):184-213.


L-Thongkum, Theraphan, and Intajamornrak Chommanad. 2008. Tonal evolution induced by language contact: a case study of the Tin (Lua’) language of Nan province,


26 The areal linguistics of Australia

Luisa Miceli and Alan Dench

1. Introduction

Australian languages are particularly interesting from the point of view of areal linguistics as there is little consensus on which historical processes are responsible for the current distribution of shared features. The mainstream interpretation is that much of the similarity reflects genetic relationship – retentions of considerable time depth if widespread features, or more recent common innovations in a subgroup of languages when confined to a region (see Koch 2014 for a review). The picture is undeniably complex due to the large number of languages involved (at least 250 at the time of European colonization\(^1\)) over a small population\(^2\), and also due to the fact that indigenous societies were often highly multilingual, with as many as 4-6 languages spoken within a group.

This chapter aims to present an overview of the difficulties involved in interpreting the history of these languages, whether on the wide or small scale. Section 2 deals with the macro-scale picture while section 3 presents a regional case study. Section 4 argues that similar difficulties are in fact encountered at both levels of analysis, and that the same ‘overall’ comparative pattern emerges regardless of scope. This pattern is characterized by a ‘mismatch’ between structure and form and is highly suggestive of unbroken transmission unfolding in a multilingual context – which is indeed the kind of sociolinguistic setting we can assume to have been in place for much of Australia. For language histories of this type, the internal/external distinction is problematic since the usual theoretical assumptions do not hold. In a speech community that is mostly bilingual, speaker internal motivations for change are not necessarily also language internal. Current theory therefore provides few answers on how to interpret language relationships in a historical context of this type, so it is not surprising that there is a lack of consensus among Australianists. The final section of the chapter summarises previous discussion and outlines possible directions for future research.

2. Macro-level patterns

\(^1\)Research by Bowern suggests that this figure may be considerably higher, exceeding 360. Her updated language list can be accessed at [http://pamanyungan.sites.yale.edu/master-list-australian-languages-v12](http://pamanyungan.sites.yale.edu/master-list-australian-languages-v12).

\(^2\)Koch and Nordlinger (2014) cite an estimate of around one million.
Australian languages show a number of recurrent similarities. The degree of similarity in phonemic inventories is particularly remarkable (Busby 1980), as is a similarity in the constraints on permissible phonotactic patterns (Dixon 1980, Hamilton 1996) and of patterns of prosody (Baker 2014). We also find recurrence of similar patterns of semantic structure (Evans and Wilkins 2000, Dench 1997; see Gaby and Singer 2014 for a recent overview), and a number of short forms, both lexical and grammatical, that have widespread distribution. It is therefore not surprising that there is a longstanding assumption that Australian languages must all be ultimately related. Furthermore, a large subset of languages (about three quarters of the total number) shares a more extensive set of similarities. There are widespread lexical forms for around 200 concepts confined to this group, these languages are of the suffixing type and they lack number segmentable non-singular pronouns typical of northern languages. The received view is that they constitute a language family known as Pama-Nyungan (see Map 1) – though some treat it as a subgroup of the larger assumed Australian family/phylum. The remaining languages have been grouped into 24 separate language families, some of which are thought possibly to share a more recent common ancestor than the assumed Proto Australian. Despite the existence of obvious similarities, it is surprisingly difficult to support most hypotheses of genetic relationship with the kind of evidence that is expected to emerge when languages are related as a family.

That the Australian comparative pattern is unusual was already being discussed at the time of publication of O’Grady, Voegelin and Voegelin’s (1966) lexicostatistical study that provides the general classificatory schema still in general use today. The problem lies in the lack of phonological differentiation given the densities of cognates (in the lexicostatistical sense) observed. Voegelin et al. (1963) explain that Pama-Nyungan, which they classify as the largest family in the Australian phylum, has a cognate density more like that of a phylum in other parts of the world, and that this is then inconsistent with the degree of phonological differentiation, which matches that of the least differentiated families in their world-wide sample. For this reason, and in order to avoid being ‘misleading’ (O’Grady et al. 1966: 11), they adapt the classificatory labels used in the Australian case: the Australian phylum is referred to as a macro-phylum, families like Pama-Nyungan are ‘phylum-like language families’, and the label ‘family-like language’ is used in those cases where dialects of the same language display similar cognate densities to a language family elsewhere. So, when looking at Australian data there is a constant feeling that a greater time-depth must be involved than the near-identical form of potential cognate items implies.

3Dixon (2002) assumes areality for the majority of recurrent features. The reader is referred to this work for a detailed discussion of the geographical distribution of such features. For a more succinct overview see also Dixon (2001). Table 1 (2001: 67-69) lists twenty-three of these recurrent features that have very widespread distribution – many of these being found in over 70% of Australian languages.

4An important parameter of variation in Australia involves the position of morphological markers relative to the root/stem. Some languages use suffixes only (and all but one of the languages currently classified as Pama-Nyungan belong to this typological group), some use both prefixes and suffixes.

5Note here that we are careful to distinguish the use of the term ‘cognate’ as used in lexicostatistical studies (similar forms with similar meaning) from its more precise sense used in comparative reconstruction (forms derived from a single reconstructable ancestor form). The two
A potential explanation for the poverty of phonological diversity lies in the nature of the typical Australian segment inventory, which is cross-linguistically unusual (see Miceli 2015a: 718-720 for a summary). Going against global trends, Australian languages have twice as many sonorants as obstruents. Butcher (2012) describes the system as ‘long and thin’ because, very often, five places of articulation are distinguished but there is no manner or voicing contrast. This over-reliance on place of articulation results in phonemic contrasts that are often acoustically subtle, with the most important spectral cues occurring in VC transitions. Butcher (2012) argues that this may inhibit speakers from partaking in cross-linguistically common connected speech processes such as anticipatory assimilation since such consonants are in prosodically strong positions (see also Butcher and Fletcher 2014; Baker 2014). Furthermore, Round’s (2013) investigation of dynamic features of Australian phonologies suggests that the most widespread morpho-phonological alternations occurring in Australian languages, such as those that involve the lenition of stops to glides or zero, actually imply changes that would have a stabilizing or reinforcing influence on the static phonological patterns described above (Miceli and Round 2014). Miceli and Round (2014) argue that, given their typical shape, Pama-Nyungan roots comprised segments not prone to the most typical Australian sound changes. It follows that the near-identical forms we see in the comparative data do not necessarily imply shallow time-depth.

Nevertheless, this poverty of sound change remains problematic – regardless of whether or not an explanation can be provided. This is because demonstrations of cognacy, which allow us to identify those features that are shared as a result of inheritance, rely heavily on the presence of sound change. Although not usually explicitly specified, it is crucial that a significant number of the regular sound correspondences that we identify via application of the comparative method involve non-identical sounds (see Miceli and Round 2014). This is because only regular, non-identical sound correspondences demonstrate the persistence of a system – that the phonology, word forms (including any associated morphology) and semantics of the languages involved are linked and have been transmitted as a whole. Identical/near-identical correspondences can also arise between languages that have borrowed heavily from each other. The crucial importance of non-identical lexical correspondences is partly obscured by the fact that both types of correspondence occur in all language families. However, in well-established language families, the predominance of non-identical lexical correspondences makes it possible for the comparativist to be confident in assuming that identical lexical correspondences must also reflect inheritance. In cases where lexical correspondences are all (near)-identical, there is no sub-set of words that allows such an assumption to be made, and as a result arguments of proposed cognacy are considerably weaker. Miceli and Round (2014) present Pama-Nyungan as an extreme case of such a situation.

Due to the fact that morphemes/words are either very similar or completely different, many Australianists have proceeded to do comparisons of pronouns and other items in paradigmatic relationships, on the basis of formal and functional similarity, relying on the assumption that such morphology is more resistant to borrowing and that shared irregularities in paradigms must reflect inheritance (see papers in Evans 2003 for examples). Campbell and Poser (2008:160) point out that Australian reconstructions are therefore not reconstructions in the traditional sense. They write: “Australianists call what
Miceli and Dench -- page 280 of 512

they see in common in this kind of superficial morphological comparison a “reconstruction”, even though this notion of reconstruction is not what scholars outside the Australian field do in comparative reconstruction … such a procedure is unlikely to convince outsiders” (emphasis ours). Furthermore, detailed studies such as that presented below give us an interesting insight into how different paradigms may have a life of their own, with different reconstructed paradigms suggesting alternative groupings of languages. We must therefore be cautious with genetic hypotheses that stand solely on the basis of this type of evidence.

Dixon has tried to resolve the problems encountered in doing comparative diachronic work on Australian languages by taking a very different approach. While in his first extensive study on the topic (Dixon 1980) he presented a number of reconstructions attributed to Proto-Australian, he has since written the following about his attempts at reconstruction (Dixon 2002: xvii-xviii):

There was, we assumed, likely to have been an ancestor language, proto-Australian. LoA [Dixon 1980] was the first serious attempt to put forward a hypothesis concerning proto-Australian. But the procedure was flawed … The method was selective; by comparing similar paradigms in a number of languages, I reconstructed proto-paradigms, which were certainly sound and valid with respect to the data employed. However, they did not justify the label proto-Australian.

He has proposed an alternative model to explain how languages are likely to be related. Passages such as the one just quoted and further comments on the ease of applying the comparative method to Amazonian languages in comparison to work on Australian languages, reveal that the unusual comparative pattern is his main motivation for seeking an alternative model – and this motivation is quite justifiable. Unfortunately, this quite valid motivation has been overshadowed by the model itself.

Dixon (1997, 2001) explains what he sees as a lack of evidence for an articulated family tree in terms of an adapted version of the Punctuated Equilibrium model (Eldridge and Gould 1976). The main premise of this model is that for much of human history peoples and languages have been in a state of equilibrium, occasionally disrupted by punctuation events. Language splitting events, as modeled in family trees, are the outcome of periods of punctuation, when settlement patterns are disrupted and languages spread. During periods of equilibrium, there are no major changes to settlement patterns, neighbouring languages borrow back and forth from each other and converge, so that evidence of tree-like relationships is slowly erased. Dixon says that on the basis of the evidence available, Australia is best understood as a linguistic area of great time-depth resulting from a long period of equilibrium. He argues that following the initial population dispersal at time of colonisation, there would have been no other major punctuation events until the period of European colonisation, making Australia the only part of the world where the effects of a long period of equilibrium are still evident.

We do not provide an in-depth assessment of Dixon’s model here (see e.g. Kuteva, 1999; papers in Aikhenvald and Dixon 2001; Bowern 2003, 2006 for discussion). We simply identify a few of the problematic issues with Dixon’s use of Punctuated Equilibrium as an explanatory model in the Australian context. Firstly, it is unclear whether it can be argued that an equilibrium state dominated the full 40,000 years (conservatively) of human occupation. The Australian archaeological record suggests that there were major demographic changes in the Holocene (the period following the last glacial maximum, i.e. the last 10,000 or so years) and these would have to be interpreted
as punctuation events (see e.g. Ulm 2013, Williams et al., submitted). Secondly, although Dixon is quite justified in concluding that the evidence in support of genetic relationships remains considerably weak, his alternative conclusion that the observed recurrent similarities are clearly diffusional is not based on strong evidence. As well as concluding that we will never know whether all Australian languages go back to a common ancestor due to the time depth he thinks is involved, Dixon also rejects the other hypothesised major genetic grouping; Pama-Nyungan. He argues against it on non-linguistic grounds by concluding that there is no punctuation event that could have caused its spread (others would link it to the Holocene demographic changes that were just mentioned, see e.g. Sutton and Koch 2008: 497-500). On linguistic grounds he points to the lack of clear isogloss bundling, which he says would be needed in order to identify Pama-Nyungan as a genetic group. He writes that only one typological feature has an areal distribution that corresponds exactly to the full set of languages currently classified as Pama-Nyungan. They all lack number segmentable non-singular pronouns. He dismisses the feature that many (see e.g. Evans 2005, pp. 264-267) consider to be the most important innovation/form diagnostic of Pama-Nyungan, the first person dual pronoun *ngali*, by pointing out that although it does not occur in any non-Pama-Nyungan languages, it is also missing in about one fifth of Pama-Nyungan languages, all on the geographical fringe. He argues that it is simpler to assume that the form diffused over a contiguous area, than that it was independently lost in eight separate regions. Both Sutton and Koch (2008) as well as Evans (2005) take up this point in their reviews of Dixon (2002) and argue that while loss is to be expected in all linguistic systems, Dixon’s account does not explain why this proposed diffusional feature, like many of the other features that could be reconstructed for Proto Pama-Nyungan, always stop at the Pama-Nyungan border. Evans (2005:270) argues that Dixon is literally correct when he says that isoglosses do not coincide, but that considering the number of forms that never go beyond Pama-Nyungan, and that have quite wide distributions within it, “… we have a very significant clustering indeed”.

What emerges from this discussion is a tendency for scholars to have a “default position”. Evans himself suggests that disagreements between Dixon and ‘Pama-Nyunganists’ are not really about data but rather about whether one considers a form “diffused until proven inherited, or inherited until proven diffused” (2005: 264). However, holding a default position, whichever it may be, is not methodologically defensible. One should always be considering and comparing the likelihood of both hypotheses, and in some cases there will be no principled way of deciding which is the most likely scenario. We are then faced with and must accept possible indeterminacy. In the following section, we present a detailed regional study that shows that many of the problems found on the wider scale are just as evident at the lower level.

Map 1. Geographical extent of Pama-Nyungan and non-Pama-Nyungan (shaded) languages

---

As already mentioned, due to the widely held assumption that all Australian languages must be related, Pama-Nyungan has often been treated as a subgroup and evidence in its favour has generally been presented in the form of a subgrouping argument (see Miceli 2004).
3. Regional patterns: the Pilbara

There are around 20 named languages recognised (by their speakers) in the larger Pilbara region. The received view is that the languages fall into a number of named ‘subgroups’ of Pama-Nyungan. While the labeling of these groups is well established in the Australian literature their status as genetic subgroups has not (yet) been strongly supported by detailed comparative reconstruction. It is reasonable to assume that the

---

7We have chosen to present a single case study in order to provide a more detailed analysis. We consider this case study to be representative of the problems faced in unraveling the history of shared features in the Australian context, especially within Pama-Nyungan. For a detailed case study of areal features across the Pama-Nyungan/non-Pama-Nyungan boundary see Heath’s (1978) study of languages in Arnhem Land. For an overview of outcomes of language contact in Australia see also McConvell (2010).

8Named language varieties within the wider area that are not discussed here include Warriyangka, Pinikura, Thiin, Jurru, Nhuwala, and Kariyarra. The materials available for these languages are especially limited and do not illuminate the issues. Sources used for the languages are: Yingkarta (Dench 1998a); Wajarri (Marmion 1996); Thalanyji, Payungu, Burduna (Austin 1994b); Jiw artillery and Tharrkari (Austin 1994a); Martuthunira (Dench 1995); Kurrama (Hill 2011); Panyjima (Dench 1991); Yinhawangka (Dench fieldnotes); Ngarluma (Kohn 1994); Yindjibarndi (Wordick 1982); Ngarla (Westerlund 2007, 2013); Nyamal (Dench fieldnotes); Nyiyaparli (Kohn fieldnotes).
languages of the region are related, at some level, but there is a degree of typological and morphosyntactic variation across the region that might initially suggest some depth to these connections. Map 2 (below) shows the approximate original location of the languages referred to in this summary survey and the received classification into the named Ngayarta, Kanyara, Mantharta and Kartu subgroups based on O’Grady et al’s (1966) original lexico-statistical classification, reinforced by O’Grady (1966) and Austin (1988) (see Dench 2001 and Koch 2004 for reviews). More recent comparative work calls some aspects of this into question (Dench 1998b, 2001). Note that initial documentation of Yinhawangka postdates these classifications.

Map 2. Approximate locations of languages and classification following O’Grady et al (1966)

The languages of the southern Pilbara (including Jiwarli, Tharrkari, Thalanyji) and the northern Pilbara (Nyiyaparli, Nyamal, Ngarla) have an essentially ergative pattern of alignment with varying degrees of tripartite case marking within the nominal paradigm and some differentiation in case assignment by predicate type and clause type (see Austin 1981c, Dench 2006, 2008). By contrast, languages in the central Pilbara (Panyjima, Yinhawangka, Kurrama, Yindjibarndi, Martuthunira, Ngarluma) have innovated a consistent accusative alignment and a productive passive voice (Dench 1982). While there is some similarity in nominal morphology across the region, the verbal inflectional paradigms show a high degree of diversity both in the number of categories represented
and the inflectional forms employed (Stirling and Dench 2012). Languages in the southern Pilbara share patterns of switch-reference in subordinate clause with languages to the east (Austin 1981a), but these patterns are not diffused more generally within the Pilbara.

Despite this diversity and the relatively secure assumption of an ultimately shared origin, grouping the languages is problematic. The phonological changes that have occurred in a number of the languages (and the region is a little unusual in Australia in showing evidence of sound change) are restricted in extent and provide little basis for anything more than very low-level subgrouping (O’Grady 1966, Austin 1981b). At the same time, the phonetic bases of these changes are reflected in a range of languages in ways that suggest a tendency to genetic drift, or diffusion of phonetic traits (Dench 2001). The diachronic investigation of paradigms and forms confronts a not uncommon problem in the Australian context; the reconstructions arrived at differ little from those posited for the presumed highest order grouping of the languages; Pama-Nyungan.

The resulting picture is one of significant diversity arising from some deep differences (verbal systems) and a range of perhaps more shallow shared similarities with different regional distribution. For these latter cases especially, it is difficult to determine whether they have arisen from a shared inheritance or through processes of diffusion. As we have suggested above, the ultimate explanation must lie in particular patterns of multilingualism across the region (both in the past and in recent times), the relative salience for their speakers of particular linguistic features as markers of language difference, and associated with this the appetite for variation within the community of speakers of any identified linguistic variety.

What can we say then of patterns of cultural and social connection that might inform these questions? The traditional language communities of the region can be grouped by some broad cultural criteria, the most salient features of which include the westward extent of particular male initiation rites (their linguistic significance first being suggested by O’Grady 1958) and the patterning of kinship systems (first discussed by Radcliffe-Brown 1913). However, these traits do not pattern in a way that immediately suggests distinct cultural clusters and there is no evidence that the different practices inhibited contact and extensive interaction between groups.

Some evidence of historical connections between particular communities of speakers can be gleaned from traditional narrative texts. For example, a set of traditional Martuthunira stories (Dench 1995) describes the introduction of the returning boomerang and hafted stone axe in a context of conflict between inland and coastal peoples. Interwoven stories also describe the establishment of coastal trading routes and the seemingly more peaceful introduction from the east of those patterns of male initiation mentioned above. What emerges in these stories is a received understanding of a historical distinction between coastal people – who subsisted on marine, mangrove

---

9Earliest dates for these artifacts is between 12-15,000 BP, at a time of relatively fast rising sea levels (Peter Veth, personal communication) and thus what was likely to have been a time of demographic pressure and social upheaval on the coast. If the stories do indeed date from that time, and if they do relate relatively local events, then it would assume an extreme degree of linguistic conservatism to expect linguistic patterns from those times to have survived into the modern languages. The introduction of initiation rites, woven into these same stories, is likely to be of more recent origin and is thus more likely to have linguistic correlates.
estuarine coastal hinterland resources – and people who inhabited the inland river valleys and plateaus.

As in most other parts of Australia, multilingual contact was the norm (Sutton 1997). In the Pilbara, traditional marriage patterns were exogamous and this exogamy was encouraged by a system of promised marriage established through male initiation arrangements that brought with them extensive patterns of kin avoidance. There was thus some incentive to establish important relationships with people already at some social distance. Children naturally grew up speaking at least their mother’s and their father’s languages.

Colonial settlement in the 1870s severely and irrevocably disrupted historical social patterns. European contact and settlement resulted in the partial collapse of coastal populations from a range of introduced diseases (many likely preceding European settlement) and the ravages of an enforced labor program in the early pearling industry. The more extended post-colonial period was characterised by the apportionment of much of the land into pastoral holdings on which introduced stock were tended by a primarily indigenous workforce who thus mostly remained on their traditional country or were moved as largely intact extended family groups from traditional inland country to the by now less populous coastal regions. General labour strikes among the indigenous pastoral workers in the mid-twentieth century led to further forced displacements (from pastoral properties and traditional homelands) and generally greater daily contact between speaker groups in government ‘ration camps’ or town fringe camps. The changes have meant greater prolonged contact between the languages originally spoken in the inland ranges (Yindjibarndi, Kurrama, Panyjima) with those of the coast and southern Pilbara than was likely to have occurred prior to European settlement. Other than a few early wordlists dating from the late nineteenth century, there are no descriptions of the languages before O’Grady’s initial short sketches based on data collected in 1958. In the absence of any substantial records of languages from the time of first settlement it is now very difficult to determine to what extent patterns of similarity we see in the descriptions of Pilbara languages that may be due to contact reflect relatively recent conditions of social interaction rather than the longstanding traditional multilingualism engendered by exogamy.

Received traditions of linguistic integrity apply more to country than to (communities of) speakers in a similar way to that described by Merlan (1981) and Sutton (1997). Languages belong to country and speakers traditionally identified both with a language and with a country (see also descriptions in Koch and Nordlinger 2014, Harvey 2011). While language identity was perhaps primarily mediated by affiliation to country historically, in the modern context this is less clear and an affiliation to language is often constructed as a claim to country. We see some contestation over which linguistic variety is properly associated with particular country and also negotiated succession in cases of language loss. There is also some interrogation of what properly constitutes the named variety. It is difficult to know whether this has arisen as a result of more than three decades of legal struggle for rights to traditional land and the imperative to identify legitimate claimants, or whether this contestation has always occurred and has simply been exacerbated by modern circumstances. Argument about the legitimacy of association with country based on linguistic affiliation naturally brings with it a particular metalinguistic awareness.

Folk descriptions of dialect and language differentiation now very often cite lexical differences between varieties (sometimes with strong assertions to the exclusive
provenance of particular forms despite evidence that these have been common to a number of languages for at least a number of generations). Thirty years ago, linguists’ questions about perceived differences between varieties often produced statements about phonological differences. Speakers described general impressions but also recognised the sometimes subtle phonotactic differences among languages with similar inventories – using particular (apparently cognate) forms to exemplify. Lexical differences were certainly recognised, usually simply taken for granted, often with reference to received shibboleths.10

The speakers Dench worked with in collecting materials in Panyjima, Kurrama, Martuthunira, Nyamal and Ngarla would generally seek to achieve a linguistic purity in elicitation sessions – sometimes to the point of resorting to correspondence mimicry to construct a missing of forgotten word (Dench 2001). The use of alternative lexical items in freely offered narrative discourse was explained as stylistic effect or as generosity to protagonists with a different language background, but it was sometimes maintained that the alternatives were not ‘true’ native forms. At other times speakers described the use of a range of ‘synonymous’ forms (see Hansen 1984 for patterns of lexical use in the Western Desert).

We noted in Section 1 that the comparison and reconstruction of pronoun paradigms has been given particular significance in the Australian context. The discussion that follows focuses on patterns of innovation in pronoun and demonstrative paradigms in the Pilbara languages. Dench (1994) presents an initial reconstruction of pronominal paradigms in the Pilbara and argues that it is particularly difficult to decide for any innovation whether it is the result of a shared inheritance or is the result of contact. The most extensive restructuring of the pronoun paradigm has taken place in the accusative Pilbara languages – Panyjima, Yindjibarndi, Kurrama, Ngarluma, Martuthunira – and in Jiwarli and Tharrkari. The remaining languages are relatively conservative. General changes within the paradigm involve the replacement of forms through the analogical extension of paradigm mates. Only the 1du.inc, ngali, remains immune to analogical restructuring within the non-singular paradigm, and it often serves as the base for other forms. While the analogical leveling of the paradigms in the different languages produces similar results, the base forms selected and the morphological forms employed differ and thus the languages remain very clearly distinct.

Nevertheless, it is possible to identify some low level shared innovations that are the most likely to reflect a shared inheritance. These include (see Figure 1):

1) Replacement innovation of 1pl.exc ngana-rtu (and ngantu-rtu) (Ngarla, Nyamal)
2) Replacement innovation of 1pl.inc nganyju-la (Ngarla, Nyamal)
3) Innovation of 3du piyalu (Ngarla, Nyamal, Nyiyaparli)
4) Compound oblique plurals with *karrangu (Yindjibarndi, Kurrama)

10The naming of varieties by reference to a particular lexical feature, or isogloss, is common across Australia and is perhaps most interestingly described for the Western Desert region (see Miller 1971). In the Western Desert, names may consist of the local variant of a nominalisation of ‘go’, or a demonstrative form, followed by the proprietive or ‘having’ affix. Thus, Pitjantja-tjarra, Yankuntja-tjara, Nggaanja-tjarra. Similar name constructions appear to exist in the Pilbara but are, interestingly, fossilised. Thus while the names Nyiya+parli, Panyji+ma, Yinha-wangka, Yindji+barndi, are all apparently based on demonstrative forms, the names are either no longer morphologically transparent or the root (and/or suffix) does not (now) occur in the language bearing the name.
5) Replacement of 2pl based on 2sg: *nyinta-kuru / *nyinku-(ku)ru (Yindjibarndi, Kurrama, Ngarluma)

6) The 3pl *thana shifts to a restricted demonstrative function (‘singular’) (Panyjima, Yinhawangka)

7) Replacement of 1pl.inc based on 1du.inc: ngali-kuru (Yindjibarndi, Kurrama, Ngarluma, Martuthunira, Panyjima, Yinhawangka)

8) Shift of *2pl (nhurra) to 2sg (Tharrkari, Jiwarli)

9) Innovation of exclusive pronouns in –ju (-1sgDAT) (Tharrkari, Jiwarli, Wajarri)

10) Innovated 1sg.DAT/GEN nganaju ( < *nganu–ju) (Tharrkari, Jiwarli, Wajarri, Martuthunira)

Figure 1. Innovations in pronoun paradigms most likely reflecting shared inheritance

The patterns shown here generally coincide with the received classifications, but with the following qualifications: Martuthunira is placed with the central Pilbara languages on feature (7) but with southern Pilbara languages on feature (10); Nyiyaparli (Palyku) is grouped with Ngarla and Nyamal rather than with Panyjima (the lexicostatistical studies, treat Palyku as in a ‘dialectal’ relationship with Panyjima).

While these pronominal innovations are likely to have arisen through a shared inheritance, other innovations are more clearly diffusional. For example, there is widespread replacement of the original 1sgNOM pronoun form *ngayi by ngatha (a reflex of an original 1sgERG), with just Martuthunira, Yindjibarndi, Kurrama, Ngarluma and Ngarla retaining reflexes of the original (Feature 11 in Figure 2 below).
The diachronic development of demonstrative paradigms is complicated by the complex pragmatic functions of demonstrative systems. In the Australian context we see a range of innovations of category and form including the development and loss of adverbial forms, shifts from three-term to two-term systems, innovation of third person pronouns from demonstratives (and possibly vice versa), and the denaturing of paradigm sets into a small number of isolated non-inflecting particles (Koch 2009). Indeed, it is not necessarily the case that the demonstratives form a symmetrical or coherent system. In the Pilbara the reconstructed mid-distal (‘near you’) has a complex history that differs markedly from the proximal (‘near me’), and generally conservative distal. Table 1 lists demonstrative forms reconstructed from a comparison of Pilbara languages (note that no claim is made here about the deeper history of these forms within Pama-Nyungan).

Table 1. Reconstructed demonstrative stems

<table>
<thead>
<tr>
<th></th>
<th>proximal</th>
<th>mid-distal</th>
<th>distal</th>
</tr>
</thead>
<tbody>
<tr>
<td>ergative</td>
<td>*nhulu</td>
<td>*palu</td>
<td>*ngulu</td>
</tr>
<tr>
<td>nominative</td>
<td>*nhiya</td>
<td>*panha</td>
<td>*ngunha</td>
</tr>
<tr>
<td>accusative</td>
<td>*nhinha</td>
<td>*panha</td>
<td>*ngunha</td>
</tr>
<tr>
<td>dative/oblique</td>
<td>*nhurnu</td>
<td>*parnu</td>
<td>*ngurnu</td>
</tr>
<tr>
<td>locative</td>
<td>*nhula</td>
<td>*pala</td>
<td>*ngula</td>
</tr>
<tr>
<td>locational</td>
<td>*nhungu</td>
<td>—</td>
<td>*ngunthi</td>
</tr>
</tbody>
</table>

The mapping of diachronic pathways for these demonstrative forms shows patterns of historical connection between languages that cut across other evidence. Thus, reconstruction of the proximal and distal demonstratives yields the following general patterns: In a set of southwestern languages including (at least) Jiwarli, Tharrkari, Thalanyji, Payungu and Purduna an original accusative form *nhinha, became yinha. In Martuthunira, the direct reflex of *nhinha has been lost with the shift to an accusative alignment system but the fuller paradigm reveals evidence of the earlier change (Feature 12 in Figure 2 below). The original nominative proximal stem has been replaced by a reflex of the original accusative in a set of languages including Payungu, Purduna, Jiwarli, Tharrkari, Yingkarta, Wajarri and Yinhawangka. This change is not co-extensive with the aforementioned innovation. In particular it does not affect Thalanyji and Martuthunira (which thus remain conservative) and it extends to Yinhawangka, which by other criteria is closer to Panyjima (Feature 13 in Figure 2 below). Both the proximal and distal paradigms have been leveled on the nominative stem in a set of eastern languages including Ngarla, Nyamal, Nyiyaparli, Panyjima, Yinhawangka and Wajarri. In the latter two languages the proximal nominative is a reflex of the original accusative and thus this change follows the previously noted innovation (Feature 13). In Wajarri, the distal nominative stem is a reflex of an original mid-distal accusative (Feature 14 in Figure 2 below).

Figure 2. Innovations in 1sg pronoun (11), proximal and distal demonstratives (12-14), and the leveling of irregular verbs (15, 16)
The history of the mid-distal is much more elaborate and involves the splitting of the original paradigm (Dench 2007). The result is the development of distinct paradigms with different functions built from original paradigm mates.

The ergative mid-distal form descends as a third person pronoun in Nyiyaparli, Nyamal, Ngarla, Ngarluma and Wajarri. In Nyamal, Ngarla and Ngarluma the original genitive also survives in this paradigm, with the other forms being elaborations of the ergative. The original nominative descends as a mid-distal demonstrative in Panyjima and Yinawangka with the locative surviving as member of this paradigm. Yinawangkarta shares the same pattern but the paradigm is described as a third person pronoun. In Wajarri, the nominative and locative forms survive and serve as the basis for separate demonstrative paradigms (‘distal anaphoric’ and ‘distal’ respectively – Marmion 1996). In Nyiyaparli, Nyamal, Yindjibarndi and Kurrama, the modern mid-distal is based on the original locative stem. In Thalanyji, the paradigm based on the old locative is a third person pronoun. In Martuthunira, only the locative form survives, as a non-inflecting ‘definite’ particle.

Table 2. Paradigm reflexes of the *pa- mid-distal demonstrative (see Dench 2007)

11 Western Desert languages also have paradigms based on the *palu stem, for example the ‘definite nominal’ in Yankunytjatjara (Goddard 1985).
Table 2 maps the extent of surviving reflexes of the original paradigm and their functional alignment with the pronoun paradigm (3sg), demonstrative paradigm (Dem) or as a non-inflecting particle (Part). The innovations thus include paradigm splits and functional shifts, the creation of new forms to fill gaps in innovated paradigms, and varying degrees of analogical leveling – leveling which has proceeded differently depending on whether or not the bases for the analogy are inflected members of the pronoun paradigm or of the extant proximal and/or distal demonstrative paradigm.

The resulting picture is something of a patchwork of difference across the array of languages, in which the innovations shared and also not shared are changes that have arisen through analogy. The demonstrative forms are essentially transparent; they are unaffected by regular phonological change (though there are some sporadic changes). Given the general similarities of form and some degree of functional congruity there is little to prevent analogy operating both within and between languages (where these languages are spoken by bilinguals).

A number of additional features can be added to the picture emerging here. The languages share a general reduction of an inherited conjugational complexity in the verb paradigm. The simplification has proceeded through the leveling of irregular verb forms on selected stems. In an arc of contiguous languages extending from the northern coast, across the inland ranges, to the southwestern coast of the region – in Ngarla, Nyamal, Nyiyaparli, Panyjima, Yinhawangka Jiwarli, Purduna, Payungu, Yingkarta – the ‘past’ stem of a small set of originally irregular verbs serves as the new stem form (Feature 15 in Figure 2 above). Within this arc – in Thalanyji, Martuthunira, Ngarluma, Kurrama and Yindjibarndi – the new forms are built on the ‘purposive’ stems; a pattern also found in the Western Desert languages (Dench 1998b) (Feature 16 in Figure 2 above).

As noted earlier, there are interesting differences in morphosyntactic patterning across the Pilbara including the innovation of an accusative alignment system and a productive active/passive voice alternation (Features 17 and 18 in Figure 3 below respectively), and the development of switch-reference in relative and purpose subordinate clauses (Feature 19 in Figure 3).

Figure 3. Innovation of accusative alignment (17), passive voice (18) and switch-reference in subordinate clauses (19)
The patterns in Figure 3 are most like those in Figure 1, perhaps suggesting that the case marking and voice innovations are inherited innovations in a central Pilbara group. Martuthunira then presents a particularly intriguing case in that it shares quite marked typological features with Yindjibarndi, Kurrama, Panyjima, Yinhawangka and Ngarluma and has participated in a number of pronominal innovations that align it with this set of languages, yet on the other hand other innovations in the pronoun and demonstrative paradigms and the presence of a switch-reference system suggests a strong connection to southern Pilbara languages, including both Thalanyji and Jiwarli. Martuthunira is unusual in having both a switch-reference system and an interacting voice system (Dench 1987, 1995) and this begs the question of how this level of syntactic complexity might have arisen.

Austin (1981a) demonstrates convincingly that switch-reference is a diffusional feature in Australia and thus it might initially be assumed that Martuthunira developed its patterns under diffusional pressure from neighbouring languages to its south and east. That is, switch-reference emerged into a system that had already innovated accusative alignment and a passive voice. However, a number of factors argue against this. First, it is difficult to imagine what kind of pressure might have led to the innovation of switch-reference (same- or different-subject marking) in a consistently nominative-accusative language that has the resources (a passive voice) to place patient and theme arguments into the morphologically transparent nominative subject relation. Second, the Martuthunira subordinate clause verb morphology that marks switch-reference is formally quite similar to that of Thalanyji and Jiwarli, suggesting that either the switch-reference constructions were borrowed (supplanting existing subordinate clause
inflections) or Martuthunira shares a history with these languages.\footnote{12} The latter seems more plausible, which might suggest instead that it is accusative alignment and the passive voice that are the later developments in Martuthunira.

The general alignment shift in this context is again an instance of leveling; the extension a nominative-dative pattern for transitive verbs into contexts where such verbs originally selected an ergative-accusative/absolutive pattern of marking (see Dench 1982).\footnote{13} Case marking patterns, like patterns of general leveling in the verb paradigm or the analogical creation of new pronoun and demonstrative forms are subject to indirect pattern diffusion and we return to the essential Australian problem. In the absence of extensive regular sound change as a guide to the stratification of genetic relationships among languages, we are forced to consider instead altogether more psychologically salient features as potential evidence – lexical similarity and differentiation, shared innovations and irregularities within paradigms, and interacting morpho-syntactic patterns.

What is still missing from this discussion is some measure of relative salience. Unless we know in more detail which aspects of a language variety are salient markers of language distinctiveness for the multilingual speakers themselves, we cannot develop a complete understanding of which features may be subject to (relatively unconscious) convergence and which may be subject to (a perhaps subconscious) divergence. Ideally, we would like to be able to approach situations like that presented by the Pilbara languages with some knowledge of individual speaker variation and levels of tolerance for the kind of calquing of pattern and possible diffusion of form that might be expected in this multilingual context. Unfortunately, many of the language descriptions we have are based on data mostly salvaged from a few remaining speakers.

4. Discussion and concluding remarks

The two previous sections reveal the same problematic issues. The comparative Australian pattern at both wide and regional levels is unusual and detailed analysis fails to clarify the indeterminacy encountered when considering competing hypotheses. As already mentioned, the pattern can be described as a form/structure ‘mismatch’: we find much structural similarity – near identical sound systems, recurring patterns in morpho-syntax, recurring patterns of semantic structure – but cognate densities are often lower than expected, with a variety of quite different forms often associated with what are essentially the same semantics (whether lexical or grammatical categories are involved). The significant clustering of features that we often find, such as in the case of Pama-Nyungan, suggests that we are dealing with genetically related languages, but the overall pattern is not consistent with family-tree like transmission.

Although the Australian situation is rendered more complex by the paucity of sound change, which usually provides essential clues on stratigraphy, comparative patterns that reveal a form/structure mismatch are not confined to Australia and are not

\footnote{12} Though determining this would depend on a detailed comparative reconstruction of verb morphology in the region, and that work remains in its infancy.

\footnote{13} The Martuthunira passive might then involve the borrowing of a construction (in the case of the derivational passive) or the reanalysis of existing morphology (in the case of inflectional passives).
uncommon. They are commonly found in linguistic situations characterised by a high degree of multilingualism. One detailed study of languages that display this type of pattern is François (2011), which focuses on the Torres and Banks islands of Northern Vanuatu. Here a group of seventeen closely related Oceanic languages have developed morpheme for morpheme inter-translatability (or structural isomorphism) but where the actual phonological forms involved are often completely distinct. An example sentence from two of these languages, Lemereig and Koro, illustrating the very obvious mis-match (data from François 2011: Table 1) is given below:

Table 3. Structural isomorphism in Lemereig and Koro

<table>
<thead>
<tr>
<th>LEMERIG</th>
<th>KORO</th>
</tr>
</thead>
<tbody>
<tr>
<td>t@r</td>
<td>nTr</td>
</tr>
<tr>
<td>r@l</td>
<td>ry</td>
</tr>
<tr>
<td>?orma?</td>
<td>tafal</td>
</tr>
<tr>
<td>?@ki?is</td>
<td>wUSmE@E</td>
</tr>
<tr>
<td>n</td>
<td>?ek@ek</td>
</tr>
<tr>
<td>mUYrol</td>
<td>namTYn</td>
</tr>
</tbody>
</table>

Abbreviations:
- 3pl - third person plural; ART – article; POSS:1incl.pl - possessive, first person inclusive plural

Ross (2001) describes the development of a similar mismatch in Takia, an Oceanic language whose speakers are currently bilingual in the Papuan language Waskia. The Takia lexicon has been semantically reorganised, along with the morpho-syntax, so that it closely matches that of Waskia, but there has been little to no borrowing of forms. When comparing Takia to its Oceanic relatives we therefore find a mismatch that goes in the opposite direction to that typical of Australia: many forms are shared, but little else – there has been a general structural divergence. From these and other studies, we can draw the conclusion that structure and form may follow quite different historical paths.

Moreover, studies of highly multilingual contexts (see also Epps and Michael, this volume), strongly suggest that phonological form is a more salient marker of language distinctiveness than semantic or grammatical structure. François (2011: 234) argues that “[i]n speakers’ meta-discourse, the distinctive identity of each particular language is typically defined by the form of its words rather than by their structural properties, which appear to be less accessible to spontaneous representations.” It follows that structure is more likely to converge while word forms are more likely to be either maintained as distinct or become distinct (in the case of related languages) due to this speaker bias. Most of these studies therefore explain the fate of word forms – whether they are maintained as distinct or are differentiated – by highlighting the emblematic role of languages in these multilingual contexts. Word forms, being highly salient, are then best at fulfilling this particular role. François (2011: 228-229) explains form differentiation in Northern Vanuatu in terms of a broader social bias for cultural differentiation in place in this part of Melanesia.

However, a recent study by Ellison and Miceli (2013) suggests that specific social factors giving languages a high emblematic function are not necessary for form differentiation to occur. The study investigates the lexical choices of Dutch/English bilinguals through the administration of a production task designed to identify any cognitive bias either for or against the selection of similar forms when language
distinctive options are available for the target meaning. Using the choices of monolingual speakers as a baseline, a statistically significant bias against similar forms is identified. Simulations of the diachronic effects of such a bias show that in a community with ongoing bilingualism and substantial numbers of bilinguals, forms will be replaced at much faster rate than they would be if those languages were to have no bilingual speakers. That is, there is an apparent tendency towards differentiation without any particular social biases favouring differentiation being in place (though these may, of course, amplify the effect of the bias if they do exist). Since we rarely know the exact sociolinguistic conditions prevailing in a given context prior to historical documentation – one of the criticisms often extended to explanations that rely on social factors – this result is particularly useful. It suggests that an assumption of ongoing active multilingualism is a sufficient basis for arguing that differentiation may have occurred.

Given the paucity of sound change in Australia and the likelihood that multilingualism extended well into the past, the presence of such a cognitive bias may help to explain why we might find the high replacement rates posited between certain pairs of neighbouring languages that are assumed both to be related and to have remained in close contact. Since cognate word forms would not be differentiated much by sound change, the bias would affect a larger portion of the vocabulary and give rise to a higher replacement rate than in situations where sound change has a more prominent role. Similarly, the bias is likely to have its strongest effect when languages that come into a relationship of bilingualism are closely related because there will be more cognate forms to be affected. The Australian pattern – at both the macro-level of Pama-Nyungan and micro-levels such as in the Pilbara – is therefore likely to reflect relatively long periods of unbroken transmission in a multilingual context.

It may be possible to do more with the findings of Ellison and Miceli’s study than just providing general explanations. For example, an investigation of the patterns of form replacement in a group of languages may be useful in gaining insights into contact relationships within a region. A study focusing on the Pilbara is currently under way (Miceli 2015b), comparing the number of expected shared replacements between pairs of languages (calculated on the basis of individual rates of replacement) to the actual observed number of replacements. Initial results are suggestive of a weak bilingual differentiation bias between Martuthunira and Yindjibarndi.

To conclude, Australian languages offer unique insights into the complex interaction between in an areal setting. They lead us to consider often ignored questions such as whether contact-induced change may actually proceed differently depending on whether languages are genetically related or not. They further establish the emerging picture that structure and form often have independent histories, which we see reflected in the history of paradigms as well as in the history of the lexicon. They highlight the problematic nature of the traditional assumption underlying comparative work that all speaker internal change is also language internal. Although many difficulties are

---

14Harvey (2011) gives specific examples of such cases where as little as 8% of the basic vocabulary is shared.

15Avoidance of word forms due to social practices such as death-taboo is one example of a social factor that could have amplified the cognitive bias identified in Ellison and Miceli’s study.

16Usually no distinction is made between change that is due to causes within the language system and change that is due to language implementation in individual speakers (i.e. cognitively motivated change). This is because although the two causes are distinct they give rise to
encountered in seeking to determine the historical processes that have given rise to the current distributions of shared features in Australia, there are valuable opportunities here. It is likely that ongoing detailed comparative study of Australian languages will continue to push traditional theoretical boundaries in interesting ways.

References


Austin, Peter K. 1994 ms.a. *A Reference Grammar of the Mantharta Languages*, Western Australia. La Trobe University, unpublished manuscript.

Austin, Peter K. 1994 ms.b. *A Reference Grammar of the Kanyara Languages*, Western Australia. La Trobe University unpublished manuscript.


indistinguishable biases in monolingual speaker behaviour. Both types of change are therefore traditionally classified as ‘internal’. However, this is a problematic working assumption when dealing with bilingual speakers or with languages spoken in a multilingual context.


Dench, Alan. ms. Field notes: Ngarla, Nyamal, Yinhawangka.


Ellison, T. Mark and Luisa Miceli (in prep.). *From Lexical Clash to Rapid Differentiation: Bilingual Cognitive Processing in Contact-Induced Change.*


Kohn, Allison. 1994. *A Morphological Description of Ngarluma*. BA (Hons) dissertation, University of Western Australia.

Kohn, Allison. ms. Nyiyaparli field notes. University of Western Australia.


Miceli and Dench -- page 298 of 512


Williams, Alan, Sean Ulm, Chris S. M. Turney, David Rohde, Gentry White (submitted). Holocene Demographic Changes and the Emergence of Complex Societies in Prehistoric Australia. *Plos One*.

27 Languages of the New Guinea region

Malcolm Ross

1 Background: geographical and socioeconomic

The island of New Guinea is linguistically the most diverse area on earth. According to Nettle (1999: 117), New Guinea has 1,109 languages and 27 stocks in an area of 786,000 km\(^2\). This is three times as many languages and stocks in proportion to area as any other continent-sized region in the world. Estimates of the number of languages in New Guinea range from 850 to 1200, and of the number of its lineages from the low twenties to sixty or more. By any figures within these ranges New Guinea displays startling diversity.\(^1\)

In order to profile New Guinea’s structural diversity, Comrie and Cysouw (2012) take a sample of 48 New Guinea languages from Haspelmath et. al. (2005) and compare it with the latter’s world-wide sample. The diversity of the New Guinea sample turns out to be a microcosm of planet-wide structural diversity.

The topic of this chapter is areality among the languages of the New Guinea Region (henceforth NGR).\(^2\) The NGR recognised by scholars (e.g. Wurm 1975, Foley 1986, 2000, Comrie and Cysouw 2012) extends beyond the island of New Guinea itself to embrace islands to its west and its east (Figure 1), and is defined as the region in which languages conventionally labelled ‘Papuan’ are found. It also contains numerous Austronesian languages. A count based on the *Ethnologue* (Lewis, Simons and Fennig 2014) gives 1365 NGR languages.

Austronesian languages belong to a well researched language family (Blust 2013). Proto-Austronesian was located in Taiwan, and archaeology suggests that from about 2200 BCE its daughters spread out to occupy the Philippines, bits of mainland Southeast Asia, most of the Indo-Malaysian archipelago, Madagascar, small mainly coastal enclaves on the island of New Guinea, and the Pacific islands of Melanesia, Micronesia and Polynesia (Bellwood et al. 2011). Austronesian speakers had reached the Bismarck Archipelago (New Britain, New Ireland and the Admiralty Islands) to the east of New Guinea by 1300 BCE.

Whereas Austronesian is a single lineage, ‘Papuan’ denotes a number of apparently unrelated lineages which share in common only that they are spoken in the NGR and are not known to be related to any lineage outside the NGR. The westernmost known Papuan language is Tambora, now extinct, on the Indonesian island of Sumbawa (Donohue 2007a). The easternmost is Savosavo of Savo Island in the middle of the Solomon Islands (Figure 1). The islands of the NGR to the west of New Guinea have acquired the

---

\(^1\) I am indebted to Andrew Pawley and Ger Reesink for their comments on an earlier draft of this chapter and to Raymond Hickey for his careful editing work on a very complex text.

\(^2\) Abbreviations for regional designations are ENus (East Nusantara), NGR (New Guinea Region), NWIM (Northwest Island Melanesia), NWNG (Northwest mainland New Guinea); for linguistic lineages ENB (East New Britain), OMTK (Orya-Mawes-Tor-Kwerba), RLS (Ramu-Lower Sepik), TAP (Timor-Alor-Pantar), TNG (Trans New Guinea).
collective label East Nusantara (ENus). Exact definitions of ENus vary, but I follow Klamer and Ewing (2010: 1) by including in it Halmahera and the Bird’s Head of New Guinea. Politically they belong – with the exception of Timor Leste (East Timor) – to Indonesia. The island of New Guinea itself (‘the mainland’) is divided between Indonesia (its western half) and Papua New Guinea (its eastern half). The islands of the NGR to New Guinea’s east and southeast as far as Bougainville also form part of Papua New Guinea, while those beyond Bougainville form the western half of the Solomon Islands. Just one Papuan language, Meryam Mir, is spoken on Australian territory in the Torres Strait Islands, just off the south coast of New Guinea, cheek by jowl with Western Torres Strait, an indigenous Australian language.

Map 1. The New Guinea Region

The western and northwestern boundary of the NGR not only marks the extent of Papuan languages, but also divides Austronesian languages of different structural types from one another. Beyond this boundary are morphosyntactically more conservative Austronesian languages. Within it are languages that share features with various Papuan lineages and with Austronesian languages throughout the western Pacific. The structural boundary coincides with the phylogenetic boundary drawn by Blust (1982, 1993, 2009) between western Malayo-Polynesian languages and those of the Central Malayo-Polynesian subgroup of Austronesian. This correspondence of boundaries is surely not fortuitous, and its historical interpretation is discussed in section 4.

From a Papuan perspective the NGR’s Austronesian languages are very recent arrivals. It is reasonably well established that human beings had entered Australia during the last Ice Age by 50,000 years ago and New Guinea by 40,000, and recent archaeology suggests that human habitation in New Guinea is perhaps as ancient as in Australia (Summerhayes et al. 2010). Indeed, this is probable, as sea levels were lower than they are today, and until just 8500 years ago New Guinea and Australia formed a single continent, known as Sahul. Figure 2 shows the coastline of Sahul 40,000 years ago when the sea level was 87m lower than it is today (Coller 2007). Nonetheless, even when the sea level was lowest, there was still a sea crossing of some 70 km between Sunda (a much extended mainland southeast Asia) and Sahul. The sea gap was enough to prevent the transfer of Sunda’s placental mammals to Sahul or of Sahul’s marsupials to Sunda.

---

3 Hence the areas labelled ‘East Nusantara’ and ‘(New Guinea) mainland’ overlap, as both include the Bird’s Head.
giving rise to the biogeographical boundary known today as the Wallace Line – and population genetics points to initial human settlement of Sahul between 50,000 and 25,000 years ago by a single population or a few already related populations, with no further genetic input until Austronesian speakers arrived (Friedlander et al. 2007; Scheinfeldt et al., 2007). Once in Sahul, however, people could island-hop southeastwards from New Guinea to the southeastern extreme of the Solomon Islands by way of short sea crossings, and had settled many of these islands by 30,000 years ago.\footnote{New Britain and New Ireland by 35,000 years ago. (Spriggs 1997: 47; Specht 2005), Buka–Bougainville by 32,000, Manus Island in the Admiralty Islands by 21,000 years ago (Kennedy 2002).}

For much of this time, there was a land bridge linking the island combining Choiseul and Santa Isabel to the island of Guadalcanal. This bridge subsumed today’s Savo Island, the location of the easternmost Papuan language.\footnote{Until Ross and Næss (2007) showed that the languages of the Reef and Santa Cruz Islands were Oceanic Austronesian, not Papuan, there was considerable discussion of how Papuan-speakers had found their way to these islands 390 km beyond the Solomons. In fact, however, it was Oceanic speakers who reached them in their ocean-going outrigger canoes around 1100 BCE.}
The geography of Sahul partly explains the extent of today’s Papuan languages. Tambora is close to the Wallace Line, and Savo is almost as far east as settlers from Sahul could travel.

Map 2. Sunda and Sahul 40,000 years ago

If this initial settlement scenario is approximately correct, it provides a partial explanation for New Guinea’s extraordinary linguistic diversity. Papuan lineages have been evolving more or less \textit{in situ} for fifty millennia. Since it is usually reckoned that the comparative method detects relationship only to a time depth of ten millennia, it is small wonder that there appear to be so many unrelated Papuan lineages, and that none of these appears related to Australian lineages, which were, after all, on the same continent. Time depth cannot be the only cause of diversity, however. Other parts of the world have also been continuously inhabited for many millennia, and show nothing like New Guinea’s phylogenetic diversity. Nettle (1999: 72–73) argues that economic interdependence...
between groups fosters social contact, hinders divergence and mediates the adoption of lingua francas. For example, pastoralists on the southern margins of the Sahara and in eastern central Asia developed social networks for mutual support in times of disaster, resulting in a single language or several closely related languages being spoken over a large area. But for perhaps forty millennia of their history Papuan speakers were foragers (hunter-gatherers). About 10,000 years ago agriculture appeared in at least the Sepik-Ramu basin in the northern lowlands (Swadling and Hide 2005: 291) and then in the Highlands (Golson 1977, Denham et al. 2003). But these neolithic farming communities, which continue today in much of the NGR, were and are small. Each extended family cultivated root crops on its own land and continued to hunt birds and mammals, remaining economically independent. In pre-contact times there were long-distance trading networks, but they tended to be single-purpose (e.g. the trading of shells from the coast to the Highlands) and the links in the network were fairly short, so that encounters with people more than 20-30 km away would have been rare. One relied on one’s neighbours mainly for wife exchange, and there was little social motivation for larger speech communities (Sahlins 1963).

This history helps us to understand the three geolinguistic subregions into which the NGR naturally falls. Two of these have already been mentioned: mainland New Guinea and ENus to its west. The third subregion, to its east, is Northwest Island Melanesia (NWIM), comprising the Admiralties, New Britain, New Ireland, Bougainville and the Solomon Islands.

The mainland is the principal locus of Papuan languages, which fall into numerous lineages (Figure 3). By the time Austronesian speakers arrived, agriculture was well established across much of its habitable territory, with a concomitant growth in population density. This evidently left little room for the newcomers, also neolithic farmers. But there is almost certainly a second reason why incoming Austronesian speakers initially bypassed much of New Guinea, namely that they were not only farmers but also seafarers and traders, as well as reef foragers. This expressed itself in a preference for small offshore islands with fringing reefs, which they found in the Bismarcks (Lepovsky 1988). Indeed, the reconstructable history of Austronesian languages spoken on the mainland indicates that all except those around the Bird’s Head and Cenderawasih Bay resulted from later Oceanic back-migrations from the Bismarcks to the mainland (Ross 1988).

Whereas the mainland is largely occupied by Papuan speakers, the converse is true of ENus and NWIM, where Austronesian speakers predominate and Papuan languages are spoken only in small enclaves. This evidently points to important socio-economic differences between the mainland on the one hand and ENus and NWIM on the other when Austronesian speakers arrived. Whereas the mainland was relatively densely populated by Papuan speaking agricultural groups, it is tempting to infer that ENus and NWIM either lacked agriculture altogether or had much less of it, resulting in lower population densities that permitted incoming Austronesian speakers gradually to occupy the larger islands from their smaller offshore habitats. The issues here are controversial, however. Archaeology suggests that pre-Austronesian NWIM populations were at least sedentary arboriculturalists, not mobile foragers, but it is unclear whether agriculture was already present (Spriggs 1996, 1997: 31–34, 61; Kennedy and Clarke 2004). Linguistics points to contact-induced change in Austronesian languages, and here the arguments concern what contact processes occurred (Donohue and Denham 2010, Ross 2013). What is reasonably clear is that ENus and NWIM were not swamped by a large immigrating
population of Austronesian speakers, but that Austronesian languages largely won out as their speakers achieved a degree of cultural dominance.

2 Background: linguistic

A linguistic area is conventionally defined as one with ‘three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor’ (Thomason 2001: 99). In the case of Papuan languages, however, it is often unclear whether two languages are related or not. Consequently, it is sometimes difficult to determine whether shared features result from contact or from inheritance.

2.1 Papuan classification

The difficulty of deciding whether two Papuan languages are related has two causes. First, some possible relationships are so old that they lie at the very limit of the comparative method. Second, Papuan historical studies are in their infancy, and attempts to catalogue relationships and lineages vary greatly in their methods. There is thus no agreed list of Papuan lineages.

The first modern attempt was Wurm (ed. 1975),\(^6\) which included survey articles by Wurm’s team members. They organised Papuan languages into eight lineages and a number of isolates (single-language lineages). They relied largely on wordlists — the collection of which was in itself a monumental achievement — to classify Papuan languages into numerous microgroups, i.e. groups of two to sometimes 30–40 languages that were obviously related. They employed lexicostatistical analysis, which, however, often failed them when they came to combine microgroups into larger groupings, because cognate percentages dropped to a level consistent with chance lexical similarities. Larger groups were then based on typological similarities, which meant that some groups (or parts of them) were based on areal features that might or might not reflect common inheritance. The largest of their groupings was the Trans New Guinea (TNG) phylum with 491 languages (Wurm, Voorhoeve and McElhanon 1975), a much smaller version of which had been proposed by McElhanon and Voorhoeve (1970). Wurm and McElhanon (1975) distinguished between (A) a ‘main section’ with 256 languages in 35 or so microgroups plus several isolates and (B) a further 20 or so microgroups and various isolates outside the main section with a TNG superstrate imposed on a non-TNG base.

Ross (2005a) used pronouns as a preliminary diagnostic for grouping Papuan languages, arriving at 23 possible lineages.\(^7\) The goal was to set up hypotheses for future work, rather than an actual catalogue, and it is unfortunate that some writers have treated it as a catalogue (e.g. Lewis, Simons and Fennig 2014). Ross’s listing retains within a 311-member TNG family all the microgroups of Wurm et al.’s main section except one (Sentani) and also includes a few microgroups external to the main section. The remaining microgroups are treated as distinct lineages.

---

\(^6\) Wurm (ed. 1975) was summarised in more manageable form as Wurm (1982). It also formed the basis for the New Guinea area maps in Wurm and Hattori (1981).

\(^7\) This methodology is criticised by Hammarström (2012).
Hammarström’s (2010) Papuan classification has a different goal, as it forms part of a worldwide listing of language families based on a consistent criterion, namely that for each family there should be a published demonstration of its integrity ‘by orthodox comparative methodology’ and ‘no convincing published attempts to demonstrate a wider affiliation.’ The classification is thus far more conservative than Wurm et al.’s or Ross’s and has 65 lineages and 46 isolates.

The point of enumerating these classifications is not to debate their merits but to highlight the uncertain state of Papuan classification. The classification adopted here of 32 lineages and numerous isolates (list below; Figures 3 and 4) is more conservative than Ross (2005a) but decidedly less conservative than Hammarström’s. It is adopted more for the sake of manageability than out of a conviction of its correctness.

A tentative listing of Papuan lineages

**East Nusantara and northwest mainland New Guinea**

1. West Papuan
2. East Bird’s Head
3. Mairasi
4. East Cenderawasih Bay (ECB)
5. Lakes Plains
6. Orya-Mawes-Tor-Kwerba (OMTK)
7. Nimboran
8. Sentani
9. Skou
10. Border
11. Pauwasi
12. Senagi
13. Kwomtari
14. Amto-Musan
15. Leonhard Schultze
16. Left May
17. Torricelli
18. Sepik
19. Ramu-Lower Sepik
20. Yuat
21. Upper Yuat
22. Bulaka (alias Yelmek-Maklew)
23. Yam (alias Morehead-Upper Maro)
24. Pahoturi
25. Eastern Trans-Fly
26. Eleman
27. Trans-New Guinea (TNG)
28. Anêm-Ata
29. East New Britain (ENB)
30. North Bougainville
31. South Bougainville
32. Central Solomons

All these classifications agree that one Papuan lineage, TNG, is much larger than any other. This means that the phylogenetic diversity is crammed into two smaller parts of the mainland. One is the northwest New Guinea (NWNG) area that stretches from the Bird’s Head to the Sepik-Ramu Basin (the latter arguably the linguistically most diverse area on earth; Foley, in prep.). The other lies on and around the Fly-Digul shelf in the south. This uneven distribution of lineages has a historical explanation. The NWNG and Fly-Digul
areas reflect a diversity that covered a larger portion of the mainland before the spread of TNG, a diversity that resulted from forty millennia of pre-agricultural populations. This ancient period ended about 10,000 years ago, when taro- and banana-based agriculture began somewhere in the east central highlands and spread across the valleys of the cordillera in both directions. These agriculturalists apparently spoke early TNG languages (Pawley 2005), and the spread of TNG is thus one of the agriculturally based spreads which Renfrew (2002) and Bellwood (2002) posit for a worldwide set of tropical and sub-tropical locations.

The history of TNG studies is tracked by Pawley (2005, 2007, 2012), while Pawley (2012) and Pawley and Hammarström (In prep.) summarise what we can reconstruct of early TNG. Regular sound correspondences, lexicon and bound morphology have been reconstructed for a few microgroups (Healey 1964, 1970, Voorhoeve 2001, Dutton 2010, Smallhorn 2011, Loughnane and Fedden 2011, Usher and Suter, forthcoming), but regular sound correspondences across microgroups are elusive, partly because the number of apparent cognates is small. One of the most promising pieces of evidence for at least a portion of TNG is Suter’s (2012) reconstruction of object prefixes on a small number of frequently used transitive verbs, a phenomenon first noted by Foley (1986: 259). Whether more detailed work will uncover sound correspondences or whether because of its antiquity TNG will remain on the list of what Nichols (2010) calls ‘macrofamilies’ – ‘hypothetical or debated older groupings posited by comparing proven families’ such as Afro-Asiatic – remains to be seen.

The second largest lineage posited in Wurm (ed. 1975) is the Sepik-Ramu phylum, the largest occupant of the Sepik–Ramu Basin. Foley (2005a) shows that there is no serious evidence for the phylum as a whole, but a small but persuasive body of evidence for two relatively large lineages, Sepik and Ramu–Lower Sepik, with 49 and 32 members respectively (Hammarström’s 2010 figures).
Map 3. Papuan lineages of East Nusantara and mainland New Guinea

Map 4. Papuan lineages of northwest island Melanesia

2.2 Studies of language contact in the New Guinea Region

If shared inheritance is one side of the areality coin, contact is the other. A category of ‘mixed’ (Oceanic/Papuan) languages arising from contact has long been recognised by
scholars working in the NGR, discussion of which has been dominated by the question, Are they the outcomes of radical discontinuity – pidginisation and creolisation – or the results of more gradual change in bilingual communities? The pidginisation hypothesis remained on the table from the 1940s to the 1970s, but has faded from view in the light of more nuanced contact studies worldwide.

In 1911 Ray and Strong had both published papers on Maisin, an alleged mixed Oceanic/Papuan language of southeast Papua New Guinea. However, neither author imagined that two sets of ingredients had been thrown into a pot and stirred. Each thought that Maisin had a single origin – Oceanic according to Strong, Papuan according to Ray – but had acquired elements from the other source through contact, prefiguring later work initiated by Weinreich (1953) on bilingually induced change.

Capell (1943, 1962), however, believed that the Austronesian languages of southeast mainland New Guinea were the outcomes of pidginisations by Papuan speakers, thereby contradicting the claim by Dempwolff (1937) and Milke (1958, 1961) that the Austronesian languages of the eastern three-quarters of the mainland belonged to the Oceanic subgroup of Austronesian, along with the languages of NWIM and the Pacific. A somewhat similar argument was made by Thurston (1982, 1987, 1989, 1992, 1994), who sought to show that Lusi, an Oceanic language on the northwest coast of New Britain, also reflected something akin to pidginisation. However, a pidginised language is likely to display irregular sound correspondences with its alleged relatives and simpler morphology than they do. In the event, the Oceanic languages of the NGR display regular sound correspondences (Ross 1988), pointing instead to bilingually induced change. In comparative perspective Lusi shows no evidence of simplification but is perhaps the outcome of language shift over a number of generations (Ross 2014).

The histories of Maisin and another ‘mixed’ language, Takia, the Oceanic language of Karkar Island, off the north coast of New Guinea, were examined by Ross (1996). Several papers have sought to analyse the effects of contact on Takia and to relate them to the neolithic social scenarios that brought them about (Ross 2001, 2003, 2013). The 1996 paper also coined the term ‘metatypy’ to label the change in morphosyntactic type, e.g. from VO to OV and from preposition to postposition, that appears quite often in Oceanic languages as the outcome of Papuan contact. Comparison across the Bel group of languages, to which Takia belongs, shows that Bel metatypy was a gradual process, not a sudden discontinuity (Ross 2008). Maisin (and to a lesser degree its Oceanic neighbours) and the Bel languages have progressively become more like local TNG languages, in some respects joining the core mainland linguistic area (sections 3.6.2, 3.6.5).

As noted in section 2.1 the controversy about mixed languages also affected Wurm and colleagues’ Papuan classification. If (most of) the languages that they attributed to TNG on typological grounds but excluded from its main section are languages of other lineages that have undergone metatypy as a result of speakers’ bilingualism in a TNG language, then they are not, on today’s criteria, members of the TNG family any more than Maisin and Takia are. They are in varying degrees members of the same linguistic area as their TNG neighbours.

2.3 Areal studies embracing the New Guinea Region

---

8 In the event, Strong proved to be right (Lynch 1977, Ross 1996).
The first largescale typological study of New Guinea’s languages was Capell (1969). Earlier scholars had mapped and briefly described numerous languages, and some also produced grammars (e.g., Ray 1932; Drabbe 1952, 1953, 1955, 1959; Dempwolff 1939, n.d.; Pilhofer 1933).

In his 1969 volume Capell appropriated and organised the knowledge of his predecessors in the first work on the typology of languages of the NGR. So soon after Joseph Greenberg had laid the foundations of modern typology this was a remarkable feat (Capell was inspired by Greenberg 1960), in contrast to his failed pidginisation hypothesis (section 2.2). Capell’s major insight was to divide NGR languages into three types: event-dominated, object-dominated and dominationally neutral. The morphology of event-dominated languages displays complexity in the verb phrase, including argument indices and elaboration of tense, aspect and/or mood markings. Object-dominated languages have genders and/or noun classes that are reflected in the morphology of either the noun phrase, the verb or both. In dominationally neutral languages verb phrase and noun phrase are equally (un)complicated.

More recent NGR typology deconstructs Capell’s types into their component constructions, as languages do not fall as neatly into types as his categories imply, a fact he himself recognised. But the features on which Capell focussed continue to loom strikingly large in NGR typology. Some of the longer sections of Foley’s (1986) Papuan languages textbook concern gender and nominal classification systems (pp. 77–91) and aspects of verbal morphology (pp. 128–166), a pattern repeated in his 2000 update (pp. 370–382).

Although Capell does not major on areality, his map (1969:129) of the distribution of domination types locates object-dominated languages in the NWNG and Fly–Digul areas, as well as in southern Bougainville, and event-dominated languages elsewhere. This prefigures the areality outlined in section 3.3.

The very fact that Wurm et al.’s TNG included groups with a non-TNG substrate points to the copying of structural features across lineage boundaries that is the hallmark of areality. However, since Capell little attention has been paid to areality on the mainland or in the NGR as a whole. Only Comrie and Cysouw (2012) come close (section 1). They find a basic typological division between (1) Austronesian and West Papuan languages (plus their one Torricelli language, Arapesh) and (2) the remaining Papuan languages in their 48-language sample. Given that it is possible to draw a line on the map around each of these divisions, this is effectively a statement about linguistic areas. However, their coverage is limited. It excludes Fly-Digul lineages and several NWNG lineages, and smaller lineages are represented by just a single language. They list the WALS values that are typical of languages in each of the two divisions in their sample, but provide contrasting values for only three variables, listed in Table 1.10

Table 1. WALS variables distinguishing Austronesian and West-Papuan from the other non-Austronesian languages.

---

9 They included, in British/Australian Papua, Sidney H. Ray (e.g., 1923, 1926, 1929, 1938); in Dutch/Indonesian New Guinea, Peter Drabbe (e.g. 1940–41, 1949a, b, 1950); and in German/Australian New Guinea, a number of missionary linguists (e.g., Klaffl and Vormann 1905, Kasprüs 1945) and Arthur Capell himself (e.g. 1948, 1949, 1952).

10 Actually they list a fourth, order of object and verb, but this is entailed by order of subject, object and verb.
Nichols (1997) looks at the typological profiles of Sahul languages as evidence of their ancient history. A ‘southeast interior’ profile evidenced in the New Guinea highlands and in Australia’s Pama-Nyungan languages is said to be associated with the initial settlement of Sahul, whilst a ‘northwest coastal’ profile, on the north coast of New Guinea, is the product of a migration occurring when the Pacific Rim, including the Americas, was colonised. However, whilst most of her northwest coastal features (presence of head-marking, low complexity, adpositions, exclusive/inclusive distinction, noun classes, numeral classifiers, tone, absence of ergativity) can be readily associated with some (not all) languages of NWNG, only a minority of her southeast interior features (presence of high complexity; absence of noun classes and numeral classifiers) are typical of New Guinea highlands (i.e. TNG) languages. The remaining features – presence of ergativity and absence of head-marking, adpositions and tones – are not features of highlands languages. TNG languages often use their instrumental postposition as a typically optional agent subject marker, and this is often labelled ‘ergative’, but the languages are in all other respects solidly accusative in alignment (Donohue 2005c). Tone, on the other hand, is now known also to be widely present in highlands languages (Donohue 2005a, Ross 2005b). An exclusive/inclusive distinction in pronouns occurs sporadically all over New Guinea, though admittedly somewhat more densely in NWNG (section 3.6.10).
Dunn et al. (2008) use the *Structure* algorithm to identify probabilistic clusters of NWIM languages based on structural features. Reesink, Singer and Dunn (2009) apply it to a more extensive sample of 121 languages, 81 of them in the NGR and 40 in adjacent areas (Borneo, the Philippines, Australia, Oceania beyond the Solomons), using 160 structural features.11 Reesink and Dunn (2012) add a more detailed study of ENus (and western Indonesia). The algorithm organises the languages into clusters, the number being determined by the user. Reesink et al. (2009) apply *Structure* iteratively, dividing their sample first into two clusters, then three, and so on up to fifteen. However, only the first nine iterations have respectable probability values. The main results for the latter, as they apply to NGR languages, are shown in Table 2, employing the nomenclature of Table 1. Numbers in the leftmost column indicate the number of groups assumed for that iteration (missing iterations only affect non-NGR languages). The table is read from top to bottom. Hence iteration 3 divides NGR languages into two clusters, A, consisting of Austronesian, West Papuan and the Timor-Alor-Pantar microgroup of TNG (Figure 5), and B, all other Papuan languages – thus reflecting the two-way division of Comrie and Cysouw (2012). At iteration 5, B is divided into a ‘rump TNG’ (TNG minus the Timor-Alor-Pantar, Marind and Kiwai microgroups (Figure 5), which each cluster with non-TNG neighbours; see Figure 5) and a large cluster that includes all other Papuan lineages. At iteration 6 cluster A is divided into Austronesian and the Papuan languages of ENus. Iteration 8 divides the large cluster of Papuan lineages from iteration 5 into a NWNG cluster and a Fly-Digul/NWIM cluster, and iteration 9 divides the latter into NWIM and Fly-Digul clusters.

11 Their sample contains one Andamanese language, 48 Austronesian languages (39 Oceanic, 9 western Austronesian) 55 Papuan languages (22 putative TNG, 33 putative non-TNG), 17 Australian (7 Pama-Ngungan, 10 non-Pama-Ngungan).
The table is a gross simplification of Reesink et al.’s (2009: 5) Figure 3, as clusters often overlap. Many languages from the sample display minor participation(s) in clusters other than their principal membership, reflecting features shared with languages of another cluster. These cluster-overlap features are due either to contact or to chance. For example, Kobon, a TNG language, also shares features with NWNG languages, presumably due to its proximity to Sepik languages. Table 2 depicts one important overlap, however: the East New Britain lineage and neighbouring isolates Kol, Sulka and Kuot share features with the languages listed at each iteration in the cells to both their left and their right.

Table 2. Typological clustering of NGR languages (after Reesink, Singer and Dunn 2009: 5)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Austronesian, W ENB, all other Papuan</td>
<td>NWNG: ECB, OMTK, Skou, Border, Sepik, Torricelli</td>
</tr>
<tr>
<td>5</td>
<td>Papuan, TNG-Timor-Alor-Pantar</td>
<td>NWNG: W Torricelli</td>
</tr>
<tr>
<td>6</td>
<td>Austronesian E Nusantara: W Papuan, TNG-TAP</td>
<td>NWNG: NWIM: Yélî Dnye, N and S Bougainville, C Solomons</td>
</tr>
<tr>
<td>8</td>
<td>NWIM: NWNG: W Torricelli</td>
<td>NWNG: NWIM: Yélî Dnye, N and S Bougainville, C Solomons</td>
</tr>
<tr>
<td>9</td>
<td>NWIM: NWNG: E Torricelli</td>
<td>NWNG: NWIM: Yélî Dnye, N and S Bougainville, C Solomons</td>
</tr>
</tbody>
</table>

Notes: Names in *italics* are isolates. The area between dashed vertical lines represents overlap between A and B.

One cannot tell *a priori* from Reesink et al.’s analysis whether a given cluster is phylogenetic or contact-induced (or both). Some lineages are represented in their sample by only one language, which may be atypical of its lineage, and a number of lineages in Table 1 are not represented at all. Nonetheless, the analysis reveals quite strong areality with numerous local overlaps. The overall pattern, however, suggests that the areas identified in Table 2 have nothing like the homogeneity of better known linguistic areas in the Balkans, India (Emeneau 1956, 1980) or mainland Southeast Asia (Enfield 2003, 2005). Instead, the NGR is a patchwork of overlapping and somewhat heterogeneous areas. This is hardly surprising, given a pattern of small-group economic independence over a long history. Indeed, largescale homogeneous areality is not expected in these circumstances.
The patterning that we do find perhaps calls into question the usefulness of the concept of ‘linguistic area’, at any rate for the NGR, in agreement with the conclusion of Campbell (2006: 21–22):

… the concept ‘linguistic area’ is not significant in itself. Instead of pursuing definitions of linguistic areas, we should attempt to account for the history of individual borrowings and diffusion, together with language change in general, in order to answer the question, ‘What happened?’

There is one respect in which one might take issue with Campbell, however. Although there is a qualitative difference between areality in, say, mainland southeast Asia and in the NGR, the concept of areality remains useful if one treats it as a continuum running from more to less homogeneous, placing NGR areas nearer to the ‘less’ end of the scale than perhaps areas in India and mainland Southeast Asia.

3. The Northern Arc, the Fly-Digul Shelf and the core mainland area

3.1. The Northern Arc

The remainder of this chapter evaluates claims about areality within the NGR, some of which are mentioned in section 2.3. The one part of the NGR that has earned the label ‘linguistic area’ is ENus (Klamer, Reesink and van Staden 2008, Klamer and Ewing 2010:12–13). The features that have given rise to this are discussed in sections 3.6.2–3.6.4, 3.6.7–3.6.9. The same features, however, also occur sporadically along the NWNG coast and into NWIM, and I label this discontinuous area the ‘Northern Arc’. It embraces

1) a. in East Nusantara:
   i. Austronesian languages;
   ii. the West Papuan lineage
   iii. marginally, the Timor-Alor-Pantar languages of TNG (TNG-TAP)

b. in northwestern mainland New Guinea:
   i. the East Bird’s Head lineage
   ii. the Torricelli lineage
   iii. (perhaps the Skou lineage)

c. in Northwest Island Melanesia
   i. most Oceanic Austronesian
   ii. the Anêm-Ata lineage
   iii. the East New Britain lineage
   iv. the isolates Kol, Sulka and Kuot.

12 Gil’s (2015) Mekong-Mamberamo linguistic area came to my notice after the NGR research was complete. It includes ENus, west Nusantara and mainland SE Asia, thus overlapping with the NGR. As the features used in Gil’s and the present work are almost mutually exclusive, the relationship between the two areas is unclear.

13 Microgroups within the large TNG family are denoted as TNG-TAP, TNG-Madang, TNG-Marind and so on.
The arc formed by these lineages was apparently an ancient voyaging corridor. It was certainly the route by which Austronesian speakers would later penetrate the NGR as far as the Bismarck Archipelago (Pawley 2008). The fact that certain commonalities occur along this corridor, but do not occur in combination elsewhere in the NGR, suggests that this is not a chance constellation of features but reflects a historical circumstance. We will perhaps never know with any certainty whether this circumstance was a common origin in the ancient period, or contact among ancient ancestors, or – more probably – both.

The Northern Arc shows up in the literature in various ways. It bears an approximate resemblance to Nichols’ (1997) northwest coastal area (section 2.3). It is reflected in Comrie and Cysouw’s (2012) Austronesian–West Papuan–Torricelli grouping, the more so when one recognises that TNG-TAP, the East Bird’s Head lineage, and the lineages and languages listed under (1)c) are not represented in their sample. But in particular it resembles Reesink’s (2003) ‘North Papuan linkage’, based on a set of variables that overlaps with the set considered below. Reesink’s variables are gender encoding, noun classification, encoding of plurality in noun morphology, quinary numeral systems, clause-final negation and possessum–possessor order.

3.2 The investigation

In order to gain a picture of NGR areality, data from 508 languages were collated. As complete coverage of NGR languages as possible was attempted, since some putative Papuan phylogenetic groups are very small and vanish if a sampling procedure is used. As a statistical comparison with other regions of the world was not intended, sampling was unnecessary. Of the 508 languages, 321 are Papuan and 187 Austronesian, and of the latter 138 are Oceanic; 290 out of 508 are represented in WALS online, but a number have values for phonological features only, which are not considered here. Due to descriptive gaps not all languages yielded a value for all variables. By way of comparison, Comrie and Cysouw (2012) use a sample of 48 languages, Reesink et al use 81 NGR languages, and Nichols 36 NGR languages.

With just the 17 variables listed in Table 3 the scope of this investigation was less ambitious than two of those mentioned above. Comrie and Cysouw (2012) use all the 140 WALS variables. Reesink, Singer and Dunn (2009) constructed a database of 160 variables. Nichols (1997) has just eleven variables.

---

14 There are today Oceanic languages in the southeast of the mainland, but they came somewhat later from the Bismarcks (Ross 1988: 211–212; Bruno et al. 2011)
15 Values for variables 1–2, 5–6, 8–11, 14–15 and 17 were imported where available from WALS online (Dryer and Haspelmath 2013), supplemented by data from the online supplement to Reesink et al. (2009). The remaining data were culled from language descriptions. Data were also collected for three further variables: the encoding of elevation distinctions in the demonstrative paradigm (Schapper 2013), the position of phasal adverbs in the clause (section 3.6.4), and gender in object indices. The first two were excluded from analysis because data were absent from language descriptions, the last because of the weight, alongside free pronoun and subject index gender (section 3.6.10), that it would have placed on gender.
16 There are considerable differences in the ways that WALS and Reesink et al. define variables, but this matter lies beyond the scope of this chapter.
Table 3. Variables used to investigate of NGR areality and their assignment to two types A and B (also a key to abbreviations used in section 3.6 and in Tables 4 and 5)

<table>
<thead>
<tr>
<th>Morphosyntactic ordering</th>
<th>Type A</th>
<th>Type B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Verb/Object order</td>
<td>VO</td>
<td>OV</td>
</tr>
<tr>
<td>2. Adposition/NP order</td>
<td>Prep</td>
<td>Postp</td>
</tr>
<tr>
<td>3. Actor/subject index on verbs</td>
<td>sV</td>
<td>_Vs</td>
</tr>
<tr>
<td>4. Undergoer/object index on verbs</td>
<td>Vo</td>
<td>oV</td>
</tr>
<tr>
<td>5. Possessor index on nouns</td>
<td>Npr</td>
<td>prN</td>
</tr>
<tr>
<td>6. Clause-final negator in declarative realis clause</td>
<td>ClNeg (clause-final negator word)</td>
<td>Negation other than a clause-final negator word</td>
</tr>
<tr>
<td>7. Clause sequences</td>
<td>–Ch (parataxis and coordination)</td>
<td>+Ch (clause-chaining)</td>
</tr>
<tr>
<td>8. Noun/Adjective order</td>
<td>NA</td>
<td>AN</td>
</tr>
<tr>
<td>9. Noun/Demonstrative order</td>
<td>ND</td>
<td>DN</td>
</tr>
<tr>
<td>10. Possessum/Possessor order</td>
<td>PmPr</td>
<td>PrPm</td>
</tr>
</tbody>
</table>

**Semantic encoding**

| 11. Alignment of verbal argument indices | Spl(it-S) | Acc(usative) |
| 12. Realis/irrealis distinction basic in verbal morphology | +Irr (realis vs irrealis basic) | –Irr (realis vs irrealis not basic) |
| 13. A least one degree of past distinguished in verbal morphology | –Past (no past tense morphology) | +Past (present vs at least one degree of past) |
| 14. Clusivity in free first person pronouns | +Inc (inclusive/exclusive distinction) | –Inc (no inclusive/exclusive distinction) |
| 15. Masculine/feminine distinction in pronominals | +Gnd (gender encoded in at least 3SG) | –Gnd (No encoding of gender) |
| 16. Distinction between two or more non-SG numbers in free pronouns | +Dual (two or more non-singular numbers) | –Dual (one or no non-singular number) |
| 17. Distinction between inalienable and alienable possession constructions | +Inal (dedicated inalienable possession construction) | –Inal (no dedicated inalienable possession construction) |

The list in Table 3 is divided into morphosyntactic ordering variables and semantic encoding variables. The latter concern how morphology partitions conceptual space.
division between the two sets of variables is discussed briefly in section 3.3. Commentary on each variable is given in sections 3.6.1–3.6.11.

Each of the 17 variables has two or three contrasting values common in the NGR, and on the basis of known collocations, the values are organised into two canonical language types A and B, somewhat in the style of Capell (1969). These represent the typological extremes found in the NGR, and have no theoretical status beyond being a convenient presentational tool: a language is more A-like or more B-like.

Table 4 summarises the values of the morphosyntactic ordering variables in the various lineages and in certain isolates for which sufficient data are available. Table 5 does the same for the semantic encoding variables.

The values shown in the tables are mostly the A and B values of Table 3. but for some variables an M, 'mixed', or 0 (zero) value appears. For example, variable 9, order of demonstrative and noun, has the value M when both orders appear in a single language. The zero value of variable 4 occurs when a language has no undergoer affix, and of variable 11 when the verb has no argument-indexing affixes.

For each variable Tables 4 and 5 show its dominant value in each lineage, where ‘dominant’ is defined as ‘present in at least 75% of the lineage’s members’. Where there is no dominant value, two values are shown. If there is only one representative of a lineage for a given variable, or two representatives that disagree in value, then this is signalled by italics. Also shown in italics are all values of isolates and of the East Cenderawasih Bay and Lakes Plain lineages, each represented by a single language.

In both tables an A-ness index is created for each lineage by summing the row’s values: A = 1; AB, AM or A0 = 0.5; any value not containing A = 0. The B(A) value of variable 11 is weighted 0.25 (it is described in section 3.6.8). A B-ness index was similarly calculated. Lineages are arranged in each table in rank index order with the most A-like at the top, the most B-like at the bottom. Kwomtari and Bulaka are excluded from the ranking as values are missing. The three TNG microgroups (TAP, Marind and Kiwai) that on Reesink et al.’s (2009) analysis did not cluster with TNG are also included in both tables, but are not ranked, as they are also included within the Trans New Guinea entry. The semantic encoding ranking is copied from Table 5 to Table 4 to facilitate comparison.

3.3 The results

Table 4 allows us to distinguish three possible linguistic areas, although all three have a fuzzy boundary and a degree of internal heterogeneity. They are the Northern Arc, the Fly-Digul Shelf and surrounds, and the core area that occupies much of the mainland.

A dashed line across Table 4 below Skou marks the boundary between the more A-like Northern Arc lineages, where the A-index is higher than the B-index, and lineages with predominantly B values. Variables 8, 9 and 10 play little role in this difference, as they remain much the same for the lineages above and below the line.

The placing of a second dashed line above TNG-Kiwai is more arbitrary, but the lineages below it include all those of the Fly-Digul Shelf and surrounds (TNG-Kiwai, TNG-Marind, E Trans-Fly, Bulaka, Yam). With the exception of the Bulaka lineage, for which there are gaps in the data, the Fly-Digul lineages have B values for variables 8 and 9, unlike the lineages above the line. Whether the Fly-Digul shelf represents a linguistic area is unclear because descriptions of these languages are only now emerging from
ongoing research, but Evans (2012) points to possible areal features linking Nen, a Yam language, and Idi, a language of the Pahoturi lineage.\footnote{The Pahoturi lineage is not included in Tables 5 and 6 because adequate data were not available to me.}

The fact that the isolate Yade, the Pauwasi language Karkar-Yuri, and the North Bougainville lineage also occur here is perhaps due to chance (but Yade and Karkar-Yuri are located close to one another).
Table 4. Morphosyntactic ordering variables: dominant values by lineage

<table>
<thead>
<tr>
<th>A- rank</th>
<th>B- rank</th>
<th>Type A</th>
<th>V</th>
<th>Pre</th>
<th>sV</th>
<th>Vo</th>
<th>Np</th>
<th>CINe</th>
<th>-Ch</th>
<th>NA</th>
<th>ND</th>
<th>PmPr</th>
<th>Sem-rank, from Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>9</td>
<td>1</td>
<td>Anêm-Ata</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>8.5</td>
<td>1</td>
<td>Austroan</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>6.5</td>
</tr>
<tr>
<td>3</td>
<td>8.5</td>
<td>1</td>
<td>Torricelli</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>6.5</td>
</tr>
<tr>
<td>—</td>
<td>…</td>
<td>…</td>
<td>Kol</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Q</td>
<td>B</td>
<td>…</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>2</td>
<td>E Bird’s Head</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>2.5</td>
<td>W Papuan</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>11.5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>Sulka</td>
<td>A</td>
<td>A</td>
<td>Q</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>6.5</td>
<td>3</td>
<td>E New Britain</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>9</td>
</tr>
<tr>
<td>8.5</td>
<td>5</td>
<td>2</td>
<td>Kuot</td>
<td>A</td>
<td>A</td>
<td>M</td>
<td>M</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>8.5</td>
<td>5</td>
<td>3.5</td>
<td>Skou</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>4.5</td>
<td>4.5</td>
<td>OMTK</td>
<td>B</td>
<td>B</td>
<td>AB</td>
<td>A</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>20.5</td>
</tr>
<tr>
<td>—</td>
<td>…</td>
<td>…</td>
<td>Kwomtari</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>Q</td>
<td>Q</td>
<td>AB</td>
<td>A</td>
<td>…</td>
<td>…</td>
<td>B</td>
</tr>
<tr>
<td>11.5</td>
<td>4</td>
<td>3</td>
<td>Eleman</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>20.5</td>
</tr>
<tr>
<td>—</td>
<td>…</td>
<td>…</td>
<td>ECB (Bauzi)</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>…</td>
<td>…</td>
<td>A</td>
</tr>
<tr>
<td>—</td>
<td>4</td>
<td>4</td>
<td>TNG-TAP</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>—</td>
</tr>
<tr>
<td>11.5</td>
<td>4</td>
<td>6</td>
<td>Yélî Dnye</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>20.5</td>
</tr>
<tr>
<td>13</td>
<td>3.5</td>
<td>5.5</td>
<td>Senagi</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>17</td>
</tr>
<tr>
<td>15.5</td>
<td>3</td>
<td>4</td>
<td>Lakes Plain (Iau)</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>26</td>
</tr>
<tr>
<td>15.5</td>
<td>3</td>
<td>5</td>
<td>Taiap</td>
<td>B</td>
<td>B</td>
<td>M</td>
<td>A</td>
<td>Q</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>9</td>
</tr>
<tr>
<td>15.5</td>
<td>3</td>
<td>5.5</td>
<td>Nimboran</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>Q</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>AB</td>
<td>17</td>
</tr>
<tr>
<td>15.5</td>
<td>3</td>
<td>7</td>
<td>Mairasi</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>19.5</td>
<td>2.5</td>
<td>5</td>
<td>Sepik</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>AB</td>
<td>AB</td>
<td>AB</td>
<td>20.5</td>
</tr>
<tr>
<td>19.5</td>
<td>2.5</td>
<td>5</td>
<td>RLS</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>AB</td>
<td>AB</td>
<td>AB</td>
<td>26</td>
</tr>
<tr>
<td>19.5</td>
<td>2.5</td>
<td>6.5</td>
<td>Sentani</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>A</td>
<td>B</td>
<td>26</td>
</tr>
<tr>
<td>19.5</td>
<td>2.5</td>
<td>7.5</td>
<td>S Bougainville</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>AB</td>
<td>B</td>
<td>11.5</td>
</tr>
<tr>
<td>23.5</td>
<td>2</td>
<td>5.5</td>
<td>Border</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>B</td>
<td>A</td>
<td>AB</td>
<td>29</td>
</tr>
<tr>
<td>23.5</td>
<td>2</td>
<td>6</td>
<td>C Solomons</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>Q</td>
<td>Q</td>
<td>A</td>
<td>AB</td>
<td>AB</td>
<td>AB</td>
<td>3.5</td>
</tr>
<tr>
<td>23.5</td>
<td>2</td>
<td>7</td>
<td>Trans New Guinea</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>Q</td>
<td>Q</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>—</td>
<td>2</td>
<td>6</td>
<td>TNG-Kiwai</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>Q</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>—</td>
</tr>
<tr>
<td>23.5</td>
<td>2</td>
<td>7</td>
<td>Yade</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>M</td>
<td>B</td>
<td>B</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7</td>
<td>TNG-Marind</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>M</td>
<td>B</td>
<td>—</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>26</td>
<td>1.5</td>
<td>7.5</td>
<td>E Trans-Fly</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>M</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>17</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>5</td>
<td>Bulaka</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>M</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>28.5</td>
<td>1</td>
<td>7</td>
<td>Pauwasi (Karkar-Yuri)</td>
<td>B</td>
<td>B</td>
<td>θ</td>
<td>θ</td>
<td>θ</td>
<td>θ</td>
<td>A</td>
<td>B</td>
<td>M</td>
<td>B</td>
</tr>
<tr>
<td>28.5</td>
<td>1</td>
<td>7.5</td>
<td>N Bougainville</td>
<td>B</td>
<td>B</td>
<td>θ</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>5</td>
</tr>
<tr>
<td>28.5</td>
<td>1</td>
<td>7.5</td>
<td>Yam</td>
<td>B</td>
<td>B</td>
<td>θ</td>
<td>B</td>
<td>θ</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>M</td>
<td>B</td>
</tr>
</tbody>
</table>

Table 5. Semantic encoding variables: dominant values by lineage

<table>
<thead>
<tr>
<th>Sem-rank</th>
<th>A-index (max=7)</th>
<th>B-index (max=7)</th>
<th>Type A</th>
<th>Type B</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0</td>
<td>Kuot</td>
<td></td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>1.5</td>
<td>Anêm-Ata</td>
<td></td>
<td>AB</td>
<td>A</td>
<td>AB</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>AB</td>
</tr>
<tr>
<td>3.5</td>
<td>5</td>
<td>2</td>
<td>E Bird’s Head</td>
<td></td>
<td>AB</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>—</td>
<td>5</td>
<td>2</td>
<td>TNG-TAP</td>
<td></td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>—</td>
<td>5</td>
<td>2</td>
<td>Kol</td>
<td></td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>3.5</td>
<td>5</td>
<td>2</td>
<td>C Solomons</td>
<td></td>
<td>AB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>AB</td>
</tr>
<tr>
<td>5</td>
<td>4.5</td>
<td>2.5</td>
<td>N Bougainville</td>
<td></td>
<td>A</td>
<td>B</td>
<td>AB</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>6.5</td>
<td>4.25</td>
<td>2.75</td>
<td>Austronesian</td>
<td></td>
<td>B(A)</td>
<td>A</td>
<td>AB</td>
<td>A</td>
<td>B</td>
<td>AB</td>
<td>A</td>
</tr>
<tr>
<td>6.5</td>
<td>4.25</td>
<td>2.75</td>
<td>Torricelli</td>
<td></td>
<td>B(A)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>AB</td>
<td>AB</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>Taiap</td>
<td></td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>E New Britain</td>
<td></td>
<td>AB</td>
<td>AB</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>Skou</td>
<td></td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>11.5</td>
<td>3.25</td>
<td>3.75</td>
<td>W Papuan</td>
<td></td>
<td>B(A)</td>
<td>A</td>
<td>B</td>
<td>AB</td>
<td>A</td>
<td>B</td>
<td>AB</td>
</tr>
<tr>
<td>11.5</td>
<td>3.25</td>
<td>3.75</td>
<td>S Bougainville</td>
<td></td>
<td>B(A)</td>
<td>A</td>
<td>B</td>
<td>AB</td>
<td>A</td>
<td>B</td>
<td>AB</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4</td>
<td>Mairasi</td>
<td></td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4</td>
<td>Salka</td>
<td></td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4</td>
<td>Pauwasi (Karkar-Yuri)</td>
<td></td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>2.5</td>
<td>4.5</td>
<td>Senagi</td>
<td></td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>AB</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>17</td>
<td>2.5</td>
<td>4</td>
<td>Nimboran</td>
<td></td>
<td>Bθ</td>
<td>AB</td>
<td>B</td>
<td>A</td>
<td>AB</td>
<td>AB</td>
<td>B</td>
</tr>
<tr>
<td>17</td>
<td>2.5</td>
<td>4.5</td>
<td>E Trans-Fly</td>
<td></td>
<td>B</td>
<td>B</td>
<td>AB</td>
<td>A</td>
<td>AB</td>
<td>AB</td>
<td>B</td>
</tr>
</tbody>
</table>
The languages between the two dashed lines in Table 4 constitute the core mainland area. It consists of languages that are B-like on variables 1–7 but A-like on variables 8 and 9. One can infer that much of this area has developed in the past ten millennia through the spread of TNG and through metatypy on TNG models in languages around its periphery. But caution is needed: according to Table 4 the South Bougainville and Central Solomons lineages fall into this ‘area’, but because of their geographic separation from the mainland any phylogenetic relationship or contact between them and other members of the core area would be very ancient indeed, perhaps thirty millennia ago. It is likely therefore that their presence in this part of Table 4 is due to chance – and if their presence is due to chance, then so perhaps too is the presence of some mainland lineages.

The rank order in the leftmost column of Table 4 refers only to the dominant lineage values for the morphosyntactic ordering variables shown in the table. The rank order of Table 5, which shows dominant lineage values for the semantic encoding variables, is replicated in the rightmost column of Table 4. The differences between the two rank orders are striking, but the ten more A-like lineages of Table 4 (except for the isolate Sulka) are also found in the top 12 lineages of Table 5, whereas the next lineage (OMTK) in Table 4 is at rank 20.5 in Table 5. A difference in behaviour between the two sets of variables is not surprising, as bilingually induced grammatical calquing typically precedes metatypy (Ross 2007, 2013), so a language’s semantic encodings will usually change before its linear orderings do. A language that is higher in rank in Table 5 than in Table 4 may be one that has undergone some grammatical calquing through contact with a neighbour but has yet to undergo metatypic change. This may explain the presence of the North Bougainville and Central Solomons lineages high in the rank order of Table 5 but low in Table 4. Both have been in contact with Oceanic languages with features of
type A for perhaps 3000 years (a relatively short period on a Papuan timescale). But this is an issue that needs targeted comparative research.

The results for a number of the variables in Tables 4 and 5 are mapped in Figures 6–12, where each symbol represents a language. The lineage to which the language belongs is indicated by its symbol shape, to which Table 6 provides a key.

Table 6. Key to lineage symbols in Figures 6–12

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>△</td>
<td>Anêm-Ata</td>
</tr>
<tr>
<td>○</td>
<td>Austronesian</td>
</tr>
<tr>
<td>◐</td>
<td>Border</td>
</tr>
<tr>
<td>□</td>
<td>Bulaka (alias Yelmek-Maklew)</td>
</tr>
<tr>
<td>▲</td>
<td>Central Solomons</td>
</tr>
<tr>
<td>★</td>
<td>East Bird’s Head</td>
</tr>
<tr>
<td>◊</td>
<td>Eastern Trans-Fly</td>
</tr>
<tr>
<td>★</td>
<td>East New Britain (ENB)</td>
</tr>
<tr>
<td>○</td>
<td>Eleman</td>
</tr>
<tr>
<td>○</td>
<td>Kwomtari</td>
</tr>
<tr>
<td>○</td>
<td>Nimboran</td>
</tr>
<tr>
<td>★</td>
<td>North Bougainville</td>
</tr>
<tr>
<td>♦</td>
<td>Orya-Mawes-Tor-Kwerba (OMTK)</td>
</tr>
<tr>
<td>♦</td>
<td>Ramu-Lower Sepik</td>
</tr>
<tr>
<td>◐</td>
<td>Senagi</td>
</tr>
<tr>
<td>□</td>
<td>Sentani</td>
</tr>
<tr>
<td>▼</td>
<td>Sepik</td>
</tr>
<tr>
<td>◐</td>
<td>Skou</td>
</tr>
<tr>
<td>★</td>
<td>South Bougainville</td>
</tr>
<tr>
<td>◐</td>
<td>Torricelli</td>
</tr>
<tr>
<td>□</td>
<td>Trans-New Guinea (TNG)</td>
</tr>
<tr>
<td>□</td>
<td>West Papuan</td>
</tr>
<tr>
<td>○</td>
<td>Yam (alias Morehead-Upper Maro)</td>
</tr>
<tr>
<td>♦</td>
<td>Other Papuan languages, including isolates</td>
</tr>
</tbody>
</table>

The shading of each symbol indicates the value(s) of the relevant variable(s) in that language, as indicated in the key that accompanies the map. Where two variables are mapped together, the symbol for a language in which both variables have A values is dark gray, and for one in which both have B values it is white. Where a symbol is split into two different shades, the values differ. In Figure 6, for example, the lefthand part of each symbol indicates OV or VO order, the righthand part postpositions or prepositions. If a language has VO order and prepositions (both A values) its symbol is dark grey, but if it has OV order and postpositions it is white. (In this instance the adposition variable has two additional values, as indicated in the map key.)

These and other similar maps were created as a heuristic in the search for areality. The islands at the western and eastern extremes of the NGR appear in insets, retaining their latitude and the scale of the rest of the map.

---

18 The maps were created with the mapping application *TileMill*, which reads a table of values and locations and interprets them in accordance with a script. The tables were produced from the typological database briefly described in section 3.2.
3.4 An example of contact-induced change

Before discussing each variable, it may be useful to provide an example of the contact-induced change that is at least partly responsible for today’s areality. Unlike many such changes, the example is one where fairly detailed inferences can be made about what has happened.

It is reasonably certain that TNG-TAP languages have undergone contact-induced structural change to join the Northern Arc (Table 4). Recall that Reesink et al. (2009) find that TNG-TAP belongs to their West Papuan and Austronesian cluster (Table 2). Yet it is as certain as it can be in the present state of TNG studies (section 2.1) that this group is descended from a TNG language whose speakers migrated west from the mainland. The integrity of TNG-TAP is demonstrated by Schapper, Huber and van Engelenhoven (forthcoming). Its languages maintain certain TNG features, as comparison of the Teiwa (TNG-TAP) and Nggem (mainland) examples below shows: OV clause order (2a), (3a), (3b), object-indexing prefixes, some of them cognate with mainland counterparts (2a), (2b), (2c), (3a), (3b) and possessor prefixes on inalienably possessed nouns (2a), (3b). But other TNG values have been replaced by values found in West Papuan and local Austronesian languages. TNG verbal suffixation is lost. This entails loss of subject indexing suffixes (3a), (3b), loss of a distinction between present and (often multiple) past tenses (3a), (3b) and loss of clause-chaining morphemes (see section 3.6.5). Negation is encoded by a clause-final negator word (2b) instead of a preverbal or affixal negator (3c).

2) Teiwa (TAP, TNG)
   a. Nome ha’an n-oqai g-un-ba’?
      Sir you 1sP-child 3sO-APPL-fall
      ‘Sir, did you see/meet my child?’
   b. Maan, na g-un-ba’ maan.
      NEG 1sF 3sO-APPL-fall NEG
      ‘No, I haven’t seen/met him’ (Klamer 2010:275)
   c. Ha wa ni-paxai!
      2sF come 1peO-divide
      ‘You divide us!’ [e.g., in groups]. (Klamer 2010:55)

3) Nggem (Dani, TNG)
   a. Ap andi en dirup nini-t-as
      man that AGT syringe 1pO-hit-R.3sS.RPST
      ‘That man “syringe hit” (injected) us.’ (Etherington 2002:117)
   b. Ap ako at aninis en o-gwa ako at… w-at-igi.
      man DEM the 3sP.rage POSTP 3sP-wife DEM the… 3sO-hit-R.FPST.3sS
      ‘That man hit his wife … in a rage.’ (Etherington 2002:117, edited)
   c. Nit yup ya-g-atek
      1pS cry plant-R-NEG
      ‘We did not cry.’ (Etherington 2002:110)

---

19 Abbreviations in interlinear follows the Leipzig Glossing Rules, plus DS (different subject), FPST (far past), R (realis), RPST (recent past), SS (same subject). Pronominal functions are A (actor), F (free), O (object), P (possessor), S (subject), U (undergoer).
In some TNG-TAP languages alignment of verbal argument indices has shifted from accusative to split-S, so that the erstwhile object prefixes like *ge-* ‘3s’ in (4a) now also serve as subject prefixes (4b) with certain lexically specified intransitives.

4) Klon (TAP, TNG) (Baird 2008:73)
   a. ge-uur ‘see him/her’
      ge-eek ‘rebuke him/her’
      ge-moi ‘help him/her’
   b. ge-ampi ‘s/he follows’
      ge-eten ‘it’s ripe’
      ge-wet ‘s/he urinates’

Larike illustrates the split-S pattern present in certain ENus Austronesian languages.

5) Larike (Austronesian, ENus) (Laidig and Laidig 1995)
   a. pese-ne ‘hold you (SG)’
   b. hanahu-ne ‘you (SG) fell’
      lopo-ne ‘you (SG) are wet’
      duarene-ne ‘you (SG) are hungry’

It would be easy to assume that the new features in TAP reflect contact with Austronesian languages, but their history is probably more complex. It is a reasonable assumption that on their arrival in ENus TNG-TAP languages were, like most TNG languages, accusative. At some point, however, they lost their subject suffixes, retained their object prefixes and become split-S. Split-S is a local areal feature of ENus, where it is fairly widespread in scattered Austronesian and West Papuan languages as well as TNG-TAP (section 3.6.2) and it is very reasonable to infer that this is an outcome of contact. But if it is, where was the model from which speakers copied? Although split-S appears sporadically in western Austronesian languages, apparently ancestors of ENus Austronesian and Oceanic languages had accusative alignment. No Oceanic language has split-S verbal morphology and a number of ENus Austronesian languages that are conservative in other respects are also accusative. It is fairly safe to assume that ENus Austronesian and TNG-TAP languages both reorganised their alignment on an indigenous Papuan model, which may or may not have been related to the West Papuan lineage, where split-S alignment already occurred. Such a substrate is apparently also responsible for the innovation of clause-final negation (section 3.6.4), the loss of inherited mood-prominence in many ENus Austronesian languages (section 3.6.9), and the innovation of a distinction between alienable and inalienable possession constructions (section 3.5.10).
Figure 6 shows that in the NGR, reflecting a worldwide trend (Dryer 2013c), there is a strong tendency for verb/object and adposition/NP in a given language to share a single head/complement order. Given the existence of a common grammaticisation path from verb to adposition, this is unsurprising. Most instances in Figure 6 where head/complement order is not shared are languages that lack adpositions altogether. This is true of a few Sepik and western Torricelli languages, and of a cluster of TNG-TAP languages on Alor and Pantar – OV languages that have evidently lost their postpositions, as their immediate relatives in Timor retain them. Figure 6 also highlights the Northern Arc, as Austronesian, West Papuan, East Bird’s Head, western Torricelli, and the Papuan languages of New Britain and New Ireland all display VO&Prep, in contrast with the OV&Postp pattern of the core mainland area and the Fly-Digul Shelf. The Skou languages are missing from the Northern Arc on this variable as they have OV order and display various values for adpositions. In fact, the map exaggerates the Northern Arc because one lineage, Austronesian, stretches its full length.

Within OV languages of the NGR there is a seemingly sharp difference between those that allow a postverbal adjunct and those that do not, i.e. between OVX and OXV. Many TNG languages appear to insist on OXV, whereas many languages of the Sepik-Ramu Basin allow OVX (Foley, In prep.).

An important feature of Figure 6 is that where Oceanic and eastern Torricelli languages are in close proximity to OV&Postp languages, they tend to replace inherited VO&Prep with OV&Postp. Because much more is known about the histories of Oceanic

---

Abbreviations for typological features follow those in Table 4.
languages than of Papuan, recognising these changes in Oceanic is fairly straightforward. We find clusters of OV&Postp Oceanic languages in two locations: on the coasts and offshore islands of the southeast mainland (Ross 1996) and on the north mainland coast, the latter including the Bel languages, whose history has been studied in some detail (Ross 2008). In both cases there has been intense contact with TNG languages, and Oceanic languages have been remodelled on TNG patterns. In central and south Bougainville we find Oceanic languages that are headed in the same direction. Papapana and Mono-Alu vary between SVO and SOV and between pre- and postpositions, whilst Torau is now SOV but has mixed adpositions (Evans and Palmer 2011, Smith, forthcoming). It is a reasonable inference that the eastern Torricelli languages, with OV&Postp, have undergone a similar set of changes thanks to their isolation from the more conservative (VO&Prep) western part of the lineage, and to their interaction with neighbouring Sepik and RLS languages with an OV&Postp template.

Figure 6 shows that the Oceanic languages around the Vitiaz Strait between the mainland and New Britain have VO clause order but a mixture of prepositions and postpositions – but no neighbouring Papuan language is an evident source of this value combination. One may reasonably infer, though, that the innovation of postpositions in these languages reflects contact with now extinct Papuan languages of New Britain. The languages of New Britain’s south coast have other quite un-Austronesian features, among them pronominal gender, also a Northern Arc feature (section 3.6.10) which can only be reasonably explained by contact (Ross 2013).

3.5.2 Variables 3 and 4, subject and object indices on verbs

In many languages of the NGR the subject of a clause is indexed on the verb (Ewing 2010, Reesink 2010). In fewer languages, the object is also indexed, and in fewer still only the object is indexed (Figure 7). An index is either an affix or a clitic, but sources often do not distinguish between them, and ‘prefix’ and ‘suffix’ should here respectively be read as including proclitics and enclitics.
Map 7. Positions of verbal subject and object indices in New Guinea Region languages

‘s’ and ‘object’ imply accusative alignment, but a minority of NGR languages have split-S alignment (section 3.6.8). For the purposes of variables 3 and 4, if there is split-S alignment the index of the actor argument of both transitive (A) and intransitive (S_A) verbs is treated as subject, and the index of the undergoer argument (U, S_U) as object.

An index has three possible values – prefix, suffix, absence – giving nine possibilities, listed below with their distributions (where \{s,o\} indicates a sequence of subject and object indices in either order).

6) Distribution of subject and object indices

\( sV_{o} \) some ENus Austronesian, widespread Oceanic, one East Bird’s Head, some Skou, some Torricelli, Anèm-Ata, East New Britain, Bilua (C Solomons)

\( sV \) some W Papuan, East Bird’s Head, some ENus Austronesian, some Oceanic, some Skou, some Torricelli,

\( V_{o} \) Some Oceanic in New Britain and New Georgia (Solomons), Touo and Savosavo (C Solomons)

\( \{s,o\}V \) some TNG-TAP, TNG-Marind, some West Papuan

\( oV_{s} \) scattered TNG

\( oV \) some TNG-TAP

\( V_{s} \) scattered TNG, North Bougainville

\( V_{s} \{s,o\} \) few TNG, Nimboran, Sentani, Senagi, Kwomtari, some Sepik, South Bougainville

\( V \) TNG south of cordillera, Border, Sepik, LSR

From this distribution, \( sV_{o} \) emerges as a Northern Arc feature, although it is underrepresented in ENus, where TNG-TAP and West Papuan languages have other configurations. Examples are:

7) a. Allang (Austronesian, Ambon, ENus)
    mane me=patina=wa.
    3sF 3sA=know=2sU ‘He knows you.’ (Ewing 2010:121)

    a. Manam (Oceanic, NWNG)

    \( \# u\)-lele-\( \# a \).ma.
    2sS-look.for-1peO ‘You looked for us.’ (Lichtenberk 1983:125)

    b. Meyah (East Bird’s Head)
    Memef me-agob-ir.
    1peF 1peA-strike-3pU ‘We strike them.’ (Gravelle 2002:138)

    c. Barupu (Skou, NWNG)
    /k-opu-jara-\( n \)/ LH ---> [kòpìjäräńf]
    R-2pmS-see-1sfO ‘You see me.’ (Corris 2005:77)

    d. Bukiyip (Toriccelli, NWNG)
    n-a-hw-\( \# k \)
3smS-R-hold-3sfO  ‘He married her’ (Conrad and Wogiga 1991:24)
e.  Anêm (Anêm-Ata, NWIM)

da-b-ɯ
IRR.1sS-hit-2sO  ‘I’ll hit you.’ (Thurston 1982:48)

The Oceanic evidence indicates that sV often reflects earlier sVo, whilst Vo does so more occasionally. Apart from the Central Solomons languages Bilua, Touo and Savosavo, all the languages that manifest sVo, sV or Vo belong to the Northern Arc.

The order oV is found only in TNG languages, whilst oVs, {s,o}V, Vs, V{s,o} and V are found in TNG and other Papuan lineages. On the evidence of its distribution and of morphological reconstruction by Suter (2012) oVs was evidently a Proto TNG feature, which has undergone a variety of erosions and transformations.

The common grammaticisation path free pronoun > bound pronoun gives rise to a prediction that sVo will occur in VO&Prep languages and oVs in OV&Postp languages, and this is confirmed for the NGR.

3.5.3 Variable 5, possessor index on nouns

Possessor indices occur on possessum nouns in 296 of the 387 languages for which there is relevant information in the database. In the vast majority of these, indices occur only on inalienably possessed nouns, or they are obligatory on inalienably possessed nouns and optional elsewhere (cf section 3.6.11). There is a strong tendency throughout the NGR for possessor indices to complement subject indices in their placement. That is, if a language has both possessor indices on nouns and subject indices on verbs, either the possessor index is a prefix and the subject index a suffix, or vice versa. This is tantamount to saying that possessor indices usually match object indices in their position. This is true of languages that have both possessor and object indices, but there are enough languages that lack one or the other for the correlation not to be obvious.

Npr&Vs(s) occurs in a good many ENus Austronesian languages, in many Oceanic languages, and in Anêm-Ata. The prN&oV(s) pattern is found in a good number of scattered TNG languages, and this was evidently the Proto-TNG pattern. It also occurs in South Bougainville.

However, there are also mismatches, i.e. prN&sV or Npr&Vs, a number of which are attributable to a historically reconstructable shift in the position of the possessor index. Thus in the Austronesian languages Tugun and Tulehu (ENus) and Waropen and Warembori (Cenderawasih Bay area, mainland) inherited Npr is replaced by prN, giving prN&sV. Since this pattern is also reflected in Kamang and Wersing of eastern Alor (TNG-TAP), and in the West Papuan and Eastern Bird’s Head lineages, one may infer that Papuan contact has played a role in the change. The usual Torricelli pattern is subject prefixes and no possessor affixes, but the eastern Torricelli languages Kamasau and Monumbo have apparently acquired possessor prefixes. The TNG-Marind microgroup also manifests prN&sV(o), but this reflects a shift from subject suffixes to subject prefixes (section 3.6.2).

Apparently the converse change in possessor affixes, from prN to Npr, has occurred in a number of TNG microgroups, giving rise to Npr&Vs in the Gogodala-Suki, Koiari, Managalasi, Dagan, Finisterre-Huon and Chimbu-Wahgi microgroups, a distribution wide enough to imply that possessor suffixes occurred alongside prefixes in early TNG.
Map 8. Clause-final negators in declarative realis verbal clauses in New Guinea Region languages

3.5.4. Variable 6, negator position in declarative realis clause

Clause-final negation is commonly mentioned as an ENus areal feature (Reesink 2002, Klamer, Reesink and van Staden 2008), and Reesink (2003) shows that it stretches from ENus along the north coast to New Britain and also occurs in the Oceanic languages of the Admiralties. Figure 8 shows that negators occur clause-finally in declarative realis clauses in all the Northern Arc languages, other than some Torricelli languages and the East New Britain lineage. The languages in (8) have VO order, those in (9) OV.

8) a. Taba (Austronesian, ENus)
   nik calana kuda-k asfal te my trousers be.black-APPL asphalt NEG
   ‘My trousers are not blacked with asphalt.’ (Bowden 2001:336)

   b. Arop-Lokep (Oceanic, mainland north coast)
   i-yiri ookoo tiap. 3sS-board canoeNEG
   ‘I did not board (the) canoe.’ (D’Jernes 1990)

   c. Maybrat (West Papuan, Bird’s Head)
   ana m-amo Kumurkek fe they 3pS-go Kumurkek NEG
   ‘They do not go to Kumurkek’ (Reesink 2002:246)

   d. Meyah (East Bird’s Head)
   me-en-et mat guru 1peA-DUR-eat food NEG
   ‘We have not eaten food.’ (Gravelle 2002:168)

---

21 For a slightly different view, see Lichtenberk (2013).
e. Bukiyp (Torricelli, NWNG)

… m-u-wu_ kakwich e
2pS-IRR-plant garden.food NEG.PAST
‘… we would not have planted garden food.’ (Conrad and Wogiga 1991:131)

f. Anêm (Anêm-Ata, NWIM)

Paulus u-tl-wu aba mantu
Paulus 3sS-spear-3sO pig NEG
‘Paulus didn’t spear a pig.’ (Thurston 1987:75)

9) a. Bunaq (TNG-TAP, ENus)

hot baq no zapal ga-sasi niq
sun noon OBL folktale U3-say NEG
‘During the day (we) don’t tell folktales.’ (Schapper 2009:181)

b. I’saka (Skou, NWNG)

… tâu d-ei mi
noise 1sS-do NEG
… I didn’t make any noise.’ (Donohue and San Roque 2004:100)

There are two possible objections to a claim of areality, however. First, most languages with clause-final negation have VO order, but some (TNG-TAP, West Papuan, Skou) have OV order. Are these the same value of a single variable? VONeg is clause-final negation, but OVNeg could also be interpreted as postverbal negation, corresponding with VNegO. Second, OVNeg crops up in places outside the Northern Arc: in scattered TNG languages – Mbaham (Bomberai Peninsula), Ekari (Wissel Lakes), Asmat (Asmat-Kamoro), Pisa (Awyu-Dumut), Kati and Tifal (Ok), Pawaian (Teberan-Pawaian), Koriki (Purari) – and in the Eleman languages (which may be phonologically unusual TNG languages) and Lavukaleve (Central Solomons). However, there are no cases of VONeg outside the Northern Arc. This suggests that OVNeg can develop fairly spontaneously, initially as an alternative to the morphological negation that is widespread in OV languages. It does so perhaps because the negator is adjacent to the verb.

Reesink (2002. 2003; Klamer et al. 2008) makes the well-supported proposal that clause-final negation, at least in ENus, has a Papuan origin. It appears more consistently in ENus Papuan (TNG-TAP, West Papuan) languages than in Austronesian. Indeed TNG-TAP’s Austronesian neighbours do not have clause-final negation, so they are not the source of its presence in TAP languages. On the other hand, the Papuan languages of ENus (TNG-TAP and West Papuan) are OVNeg, so clause-final negation may well have developed in one or more Papuan groups, subsequently infecting their Austronesian neighbours.

Support for this comes from the worldwide distribution of SVONeg languages. They occur particularly in two areas, one centred around Cameroon, the other around New Guinea and especially along its north coast (Dryer 2013e). This implies a contact-induced origin for VONeg and that clause-final negation is indeed an areal feature of the Northern Arc.

Reesink (2009) claims that clause-final phasal adverbs (‘still’, ‘no longer’, ‘already’, ‘not yet’) are also a feature of an area that roughly corresponds to the Northern Arc. It is difficult to check this claim in detail as so few language descriptions recognise
the category, but a survey of languages where information is available suggests that these are usually also languages with clause-final negation.

3.5.5 Variable 7, clause sequences

One of the signature structures of Papuan languages is the clause chain (Longacre 1972, Foley 1986: 175–198). This consists of two or more OV clauses, of which only the final clause verb is independent, i.e. fully specified with suffixes for tense, aspect, mood and indexing the subject (see the suffixes of toto-iʔa ‘they cut’ in 10)). Medial (= non-final) clauses are dependent, i.e. cannot normally end a sentence. If they have the same subject as the final verb, they may be morphologically simple. But a medial verb may also have morphological complexity, indicating a temporal or causal relationship (e.g. sequence, simultaneity, reason) with the following clause, and indicating whether the subject of the medial verb is the same as or different from that of the final verb. A morphological indication that their subjects are different is known in the literature as switch-reference, a very common feature of clause chains (Roberts 1997) but not an invariable one, and Figure 9 distinguishes between clause-chaining with and without switch-reference. Foley and Van Valin (1984: 242) and Foley (1986: 177) characterise medial verbs as dependent but coordinate.²²

In the Tauya example in (10)), switch-reference is morphologically simple. If the next clause has the same subject (SS), the medial verb takes the suffix -pa. If it has a different subject, the suffix is -te. The morphology of the chain-final verb is italicised.

10) Tauya (TNG-Madang)

```
nono ʔ-imai-te-ʔa mai mene-a-te pai aʔate-ʔa nono wi nen-fe-ʔa
child 3sO-carry-get-SS come.up stay-3sS-DS pig hit-SS child show
3pO-TR-SS
yene wawi wi nen-fe-ʔa mene-ʔa pai aʔate-ti tefe-ʔa ʔeʔeri-ʔa
toto-iʔa
sacred flute show 3pO-TR-SS stay-SS pig hit-CONJ put-SS
dance-SS cut-3pS-IND
```

‘She carried the child, arrived and stayed, and they killed the pigs and showed them to the children, and showed them the sacred flutes and stayed, and killed the pigs and put them (down), and danced and cut (the pigs up).’ (MacDonald 1990: 218)

Clause-chaining is absent from languages of the Northern Arc and from the Fly-Digul shelf and its surroundings. It is present in much of the rest of the NGR: in many TNG languages, in a number of Sepik and South Bougainville languages and in Savosavo (C Solomons) and without switch-reference in Ramu-Lower Sepik and North Bougainville languages and in Lavukaleve (C Solomons). Clause-chaining without switch-reference is also found in Oceanic languages that have intense contact with TNG languages: the Bel group on the north coast of the mainland (Ross 1987, 2008) and in Maisin in southeast mainland New Guinea (Ross 1984, 1996). These are clearly instances of metatypy.

²² Foley (2010) offers a revision of this view.
3.5.6 Variable 8 and 9, noun/adjective order and noun/demonstrative order

NA order is an areal feature of the NGR as a whole. Many linguists postulate headedness harmony, i.e. VO&NA or OV&AN, on the basis of Eurasian languages, but neither worldwide nor NGR data support it (Dryer 1988, 1992, 2013a, 2013b). In fact OV&NA is common among the world’s languages. In Dryer’s (2013a) sample there are 216 OV&AN languages and 332 OV&NA. He notes that New Guinea is a region where OV&NA is common, but that OV&AN intervenes in the Sepik lineage and the TNG-Kainantu-Goroka microgroup. It also occurs in the Fly-Digul lineages.

Almost all VO languages in the NGR have the pattern VO&NA. Only one, Bilua, a Papuan language surrounded by Oceanic, displays VO&AN, and this presumably reflects a contact-induced switch from OV to VO, as the other three scattered members of the putative Central Solomons lineage have OV order.

Throughout the NGR order of noun and demonstrative almost always matches order of noun and adjective. ND is thus also a NGR areal feature. The results of combining the two variables in WALS online are: NA&ND = 476 languages; NA&DN = 25; AN&ND = 266; AN&DN = 184. There is thus a strong worldwide tendency for NA to be associated with ND (but not for AN to be associated with DN).

3.5.7. Variable 10, possessum/possessor order

It is usually assumed that verb/object and possessum/possessor orders will agree in headedness direction, but, as Dryer (2013d) remarks, things are not that simple, as VO languages that are near to OV&PrPm languages are also likely to have PrPm order. This implies that PrPm order spreads through contact faster than OV order. This is relevant to the NGR, as it has often been observed (since Brandes 1884) that PrPm order is a widely distributed feature of NGR VO Austronesian languages, and a little less often asserted that this is due to contact with Papuan languages (Grimes 1991:282; Klamer et al 2008: 128). In light of Dryer’s generalisation, this is probably true.

Thus PrPm order is a NGR feature, regardless of whether a language has OV or VO order. However, as Figure 10 shows, there are parts of the NGR where PmPr order occurs. The Austronesian languages of the extreme west of ENus (on Flores and Sumbawa) and of the Bismarcks, Bougainville and the western Solomons all have the inherited Austronesian PmPr order. Significantly, many of these languages are well away from Papuan languages, and the likelihood of contact-induced change is lower than elsewhere in the NGR. But Bougainville and the Solomons are an exceptional area in this regard: their Papuan languages have OV&PrPm, but only a few of their Oceanic neighbours (Papapana, Torau, Mono-Alu) have shifted to the Papuan pattern.

Combining the adposition/NP and possessum/possessor variables in WALS online gives these results: Prep&PmPr=351; Prep&PrPm=54; Postp&PmPr=13; Postp&PrPm=442. In words, there is a strong likelihood worldwide that prepositional phrases will cooccur with postposed possessors, and postpositional phrases with preposed possessors. Again this is unsurprising, as adpositions are sometimes grammaticised from possessum nouns (‘inside of the house’ becomes ‘inside the house’). The VO&PrPm languages of the NGR are thus exceptional, displaying Prep&PrPm.
Map 9. Clause-chaining with and without switch-reference in New Guinea Region languages

Map 10. Possessum/possessor order in New Guinea Region languages
Map 11. Alignment of argument indices in New Guinea Region languages
3.5.8 Variable 11, alignment of verbal argument indices

In a large majority of NGR languages subject and object indices are accusatively aligned. In the few that are not, there is a split-S pattern. In split-S alignment, if the single argument of an intransitive verb is semantically actor- than undergoer-like, then its index is identical to the transitive subject index. If, on the other hand, it is semantically more undergoer-like, then its index is the same as the transitive object index. Where the line is drawn between actor-like and undergoer-like arguments of intransitives differs from language to language (Foley 2005b), and there are even cases in ENus where the choice is lexically determined, with no obvious semantic basis (Holton 2010).

Klamer (2008) shows that split-S is an ENus areal feature embracing Austronesian and Papuan languages (also Ewing 2010 on Austronesian). Examples (2) and (4) illustrate split-S in the TNG-TAP microgroup, replacing the TNG accusative alignment of (3). However, the domain of split-S languages in the NGR extends beyond ENus along the Northern Arc. Split-S also turns up in the Fly-Digul lineages.

As Figure 11 shows, in the Northern Arc split-S is found alongside accusative in West Papuan, the Torricelli language Mufian, Anêm-Ata and East New Britain (and in the isolate Taiap in the Sepik-Ramu basin). Split-S also occurs in Lavukaleve in the Canetral Solomons. There are two possible explanations for this distribution. Either some languages in Northern Arc groups have shifted independently from accusative to split-S, or these languages reflect an ancient Northern Arc split-S pattern that has been masked by a general shift to accusative alignment. Only the latter possibility explains the distribution of split-S languages across the Northern Arc.

Further evidence for this hypothesis lies in languages that are accusatively aligned but have only object indices. They lack subject indices. Lineages that include such a language are notated B(A) in Table 5. Such languages are rare in the NGR, but they are located in the neighbourhoods of split-S languages. They are: a TNG-TAP cluster on Pantar, the Torricelli language Urim (near Mufian), the Oceanic language Naknai (neighbouring Ata), Touo and Savosavo (putative sisters of Lavukaleve in the Central Solomons) and the Oceanic languages Hoava, Marovo and Roviana (adjacent to Touo). This distribution is too good to be coincidental. In accusative languages, the privileged argument (Van Valin and LaPolla 1997: 176) is the subject, i.e. A + S. The fact that the undergoer index survives, rather than the actor, goes back to a time when U+S was a privileged argument in either an ergatively aligned pattern or, in these instances, a split-S pattern.

3.5.9 Variables 12 and 13, basic tense/mood distinctions in verbal morphology

Among NGR languages that encode tense, aspect and/or mood there are two contrasting types. In one the primary contrast is between present tense and past tense, often with two or three different past tenses (today, yesterday, remote). If one tense is morphologically unmarked, it is the present. In the other type the primary contrast is between realis and irrealis, where realis is understood to denote actual events, past or present (both often morphologically unmarked), and irrealis is understood to denote future events,
expressions of ability or obligation, and hypothetical events present and past. Exactly where a language draws the line between realis and irrealis varies (Bugenhagen 1994). For example, some of these languages mark an imperative as irrealis, and some don’t.

If a language belongs to one of these two classes, this is usually plain from the description of its verbal morphology, as the first statement about its tense/aspect/mood morphology concerns one of these two basic distinctions. One might describe these two types as ‘tense-prominent’ and ‘mood-prominent’. There is also a less common class of ‘aspect-prominent’ languages, but this variable awaits investigation.

Map 12. Tense- and mood-prominence in New Guinea Region languages

Mood-prominence is a feature of the Northern Arc, represented in scattered ENus Austronesian and many Oceanic languages, in the East Bird’s Head, Skou and Torricelli lineages, in Anêm (Anêm-Ata) and the isolates Sulka and Kuot. Apart from Skou, these are VO languages. Mood-prominence is illustrated in (11) – (15). The encoding of the realis/irrealis contrast is almost always preverbal, but the encoding patterns vary. In Larike there are distinct mood morphemes. In Manam and Anêm, there are different subject indexing prefixes. In I’saka irrealis is encoded by reduplication of a syllable consisting of the subject indexing prefix and the root-initial vowel. In Urat there is a preverbal irrealis morpheme and the verb stem itself undergoes change.

11) Larike (Austronesian, ENus)
   a. arua-i-lena
   b. arua-na-lena
   1deA-R-walk 1deA-IRR-walk
   ‘we two are walking’ ‘we two will walk’ (Laidig and Laidig 1995)

12) Manam (Oceanic, NWNG)

---

23 It is important to differentiate between this usage of ‘irrealis’ and the much narrower usage in some Papuan language descriptions that excludes the encoding of future events.
336

Figure 12 shows, however, that ENus languages are mostly exceptions to the Northern Arc pattern, being neither mood-prominent nor tense-prominent. Mood-prominence is an inherited Austronesian feature, so the scattered mood-prominent ENus Austronesian languages are conservative, while most have lost mood-prominence, bringing them into line with TNG-TAP and West Papuan.

TNG, OMTK, Border, Sepik, the Fly-Digul lineages, Eleman and North and South Bougainville are tense-prominent. They all have OV order and most index the subject with a suffix that simultaneously encodes tense, as in (16)).

16) Kyaka Enga (TNG-Engan)
   a. kande-lyo  
      see-PRS.1sS  
      ‘I see/look’
   b. kanda-pu  
      see-PST.1sS  
      ‘I saw/have seen’
   c. kande-yo  
      see-FPST.1sS  
      ‘I saw’
   d. kanda-lono  
      see-RPST.1sS  
      ‘I have just seen’
   e. kanda-ro  
      see-FUT.1sS  
      ‘I shall see’
   f. kanda-rono  
      see-IMMEDIATE.FUT.1sS  
      ‘I am about to see’

   (Draper and Draper 2002:28–31)

Generally tense- and mood-prominent languages are in complementary distribution in the NGR, but occasionally languages crop up that are both, like Sentani, Senagi and Kwomtari (the two last are contiguous) and the Oceanic languages of the southeast mainland coast. The latter are also mood-prominent by inheritance, but have added a past tense through speakers’ bilingualism in a TNG language.
3.5.10 Variables 14–16, clusivity and number in free pronouns and gender in pronominals

These three variables concern pronominals. Clusivity makes a distinction between first person exclusive (‘we’ = ‘I’ + (s)he/they’) and inclusive (‘we’ = ‘I’ + ‘you’) free pronouns. Number concerns whether free pronouns distinguish more than one non-singular number (i.e. dual, trial, quadral or paucal). Gender is defined as a masculine/feminine distinction in free pronouns and/or subject indices (a human/non-human distinction, quite common in ENUs, is ignored). The three variables were selected because of an impression that they were typical of the Northern Arc. In the event, the impression proved wrong, as Table 5 shows.

Gender overlaps with noun class, which is important in various Papuan lineages other than TNG, but has so many variants that it lay beyond the scope of this investigation.

Inclusive forms that are otherwise atypical of their lineages appear in a broad scattering of TNG languages and in a few Sepik and Torricelli languages, implying perhaps that the distinction is quite easily innovated independently.

A cursory look at these distributions shows that they do not coincide. One can hardly argue that the patchy presence of these features is a Northern Arc characteristic, first because there are too many absences of each one from Northern Arc lineages, and secondly because there are numerous non-Northern Arc languages in which these features are present. There may come a time when some historical sense can be made of these distributions, e.g. when an account can be given of absences from the Northern Arc, but such an account is at present fragmentary. Thus one can argue that the presence of only one non-singular number in the free pronouns of various Oceanic microgroups is due to Papuan contact, but one then has to explain why gender has entered Oceanic languages through Papuan contact only in southwestern New Britain.

The way that these values are interleaved geographically, and across lineage boundaries in local areas, suggests that they are quite sensitive to contact-induced change and perhaps to spontaneous innovation. Interestingly, Oceanic languages of the southeast mainland have only one non-singular number, whilst many of their TNG neighbours have both dual and plural.

3.5.11 Variable 17, distinction between inalienable and alienable possession constructions

A distinction between constructions encoding inalienable and alienable possession (‘(in)alienability’) is common in the NGR, but it is near-universal only in Oceanic languages of the NGR. In most languages with the distinction, the possessor is indicated by an obligatory affix on an inalienably possessed noun (section 3.6.3), while possession of other nouns is marked either by the same affix, but optionally, or by some other means. The distinction is common but not always present in ENUs (in Austronesian, TNG-TAP and West Papuan). It also occurs in East Bird’s Head, East New Britain and South Bougainville languages. Its occurrence is sporadic in TNG languages, but seems to have been present in pTNG, encoded by possessor prefixes. It is also sporadically present in the Sepik, and Torricelli lineages. However, it is absent from Lakes Plain, Border,
Skou, RLS, Bulaka, Eastern Trans Fly, Eleman and from Lavukaleve and Savosavo of the Central Solomons lineage.

Whilst (in)alienability is present in all Northern Arc groups except Skou, the fact that in most lineages there are some languages with it and others without suggests that the inalienable construction is quite easily lost. Moreover, the fact that the distinction is not exclusive to the Northern Arc groups and is present in TNG means that it is not really diagnostic of areality.

Ross (2001:138) suggested that (in)alienability had probably been copied into Proto-Oceanic or its immediate precursor from a Papuan language, and Donohue and Schapper (2008) show that its introduction into both ENus Austronesian and Oceanic can indeed be attributed to Papuan contact. As they note, it was the alienable construction that was innovated. The inalienable construction continues the inherited general possession construction.

4 Discussion

What emerges from the distributions of the A and B values of the variables discussed in section 3.6 in each of the three areas of Table 4 (section 3.3)?

The Northern Arc, despite its somewhat chimerical appearance, stands up to scrutiny, and the various claims of such an area, especially Reesink’s (2003) North Papuan linkage hypothesis, appear to be supported by the features in (17).

17) Northern Arc features:
   a. VO&Prep (section 3.6.1)
   b. sVo (section 3.6.2)
   c. clause-final negation (section 3.6.4)
   d. split-S (section 3.6.8)
   e. mood-prominence (section 3.6.9)
   f. (in)alienability (section 3.6.11)

The only lineages that participate in all these values are Torricelli and Anêm-Ata, and even here definition is weakened by the fact that not all languages in each lineage display all values. Only one Torricelli language is known to have split-S alignment (section 3.6.8) and only some display (in)alienability. Of Anêm and Ata, only Anêm is mood-prominent and only Ata has split-S alignment and (in)alienability. Classifying their lineages as representing these values assumes that the values reflect history in some way.

The features in (17) are affected by contact in different ways. A number of Oceanic languages, by inheritance VO&Prep, have so to speak departed from the Northern Arc by copying OV&Postp from their Papuan neighbours, yet these same languages have retained sVo. That is, bound morphology is less readily affected by contact than is clause syntax. A similar history can probably be reconstructed for eastern Torricelli languages.

The predominance of B values, especially in the orders AN and DN, in the Fly-Digul Shelf area also suggests areality, but more data will be needed before this can be confirmed. The fact that the TNG microgroups Marind and Kiwai occur almost in a block with the Bulaka, Yam and Eastern Trans-Fly lineages in Table 4 also suggests areality.

The core ‘area’, however, is neither a geographic nor a typological area. The North and South Bougainville and Central Solomons lineages hardly form a typological area.
with the mainland, nor for that matter with each other. The non-Northern Arc languages of the mainland share OV&Postp, tense-prominence and accusativity, but lack consistency with regard to object indexing, pronominal gender and (in)alienability. Some core languages famously chain clauses (TNG, Sentani, some Border, Senagi, some Sepik, RLS, North and South Bougainville, and Lavukaleve in the Central Solomons), but others do not (OMTK, Nimboran, Kwomtari, Yam, Eastern Trans-Fly, Eleman). Areality here is weak yet complex, to put it mildly.

Because so many exceptions to core area tendencies are found among lineages in the area of extreme diversity towards the northern end of the border between Indonesia and Papua New Guinea, I have looked for signs of a separate areality embracing the Nimboran, Sentani, Border, Senagi, and Kwomtari lineages. They agree on oV, on absence of both prN, Npr and inalienability, and except for Nimboran they are mood-prominent. However, the single members of lineages omitted from the investigation – Ama (Left May), Karkar Yuri (West Pauwasi) – do not share these features, and further data from the languages of the region are needed before one can discern its areality, if any.

Thus only the Northern Arc shows a constellation of values that point to areality, and this is fairly blurred. It does, however, display values that are distinct from those of much of the core mainland area, and its languages are arranged along what by Austronesian times was a maritime corridor. The lineages along the Arc are not related to each other at any time depth that the comparative method can detect. Some of the typological similarities among them may reflect ancient relationships. Others perhaps reflect ancient contact.

5 Concluding thoughts

Scholars have often assumed that one can distinguish Austronesian from Papuan languages not just phylogenetically but also typologically, implying that there is a common Papuan type (Klamer, Reesink and van Staden 2008: 113–114; Klamer and Ewing 2010: 11; Reesink 2010: 71), but this is not well supported by the data.

In section 1 I noted that the western and northwestern boundary of the NGR marks a typological boundary within the Austronesian lineage. NGR Austronesian languages are typologically more similar to their Papuan neighbours than to their Austronesian sisters to the west and northwest. As Donohue (2007b) argues, this is suggestive of contact – from an earlier time than the localised contact effects mentioned in sections 3.6.1 and 3.6.5. It has long been inferred that PrPm order in NGR Austronesian languages is due to its presence in one or more substrate languages (section 3.6.7). Clause-final negation (section 3.6.4) and the (in)alienability (section 3.6.11) are attributed to contact with a Papuan source. Donohue (2005b) justifiably questions the claim that languages of, for example, the East Bird’s Head and Torricelli lineages have acquired their SVO order by contact with Austronesian languages when they display few other signs, if any, of Austronesian contact. Rather, as SVO was not the order inherited by ENus Austronesian and Oceanic languages from earlier verb-initial Austronesian, he suggests that they acquired it from one or more Papuan substrates, whilst Papuan SVO languages have either done the same or continue the substrate’s SVO order. One can also argue that the presence of subject indexing prefixes on verbs in most NGR Austronesian languages and the presence of object indexing suffixes on many of them is an outcome of contact.
(section 3.6.2). To be sure, the indexing morphemes had been present in Austronesian at earlier stages, but there was a change in distribution and function with the shift to SVO. In this scenario, Papuan languages did not change under Austronesian influence, but early Austronesian entrants to the NGR were recruited into the Northern Arc under Papuan influence.

However, PrPm, clause-final negation, (in)alienability, sVo and SVO have different distributions, and just what this tells us about the history of Austronesian in the NGR remains to be worked out. It is, of course, tempting to look around for modern descendants of ancient substrates, but these may well have disappeared. Nonetheless, as the West Papuan lineage displays features common to ENus and in considerable measure to the Northern Arc, one can reasonably suggest that at least one ENus substrate was related to West Papuan. How closely related we will never know.

Occasionally one encounters mild scepticism about positing extinct substrates (e.g. Florey 2010), but with an area whose languages have been more or less in place for as long as in the NGR, they are a reasonable hypothesis.

It is just possible that a substrate might survive by accident. One such candidate is the isolate Taiap, spoken in the Sepik Basin. Its location happens to coincide with the one island in what 6,000 years ago was an inland sea (Ross 2005a), and one wonders if it is a relic from the deep past. However, Taiap’s features group it with no known lineage. It displays OV&Postp, mixed sV and Vs, Vo, clause-final negation, no clause chaining, split-S alignment and mood-prominence, a combination which, but for OV&Postp, places it within the Northern Arc – yet pronominal modifiers put it with the Fly-Digul languages of the south!

One of the features claimed for the Northern Arc is sVo, a feature that entails bound morphology. There is, however, no clear sign that this morphology is cognate across lineages. The claim is rather that where an sVo pattern occurs in one of a bilingual community’s languages it may be copied into their other language by cliticising free pronominals to the verb. With these clitics become affixes. Once an sVo pattern is established it tends to be maintained in the face of phonetic erosion by new cliticisations, a process that can readily be documented in Austronesian languages of ENus and Oceania that share sV but not the forms of subject prefixes.

A challenge of much work on areality is choice of variables. Bickel and Nichols (2006) propose defining areas on the basis of various approaches to population history and seeking statistically significant differences in the frequency of features inside versus outside the area. An obvious starting point is then variables known to be distinctive of or within the area, and this is where the investigation described in section 3 began. There are, however, a number of almost equally distinctive variables that could have been chosen: the OVX vs OXV distinction (section 3.6.1), phasal adverb position (Reesink 2009), relative clause position, noun classification (Nichols 1997b, Foley 2000, Terrill 2002, Reesink 2003), numeral classifiers (Nichols 1997b), encoding of plurality in noun morphology (Foley 2000, Reesink 2003), casemarking and alignment of noun phrases, aspect prominence (section 3.6.9), gender in argument indices (Klamer et al. 2008), verbal morphology indexing the number but not the person of an argument, indirect object indices on verbs, and valency increasing morphology.

All 17 variables in section 3.6 are morphosyntactic. Lexical and phonological variables could in theory also be collected. Since bilingually induced calquing gives rise to similar lexical polysemy and collocations, lexical variables would probably be revealing, but relevant data across enough NGR languages are not yet available. Comrie
and Cysouw (2012) identify just one phonological variable as common in Papuan languages, an absence of lateral consonants in languages other than West Papuan. Yet a number of highlands TNG languages have laterals at more than one point of articulation (Ladefoged, Cochran and Disner 1977, Ladefoged and Maddieson 1996:190). Also of phonological interest would be the distribution of tone systems across NGR languages, as tone plays a role especially in the TNG, Skou and Lakes Plain lineages (Donohue 1997, 2005a), and tonogenesis has also happened in some NGR Austronesian languages (Ross 1993, Remijsen 2001, 2002, Cahill 2011). However, so many descriptions of TNG languages omit mention of tone that this variable cannot yet be systematically evaluated.

What does the NGR have to teach us about areality? Perhaps that even where areas refuse to come into focus because the features that almost define them are not quite coterminous, it is possible to observe constellations of variable values that are unlikely to be due to chance and perhaps tell us something about deep history. We also need to understand better which features are more stable and which not, and why and under what circumstances some features change and others do not.

References


Dryer, Matthew S., 2013c. Relationship between the order of object and verb and the order of adposition and noun phrase. In Dryer and Haspelmath 2013:ch.95. http://wals.info/chapter/95

343


28 Languages of Eastern Melanesia

Paul Geraghty

1 Introduction

The purpose of this chapter is to review the history of the study of the languages of Eastern Melanesia, with a view to assessing the above controversial claim by Robert Blust, the doyen of Austronesian linguistics, in the light of support from Donohue and Denham (2008, 2012) and adverse reactions from Pawley (2006: 243-248), Ross and Næss (2007: 460-461) and others.

Note that ‘Melanesia’ is a geographical term for the area of the Western South Pacific comprising Papua New Guinea, the Solomon Islands, Vanuatu, New Caledonia and (usually) Fiji. ‘Melanesian’ is not normally used as a linguistic term. Within Melanesia, there are two major language ‘families’. ‘Papuan’2 languages are only found in Western Melanesia, i.e. Papua New Guinea and the western Solomon Islands, while Austronesian languages, belonging to the Western Oceanic and the western part of the Eastern Oceanic subgroups, are found throughout the area. Blust (2008: 445-446) divides Eastern Melanesia further, using the term ‘Remote Melanesia’ for Vanuatu and New Caledonia, his main area of interest, and ‘Southern Melanesia’ for New Caledonia, including the Loyalty Islands.

‘Eastern Oceania’, on the other hand, is a linguistic term, referring to all Pacific islands where Eastern Oceanic languages are spoken. Eastern Oceanic is the easternmost subgroup of Oceanic, itself the easternmost subgroup of the Austronesian language family. Although its precise nature and membership are still being debated (Lynch, Ross and Crowley 2002: 108, Pawley 2009: 535), I use the term to cover all the languages of the Pacific islands excluding (1) those of Papua New Guinea and the Western Solomon Islands, for which see Malcolm Ross’s contribution to this volume, and (2) those of the westernmost Micronesian islands of Belau (formerly Palau) and the Marianas, which are Western Austronesian (see chapter by Grant, this volume). Hence Eastern Oceanic comprises all the languages of the eastern Solomon Islands (from Bugotu on Santa Isabel eastwards), the Santa Cruz group, Vanuatu, New Caledonia, Fiji (including Rotuma), Polynesia and most of Micronesia, encompassing a vast area from the Solomon Islands in

---

1 My thanks to Robert Early, Rod and Beverley Ewins, Alexandre François, Raymond Hickey, John Lynch, Peter Moore, Patrick Nunn, Bill Palmer, Elizabeth Pascal, Andrew Pawley, and Jeff Siegel who helped me in various ways while I was writing this chapter; the usual disclaimers apply.

2 Note that I follow the tradition (Ray 1926: 24) of using ‘Papuan’ as a term for all non-Austronesian languages of Oceania – though some authors prefer the term ‘non-Austronesian’, abbreviated to NAn or NAN. Papuan languages do not constitute a single language family, but belong to some twenty-odd families and isolates (Ross 2005). For the argument that typological comparison indicates that Papuan and Austronesian languages share a remote common ancestor, see Dunn et al 2005 and Dunn et al 2007, and the counter-argument from Donohue and Musgrave 2007, who attribute what there is in common, e.g. inclusive-exclusive pronouns, to areal diffusion.
the west to Easter Island (Rapanui) in the east, and from Hawai‘i in the north to New Zealand (Aotearoa) in the south.

Eastern Oceanic was proposed as a subgroup by Biggs (1965) then defined more precisely by Pawley (1972) and modified by Lynch and Tryon (1985). It corresponds closely to what pre-historians have called Remote Oceania (Green 1991), being the islands that lie beyond Near Oceania - the chain of inter-visible islands that are believed to have been settled by speakers of Papuan (non-Austronesian) languages perhaps as many as 40,000 years ago. The current interpretation (e.g., Pawley 2009: 517) is that the speakers of Papuan languages did not have the maritime technology to expand further, so remained largely confined to Near Oceania. The earliest speakers of Oceanic languages, however, now known as the Lapita people and believed to have originated from South-East Asia, were skilful canoe-builders, sailors and navigators, and sailed eastwards to become the first occupants of Remote Oceania over 3,000 years ago.

In this chapter, I will not be using the terms ‘Western/Eastern Oceanic’ or ‘Remote/Southern Melanesia’, since my main concern is to compare and contrast Western Melanesia - the islands of Melanesia where Papuan languages are still spoken – with Eastern Melanesia, where currently no Papuan languages are found.

By ‘linguistic area’ I mean an area where different languages share features as a result not of common inheritance but of borrowing (i.e., the spread or diffusion of ‘areal features’), taken to be due to widespread bilingualism, indeed multilingualism. Classic examples of linguistic areas outside of Oceania include the Balkans (see Friedmand and Joseph, this volume), where numerous features are shared exclusively by Albanian, Romanian, Bulgarian and Greek, languages that are only distantly related, as well as by unrelated Turkish dialects; and India, where languages of Indo-European and Dravidian ancestry have come to share features such as retroflex consonants, and in extreme cases appear to even share a common grammar, with only the lexicon being distinct. Closer to the Pacific, the mainland Southeast Asian language area (Enfield, this volume) includes a number of Austronesian languages spoken in Vietnam.3

I will introduce in this chapter the term ‘post-linguistic area’, meaning an area comprising languages that share exclusively features diffused from a common unrelated language (or group of similar languages) that has disappeared – in other words, that share a common substrate that is now extinct. Thus, for example, for those who believe that the French dialects have a common Gaulish substrate, France would be a ‘post-linguistic area’.

There are four separate areas in Eastern Melanesia that have been considered possible post-linguistic areas, similar to the linguistic areas described by Ross (1996 and this volume) for North-West Melanesia, where seemingly aberrant features, from the standpoint of canonical Oceanic languages, have been attributed to contact with non-Austronesian languages that are no longer spoken there. These areas are: (1) the Reefs Santa Cruz group in the far eastern Solomon Islands, including Santa Cruz, Utupua and Vanikoro, where some linguists have claimed that the languages either are Papuan or have been influenced by a Papuan substrate; (2) the islands of Santo and Malakula (and

3 Grace (1981) proposed a rather different use of the term ‘linguistic area’ to refer to a community of languages in New Caledonia that are clearly related and share a single grammatical system but appear to have borrowed freely among themselves, even in core vocabulary, thus making reconstruction by the comparative method difficult if not impossible; see also discussion by Harrison (1981: 220-224).
perhaps others such as Ambrym and Efate) in north-western and central Vanuatu, in
which a former linguistic area comprising Oceanic and Papuan languages has been
proposed; (3) Southern Vanuatu, comprising the languages of Erromango, Tanna and
Anejom, where a Papuan substrate has also been proposed; and (4) New Caledonia
(including the Loyalty Islands), where the non-Oceanic component of the post-linguistic
area has been proposed as Papuan or Australian.

Map 1. The three main cultural and linguistic areas of the Pacific

A different type of linguistic area is found in those parts of Eastern Melanesia where
Polynesians have settled, probably within the last eight hundred years or so, in some
cases in close proximity to non-Polynesian communities, and retained their languages. I
will attempt below to demonstrate that the study of the changes resulting from contact
between these Polynesian ‘outliers’ and their Melanesian neighbours may provide
indications of the results of language contact among Oceanic peoples, and hence clues as
to what might have happened in such situations in the more distant past.

Further east, the distances between some of the islands of the Central Pacific
subgroup of Eastern Oceania (comprising Fijian, Rotuman and Polynesian languages)
may have inhibited the diffusion that is necessary for the development of linguistic areas.
Indeed, Polynesia was until quite recently viewed by some historical linguists as a perfect
‘laboratory’ of language change, with each island or island group being almost totally
isolated, so that borrowing or other external influence could almost be ruled out as factors
in language change. However this view has recently lost favour, as evidence from
archaeology, linguistics, oral tradition, toponymy and other disciplines (see, for example,
Geraghty 2004 and references therein) has painted a picture of long-distance voyaging
across much of the Pacific, at least during certain periods of prehistory, so that perhaps
only Easter Island appears to have been totally isolated since initial occupation - and even there a loanword or two from South America appears to have made it through, most notably *kumara/kumala* ‘sweet potato, *Ipomoea batatas*’. Arguments have been made for the existence in the past of linguistic areas in Western Polynesia (Tonga, Samoa and adjacent islands) and in what is now French Polynesia, but I will neither relate nor pursue those arguments here.

2 The problem of ‘aberrant’ Oceanic languages

It has long been recognised that there are major differences among the Oceanic languages of Melanesia with regard to their ‘Oceanic-ness’ \(^4\). While there are indeterminate cases, by and large the Oceanic languages of Melanesia can be classified into two groups: those that are similar to Proto-Oceanic (POc), usually labelled ‘exemplary’, and those that are very different from Proto-Oceanic, usually labelled ‘aberrant’. There has been much discussion regarding the cause of this aberrancy – in particular, whether it is due to internal factors or contact-induced change.

Aberrant languages are found in all subgroups of Oceanic except Micronesian and Central Pacific (which comprises the languages of Fiji, Rotuma and Polynesia). In particular, they are found in the western Solomon Islands, the Santa Cruz group, western and southern Vanuatu and New Caledonia and the Loyalties. Perhaps the first linguist to recognise this distribution was Ray (1926: 37, 594), who further pointed out that exemplary languages (in his terminology those with a large ‘Indonesian’ content) tend to be found on small, often off-shore islands, while aberrant languages tend to be found on large islands – a distribution parallelling that of Austronesian versus Papuan languages in Papua New Guinea and western Melanesia.

The question of aberrant languages raises the broader question of the origins of the Melanesian people. As has been pointed out by anthropologists, the fact that the Melanesians are physically distinct from other speakers of Austronesian languages to the west (e.g., Indonesia) and north and east (Polynesia and Micronesia) requires an explanation. Hocart (1923) put it thus: “There are other languages which are descended from the same original as Melanesian, namely, Polynesian, Malay, Malagasy. The people who use these languages are nearly all light in colour, straight-haired and somewhat Mongoloid, very unlike the Melanesian type. .. Either, therefore, the Melanesian type borrowed its speech from the Malay-Polynesian or *vice versa*. The map most decidedly lends its support to the first view.”

Ray’s (1926) monumental survey of Melanesian languages was, in part, a response to such observations. He viewed all the Melanesian languages as resulting from borrowing from an Austronesian language or languages, which he called ‘Indonesian’, spoken by colonisers from the west and adopted in varying degrees by speakers of Papuan languages.

The presence of such contact-induced change in Western Melanesia (Papua New Guinea and the western Solomon islands) has never been in dispute, since Papuan languages are there in great numbers and there are linguistic areas where diffusion continues to take place in both directions. It is in Eastern Melanesia that the historical

\(^4\) For a summary of grammatical features of Proto Oceanic, see Lynch, Ross and Crowley (2002: 54-91).
interpretation is disputed, with some attributing the changes from Proto-Oceanic seen in aberrant languages solely to internal factors, and others following Ray in attributing the changes at least in part to external factors, specifically features transferred from non-Oceanic languages that have since become extinct.

The more traditional view of Ray was espoused by Dempwolff (1937: 142, 160, 193), Capell (1943, 1962, 1971), Wurm (1954) and Cowan (1962) (see discussion in Pawley 2006: 224-227). As summarised by a more recent adherent (Thurston 1982: 77), the Austronesian languages of Melanesia are “Austronesian languages with NAN substrate of various kinds and to varying degrees.” Siegel (forthcoming) has also pointed out that one of the objections to this ‘pidginisation’ hypothesis, as it is sometimes called – the relative lack of Papuan loans in aberrant languages – is not valid, since a high degree of lexical loans is not expected in pidgins and creoles of this nature. He further argues that “some of the grammatical features they [Austronesian languages in Melanesia] display could have been the result of one of the processes involved in the development of pidgin and creole languages – i.e. language transfer. This could have been Papuan speakers transferring features into the Austronesian languages they were acquiring ... in most cases they were entirely absorbed into Austronesian communities, and their original languages no longer exist”.

Grace (1961, 1962: 409-410) was perhaps the first to seriously question the ‘pidginisation’ hypothesis. He pointed out that there is a tendency to uniformity in the Austronesian vocabulary shown by aberrant languages, which one would not expect if people from different parts of Indonesia had settled different parts of Melanesia, as was, according to Grace, implicit in the ‘pidginisation’ hypothesis. He added that no-one had demonstrated that any of the supposed Papuan words of aberrant languages could be sourced to any known Papuan language – to which, of course, the response is that the source languages are now all extinct, in Eastern Melanesia at any rate. As Capell (1962: 423) pointed out, such languages would have been very disparate, judging by those Papuan languages that remain in Western Melanesia. Nevertheless, Grace acknowledged that he was by no means certain that pidginisation had played no role in the linguistic history of Melanesia.

The more recent orthodox view (Pawley 1981, 2006) concurs with Grace’s doubts about pidginisation and holds that aberrant languages are so because of rapid internal change. It points to the fact that no Papuan language is found in Eastern Oceania and therefore the hypothesis that such Papuan languages have become extinct is ‘a pretty uneconomical hypothesis’ (Pawley 2006: 226). This is despite the fact that, as Pawley acknowledges, non-Austronesian languages have become extinct in many parts of Western Melanesia, such as the Admiralties (Blust 2008: 456) and New Ireland (Pawley 2006: 246)⁵; and Pawley has stated elsewhere (2009: 531) his belief that there were Papuan languages also in the eastern Solomon Islands, where there are none today, that have since become extinct.

In this chapter, I will argue with the former camp: that at least some of the features of aberrant languages of Eastern Melanesia can be attributed to the influence of non-Oceanic languages that were spoken previously but have since become extinct. The essential question which I will attempt to answer is: how can one demonstrate that an area is a post-linguistic area where substrate languages once existed and where features

⁵ Note also the original languages of the Negrritos of the Philippines, all now extinct and replaced by Austronesian languages (Reid 1994: 449).
were transferred to the surviving languages when the donor languages were never recorded and are all now extinct?

3 Papua New Guinea and the Western Solomon Islands

It has never been disputed that Oceanic and Papuan languages have been in contact in Papua New Guinea and the Western Solomons. As a result, linguistic areas have developed, for example, Papua Tip and the Western Solomons, where both Oceanic and non-Oceanic languages show OV order and postpositions, presumable replacing the VO order and prepositions of the earlier Oceanic languages (Lynch, Ross and Crowley 2002: 15, Evans and Palmer 2011; see also Ross this volume). Ball (2007: 140) also claims that the Papua Tip language Hula may have picked up ergativity from contact. In other areas, Papuan languages have acquired SVO order and other features from their Oceanic neighbours⁶, and there has been considerable lexical replacement in both directions. A similar situation obtains in north-western New Britain, for which Thurston (1994) has meticulously documented the results of areal diffusion among Austronesian and Papuan languages.⁷

It has also been claimed that the ancestral language, Proto-Oceanic, added a series of labiovelars (*bw, *pw and *mw) to its phonemic inventory as a result of extensive borrowing from Papuan languages (Dempwolff 1920: 91-92, Blust 1981, Lynch, Ross and Crowley 2002: 65, Lynch 2002: 320-322). It is possible too that the Oceanic innovation of indirect possessive classifiers (Lynch, Ross and Crowley 2002: 77-80) was the result of the influence of Papuan nominal classifier systems⁸. This would suggest that a linguistic area already existed, probably in the Bismarck archipelago, during the early development of Proto-Oceanic over three thousand years ago.⁹

The addition of a final echo vowel in the Western Solomons and Vanikoro, e.g. POc *onom ‘six’ > PWS *onomo, may also have resulted from substrate influence, to match the open syllabic structure of the Papuan languages.

In the Western Solomons, where formerly there were many Papuan languages, only a small number of which survive, it is believed that these languages affected, and were affected by, the incoming Austronesian languages in many ways (Pawley 2009: 527, 536, 559).

---

⁶ But see Thurston (1994: 586-587), Donohue (2005) and others cited therein for the argument that SVO is itself an innovation in some Austronesian languages, which had previously been VSO, resulting from the substrate influence of some non-Austronesian languages, e.g. in Bird’s Head. Lynch, Ross and Crowley (2002: 86-87) also believe that Proto Oceanic was verb-initial.

⁷ While some have assumed that these linguistic areas involved Austronesian speakers acquiring fluency in Papuan languages, see Pawley (2007: 40) and Siegel (forthcoming) for the argument that the reverse was the case. Undoubtedly diffusion occurred in both directions.

⁸ Donohue and Schapper (2008) have pointed out that similar possessive constructions are found in languages of eastern Indonesia, and claim that they originate from contact with Papuan languages there.

⁹ Indeed, Ross (2001: 312) has argued that the Papuan languages of this area may have earlier formed a linguistic area, as witnessed by the common feature of gender marking in the pronoun systems.
Terrill 2011), notably through lexical replacement. After demonstrating the relatively high rate of replacement of the Proto-Oceanic lexicon in the Western Solomons, Pawley (2009: 531) comments that a “reasonable inference is that .. the speakers of incoming NW Solomonic languages encountered substantial populations of non-Austronesian languages and that sustained bilingualism, especially on Choiseul and Santa Isabel but also in the New Georgia group, led to many non-Austronesian loanwords entering the basic vocabulary of the NW Solomonic languages.”

4 Eastern Solomon Islands

In the eastern Solomons (Guadalcanal, Malaita, Makira and adjoining islands), by contrast, there are now no Papuan languages, though they were “presumably .. once spoken on all the main Solomon islands at least as far east as Guadalcanal and possibly on Malaita and Makira as well” (Pawley 2009: 521). However, there is little evidence of any substrate influence from Papuan languages – for example, there are no quinary numeral systems - because the “non-Austronesian speaking populations in this region were small and were easily absorbed or replaced” (Pawley 2009: 531), so that Eastern Solomonic languages appear to be on the whole more conservative than their western counterparts (Pawley 2009: 530).

5 The Reefs Santa Cruz group

The three main languages of the Reefs Santa Cruz area in Temotu Province, the most easterly province in the Solomon Islands, have caused problems in their classification ever since they were first described by Codrington (1885: 16). The more traditional view has been that they are essentially Papuan (Capell 1962: 371, 382, who writes that they have “such a thin veneer of AN [Austronesian] as to be practically still NAN [non-Austronesian]”, Davenport 1962, Wurm 1978, 1985, Tryon and Hackman 1983: 41, 43-45). Tryon and Hackman (1983: 44) summarise their history thus: “In view of the complexities of Reefs-Santa Cruz noun and verb morphology so central to the languages and typical of Papuan languages in general, it seems unlikely that such features would have been borrowed by Austronesian speakers, but would much more probably represent retentions from earlier forms of the Reefs-Santa Cruz languages which remained fully functional when their speakers took over many features of an Austronesian language.” Tryon (1994: 635-637) also stated that they “have undergone contact-induced change to such an extent that they are difficult to classify .. they appear to share a core structure with obvious Papuan affiliations, while sharing just as obvious Austronesian features”, and pointed out specific links between these languages and the Papuan languages of the Solomons and Bougainville.

More recently, however, the pendulum has swung in the other direction, beginning with Lincoln (1978) arguing that they are Oceanic. Terrill (2002: 84) and Næss (2006: 288-290) have since demonstrated convincingly that at least one of the features that Wurm had attributed to Papuan influence, the multiple noun classes of Äiwoo, bears no

10 Terrill (2011: 313) demonstrates, however, that in the Western Solomons “long-term contact does not necessarily entail major structural borrowing, even over … thousands of years.”
similarity to Papuan structures, that many of the markers derive from Austronesian roots, and the system has parallels in some languages of Vanuatu. While Næss (2006: 293) has pointed out striking parallels between nominalising affix systems of Reefs Santa Cruz and those of Nasiioi and Buin, Papuan languages of Bougainville, and while there do remain questions about the large quantity of non-Oceanic vocabulary (Ross and Næss 2007: 466) and other features of these languages, such as their highly complex agglutinative verb structure and ergative verb phrase (Næss 2013), Ross and Næss (2007) nevertheless conclude that the Reefs Santa Cruz languages are Austronesian, probably constituting a first-order subgroup of Oceanic. This conclusion is supported on typological grounds by Dunn et al (2007: 398-499).

François (2006) has also queried whether the languages of nearby Vanikoro are truly Austronesian, and demonstrated (2009) that the three languages share a common grammar with distinct lexicons – a typical outcome of prolonged contact and bilingualism in a linguistic area. 11

6 Vanuatu

While it is generally accepted that there are no Papuan languages in Vanuatu, and archaeologists have so far failed to find traces of any pre-Lapita (Oceanic-speaking) population, a number of linguists have speculated that there may once have been. Codrington, the compiler of the first major work on the languages of Melanesia, while at pains to affirm that Melanesian languages were as Oceanic as Polynesian languages, nevertheless believed that Melanesians may have once spoken a non-Melanesian language which is no longer recoverable (1885: 14-15, 17, 30-33). Capell (1962: 394) was of the opinion that Vanuatu was “probably .. inhabited before the coming of the Austronesians”, while Lynch (1981: 111, 119-120), Wurm (1982), and Tryon (1982: 245) have all raised the possibility of certain linguistic features being due to borrowing from now-extinct Papuan languages. 12

More recently, Blust (2005: 551-556) has drawn attention to a number of features that are found in parts of Vanuatu, especially the westernmost islands of Santo and Malakula (and, in some cases, New Caledonia) and in many Papuan languages, but not among Oceanic speakers elsewhere: (1) quinary13 (five-based) numeral systems, (2) ‘twenty’ being represented by ‘one man’ in some languages, e.g. Paamese14, and (3) extensive verb serialisation. He also pointed out a number of cultural traits that arguably reinforce the linguistic evidence: (1) penis sheaths (more correctly penis wrappers or

---

11 However François (p.c. July 2014) has now re-assessed the evidence and believes that Vanikoro languages may well be Oceanic, but are remarkably aberrant, showing many lexical and morphological innovations.

12 Lynch (p.c. July 2014) no longer believes that Papuan languages were ever spoken in Vanuatu or New Caledonia, largely because of the absence of archaeological evidence for pre-Lapita occupation.

13 Following Blust (2008: 446), I use ‘quinary’ here to mean any system that uses additive numerals for ‘six’ to ‘nine’, regardless of whether the word for ‘ten’ is monomorphic or not.

14 Lynch (2009: 395) lists Southeast Ambrym and Lenakel in Vanuatu and Nemi and Xaracuu in New Caledonia as languages expressing twenty as a phrase incorporating the word for ‘man’, ‘person’ or ‘fingers/toes/digits’.
nambas), (2) insertion of large ornaments into the pierced septum (these two features are elaborated on in Blust 2008: 453), and (3) physical resemblance: “In Malakula and Espiritu Santo, the prominent noses and full beards of many men are strikingly similar to features common among New Guinea highlanders” (Blust 2005: 554).

Another physical trait, not mentioned by Blust, is short stature. Like the negritos of the Philippines, who are believed to have originally spoken a non-Austronesian language or languages (Reid 1994), some of the inland inhabitants of Santo and Malakula are significantly shorter than average (Speiser 1991: 54-55). There is, as observed by the anthropologist John Layard (1942: 9), a “definite racial distinction between the coastal people .. all of whom share certain physical characteristics, and the small, almost pygmyoid inhabitants of the mountainous interior of South Malekula who represent the last remnants of an earlier racial stock similar to that found in the interior of certain other larger land-masses in the Western Pacific.”

Later, Blust (2008: 453) added other cultural traits: “wide girdles of rattan cane used by men in some highland New Guinea groups, and similarly wide waistbands of other materials among the Big Nambas of Malakula .. [and] the thick, mop-like headdresses of red fibers worn by Big Nambas women .. strikingly similar to those used by women in the eastern highlands of New Guinea.” Nevertheless, Blust (2008: 454) adds a note of caution: “cultural comparison of such traits is difficult because little information is available about them, even in works that are otherwise fairly thorough.”

Another pertinent cultural trait – again, not mentioned by Blust – is tooth avulsion (also known as tooth ablation), specifically the removal of the upper incisors, found in parts of Santo and Malakula as part of girls’ and women’s initiation ceremonies (Speiser 1990: 162, Muller 1972: 64-67). This practice appears to be otherwise absent in Oceania, the sole exception being parts of Hawai’i (Pietresewsky and Douglas 1993) where it was a sign of mourning, but relatively common in Australia, at least in parts of northern South Australia, Victoria and New South Wales, where it was part of boys’ maturity rites (Arthur and Morphy 2005: 96, Attenbrow 2010: 131, Horton 1994: 1086). Incidentally this custom may have given rise to the presence of labiolingual consonants replacing bilabials in south Santo and north Malakula (Lynch 2005: 403).

Donohue and Denham (2008: 435), encouraged by Blust’s proposals, have added grist to the Papuan mill, largely with phonological arguments. First they point out the distribution of labiovelars in Oceania: they are largely confined to the north coast of New Guinea, close to where presumably they were first added to the phonemic inventory of Proto-Oceanic, and Southeast Solomons, Vanuatu and New Caledonia. Elsewhere in Oceania they have largely disappeared. They suggest that “the labiovelars found greater support in Vanuatu because of a stronger non-Austronesian substrate.” Similarly, rounded labial fricatives, while not reconstructed for Proto-Oceanic, are found only along the north New Guinea coast and in Vanuatu.

Donohue and Denham further point out (2008: 438) that the lack of /p/ in phoneme inventories is relatively rare, occurring in the Pacific only in a small group of languages in Papua New Guinea and in Proto North-Central Vanuatu and Fiji. Of the 360 Oceanic languages the authors surveyed, only 11 percent lack /p/, while in Vanuatu the figure is 27 percent. Furthermore, languages lacking both /p/ and /c/ (voiceless palatal stop) are only found among Papuan languages of Papua New Guinea and in North-Central Vanuatu. They conclude that the “appearance of a double gap in older Vanuatu is an

\[15\text{ See also Deacon (1934: 8).}\]
enigma, unless we posit an earlier Papuan substrate in the area that influenced the development of the Austronesian languages that later arrived in Vanuatu.”

Another new strand of evidence introduced by Donohue and Denham (2008: 440) is genetic. They cite a number of recent works that show that “the people of Vanuatu share most molecular markers with the Melanesians of northeast Papua New Guinea, and not with the peoples of insular Southeast Asia or Polynesia.”

Even though such arguments have been put forth for the former existence of linguistic areas in Eastern Melanesia, as in present-day Western Melanesia, involving Papuan and Austronesian languages, proponents of this theory accept that there is no archaeological evidence for any human occupation prior to the arrival of the Lapita culture approximately three thousand years ago. As reported by Blust (2008: 454), Matthew Spriggs, a leading archaeologist of the region, has stated that, with regard to Vanuatu, “archaeologically there is no evidence of pre-Lapita occupation despite a targeted campaign to look for such over the last 14 years.” So if the speakers of Papuan languages were not in Eastern Melanesia before the arrival of the Lapita culture, whose bearers were supposedly Austronesian speakers, what scenario can best explain the facts?

Donohue and Denham (2008: 440-441, 2012) argue that the bearers of Lapita pottery and its associated culture were not necessarily, as has been commonly assumed, only speakers of Proto-Oceanic: “We suggest that it is just as likely that Papuan-speaking Melanesians, occupying the islands out to the end of the main Solomons chain for at least 30,000 years and having settled Manus by 20,000 years ago .. would have participated in the maritime expansion that accompanied the spread of Lapita.” They then offer two possible historical scenarios that are compatible with the linguistic and culture-historical evidence. First, that the first settlers in Vanuatu were “people whose ancestors had been resident in Melanesia for many millennia, and who adopted various aspects of a new immigrant culture, including an Austronesian language [which] exhibited a number of structural features inherited from pre-Austronesian language(s) in New Guinea, and these features were borne, more-or-less undiluted, to Vanuatu.” This is essentially the theory of Speiser (1946, cited in Schmitz 1962: 418).

The second scenario offered by Donohue and Denham (2008: 441) is that “the settlers were Papuan-speaking peoples who arrived with Lapita pottery (and other technologies), but without an Austronesian language.. who were later swamped by the expansion of Austronesian-language speaking groups and their languages. .. they were relatively quickly submerged in the developing Austronesian linguistic milieu, though not without leaving traces of their passing in the phonology and semantic organization of the languages .. as well as preserving many aspects of their original ethnic identities.” They point out that this latter scenario is almost identical to the theory of Codrington (1885: 32-35) and deem it the more likely of the two because “the ‘Papuan’ traits that are found are not uniformly represented in the languages of Vanuatu. They are attested among other languages with a more typical Austronesian profile, suggesting either a linguistically complex initial settlement or a complex post-settlement linguistic ecology (or both).”

Donohue and Denham’s hypothesis also includes (2008: 441-442) an explanation for some features in which Southern Vanuatu differs from North-Central Vanuatu, claiming that “the appearance of echo subject agreement mirrors the switch reference systems that proliferate in mainland New Guinea, especially in the east” and that verb-initial clause order is found “only in the far south, indicating that a later spread of SVO order was less successful in the south than in the rest of Vanuatu.”
An opposite stance is taken by Blust (2008: 454-456), following Spriggs (1997: 158-159), offering a “linguistic Melanesianisation” theory, i.e., that the original inhabitants of Eastern Melanesia were indeed Proto-Oceanic speakers, of what we would now judge to be Polynesian appearance, and that a second, more numerous, wave of migrants speaking Papuan (or Papuan-influenced Austronesian) languages followed soon after, most likely spread over a long period (the “trickle in” model), since they appear to have had no effect on material culture. However, Blust agrees with Donohue and Denham that, if the archaeological evidence is put to one side, a more likely scenario is one in which the “‘melanesianisation’ of Remote Melanesia was carried out by a population of largely PM [Papuan] physical type that had acquired certain aspects of material culture, including the outrigger canoe complex and pottery, from Proto-Oceanic speakers in Near [West] Melanesia, but that still spoke Papuan languages and retained some distinctively Papuan cultural traits when they arrived in Vanuatu.”

Regarding lexical replacement in Vanuatu, it was noted above with respect to the Solomon Islands that Pawley (2009) demonstrated that all Western Solomonic languages in a sample he surveyed had a higher rate of lexical replacement that any Eastern Solomonic language in the survey; from this he inferred that the speakers of Papuan languages were far more numerous in the west than in the east during the period of eastward expansion of Austronesian languages. In the same study, Pawley (2009: 531-532) extended the survey to a small number of languages beyond the Solomons to demonstrate that they “probably had no direct contact with non-Austronesian languages after these islands were settled.” However he reported that in Erromangan, a language of South Vanuatu, the retention rate on the same list was 61%, only 2% higher than that of Roviana, a western Solomonic language, and lower than the lowest Eastern Solomonic language (65%). By this reasoning, then, the extent of lexical replacement in Erromangan could well indicate that it was in contact with a Papuan language. Note also that no language of Reefs Santa Cruz, Santo, Malakula, or the Loyalties or New Caledonia was included in this survey. Had some of these been included, their retention rates would certainly have been found to be as low as, if not lower than, those of the western Solomonic languages.

7 New Caledonia

The languages of New Caledonia and the Loyalty Islands have also been viewed as difficult to classify in the past (Codrington 1885: 16, Ray 1926: 76, Capell 1962: 377), but the recent increase in the quantity and quality of descriptions has now persuaded most linguists that, as with Vanuatu, these languages are essentially Oceanic (Geraghty 1989) and archaeologists have failed so far to find convincing evidence of a pre-Lapita population. However, Lynch (1981: 119-120) suggested the possibility of a Papuan substrate (but see footnote 11 above). Blust (2005: 551-556) then pointed to the following features (some also found in Vanuatu) which also suggest a Papuan substrate: (1) quinary numeral systems, (2) twenty being formed by ‘one man’ in a number of languages in New Caledonia and the Loyalties (Iaai, Xaracuu, Paici, Nyelayu, Canala, Nengone etc), and (3) some verb serialisation. Blust (2008: 456) also noted – as have many before him – the physical resemblance of some New Caledonians to aboriginal Australians. To this
may be added a cultural trait not mentioned by Blust for New Caledonia, the use of penis sheaths or wrappers.16

The current standard account of the prehistory of New Caledonia (e.g. Sand 1996: 46, Kirch 1997: 72-73) holds that there is no evidence of any human occupation prior to the arrival of the ‘Lapita people’, making Lapita pottery and speaking Oceanic languages, some three thousand years ago. Yet up until the 1980s, many authorities (e.g. Howells 1973: 35, 168, Dubois 1976: 20, 1981: 5, 1982: 13) agreed that there was evidence for a pre-Lapita population from various disciplines, including oral traditions, material culture, and physical anthropology. Also pertinent is the fact that it would have been possible 20,000 years ago to walk from Australia to New Caledonia with no sea crossing greater than 30km, raising the possibility that in the case of New Caledonia the substrate language may have been not Papuan but Australian.

Given the extreme difficulty of discovering anything about an Australian aboriginal language that may have been spoken in New Caledonia some 20,000 years ago, Geraghty and Nunn (2013) have proposed an innovative method of demonstrating, at least, that a substrate language may have existed. They argue that there is linguistic evidence for a non-Oceanic substratum, particularly in the names given to novel flora and fauna – that is, natural species that were found in New Caledonia but had not been present in the Vanuatu homeland of the Lapita colonisers. The evidence does not take the form of identifying borrowings from Australian languages since, given that the presumed donor languages are now long extinct, and in any case of uncertain relationship to contemporary Australian languages, such would be difficult to find. However, given that the usual response of Pacific colonisers of uninhabited lands was to create names for newly-discovered flora and fauna by semantic expansion and compounding, as has been demonstrated for New Zealand Maori by Biggs (1991: 67-68), the fact that most of the reconstructed names for novel natural species in New Caledonia appear not to have been coined using these methods suggests that there was a now vanished language that served as a source of these new terms.17

Unusually for Oceanic languages, those of New Caledonia and the Loyalty Islands are largely ergative. Ball (2007: 141) has argued that the “New Caledonian languages likely developed ergativity through some shared process.. The development of ergativity in New Caledonia was also likely not a direct inheritance from Proto Oceanic .. the Loyalty Island languages, too, suggest that there is not a single Proto Oceanic source of ergativity, because the distribution of and the form for marking ergative NPs does not match anything else in the Oceanic family.” Since ergativity is relatively common in both Papuan (Foley 1986) and Australian (Dixon 1980) languages, either of these could be sources of the unexpected ergativity of New Caledonia and the Loyalties.

As in Vanuatu, certain non-linguistic features of New Caledonia and the Loyalty Islands can be adduced to suggest the possibility of a non-Austronesian, in some cases specifically Australian, substrate. These include: oral traditions of a very different earlier population with no agriculture, chiefs, houses or mats (Dubois 1976: 20); presence of

---

16 Captain Cook noted in 1774 that they “had no other covering than a little case to the Penis which was suffered to hang down” (Beaglehole 1969: 530).
17 Note that this argument cannot be applied to Vanuatu, since the colonisers would have been familiar with most of the natural species there from their earlier home in the Solomon Islands. It is only in New Caledonia that many new species would have been encountered, some with close relatives in Australia.
spear-throwers (Labillardière 1800: 424) and boomerangs (Dubois 1976, 1981: 10); and, as also pointed out by Blust and others, physical appearance (Labillardière 1800: 388, Howells 1973: 35, 168).

Finally, the presence in the Nengone language spoken on Mare in the Loyalty Islands of the typologically unusual change of bilabials to apicals (*b > d, *m > n), presumably by way of apicolabials (see Lynch 2005 for discussion of a similar phenomenon in Vanuatu), is possibly evidence of the former practice of tooth avulsion, which would provide a physiological motivation for such a marked change. As noted above, tooth avulsion was widely practised in Australian aboriginal society but practically non-existent in Oceania.

In the following sections, I will focus on certain linguistic features that may indicate Papuan substrate influence on Oceanic languages of the Reef Santa Cruz Islands, Vanuatu and New Caledonia: serial verb constructions and quinary numeral systems (as proposed by Blust), then toponyms and pot-and-handle borrowings.

8 Serial verb constructions

Extensive verb serialisation is one of the typological features found in Eastern Melanesia that Blust (2005: 553) attributed to a Papuan substrate. This attracted a riposte from Pawley (2006: 247) to the effect that SVCs (serial verb constructions) are not particularly rare in Austronesian languages outside of Oceania, and indeed must be reconstructed for Proto-Oceanic (Ross 2004). He added that the kinds of SVCs found in Vanuatu and southern Melanesia are structurally unlike those found in the Papuan languages of New Guinea.

Blust (2008: 447) responds to this, first, with a novel method that I will call ‘bibliostatistics’. He presents a statistical analysis of available grammars of a sample of Austronesian languages to determine what percentage of the text is devoted to verb serialisation. Among the grammars of Austronesian languages in Melanesia, the percentage ranges from 2.4% to 12.1%. Among the grammars of Austronesian languages outside Melanesia, 11 of the 14 selected have no discussion at all of verb serialisation – the inference being that it does not exist – while the three that do discuss verb serialisation are of languages, such as Tetun of East Timor, which have long been in contact with Papuan languages or, as in the case of Buru, where verb serialisation appears to be of minor importance. Whilst there are scattered languages in Western Austronesia that show verb serialisation to varying degrees, it is totally absent in the Philippines, Borneo, Sumatra, Sulawesi and most of Western Indonesia (Blust 2008: 449). Turning to Oceanic languages, Blust (2008: 448-449) points out that verb serialisation is also absent in Micronesia, and further argues that the few recent descriptions of verb serialisation in Polynesian languages are highly contrived, due to the fact that verb serialisation was a linguistic fad of the 1990s, and such analyses can only be viewed as valid by “diluting the definition of verb serialization to the point that it ceases to be distinctive.”

In response to Pawley’s claim that verb serialisation must be reconstructed for Proto-Oceanic, Blust (2008: 449-450) points out that such reasoning is invalid, since the currently accepted sub-grouping of Oceanic is such that a feature can be reconstructed for POc if it is found only in certain languages of western Melanesia – precisely the area where Papuan languages are found, or are believed to have formerly existed. Blust concludes that the data suggest strongly that “the wide distribution of serial verbs in the
AN languages of Melanesia is a product of recurrent acquisition after the breakup of POc.”

Finally, Blust rejects Pawley’s assertion that “the types of SVC found in Vanuatu and Southern Melanesia are structurally unlike those found in the Papuan languages of New Guinea”, countering that there is no such thing as a typically Papuan serial verb construction, so without more detailed study it is premature to conclude that Papuan and Melanesian SVCs are different. He sums up (Blust 2008: 450) by making the simple observation that published grammars tell us that verb serialisation is very common in most Papuan languages and many Austronesian languages of Melanesia, but absent in most other Austronesian languages.

9 Quinary numeral systems

The historical implications of the presence of quinary numeral systems in parts of Melanesia has been commented on since at least the ground-breaking survey of Codrington (1885: 220-251). He concluded that “the oldest method is the quinary, and it is pretty certain that the decimal notation in Melanesia is comparatively recent there and introduced” (Codrington 1885: 222, see also 229-230, 246-247). This view has been challenged by Lynch (2009) who considers the quinary systems ‘innovative’. These differing interpretations result from Codrington’s view that the earliest languages of Vanuatu were not Austronesian, as opposed to Lynch’s view that they were.

Blust (2005: 552-554) also views the quinary systems as innovative, but in the sense that they are an innovation in Oceanic languages. He argues that only a decimal system has been reconstructed for Proto-Oceanic, therefore the fact that quinary systems are found in many Oceanic languages in New Guinea, much of Vanuatu and New Caledonia and the Loyalty Islands, but nowhere else in Oceania, requires an explanation, and a plausible explanation is that the quinary systems of Eastern Melanesia were acquired from Papuan languages spoken there at an early date when all of Melanesia, according to Blust, was a chain of linguistic areas.

Pawley (2006: 247) countered that Papuan languages typically have ‘one, two, many’ systems, or systems based on body parts, and offered two alternative explanations: first, that quinary systems existed in POc along with decimal systems; and second, that quinary systems spread into parts of Vanuatu and southern Melanesia some time after Lapita settlement of the region. This echoes in part the proposal of Lynch, Ross and Crowley (2002: 72) who reconstructed a decimal system for Proto-Oceanic, but commented, contrary to Codrington: “to judge from today’s traditional cultures, numerals above ten were not much used in parts of early Oceania. Indeed, so widespread are quinary systems (but with the term for 10 preserved) that one suspects that the numerals 6-9 were dropping out of use among some early Oceanic speakers.”

Blust (2008: 450) does not comment on Pawley’s objection regarding typical Papuan counting systems, though he could have pointed out that none of the Papuan languages of the Western Solomons has either of the systems Pawley mentions (Tryon and Hackman 1983, Terrill 2011: 318)18. He rejects the first proposed explanation by

18 Lynch (2009: 406) cites this personal communication from Michael Dunn: “All the Solomons Papuan languages have productive decimal systems, but there’s internal evidence .. that they developed out of a quinary system.”
stating that no quinary system has been reconstructed for Proto-Oceanic, the inference being that it is impossible to do so, the existing quinary systems being so morphologically and formally disparate. Regarding the second explanation, Blust responds that such a spread of counting systems is simply implausible. He then presents a synopsis of the quinary systems of Vanuatu by subgroup (Blust 2008: 451-452), demonstrating that, under the current subgrouping hypothesis, the data require us to posit at least fourteen historically distinct replacements of decimal by quinary systems in Vanuatu – something which has happened only twice in the history of all Austronesian languages outside Melanesia.

Lynch (2009) has also provided a response to Blust, in the form of a detailed study of how the current quinary systems of Vanuatu and New Caledonia can be derived by regular processes from previous decimal systems. Blust (2008: 452) accepts the mechanisms proposed by Lynch, but points out that the question of actuation remains, since Lynch’s explanation, as indeed Lynch (2009: 408) acknowledges, “does not broach the question why the languages of this region have such atypically high rates of structural innovation in the numeral system as compared to Austronesian languages generally.”

10 Toponyms

Given that toponyms are an area of the lexicon in which traces of substrate languages are often found in post-linguistic areas – such as the Celtic Avon as a river name in England, or many Gaulish names of towns in France – it is surprising that no-one has ever, to my knowledge, attempted a serious study of the origin of place-names in Melanesia. Codrington (1885 passim, 1891: 4-8) attempted to identify the indigenous names of the main islands of central Melanesia – cautioning that “[l]arge islands seldom have a name” – but was not concerned with their etymologies. Ray (1926 passim) similarly listed place-names that had been recorded for the islands whose languages he studied, again largely without comment, and many dictionaries of Melanesian languages include at least some place-names. Here, then, is an area where it might be possible to associate an island name with a cognate in a Papuan language, though the probability that the source language has become extinct does indeed reduce the likelihood of identifying a cognate. Obvious names to start with are Gera (Guadalcanal), Mala (Malaita), Ulawa, Ndeni (Santa Cruz), Mota, Malakula, Maewo, Raga, Oba, Efate, Erromanga and Anejom/Aneityum, since all of these appear to be genuine indigenous names, and none has an obvious Oceanic etymology. Here again, the contrast with Micronesia and the Central Pacific is striking: in these areas, with their relatively shallow linguistic history and perhaps lack of substrate languages, many island and archipelago names have obvious etymologies, such as Viti (Fiji) and Tahiti both meaning ‘sunrise, east’, Tonga ‘south’, Tokelau ‘north’, Futuna ‘land of Barringtonia trees’, Niua ‘land of coconuts’ (see Geraghty 2001 for more examples).

11 Pot-and-handle loanwords

To fill what is, I believe, a lexical gap in English, I will use the term ‘pot-and-handle’ loanword to refer to a loanword that is borrowed together with a minor morpheme with which it typically occurs, and reanalysed in the recipient language as a single morpheme.
In the spirit of Blust’s survey of serial verb constructions in Oceanic grammars, as mentioned above, I now offer some bibliostatistics. Below is a sample of dictionaries of typical exemplary Oceanic languages indicating the percentage of the dictionary containing words beginning with /a/, /e/, /o/ and /n/:

Table 1. Oceanic languages with words beginning in /a/, /e/, /o/ and /n/

<table>
<thead>
<tr>
<th>Language</th>
<th>Source</th>
<th>Total pages</th>
<th>a/e/o</th>
<th>%</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuvaluan</td>
<td>Jackson 2001</td>
<td>278</td>
<td>16/2/5</td>
<td>6/1/2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Tongan</td>
<td>Churchward 1959</td>
<td>574</td>
<td>15/2/4</td>
<td>3/1/1</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Fijian</td>
<td>Capell 1941</td>
<td>342</td>
<td>1/2/3</td>
<td>0/1/1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Nggela</td>
<td>Fox 1955</td>
<td>265</td>
<td>9/1/2</td>
<td>3/1/1</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Now compare these figures from dictionaries of some aberrant languages of Southern Vanuatu.

Table 2. Southern Vanuatu languages with words beginning in /a/, /e/, /o/ and /n/

<table>
<thead>
<tr>
<th>Language</th>
<th>Source</th>
<th>Total pages</th>
<th>a/e/o</th>
<th>%</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erromanga</td>
<td>Crowley 2000</td>
<td>155</td>
<td>1/1/3</td>
<td>1/1/2</td>
<td>84</td>
<td>54</td>
</tr>
<tr>
<td>Kwamera</td>
<td>Lindstrom 1986</td>
<td>137</td>
<td>26/5/5</td>
<td>20/4/4</td>
<td>56</td>
<td>4</td>
</tr>
<tr>
<td>Lenakel</td>
<td>Lynch 1977</td>
<td>104</td>
<td>21/5/5</td>
<td>20/5/5</td>
<td>56</td>
<td>4</td>
</tr>
<tr>
<td>Anejom</td>
<td>Lynch and Tepahae 2001</td>
<td>279</td>
<td>65/25/5</td>
<td>24/9/2</td>
<td>87</td>
<td>31</td>
</tr>
</tbody>
</table>

It is clear from the above that exemplary languages have a very small proportion of words beginning with ‘n’, but some aberrant languages have considerably more, in the case of Erromanga over half. I do not have access to a dictionary of a Reefs Santa Cruz language, but a perusal of entries in Tryon and Hackman (1983) suggests that they also have a large number of words beginning with ‘n’. It is also clear that words beginning with a non-high vowel are relatively common in Anejom, Kwamera and Lenakel, but not in exemplary languages.

The reason for the large number of words with initial ‘n’ is that common nouns in these languages have undergone a process of article fusion (or accretion), with the common article na becoming fused to the following noun. This process is described for Southern Vanuatu by Lynch (2001: 109-114) but, as with the change to quinary numeral systems, no motivation for the change is offered. Other languages of Vanuatu that appear to have undergone article fusion, at least to some extent, are found in Efate, Nguna, Paama, Malakula, Santo and the Banks Islands (for the last, see François 2007: 315, 317-318). In Reefs Santa Cruz, languages of Utupua and Vanikolo also show some article fusion. On the basis of a cursory survey, there appears to be an implicational relationship between the phenomena of quinary systems and article fusion, in that all areas that show article fusion also show quinary systems, though not vice versa.

A possible motivation for article fusion is found in borrowing. When nouns are borrowed, it is not uncommon for them to be borrowed with the article they are typically accompanied by. Examples in the Pacific include Bislama from French (e.g. lafet ‘celebration’ < la fête, loto ‘truck’ < l’auto, lae ‘garlic’ < l’ail) (Crowley 1990), Fijian toa ‘Banaban canoe’ < Rabe Banaban te wā, Fiji Hindi from Fijian (e.g. nakai ‘freshwater mussel’ < na kai) and some borrowings from Polynesian languages in
Rotuman (\textit{tarau} ‘a hundred’ < \textit{ta lau}) (Churchward 1940), Pohnpeian (\textit{sakau} ‘kava’ < \textit{ta kawa}) (Geraghty 1994: 244), and Reefs Santa Cruz (e.g. Reefs \textit{toponu} ‘turtle’ < \textit{te fonu}, Santa Cruz \textit{te kutu} ‘louse’ < \textit{te kutu}) (Ray 1926: 449-451, 455-456, 462, Tryon and Hackman 1983, Tryon 1994). Of course, the fact that such article fusion is common in borrowing does not necessarily mean that this must have been the cause in the languages of Reefs Santa Cruz and Vanuatu in which it occurs, but it is perhaps significant that in Oceanic languages spoken where there is little likelihood of there having been any non-Austronesian languages – those of Micronesia, Rotuma, Fiji and Polynesia – there is no evidence for article fusion.

A similar phenomenon may underlie the preponderance of verbs beginning with non-high vowels in certain languages of Eastern Melanesia, illustrated above by Kwamera and Lenakel of Tanna. While not so common as article accretion, verbs can also be borrowed with a typically accompanying particle, as for example in Pidgin Fiji Hindi, where verbs are consistently borrowed from Fiji Hindi with the suffix –\textit{o}, which marks the imperative, e.g. Pidgin Fiji Hindi \textit{baito} ‘sit, existential, copula’ < Fiji Hindi \textit{baith-o} ‘sit!’ (imperative). Similarly, verbs in Pidgin Fijian recorded from the early nineteenth century often show fusion with the preposed aspect particle \textit{sā}, e.g. Pidgin Fijian \textit{salako} ‘go’ < Fijian \textit{sā lako} ‘has gone’ (Geraghty 1977: 59-60). Lynch (2001: 144-145) argues that the accreted non-high vowel in languages of Southern Vanuatu is the result of a process of denominalisation of verbs that had earlier been nominalised and undergone article accretion parallel to the same process in nouns. Again there is no proposed motivation for this very unusual change. I would tentatively propose, without having explored the phonological evidence, that the source of the non-high vowel accreted to verbs in Vanuatu might have been \textit{*e}, the third person singular subject marker that can be reconstructed for Proto-Oceanic (Lynch, Ross and Crowley 2002: 68) and Proto Eastern Oceanic.

My proposal, then, is that pot-and-handle borrowings provide evidence that early Oceanic languages were the lexifiers of contact languages that developed in linguistic areas in early Vanuatu (and elsewhere), with these nouns and verbs being borrowed en masse in the formation of new languages. These languages now appear to be undeniably Oceanic in their lexicon (and in some cases in other aspects, due to subsequent diffusion from more exclusively Oceanic languages) but in reality are of mixed origin, in much the same way as Melanesian pidgins appear to be Germanic in their lexicon but remain, arguably, Oceanic in their structure. This accords with the proposal of Donohue and Denham (2008: 441), cited above, that “the settlers were Papuan-speaking peoples who arrived with Lapita pottery (and other technologies), but without an Austronesian language. .. who were later swamped by the expansion of Austronesian-language speaking groups and their languages. .. relatively quickly submerged in the developing Austronesian linguistic milieu, though not without leaving traces of their passing in the phonology and semantic organization of the languages .. as well as preserving many aspects of their original ethnic identities.” This in turn echoes the rather less prosaic analogy of Codrington (1885: 33): “We may conceive of the peopling of Melanesia and the settlement of its languages as of the filling with the rising tide of one of the island reefs. It is not a single simultaneous advance of the flowing tide upon an open beach, but it comes in gradually and circuitously by sinuous channels and unseen passages among the coral, filling up one pool while another neighbouring one is dry, apparently running out and ebbing here and there while generally rising.”.
12 Polynesian outliers in Melanesia

The Polynesian outliers in Melanesia present a relatively recent case of the establishment of linguistic areas in parts of Melanesia (Clark 1986, 1994). They are colonies – 11 in Papua New Guinea and the Solomons, 4 in Vanuatu and 1 in New Caledonia - set up by Polynesians on small islands hundreds of years ago, and the resulting physical, cultural and linguistic diffusion has been largely a function of distance from the nearest Melanesian settlements. Codrington (1885: 8) noted this with regard to physical appearance, while Clark (1994: 113) similarly classed the outliers linguistically into two groups, those that have experienced only ‘cultural borrowing’ versus ‘intimate borrowing’, again largely a function of their relative proximity to Melanesian neighbours: cultural borrowers are remote (Takuu, Luangiua, Sikaiana, Tikopia, Anuta), while intimate borrowers are close to Melanesian communities (Rennell, Pileni and all those in Vanuatu and New Caledonia). Among the grammatical changes listed by Clark (1994: 119) as occurring in Polynesian languages as a result of this contact is the change to strict SVO order in Mele-Fila, an echo perhaps of similar changes in more remote times in western Melanesia.

Clark has also observed that the extent of borrowing is not reciprocal, that is, while there are Polynesian outlier languages that have borrowed extensively from their Melanesian neighbours, all borrowing by Melanesian languages has been of the limited ‘cultural’ variety – similar to the situation in the western Solomons observed by Terrill (2011: 313) where long-term contact does not necessarily result in heavy borrowing. Clark (1994: 120-121) proposes a historical explanation based on the small numbers of founding settlers of outliers necessitating the taking of wives from Melanesian communities.

Remarkably, there is at least one relatively clear example of a post-linguistic area among the outliers: as Blust (1987) has demonstrated, the Melanesian language, which was the source of two new phonemes and extensive vocabulary in Rennellese, cannot be any of the extant languages of the Solomons, and is most plausibly explained as a Solomonese language that was spoken by an aboriginal population on Rennell, but is now extinct, an explanation that is supported by Rennellese oral traditions. This, then, is arguably a relatively modern example of the kind of language contact event that occurred three thousand years ago in the Solomon Islands, and perhaps elsewhere in Melanesia, in which an aboriginal language and an immigrant language functioned as a linguistic area, followed by the extinction of the aboriginal language, but not before the immigrant language had acquired many features of the aboriginal language.

Finally, the argument that the change from a decimal to a quinary system of numbers could be the result of calquing in a linguistic area is strengthened by evidence that precisely this has happened in Fagauvea, the sole outlier language of New Caledonia, which has abandoned the decimal system of Proto-Polynesian to adopt a quinary system modelled on that of its Melanesian neighbour Iaai (Clark 1994: 117-118).

---

19 There are also two Polynesian outliers in Micronesia, Kapingamarangi and Nukuoro.
20 It is of course possible that other Polynesian outlier communities existed in the past but have shifted to a Melanesian language.
13 Linguistic ‘necessity’ and historical reality in post-linguistic areas

Ross and Næss (2007: 456) refute the notion of a non-Austronesian speaking population having lived in the Reefs Santa Cruz group by claiming that “there is no need to posit a Papuan element” to understand the origin of the Reefs Santa Cruz languages. Indeed, but it is equally true that there is no need to posit a Celtic element to understand the history of the languages now spoken in England – but that does not alter the historical fact that Celtic languages were spoken throughout England two thousand years ago, and have since become extinct.21

Of course in Eastern Melanesia we do not have the written records that would permit us to identify what languages, if any, were spoken there before the arrival of Austronesian speakers, but the arguments proposed by Blust, and others presented in this chapter, do offer some clues, and by not burying our heads in the golden sands of Melanesia, and by expanding our search, as Blust and others have done to some extent, to cultural and other non-linguistic traits, I believe we may arrive at a more correct account of the prehistory of these Melanesian islands.

14 Conclusion

In conclusion, I return to the question I posed above: how can one demonstrate that an area is a post-linguistic area – i.e. that substrate languages once existed there and that features were transferred to the surviving languages – when the donor languages were never recorded and are all now extinct? Clearly it is impossible to prove conclusively, and it is possible that the transfer of some features occurred in Western Melanesia before the settlement of Eastern Melanesia, but I tentatively propose the following as indicators of post-linguistic areas in Eastern Melanesia, and propose that the more of these features that are shown by a language now classed as Oceanic, the more likely it is that it was influenced by a Papuan (or Australian) language.

Table 3. Indicators of post-linguistic areas in Eastern Melanesia

<table>
<thead>
<tr>
<th>Indeterminate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-linguistic: located on large islands, large ornaments in pierced septum, oral traditions of earlier population</td>
<td></td>
</tr>
<tr>
<td>Linguistic: aberrant with respect to Proto-Oceanic, lexical replacement, ergativity, pot-and-handle borrowings (article fusion, accretion of verbal morphology)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Papuan</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-linguistic: physical resemblance, short stature, genetic markers, penis wrappers, wide waistbands worn by men, red fibre headdresses worn by women</td>
<td></td>
</tr>
<tr>
<td>Linguistic: echo vowels, labiovelars, rounded labial fricatives, lack of /p/, lack of /c/; ‘quinary’ numeral systems, ‘twenty’ expressed by one</td>
<td></td>
</tr>
</tbody>
</table>

---

21 This is not to deny that certain features may be best understood with reference to a Celtic substrate, see Hickey (2012).
man/person’, complex morphology (especially agglutinative verb phrases), certain nominalising affixes, extensive verb serialisation, verb-initial clause order, echo subject agreement

**Australian**

Non-linguistic: physical resemblance, tooth avulsion in initiation rites, spear-throwers, boomerangs

Linguistic: bilabial > labiolingual (> dental) (resulting from tooth avulsion)

**References**


Siegel, Jeff. Forthcoming. Contact-induced grammatical change in Melanesia: who were the agents of change? To appear in Australian Journal of Linguistics.


29 The Western Micronesian Sprachbund

Anthony P. Grant

1 Introduction

Thomason (2000) has written about the difficulties involved in proving that constellations of shared features, which are found among a number of languages spoken in the same region, means that the languages in question constitute a linguistic area. Such heuristic problems will be discussed here in regard to the potential Sprachbund status of Western Micronesia in the Northwest Pacific, an area which has been characterised by extensive and intensive networks of language contact.

The languages involved here are Trukic or Chuukic, especially Western Trukic varieties (Sonsorolese-Tobian-Pulo Annian, Ulithian, Woleaian, and varieties of Saipan Carolinian, a migrant language from the Central Trukic belt; Puluwatese, Mortlockese and Lagoon Trukese or Chuukese are also Trukic languages, while the extinct language of Mapia is the westernmost Trukic language of all), Yapese of the island of Yap, now part of the Federated States of Micronesia, Palauan of Palau (now The Republic of Belau), and – peripherally – Chamorro of the Marianas. Mapian, a Trukic variety, was spoken a little north of Irian Jaya (see Map 1 for locations in Micronesia). Like Sonsorolese-Tobian and Pulo Annian it was spoken considerably to the west even of Palauan and Yapese, so that the spread of Trukic languages is the greatest of that of all the languages discussed in this paper. The westernmost and easternmost languages of Western Micronesia that are discussed here, Chuukese and Sonsorolese, are both Trukic.

Among the Trukic languages Sonsorolese-Tobian, the westernmost one which is spoken several hundred kilometres away from the epicentre of Trukic languages, is especially archaic at the phonological level (such that Sonsorolese phonological forms are closer to the phonological shapes of cognate stems in other Nuclear Micronesian languages than other Trukic forms are). Like Woleaian (though more comprehensively) it preserves on stems the word-final voiceless vowels which manifest themselves only in inflected forms in other Trukic languages. Either Sonsorolese-Tobian is phonologically conservative because it has moved little from the first point of settlement of speakers of Trukic languages, or because, surrounded by speakers of Palauan (and Yapese), it has remained isolated from waves of phonological change which have spread throughout the rest of the Trukic continuum.

All of these languages belong to the Malayo-Polynesian branch of Austronesian, so that they are ultimately related to one another, though at considerable time-depths. Most of them are Oceanic (Trukic participates in Nuclear Micronesian, Yapese is possibly part

---

1 I would like to thank the members of the NWCL, as well as Bob Blust, Francesco Goglia, the late Thomas Klein, Laurie Reid, Malcolm Ross and Sakiyama Osamu for various forms of help with this paper.

2 Evidence for a Tobian-based pidgin from the 1830s is presented in Holden (1836) and discussed in Grant (2014).
of the Admiralties group of languages, originally from northeastern Papua New Guinea, and which constitute a primordial division of Oceanic; indeed Ross 1996 suggests that these languages were probably the first languages to break away from the rest of Oceanic). Proto-Micronesian has been reconstructed by Bender et al (2003) with due attention to loanwords and other contact features. But Palauan and Chamorro belong to the Western Malayo-Polynesian ‘anti-group’ (that is, the residuum of languages which is left after the Central-Eastern Malayo-Polynesian group, which can be justified by the adduction of shared innovations, is separated off). They are not closely related to one another and have not successfully been subclassified further beyond the level of separate branches of Malayo-Polynesian (Blust 2000; the claims in Zobel 2002 that they belong, on the grounds of possessing shared morphological innovations, with many other languages in a ‘Nuclear-Malayo-Polynesian’ group, which is one rung down from Proto-Malayo-Polynesian in the Austronesian family tree, have been disproved in Reid (2000).

Map 1. Micronesia area of the Western Pacific

3 In more recent work, for instance in Crowley, Lynch and Ross (eds. 2002), Yapese is represented as the first division of Oceanic against all the other Oceanic languages, which are then divided into Admiralties languages versus all the others.

4 Map by courtesy of Holger Behr, licensed under Public Domain via Wikimedia Commons.
Thus two of the four groupings in this Sprachbund are near-‘isolates’ (in relative terms, naturally) and Yapese is an Admiralties offshoot which is spoken several hundred miles away from (and which is completely out of contact with) its nearest genetic relatives and without close relatives in Micronesia. Only the Trukic languages are part of a linguistic group which has other members that are spoken in Micronesia. ‘Western Micronesian’ is therefore a geographical term. (For the record, the languages of Eastern Micronesia – the other Trukic languages, the Ponapean languages, Kosraean, Marshallese, Kiribatese – are mostly part of the deeply-ramified Nuclear Micronesian subgroup of Oceanic, although Nukuoro and Kapingamarangi, spoken to the south of the eastern Trukic languages, are Polynesian outliers and it appears that Nauruan is coordinately related to the Nuclear Micronesian languages as their closest genetic relative.)

The part of Micronesia where Nuclear Micronesian languages are used is usually assumed to have been settled from east to west, with Kosrae as an early landfall (cf. Jackson 1983), though the centre of greatest linguistic diversity among Nuclear Micronesian languages is in the southeast, and according to Ross (1996) lexical evidence from the reconstruction of Nuclear Micronesian, naming geological and similar items, suggests that the proto-language was spoken on a high island such as Kosrae rather than an atoll such as Chuuk. Ross (1996) and Blust (2000) each suggest that the period of continued settlement of Nuclear Micronesia and the Marianas respectively took place between 3000 and 3500 years before the present, and Blust (personal communication to Anthony Grant, April 2002) has suggested that Palau may have been settled some millennia ago from the Talaud Islands off Sulawesi, where Sangiric languages (part of the Greater Central Philippine group of Austronesian languages) are now spoken, but that this does not imply that Palauan is Sangiric. We are less sure when Yap was first settled, but the overall picture is that most of the linguistic components of this potential Sprachbund were in place a few millennia ago.

2 Some intense ‘contact languages’ in Western Micronesia: contact-induced change in Nguluwan, Carolinian and Chamorro

Some languages in this area have been especially strongly shaped by the effects of language contact or contact-induced linguistic change, and as a result they seem at first hand to be ideal places to look for the bundling together of areal linguistic features.

There is, first of all, the question of Nguluwan (Sakiyama 1982), a language which is spoken by a few score people living on Ngulu(w), an island included in Yap. This is a variety of Yapese used by a population which is (or was recently) bilingual in Ulithian (some older Nguluwanese apparently still know Ulithian but younger ones do not). It contains the same lexical elements from Palauan which Yapese contains (in addition to a few others unique to Palauan), but in addition to the Western Trukic loans which are found in Yapese, it has an extra layer of both basic and non-basic loans from Ulithian which are not found elsewhere in Yapese. Furthermore, both typologically and phonologically it resembles Ulithian more and Yapese less; for instance, it has developed velarised labials out of inherited and pre-existing non-velarised forms, while shedding some of the distinctions which are characteristic of Yapese phonology. For example, the feature of glottalisation has been shed from Nguluwan phonology and Yapese glottalised consonants are unglottalised consonants in Nguluwan, although 11 pairs of Yapese...
consonants, including some continuants, distinguish glottalised and unglottalised forms. Yapese is the only language among those of Western Micronesia which has any glottalised consonants apart from the simple glottal stop, which is widely distributed within Micronesia. This loss in Nguluwan is probably the result of language contact involving Ulithian, as Ulithian lacks these sounds.

Another language which deserves attention here is Carolinian (sometimes called ‘Saipan Carolinian’), which is spoken in three dialects on Saipan in the Northern Marianas, dialects which have their origins on Satawal, Lamotrek and the Mortlocks respectively, all of them Central or (in the case of Mortlockese) Eastern Trukic lects which have been transported northwest to Saipan, where they have all been influenced by the same languages. In terms of historical and numerical precedence the major sources of loans into Carolinian are Chamorro, Japanese and English, with a handful of words also taken from German. It appears from my study of Fritz (1911) and Jackson and Marck (1991) that the bulk of Chamorro copies into Carolinian are themselves copies from Spanish, constituting a subset of the massive tranche of Spanish items which have been copied into Chamorro. It is furthermore clear that Chamorro is the source of most if not all Spanish elements in Carolinian, and that additionally most of these loans are ‘acculturational’ items inasmuch as they are phonologically Carolinianised versions of the names of introduced items from the various languages of the people who introduced these items to the Carolinians.

The degree of copying of elements from other languages into Chamorro itself deserves some analysis. The copying profile of Chamorro is interesting, since the Marianas had a different colonisation pattern from the Carolines (which include the islands on which the other languages which we are concerned with here are spoken). In fact, Guam had a different pattern of settlement from that of the Northern Marianas, and this is instantiated linguistically. The important factors here are historical: contact with Spaniards since the 1520s and settlement of the Spanish of Guam from the 1660s onwards, the decimation of the Chamorros and the relocation of all Chamorros from Saipan and other northern islands onto Guam (apart from a maroon group which remained on Rota), the landing of the first boatload of Carolinians on Saipan in 1815, the resettling of some Chamorros on Saipan in 1817, and the occupation of Guam by the Americans in 1898 but the cession of the northern Marianas by Spain to Germany from 1899 to 1915, with their post-WWI cession to Japan, which based its regional headquarters on Saipan (Japan also occupied Guam during WWII), and the transfer of all these islands to the US in 1945. Salas Palomo and Stolz (2008) is a discussion of the impact of Spanish on Chamorro and on attempts to ‘reaustronesianize’ the language. Rodríguez-Ponga (1995) is the largest study of the impact of Spanish on Chamorro, but Blust (2000) indicates that Tagalog or other Philippine languages, and also a Malayo-Polynesian language which has yet to be identified, have also shaped the language. This was noticed by Costenoble (1940). A few other forms of varied origins (Palauan, Chuukese, Pidgin English) are listed in Blust (2000).

Saipan Chamorro contains lexical elements deriving from Japanese, and these are not found in Guamanian or Rotanese forms of Chamorro, which have absorbed more forms from English (as has Saipan Chamorro). As we saw above, Saipan Chamorro has given many copies to Carolinian, many of these being of Spanish origin. Neither variety has copied much from German (there are a very few German loans in Carolinian and none in Chamorro). The reason for this is that the Chamorros had a greater and longer exposure to European cultural elements than any other Micronesians (including the
Carolinians), and they had already adopted the acculturational lexicon which they felt they needed to have from Spanish. No other Micronesian language has copied more than a handful of (acculturational) loans from Chamorro. I do not know of any loans from Carolinian that occur in Saipan Chamorro, and given the disparity of the low status of the Carolinians and their language as against that of the Chamorros on Saipan in earlier days, we might be surprised to find any such loans.

### 3 Areal features in Western Micronesia: lexical and structural elements

These languages of Western Micronesia show partial convergence over time through the sharing and spread of many structural features, especially some highly-marked phonotactic and morphosyntactic properties, which are absent in many of the languages spoken nearby. Additionally, they share further structural features (and these are mostly typological rather than morphological, transfers of pattern rather than transfers of fabric: Grant 2002, 2003) which are presumably inherited from Proto-Malayo-Polynesian.

There is also a plentiful body of shared lexicon of various kinds, which can be stratified historically: there are many items of shared inherited Malayo-Polynesian lexicon, additional items originating in one language which passed to others in the group (with words flowing especially from Palauan and Trukic to Yapese: Ross 1996), and later shared cultural loans from Philippine languages, German, Spanish, Japanese and (in most cases) two layers of loans from English, one from the pidgin English which was used by nineteenth century American whalers and traders who visited Micronesia, and the other layer is the result of the influence of post-1945 American administrators and educators. American missionaries who had previously established Protestant missions further east in Micronesia are also responsible for the loans from Mortlockese and Pohnpeian which one finds in some eastern Trukic lects.

Generally, the major direction for copying lexicon in Micronesia is from west to east. Loans, mostly labels for acculturational items, diffused from beyond Micronesia and from imperial administrative centres (e.g. Koror on Palau and Donguch/Colonia on Yap) to more easterly or less politically dominant locations. Most of these come from European languages but some are items from other ‘Western Micronesian’ languages – for instance the name of the mineral ‘lime’, which is often incorporated in betel chews (Palauan chaus /‘aus/, from PMP *qapuR; it also means ‘white’ in Yapese), diffused from Palau (which has the phonological reflex for this word which one would expect from the processes of Palauan historical phonology: Grant2002a) into Yapese and Nuclear Micronesian languages as far east as Kosrae (Ross 1996).

A few other words are widely dispersed. PIG is labelled as vaviy in Yapese, itself a copy from Tagalog baboy by way of Chamorro babui and then via Palauan babii. RICE in Woleaian is peraas, probably from Palauan beras, which has taken it from Malay beras (Chamorro has the expected reflex of this word, namely pugas, cf. Bisayan bugas). The word for PAPER in Chamorro is Spanish-derived papet, complete with the usual Chamorro reflex of syllable-final /-l/. But in Palauan it is babier, from German Papier, which has passed into Yapese as babyoor (it has clearly come into Yapese from Palauan since Yapese distinguishes between /p/ and /b/, though Palauan did not do so at this time), and thence into Woleaian as babiyoor. This is a form which indicates that some diffusion of European cultural items and their labels occurred from west to east into the era of German occupation. Palauan and Yapese speech communities had apparently not
quite lost all contact and ability to influence one another by the time the Europeans began to exert substantive influence in the area at the start of the twentieth century.

A table of major sources of loans in the various languages is presented below as Table 1. It needs to be recognised that by no means all the lexical elements in any of these languages have been fully sourced. For instance, Blust (2000) argues for the existence of an extensive stratum of phonologically distinctive but as yet unsourced etyma in the Chamorro vocabulary, in addition to an under-explored stratum of words which have been taken from one or more Philippine languages, possibly with Spanish ‘assistance’ inasmuch as they relocated many Filipinos to Guam. (On the other hand some of the Philippine borrowings, for instance, Chamorro *bibengka*, a borrowing from Tagalog *bibingka* ‘a kind of sweet cake’, may date only from the widespread arrival of Filipinos as workers on Guam, a population move which was especially strong in the twentieth century). This stratum postdates the split of these languages from PMP but is apparently pre-Spanish (and pre-Philippine contact?) in its dating.

Some common structural features, which are often cross-linguistically unusual but which are found in this area, are given in Table 2 (phonological features) and Table 3 (morphosyntactic features). In this examination I have prioritised phonological features, which are easier to identify, though there are also many cross-linguistic parallels in morphosyntax. Not all these shared features are diffused; some have simply been inherited from the proto-languages of the various languages under discussion (and thus may be shared with Admiralties languages in the case of Yapese or with other Nuclear Micronesian languages in the case of Trukic). Some others may go back to Proto-Malayo-Polynesian.

Similarly the value of these features for proving Western Micronesia as a Sprachbund varies intrinsically from feature to feature. Some features are simply stronger evidence, because the phenomenon which they instantiate is less frequent universally than others are. For instance, not all of the world’s languages have both prefixes and suffixes, but all four groups surveyed here use both prefixes and suffixes, especially in the verb complex. Is this then a strong feature with which to prove the existence of this Sprachbund? There again, what really is the probative power for the construction of a Sprachbund (if it has any)? Is it the fact that all the languages in Western Micronesia (apart from Chamorro, which uses only Spanish loans) denote ‘twelve’ by expressions which translate as ‘ten and two’? Or the fact that the pronominal systems in all four groups distinguish between inclusive and exclusive 1pl, just as PMP did? Or the fact that, like most Austronesian languages, they are all prepositional rather than postpositional languages?

Even when we have evidence of shared features which are crosslinguistically highly marked, they do not always present the uniform picture of origin, distribution and development which we may expect, and their various origins may be very different from language to language. For example, Palauan, Yapese and Trukic all permit word-initial geminate clusters. And yet these are infrequent in Palauan and are probably of recent origin there, though other initial CC- sequences involving different consonants are fairly common. Yapese has such clusters, but they are found only with sonorants and are also infrequent, while the only other CC-initial clusters in Yapese are infrequent dissimilated clusters such as that found in *bpiin* ‘woman’ (Jensen et al 1977). On the other hand, while geminate initial consonant clusters are both basic and very frequent in Trukic languages, other CC- clusters, in which the first and second consonants are different, are not attested there. Meanwhile, Chamorro, which does not permit word-initial geminate consonant
clusters (though these are common enough word-medially), still has a great proportion of CC-initial stems, but these were unknown before the impact of Spanish on Chamorro, and Spanish is the source of the vast majority of these forms.

4 On the difficulties of showing that Western Micronesia is a Sprachbund

In substantiating Western Micronesia as a potential Sprachbund we may discuss probative issues which are raised by the fact that two of the four participatory stocks, namely Chamorro and Palauan, lack close relatives, while the genetic connection of Yapese to the Admiralties languages has been obscured on both sides by millennia of separate internal development (and in the case of Yapese, also hundreds of years of intimate contact with speakers of other Austronesian languages). We should try to separate diffused from inherited features among a group of ultimately related languages. We may further discuss the paradox that many typological and especially phonological features connecting these languages appear to run in the opposite direction from those features mediating lexical influence.

An important first stage when looking for possible sources of linguistic change is to start with lexical transfers and to list the sources of the lexical loan elements in the various languages: Trukic, Palauan, apparently Malay, Philippine languages, German, Spanish, Japanese, two layers of English loans (the first layer of which was introduced by nineteenth century American whalers and thus predating German and Japanese elements in these languages) as well as Chamorro. Some languages have acquired elements from further languages (there is a small Western Malayo-Polynesian tranche of loans in Chamorro, Palauan and Yapese, and loans from other Trukic languages and Pohnpeian in Trukic languages), and there appears to have been some ‘dialect mixture’ among Trukic languages and beyond (there have been a few mission-actuated, post-European contact, acculturational borrowings from Pohnpeian and Mortlockese into Lagoon Trukese and Puluwatese, for instance, as the material in Goodenough and Sugita 1990 shows).

The effects of borrowing on each language should be noted, as it is not enough simply to count loans: with just under 600 recognised loans in a list of c. 3000 discrete lexical stems, Yapese has proportionately slightly fewer borrowings than Lagoon Trukese (Goodenough and Sugita 1990 indicate that Lagoon Trukese has about 760 recognised loans as against c. 3000 non-loan stems). But the impact of loan elements (especially from Austronesian languages) on Yapese is important out of all comparison to the impact of borrowing on Lagoon Trukese or on any other Trukic language. The everyday non-acculturational vocabulary of Chuukese (or for that matter Puluwatese) is unaffected by loans from Japanese or English, but the basic vocabulary of Yapese is full of loans (especially nouns) from Palauan and even more so from Ulithian. This has happened to such an extent that the ultimate affinities of Yapese were long in doubt; even Blust (1980: 152) assumed that Yapese was derived from a creole of which a key component was Palauan while another was an unidentified Oceanic language.

There is also the issue of lexical and morphological elements which are independent innovations and which have not arisen as the result of contact-induced language change. Some 110 items on the Chamorro translation of the Blust list\(^5\), about

\(^5\) This is a modification of the Swadesh list which has proven useful in helping to subclassify Austronesian languages because it is provided with reliable Proto-Malayo-Polynesian reconstructions.
120 on the Palauan Blust list, and maybe 2/3 of the relevant elements on the Yapese Blust list, all fall into this category.

Indeed Yapese shows a very great degree of independent innovation, as well as having received an inordinate amount of lexical borrowings from neighbouring languages (and on the Blust list these almost outnumber the number of elements which have been retained from PMP). One such innovation is the development of a separate set of glottalised consonants (both stops and continuants), which are found nowhere else in Micronesia (so that they cannot have been introduced through the transfer of numerous loans from relevant languages) but which are quite frequent in Yapese. In Yapese they represent the end-result of a process of reduction of what were originally CVC-sequences. Another innovation, which appears to be fairly recent if we are to judge from the scant philological evidence, is the development of original Yapese voiced stops into voiced fricatives in most instances. This change is not found in other Western Micronesian languages.

There are historical reasons for bracketing these languages of Western Micronesia together as a linguistic area and for positing a history of cultural exchange between several of the islands and atolls, given the role of the maritime Yap Empire in pre-modern times, the role of Ulithi as the major atoll group within this empire, and the Yap Empire’s dependence upon pieces of stone money which could only be obtained from Palau. Yapese did not apparently influence Palauan, lexically or otherwise and at a very basic lexical level, but the reverse flow of influence did take place. Meanwhile Yapese borrowed much basic vocabulary from Ulithian and donated a certain amount of cultural vocabulary to it and to Woleaian (this is alluded to in Walsh and Harui-Walsh 1979, see also the short lists of loans from Yapese into Woleaian which is given in Sohn 1975). We can additionally trace the patterns and directionality of influences, or at least explore some of the history of such influences, by doing philological work on some of these languages (this is especially feasible with Chamorro but it can also be done with Palauan, Yapese and Woleaian to a lesser extent, and there is a certain amount of older data on the dialects of Carolinian). In this light, a study of the Chamorro, Woleaian and Yapese vocabularies in Chamisso (1864), collected in 1817, shows that Yapese had already absorbed its Ulithian and Palauan elements but that Chamorro’s inherited sets of numerals were still obtainable as artefacts from some Chamorro-speakers (the others having switched to using Spanish numerals only).

5 Common phonological developments which cut across ‘genetic’ boundaries in Western Micronesia

Several important changes from the phonological system reconstructed for Proto-Malayo-Polynesian show parallel developments in more than one of the four branches in Western Micronesia: They are given here.

PMP *P  > /f/ (Chamorro, and possibly earlier in Palauan where it is now /w/),
> /w/ (Palauan; Trukic when it occurs before back vowels, Yapese)

PMP *S  > /t/ (Palauan; Woleaian),
> /θ/ (Yapese, Nguluwan; Ulithian).

PMP *B  > /p/ (Chamorro; Trukic and Oceanic languages generally where this merger occurs),
PMP *N  > /l/ (Palauan; Ulithian, also in the Tanapag form of Carolinian, though it does not occur in all Trukic languages)

PMP *T  > /ð/ (Palauan), /Ø/ (Trukic languages).

PMP *D  > /r/ (Palauan, Trukic)

There are other important phonological trends which are distributed across genetic boundaries in Western Micronesia. A rhotic/sibilant interplay, manifested in several ways, is found in some Western Micronesian languages: PMP *R > /s/ in Palauan (and in earlier Palauan loans into Yapese it was represented as /c/, presumably close to the sound which it had in Palauan at the time). The same sound is represented by trilled and assimilated rhotics in some Trukic languages (it occurs thus in Satawalese and Puluwatese) and by a retroflex sibilant or other kinds of sibilant in some others, such as Woleaian (and by palatal stops or affricates in yet others, such as Lagoon Trukese and Ulithian).

There is also a tendency for major words (especially contentives) in surface forms in Western Micronesian languages (and some beyond) to be C-initial and C-final: Palauan; Yapese, to a large extent also many Trukic languages (certainly Puluwatese). Chamorro and Palauan prefix epenthetic consonants to vowel-initial words (/gw, g/ in Chamorro and /ŋ/ in Palauan) while Yapese uses word-initial (and word-final) glottal stops to reinforce the use of the C-initial and C-final template, and Trukic languages prefix /w/ to word-initial back vowels and /y/ to front vowels. These principles are relaxed with regard to post-1885 loans in some of the languages, which permit loans that can be vowel-initial and vowel-final, although the rule had already been relaxed by the time Spanish loans flooded into Chamorro. Initial CC-clusters are only found in Chamorro (in loans only), in Palauan (where they are secondary and are usually broken up at the phonetic level by a svarabhakti schwa) and in a small number of stems in Yapese, such as /bpiin/ ‘woman’, where it looks, on the basis of comparative evidence (cf. Proto-Polynesian *fafine and modern Hawai’ian wahine), as though an intermediate vowel was recently deleted.

A third phonological tendency is for there to be essentially one stop series in these languages in terms of the lack of contrastive voicing, although allophonic conditioning through gemination (not a process present in Proto-Malayo-Polynesian phonology) and other changes can have their effect in phonetic realizations. Trukic languages have one stop series (as is the case with many other Oceanic languages), while Palauan has /b/ but no /p/ except in recent loans from Japanese and English, /k/ but no /d g/, and a voiced interdental fricative /ð/ but no voiceless one, plus /s/ but no /z/ until that was introduced in recent Japanese loans. Yapese has two sets of voiceless consonants (glottalised and unglottalised) as well as a voiced set, which is usually realised as voiced fricatives. Chamorro turned the voiced stops into voiceless ones, turned /p/ into /f/ and earlier /k/ into /h/ (though /kk/, if it is inherited from forms occurring in Proto-Malayo-Polynesian – though this is a language lacking geminate consonants - remains intact in Chamorro. But Chamorro kept /h/ as /h/, and eventually turned PMP *R into /g/, as did many Western Austronesian languages, while PMP *Z became /c/. Blust (2000) points out that /b d k/ are indicative of loanwords in Chamorro, as is /r/. Not all these loans are Spanish or Philippine in origin, though many of them are; where the remainder come from has as yet not been ascertained.

Of course there are other PMP sounds which have produced different reflexes in each of the four branches. PMP *R is the most notable of these: it goes to zero in Trukic,

6 Some other Trukic languages merge original /l/ and /n/ as /n/; this is the case with Lagoon Trukese. Sonsorolese changes original /l/ to /r/.
/r/ in Yapese. /s/ (formerly /c/) in Palauan and /g/ in Chamorro. It is also too easily forgotten that there are many more features which set these languages apart than which unite them. For instance, Chamorro has a ‘Philippine-type’ goal-focus system, part of which is also found in Palauan, and which can be reconstructed back to Proto-Austronesian. But this system is not found either in Yapese or Trukic. The structures of TMA systems are more complex in both Yapese and Trukic than in Chamorro, with Palauan standing between these poles in having a system of ‘medium complexity’.

When looking for possible Sprachbund features, we need to separate out universals and near universals from features which were inherited from the proto-language, and distinguish these from further features which have been transmitted by diffusion, and from yet other features which languages have evolved individually. The features which enable us to build a plausible Sprachbund are ‘shared diffusions’ in typology, and these consist of innovations which arose in one language and which have spread by contact to others. As such they are analogous to the shared innovations whose presence enables us to subgroup clusters of related languages in genetic classifications.

6 How did the Sprachbund come about? Earlier contact patterns

The question is how these features diffused from one language to another. Instead of looking at the whole region as a massive Sprachbund, we can usefully decompose the territory into smaller portions, in terms of the patterns of diffusion. The domain of the Yap Empire or at least a tributary system centred on Yap, with the interaction of speakers of Western Trukic, Palauan and Yapese itself, is one such subarea. Guam and the Marianas are home to Chamorro, which has had a much longer period of closer interaction with speakers of European languages than the other languages have had, and which shows very different patterns of massive borrowing from those found elsewhere in Micronesia, and which is peripheral to many of the shared areal features (Chamorro and Yapese-Nguluwan are by far the biggest and most profound ‘borrowers’ in this area, but the latter has not unambiguously borrowed free grammatical morphs, whereas Chamorro does this to a very great extent). A third factor enhancing the possibility of the geographical spread of changes is the superb knowledge of traditional techniques of navigation among speakers of Trukic languages, which are still preserved to this day on the atoll of Satawal. This enabled them to make long voyages to many distant islands, and helped linguistic innovations to diffuse very far (though some of these innovations never reached the far western Trukic languages).

We may mention here the presence of some possible Micronesian (Trukic and maybe Palauan) elements in Chamorro – can we detect language shift from these languages to the more powerful Chamorro? This stratum of forms includes apparently Trukic (or at least Nuclear Micronesian) forms such as pwengi ‘last night’, bwente ‘perhaps’ (cf. Carolinian bwete ‘although’), yalibaw ‘to thrash’, gwatu ‘there’, sirek ‘to copulate’; and a couple of forms which may (according to Reid 2000) be from Palauan: palau’an ‘woman’, possibly once ‘Palauan woman slave’, and un (an oblique form for ‘you plural’). The Micronesian-derived forms are common to all forms of Chamorro, including that of Guam; they are not recent loans into Saipan Chamorro from Carolinian. The possibility of the existence of a sizable Micronesian or Palauan component in Chamorro has never been clearly detected, let alone been explored in extenso, but the chance that some speakers of Guam Chamorro once had ancestors speaking Micronesian languages cannot quite be excluded. But even though Trukic substrate languages (or stray
speakers of Trukic languages who were washed off course during their voyages) may have influenced Chamorro, there is no evidence that this happened to Palauan. There are some speakers of Sonsorolese-type dialects who live in Chechang Village on Palau, but they are relatively recent migrants there. The reduction of the consonant inventory of Palauan appears to have been an independently-motivated phenomenon.

7 Conclusions

The arguments for the existence of a Western Micronesian Sprachbund are largely phonological and lexical in nature, and it is anomalous that the language which has most closely approximated the Trukic phonological system (especially in regard to the abolition of voicing as a distinctive feature) is Palauan, which has had the least contact with, and which is the one least influenced by, Trukic languages. The morphosyntactic characteristics which mark these languages as being similar are 1) mostly typological rather than overtly morphological in nature, instances of similar patterns rather than similar fabric (which are therefore of no real use in relating these languages to one another genetically; the actual morphs used derive from the genetic component of these languages) and 2) (probably) inherited rather than diffused.

I bracket the word ‘probably’ there because we cannot always tell whether these features were diffused or inherited. There are two reasons for this uncertainty: 1) because our records of these languages are mostly too recent for us to be able to discern this, and 2) because Chamorro, Palauan and Yapese are each of them languages with no close relatives (Yapese’ fellow-members of the Admiralties group, even moderately well-described ones such as Loniu, are very different from Yapese in surface and substance, the number of shared innovations of any kind linking Yapese to Admiralties languages is probably not large, and it is unlikely that there has been much contact between speakers of Yapese and those of other Admiralties languages for some millennia.) Many of the features which bind these languages together may have been inherited from earlier stages of Malayo-Polynesian. Syntactic influence of one language upon another is clearest where it is most discordant, thus it has clearly taken place in Nguluwan under Ulithian influence (for example in the way of handling alienable possession in nouns) and in Chamorro under Spanish influence (where it is manifested in the borrowing of many conjunctions and modal and temporal adverbs, which are sometimes semantically reinterpreted). But the unusual form of possession in Chamorro exemplified by ga-ña haguini [ANIMAL.CLASSIFIER-3SG sandcrab] ‘his sandcrab’; Topping and Ogo 1980: 289) is strongly reminiscent of the Trukic possessive constructions which are found in Nguluwan because of its Ulithian element.

We are dealing with several different kinds and degrees of influence and change here, a situation which has to be examined on a language by language basis. The impact of Spanish on Chamorro is a case of unidirectional language contact on a scale unprecedented elsewhere in Micronesia, although these do not impinge very heavily on the vocabulary of the Chamorro translation of the Blust list, where they constitute only 3-4% (they are more prominent among the elements of the Swadesh list). 2% of the items on the Blust Palauan list are acquired from other languages (Spanish words for ‘salt’ and ‘root’ and Japanese for ‘dust’ and one form for ‘egg’), and even these are anomalous as instances of relexification (rather than extension) in the Palauan vocabulary; Palauan has mostly absorbed new labels for new items, and apart from possibly borrowing the Oceanic applicative *-akin as –akl on some verb stems which cannot be analysed within Palauan morphology (Zobel 2000), there is no concrete evidence for transfer of fabric
into Palauan from Oceanic languages, although the Palauan phonological system does read a bit like a parody of a Nuclear Micronesian phonological segment roster.

On the other hand, the impact of other languages upon Yapese is profound, if we look just at the lexicon: there are about 30 borrowings from Ulithian and Palauan on the Blust Yapese list, in addition to the form giriin ‘green’ from English or just possibly German, and this impact continues to be found to a similarly high degree throughout the Yapese vocabulary. The richness of the Yapese vowel system is reminiscent in size and nature of the vowels found in that of Trukic languages such as Carolinian. Nonetheless, many of the features which make Yapese stand out from the other languages, such as the extensive glottalisation in the form of numerous ejective consonants, come neither from Palauan nor Ulithian, nor, as far as I know, are they inherited from Proto-Admiralties (if that is Yapese’s ancestor rather than its elder sister language); they are simply internal innovations. By contrast, the impact of other Western Micronesian languages (and this really means Chamorro) on Trukic languages is confined to a thin sliver of acculturational vocabulary which may predate strong Spanish contact (though Chamorro has had a stronger impact on Saipan Carolinian). This goes to show that in Western Micronesia some languages have been important donors, some have been major recipients, and some have been both.

Appendix: Statistics on the number and proportion of loan elements in some Micronesian languages

Chamorro c. 38% Spanish, 2.5% Japanese and English, 0.5% other (especially from Philippine languages), out of a total of c. 8,400 dictionary entries (Topping et al 1973). Note: all these may be an underestimate, especially regarding the Spanish elements, which are found in all form classes in Chamorro. With the possible exception of a sole form from Pidgin English that also occurs in Tok Pisin (puspus ‘to copulate’), all English loans are post-1898.7

Palauan 33 words from German, 63 from Spanish and 4 from Tagalog; over 660 from Japanese, plus 205 elements from English out of a total of c. 5,000 entries in McManus (ed. Josephs 1977) and the lists on http://tekinged.com/show_words.php. To the best of my knowledge, no loans originating in Chamorro, Trukic languages or Yapese have so far been identified in the Palauan lexicon. Palauan has been a donor but hardly a recipient language in Western Micronesia. Most English loans are post-WWII.

Yapese out of a total of c. 2,900 entries in Jensen et al (1977) which are not among the 147 personal or place names: 155+ items derive from a Nuclear Micronesian language, almost all from a Trukic language, probably Ulithian or Woleaian, 98 from Palauan, 15 from a lost language which was closely related to Palauan, 7 from an unidentified Oceanic language (possibly a Polynesian one), 102 from Japanese, 21 from German, 30 from Spanish, 157 from English, and there are a handful of early loans from Malay or Philippine languages which are often widely dispersed in other Western Micronesian languages. The same sources of elements are found in Nguluwan, which has sometimes made different loan choices from Yapese: Nguluwan lapis (< Spanish), Yapese pensel ‘pencil’ (< English). Many further loans from Palauan and Ulithian may actually be

7 An uncertain but considerable proportion of the remaining Chamorro lexicon can be shown on historical phonological grounds to be borrowed from as yet unidentified languages rather than to have been inherited from PMP.
present and may be waiting to be spotted in the Yapese vocabulary. Most of the English loans in Yapese are post-WWII. Ross (1996) deals with much of this, and provides the insights ion the Palauan-related language and the Oceanic language; other counts are the author’s.

Woleaian out of some 4000 entries in Sohn and Tawerilmang (1976), there are 300 words from Japanese, 100 from English, about 30 from Spanish, about 10 from German, a few others from other languages (peraas ‘rice’ from Malay, kaarebaw ‘water buffalo, cattle’ from Malay via Spanish or possibly Tagalog, and a few cultural loans from Yapese). Ulithian has a similar loan profile, maybe with fewer German loans but with more loans taken from Yapese (Walsh and Harui-Walsh 1979). Most of the English loans in Woleaian are post-WWII adoptions.

Puluwatese Out of 6,300 entries and c. 4,000 stems in Elbert (1972), there are c. 18 stems from Spanish (some of them via Chamorro), c. 90 from Japanese, 7 from German, 5 from Pohnpeian, 2 from Mortlockese, also hundreds of loans from English.

Chuukese (Lagoon Trukese) out of c. 3,750 stem-level elements given in Goodenough and Sugita (1990), there are c. 310 from Japanese, c. 400 from English, 6 from Pidgin English, 5 from Latin, 4 from Chamorro, 1 each from Korean, Puluwatese and Samoan, 13 from German, 13 from Spanish, 6 from Mortlockese and 11 from Pohnpeian (statistics from Goodenough and Sugita 1990).

Table 1. Sources of loans in the languages of Western Micronesia

<table>
<thead>
<tr>
<th>Language</th>
<th>Chamorro</th>
<th>Palauan</th>
<th>Yapese</th>
<th>Trukic</th>
<th>Polynesian</th>
<th>Pohnpeian</th>
<th>Philippine languages</th>
<th>Spanish</th>
<th>Pidgin English</th>
<th>Japanese</th>
<th>German</th>
<th>English</th>
<th>( \bar{U} )</th>
<th>( \bar{E} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chamorro</td>
<td>X Tiny?</td>
<td>No Tiny?</td>
<td>A couple</td>
<td>No No</td>
<td>Yes</td>
<td>Massive</td>
<td>Tiny item</td>
<td>Only on Saipan</td>
<td>no</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palauan</td>
<td>No X No</td>
<td>No? No?</td>
<td>No No</td>
<td>A few</td>
<td>A few dozen</td>
<td>c. 8 Large</td>
<td>A few</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yapese</td>
<td>No Sizeable and basic</td>
<td>X Sizeable and basic</td>
<td>A few</td>
<td>No A handful</td>
<td>C 35 A few</td>
<td>Yes</td>
<td>A few</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nguluan</td>
<td>No No X No</td>
<td>No No</td>
<td>A handful</td>
<td>Presumably</td>
<td>A few? Yes</td>
<td>A few?</td>
<td>Presumably</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sonorol</td>
<td>No Maybe now</td>
<td>No? X No No No?</td>
<td>Presumably</td>
<td>A few? Yes, a few</td>
<td>?</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulithian</td>
<td>No No Many X No No</td>
<td>No?</td>
<td>Over a score</td>
<td>Presumably</td>
<td>A few</td>
<td>Yes</td>
<td>?</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolei-an</td>
<td>No No Many X No No</td>
<td>No? c. 30</td>
<td>300 in all</td>
<td>Yes, c. 100</td>
<td>c. 10</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carolian</td>
<td>Large Maybe No X No No</td>
<td>Not directly</td>
<td>Dozen s, many</td>
<td>Some</td>
<td>Yes</td>
<td>Severa l</td>
<td>In-s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sources of items employed in partial relexification (lexical replacement) in Western Micronesian languages

Chamorro: Spanish, minimally Philippine, Oceanic and (on Saipan) Japanese
Palauan: some Japanese, Spanish to a tiny extent
Yapese: Palauan, Western Trukic, other Oceanic
Nguluwan: Palauan and especially Ulithian (at least 4 items on the 100-tem Swadesh wordlist are Nguluwan loans from Ulithian which are not also found in the sizable Ulithian tranche of loans which have entered Yapese)
Sonsorolese: ?none; maybe some nowadays from Palauan
Ulithian: ?none; some acculturational lexicon from Yapese
Woleaian: ?none; some acculturational lexicon from Yapese
Carolinian: some Chamorro
Puluwat: ?none; the sizable borrowed lexicon is all acculturational
Lagoon Trukese: ?(there are plenty of loans in the lexicon but they are not relexificational in nature)
Table 2. Some areal-typological phonological features in selected Western Micronesian languages

<table>
<thead>
<tr>
<th>Feature/ Language</th>
<th>Palauan</th>
<th>Yapese</th>
<th>Nguluwan</th>
<th>Solor-Tobi</th>
<th>Woleai/ Ulithian</th>
<th>Carolinian</th>
<th>Chamorro</th>
</tr>
</thead>
<tbody>
<tr>
<td>/n/ and /l/ are distinct phonemes</td>
<td>/n/ occurs only in loans</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>/n/ only in loans</td>
<td>Yes</td>
</tr>
<tr>
<td>/l/ phonemic</td>
<td>No (but maybe yes once)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>One stop series in native forms</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, though voiced geminates are phonemic</td>
</tr>
<tr>
<td>Consonantal gemination is phonemic</td>
<td>Yes but secondarily so</td>
<td>No</td>
<td>Yes?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Interdental fricatives are phonemic</td>
<td>Yes; voiced phonemic</td>
<td>Yes, voiceless phonemic, voiced phonetic</td>
<td>Voiceless only</td>
<td>Yes, voiceless phonemic</td>
<td>Not in W, which has substituted /t/, but the voiceless one is used in U</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>/h/ is present</td>
<td>Only in recent loans</td>
<td>Only in recent loans</td>
<td>No?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, from original /k/</td>
</tr>
<tr>
<td>Voiceless velar fricatives</td>
<td>No, but in nineteenth century yes</td>
<td>No, though a voiced velar fricative has phonetic status as allophone of /g/</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, as a single allophone of /k/</td>
</tr>
<tr>
<td>Glottal stop is present</td>
<td>Yes</td>
<td>Yes</td>
<td>No, lost under Ulithian influence</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Other</td>
<td>No</td>
<td>Yes; 11</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>such</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glottalised consonants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velarsised labial consonants</td>
<td>No</td>
<td>No</td>
<td>Yes, in Ulithian loans and also in inherited forms</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Long vowels are phonemic</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Complex vowel nuclei</td>
<td>Yes</td>
<td>Yes?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Word-final voiceless vowels</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes in W, no in U</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Number of vowel phonemes</td>
<td>5 (plus emergent schwa)</td>
<td>8 plus length for all 8</td>
<td>8</td>
<td>5</td>
<td>6 short and 7 long</td>
<td>9</td>
<td>6 (4 plus 2 loan phonemes)</td>
</tr>
<tr>
<td>Central vowels?</td>
<td>Non-phonemic schwa</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Parasitic initial glides?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (though less frequent)</td>
<td>Yes</td>
</tr>
<tr>
<td>Are there loan phones?</td>
<td>Yes, 6</td>
<td>Yes, /bw mw/ from Ulithian replace older Yapese /bm/ in many cases</td>
<td>No?</td>
<td>No?</td>
<td>Very recent borrowing of some from Chamorro</td>
<td>Yes. /r/</td>
<td></td>
</tr>
<tr>
<td>Loan phonological canons?</td>
<td>No? If so, recent</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes, CCV-</td>
</tr>
</tbody>
</table>
Table 3. Some structural parallels in Western Micronesian languages

<table>
<thead>
<tr>
<th>Feature/Language</th>
<th>Palauan</th>
<th>Yapese</th>
<th>Nguluwan</th>
<th>Sonsorol-Tobi</th>
<th>Woleaian/Ulithian</th>
<th>Carolinian</th>
<th>Chamorro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeral classifiers?</td>
<td>No, but some special number sets</td>
<td>Not now, but possibly formerly</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No, but there were formerly special numeral sets</td>
</tr>
<tr>
<td>Distinction between alienable versus inalienable possessions on nouns</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Decimal numeral system?</td>
<td>Yes</td>
<td>No, modified quinary</td>
<td>No, modified quinary</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, both Austronesian original and modern Spanish</td>
</tr>
<tr>
<td>Ablaut in noun or verb stems</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Definite article?</td>
<td>Yes, preposed</td>
<td>No, though Pre-Yapese apparently had one</td>
<td>No, modified quinary</td>
<td>Yes, postposed</td>
<td>Yes, postposed</td>
<td>Yes, postposed</td>
<td>Yes, preposed</td>
</tr>
<tr>
<td>Indefinite article?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes, from Spanish</td>
</tr>
<tr>
<td>Possession by cliticisation</td>
<td>Yes, for inalienable nouns</td>
<td>Yes, for inalienable nouns</td>
<td>Yes, for inalienable nouns</td>
<td>Yes, for inalienable nouns</td>
<td>Yes, for inalienable nouns</td>
<td>Yes, for inalienable nouns</td>
<td>Yes</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Neg-Verb</td>
<td>Neg V</td>
<td>Verb-Neg</td>
<td>Verb-Neg</td>
<td>Neg-V</td>
<td>Neg-V</td>
<td>Neg-V</td>
<td>Neg V</td>
</tr>
<tr>
<td>Basic constituent order</td>
<td>VSO</td>
<td>VSO</td>
<td>SVO</td>
<td>SVO</td>
<td>SVO</td>
<td>VSO</td>
<td></td>
</tr>
<tr>
<td>Is it ergative?</td>
<td>Split-ergative</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Split-ergative</td>
</tr>
</tbody>
</table>
What is transferred in interlingual contact in Western Micronesia

1. acculturational lexicon of all kinds, including social borrowings such as greetings in Chamorro and some central and eastern Trukic languages (this is true of all languages surveyed here)
2. non-core lexicon (in some languages only: there is some borrowing from Japanese into Palauan and massive borrowing from Spanish into Chamorro, and extensive borrowing from Palauan and Trukic into Yapese, with further Trukic loans found in Nguluwan)
3. core lexicon (strongly in Chamorro, also Yapese and Nguluwan, marginally in Palauan)
4. verbs which are incorporated into pre-existing morphological paradigms (there are many examples of these being borrowed in Chamorro, Nguluwan and Yapese)
5. loan phonemes (marginally and only recently in Yapese and Palauan, more strongly in Chamorro)
6. new loan phonological canons*
7. phonological features at a phonemic level (mid-vowels)*
8. adjectival comparative markers*
9. numerals*
10. some personal and other pronouns*
11. some conjunctions*
12. some prepositions*
13. some temporal, spatial and phasal adverbs*
14. copula* (in part)
15. indefinite article*
16. discourse markers (many in Chamorro and maybe some in Carolinian, also possibly a polite imperative marker in Nguluwan)
17. metatypy of constructions (copying of certain constructions from Ulithian to Nguluwan and from Spanish to Chamorro)
   (*) these kinds of items are only transferred in Chamorro.

What remains undiffused from one language to another in these cases

1. bound morphs indicating either inflectional or derivational processes (the Spanish derivational affixes found in the Spanish-derived lexicon in Chamorro rarely spread to pre-Spanish forms in the language)
2. free grammatical morphs (except in Chamorro, where several of these have been borrowed; these largely derive from Spanish).

Table 4. Relative degrees of retention of Proto-Malayo-Polynesian items on the modified Swadesh list used in Blust (1981) and Jackson (1983: 227)

<table>
<thead>
<tr>
<th>Language</th>
<th>Percentage</th>
<th>Language</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malay</td>
<td>56</td>
<td>Chukese</td>
<td>37.8</td>
</tr>
</tbody>
</table>

8 Standard Malay is the Austronesian language with the highest rate of retention of PMP origin on the Blust list.
References


Reid, Lawrence A. 2000. ‘Structural evidence for the place of Chamorro in the Austronesian language family.’ To be published in *Pacific Linguistics*.


30 Native North American Languages\textsuperscript{1}

Marianne Mithun

The languages indigenous to North America provide an especially fruitful arena for investigating circumstances underlying areal phenomena. Nearly 300 different languages are known to have been spoken over this vast geographical region before the arrival of Europeans, and there were surely many more. They comprise well over 50 distinct genetic groups, some spread over great distances. There are recognized culture areas, but not all correspond to linguistic areas, raising questions about the geographical, social, and linguistic factors behind the discrepancies. Intriguingly, most North American linguistic areas constitute exceptions to classical expectations about relative borrowability, the notion that vocabulary is copied first, then sounds, speech habits, sentence structure, and finally deeper grammar. In fact the strongest North American linguistic areas show surprisingly few loanwords, but often extensive parallelisms in grammatical categories and structures. Here social circumstances are described that can foster the emergence of such areas, and cognitive and communicative processes by which they might develop.

1 Contact areas in North America

There are ten commonly-recognized culture areas in North America, shown in Map 1.

(1) Culture areas of North America: Sturtevant (1988: ix)

1) Arctic  2) Subarctic  3) Northwest Coast  
4) Plateau  5) Great Basin  6) California  
7) Southwest  8) Great Plains  9) Northeast  
10) Southeast

Both culture and linguistic areas result from interactions among peoples. But most cultural traits can be transferred more easily than linguistic ones. One can adopt pottery styles, for example, with less intense contact than auxiliaries. Intensity of contact is related to both the nature of interaction and its duration.

1.1 The nature of the contact

The intensity of contact varies widely over North American culture areas. Where communities were large and population density light, as in the Northeast, speakers often had relatively little contact with outsiders. Where density was greater and communities

\textsuperscript{1} I am grateful to Wallace Chafe, Danny Hieber, Robert Rankin, and David Rood for sharing their great expertise on various languages discussed here.
were small, as in Northern California, exogamy and multilingualism were often the norm. In some areas there were differences in prestige, as in the Northwest and Southeast, but in others not, as in California and the Northeast. In some, language played a strong role in identity, as in the Southwest. In some areas contact is ancient, as along the Northwest Coast, while in others it is relatively recent, as on the Plains. The Southeast has long traditions of trade with neighbors, often via a lingua franca, either a pidgin like Mobilian Jargon, or one of the languages spoken there, like Creek in the Creek Confederacy. Still, there is much we cannot know about the social nature of prehistoric contacts, including patterns of language maintenance and shift.

Map 1. Culture areas of North America (Sturtevant 1988: ix)

1.2 Linguistic profiles

The culture areas also vary in their linguistic heterogeneity. The Arctic was occupied by speakers of a single language family, Eskimo-Aleut. On the Plains, several large families
are represented: Athabaskan, Algonquian, Siouan, Uto-Aztecan, Kiowa-Tanoan, Caddoan, and the isolate Tonkawa, but most of the shared linguistic traits there are inherited. The Northwest Coast is home to languages from a dozen distinct genetic units, many of which vary in their structures, but they share numerous features. In California, 22 different families are represented, but, as in the Northwest, numerous traits cross genetic lines.

1.3 Areal boundaries

Early large-scale investigation of North American linguistic areas was undertaken by Sherzer (1968, 1973, 1976). Observing that linguistic areas are rarely sharply delineated, he distinguished four trait types.

(2) Areal traits: Sherzer (1973: 759)

<table>
<thead>
<tr>
<th>Trait Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole areal trait</td>
<td>found in most languages of a given culture area</td>
</tr>
<tr>
<td>central areal trait</td>
<td>found in most languages of an area with locus of distribution in the center of this area</td>
</tr>
<tr>
<td>regional areal trait</td>
<td>with a continuous or almost continuous distribution within one region of a given culture area</td>
</tr>
<tr>
<td>family trait</td>
<td>in language X retained from proto-language A</td>
</tr>
</tbody>
</table>

The first three pertain to the complexity of contact situations. In the simplest situations, a linguistic area is delineated on all sides by borders which might hinder contact among groups, such as oceans, rivers, or mountain ranges. Languages on the Northwest Coast and in California have the Pacific Ocean to their west. Languages in the Southeast have the Atlantic Ocean to their east and the Gulf of Mexico to their south. But people do travel by water.

Even with strong borders, contact is rarely homogeneous within an area, and features seldom pass through all languages uniformly. They might flow from a core outward, a central trait. The Southeast core cultural and linguistic areas contain the entire Muskogean family and isolates Atakapa, Chitimacha, Natchez, and Tunica. As one moves outward, shared traits become progressively sparser. Ofo and Biloxi share many traits with the core; Dhegiha, Cherokee, Tutelo, Catawba, and Caddo some; Nottoway, Shawnee, and Timucua a few.

Traits may spread from several points, complicating the identification of basic-level areas. The Northwest Coast, stretching from the Subarctic to California, is a strong linguistic area, but some traits extend only over subareas. One is the Northern Northwest Coast, with Eyak and Tlingit, Haida (isolate); Tsimshianic; some Wakashan (Haisla, Heiltsuk, Kwak’wala), and some Salishan (Nuxalk = Bella Coola, Comox). Another extends from the Nass River in British Columbia to the Columbia between Washington and Oregon. Still others are the South Central Coast and Central Oregon. Many overlap, and some are included within others. California also contains identifiable subareas, such as Northwestern California (Wiyot, Yurok, Karuk), Clear Lake or Central California (Pomoan, Yuki, Lake Miwok, Wappo, Patwin), the South Coast Range (Esselen, Salinan, Chumashan), and perhaps Southern California-Western Arizona (Yuman, Takic Uto-Aztecan).
The distribution of traits is not necessarily continuous over even a well-defined area. Populations can move around, establishing relationships with different groups at different times, like the Apacheans (Athabaskan) in the Southwest. And the likelihood of transfer of a given trait depends on numerous other factors such as prestige, group identity, cultural practices, and language structures.

Sherzer’s final type, the family trait, plays a crucial role in the study of contact: traits shared by two languages but inherited from a common parent are not considered strong evidence of contact effects. To distinguish common inheritances, it is helpful if the languages within an area have relatives outside. A number of linguistic areas in North America present this situation. The Algic languages Wiyot and Yurok of northwestern California have Algonquian relatives all across the continent, from Montana (Blackfoot, Cheyenne) to the Atlantic (Micmac, Maliseet, Passamaquoddy). The Pacific Athabaskan languages of Oregon and California have relatives in the Southwest, western Canada, and Alaska. The Siouan languages in the Southeast have relatives on the Plains.

1.4 Inheritance versus contact

Genetic classification necessarily depends on the criteria admitted for evidence of relatedness. During the twentieth century, some remote relationships among families in North America were hypothesized on the assumption that structural similarities alone can be diagnostic of common origin. On the Northwest Coast, a “Mosan” stock was proposed consisting of the Wakashan, Salishan, and Chimakuan families. Various “Hokan” proposals have grouped families and isolates from northern California south into Mesoamerica and even South America, and from Baja California on the Pacific, into south Texas: Karuk, Chimariko, Shastan, Palaihnihan, Yana, Washo, Pomoan, Esselen, Salinan, Yuman, Cochimí, Seri, Tequistlatecan, Coahuiltecan, Jicaque, Subtiaba-Tlapaptec, Coahuilteco, Comcrudan, and Quechua. Some of these families may be very remotely related (perhaps at time depths greater than Indo-European), but many have been spoken in contiguous areas for so long that it is difficult to separate shared inheritances from ancient contact effects. The “Penutian” hypothesis presents a similar picture, merging recognized families in California (Wintun, Maidun, Yokuts, Utian), Oregon (Takelma, Coosan, Sioulawan, Alsean, Kalapuyan, Chinookan), the Plateau of eastern Oregon, Washington, and Idaho (Klamath-Madoc, Cayuse-Molala, Sahaptian) British Columbia (Tsimshianic), and Mexico (Mixe-Zoque, Huave). Again, some of these families may be very remotely related, but many observed similarities are likely the result of long-standing intensive contact. In the Southeast, a “Gulf” stock was once proposed, uniting the Muskogean family and isolates Natchez, Tunica, Chitimacha, and Atakapa. In the Southwest, an “Aztec-Tanoan” stock was proposed unifying the Uto-Aztecan and Kiowa-Tanoan families. Further details on these proposals are in Campbell (1997) and Mithun (1999). As more has been learned about the kinds of structures that can emerge from contact, these hypotheses have been reconsidered.

Of course related languages are not immune to areal effects (Epps et al. 2013). They tend to be typologically similar, facilitating bilingualism and convergence. It is easier to transfer substance when lexical, morphological, and syntactic categories are similar (Mithun 2013). Learners often reanalyze structures in a second language on the model of counterparts in their first where possible. And bilinguals often extend the functions of elements and constructions within one language on the model of the other.
1.5 Linguistic areas in North America

Not all of the culture areas listed in (1) are considered strong linguistic areas, for the kinds of reasons discussed above. The Arctic contains just one language family, Eskimo-Aleut. The Subarctic contains single branches of just two, Athabaskan and central Algonquian. The Plateau shows some genetic diversity, but most if its shared linguistic traits are also found on the Northwest Coast. The Great Basin contains just two genetic units: the Numic branch of Uto-Aztecan and the isolate Washo; most common traits are also shared with adjacent California languages. The Plains culture area was constituted recently, within the last several hundred years. Furthermore, the area was sparsely populated, so speakers of different languages were not generally in close contact. The Northeast contained branches of just three families: Siouan, Iroquoian, and Algonquian. This area too was relatively sparsely populated, and so far there is little evidence of significant contact across family lines. Most shared traits can be reconstructed for their respective parent languages.

Some culture areas, however, are also strong linguistic areas, particularly the Northwest, California, and the Southeast. All were densely populated before contact by small groups speaking a variety of genetically unrelated languages, a situation conducive to multilingualism. Contact was both intense and enduring. But none of the areas is crisply delineated or homogenous. There are core areas and peripheral areas. There are subareas of varying strengths. And a number of traits extend beyond their outer boundaries. A fourth area, the Southwest, also exhibits some shared traits presumably due to contact, but fewer than the others and of slightly different types.

1.5.1. Northwest

The Northwest linguistic area stretches from the Subarctic to California, and, with the inclusion of the Plateau, east to the Rockies (Maps 2 and 3). Generally considered one of the strongest areas in the world, it is home to 20 genetic groups, 21 with Pacific Yupik. The most thorough descriptions are in Thompson and Kinkade (1990) for the Northwest Coast, and Kinkade, Elmendorf, Rigsby, and Aoki (1998) for the Plateau.
Map 2.  Northwest Coast (Suttles 1990: ix)
Intensive contact including intermarriage extends back millennia throughout the Northwest. Suttles reports that 'There is evidence for gene flow throughout the area and beyond' (1990: 1). ‘The Tlingit, Haida, Tsimshian, and Haisla shared a system of matrilineal lineages . . . that could be equated for purposes of intergroup marriage’ (1990: 12). Texts from Boas (1921: 836-1277) indicate ‘a network of intermarriage and ceremonial relations extending from the Oowekeno to the Comox (Northern Coast Salish)’ (1990: 13). Evidence from early explorers indicates that ‘marriages, visiting,
shared access to resources, and trade in food linked tribes on the coast from the Makah to
the Alsea with the upriver Chinookans, Southwestern Coast Salish Cowlitz, and Tualatin’
(1990: 12, citing Hajda 1984: 123-132). There were similar networks in the adjacent
Plateau, as noted by Kinkade et al.

Contacts and mutual interinfluences among the languages and language families of the
Plateau are of long standing. It is clear that the majority of them participated in a common
linguistic area. Plateau ethnographic studies have demonstrated mechanisms of group
interaction that, projected into the past, would have produced these linguistic results. In this
area intermarriage, trade, and joint participation in economic and ritual activities set up social
relationships that frequently crossed linguistic boundaries. These patterns of bilingualism in
certain parts of the region obscured linguistic boundaries, and at times led to language
replacement. (Kinkade et al 1998: 69-70)

Three families, Wakashan, Chimakuan, and Salishan, constitute the core of the area.
There is relatively little lexical borrowing, but contact is so ancient that many structural
traits, otherwise rare cross-linguistically, can be reconstructed for the respective parent
languages. There are large consonant inventories, contrasting plain and ejective
obstruents and often resonants, velars and uvulars, multiple laterals, rounded and
unrounded back obstruents, and distinctive glottal stops. Most of the languages have only
three or four distinctive vowels but contrastive length. Consonant clusters can be
complex, sometimes with four or more consonants; words in some languages can consist
uniquely of consonants. There is extensive reduplication. Many languages have numeral
classifiers, and many deictic systems distinguish visible and invisible referents. The
languages are generally polysynthetic and primarily suffixed. Constituent order is
basically predicate-initial. Particularly in the core languages, the noun/verb distinction is
weak: most content words can be used to predicate.

1.5.2 Northern California

Northern California is home to a large number of typologically diverse languages (Map
4), generally spoken in small communities. Detailed discussions are in Jacobs (1954),

Active trade, intermarriage, and multilingualism were typical over at least a
millennium. Languages differed little in prestige. There was no lingua franca or code-
switching: it was polite to speak the language of the community one was in at the
moment. Bilinguals thus exerted conscious efforts to keep their languages separate.
O’Neil notes:

Intermarriage was common (see Waterman and Kroeber 1934), so growing up around
speakers of several languages was not unusual if one’s parents came from faraway places, as
they often did. Generally, the wife moved to her husband’s village after marriage, so females
were often especially multilingual. As a consequence, children often grew up in the presence
of bilingual mothers, often being exposed to a number of unrelated languages even from their
earliest days. (O’Neil 2008: 290)

There would have thus been both language maintenance and language shift.
Within California and even within sub-areas, there is considerable typological diversity. Some languages are quite polysynthetic, many are mildly synthetic, and some more analytic. Some are basically predicate-initial, and others predicate-final. Some are head marking, and others dependent marking. There are nominative/accusative, agent/patient, and hierarchical alignment patterns. Numerous shared features have been observed, however. Many extend over just a small sub-area, while others extend even beyond California. Among the traits noted are uvular stops, voiceless laterals, back apical or retroflex stops, a distinct voiced stop series, sound symbolism, pronominal dual, inclusive/exclusive first persons, nominal case, alienable/inalienable possession, verbal reduplication for distribution or repetition, means/manner prefixes, locative/directional suffixes, evidentials, and classificatory numeral systems.
1.5.3 The Southeast

The Southeast (Map 5) is home to all Muskogean languages (Choctaw, Chickasaw, Alabama, Koasati, Hitchiti, Mikasuki, Creek, Apalachee), some Siouan languages (Dhegiha, Biloxi, Ofo), isolates Natchez, Tunica, Chitimacha, and Atakapa, and more peripherally Cherokee (Iroquoian), Shawnee (Algonquian) Yuchi, Tutelo (Siouan), Catawba (related to Siouan), Timucua, and Caddo (Caddoan). Surveys are in Haas 1971, 1973, 1979, Crawford 1975, Booker 1991, and Martin 2004. Contact especially among the core groups has been intimate and long-standing.

Map 5. The Southeast of North America, sixteenth to eighteenth centuries (Fogelson 2004: ix)

Long-distance trade networks extended throughout the Southeast from late prehistoric to early historic times (Waselkov 2004: 686). Muskogean groups also maintained alliances and ‘mechanisms for incorporating autonomous local groups into intricate political structures capable of concerted action’ (Walker 2004: 375), but there was economic, political, and social inequality (Brown 2004: 677). The Creek were particularly numerous and powerful: ‘The larger, more powerful Muskogee [Creek] incorporated members from other groups, such as the Yuchi, Alabama, Hitchiti, and Shawnee, during their tenure in the Southeast’ (Innes 2004: 393). In the late seventeenth or early eighteenth century, the Creek Confederacy was formed, which included speakers not only of Creek, but also Alabama, Koasati, Apalachee, Natchez, Yuchi, Shawnee, and probably more. Creek was used as a lingua franca. The Alabama and Koasati were closely linked, residing near each
other over a long period and ‘at various times, near the Choctaw, Pakana, Biloxi, Caddo, Pascagoula, Tunica, and Ofo’ (May 2004: 413). There was also a Choctaw Confederacy, which ‘may have included speakers of Natchez, Alabama, and other languages’ (Galloway and Kidwell 2004: 499). Louisiana settlements composed of Choctaw, Tunica, and Biloxi are described by Brain, Roth, and de Reuse (2004: 589). The Natchez were also connected to the Tunica, having taken in a number as refugees, and with the Chitimacha through intermarriage (Galloway and Jackson 2004: 500). After their defeat by the French in 1731, the few Natchez survivors settled among the Creeks and Cherokees and intermarried (Kimball 2005: 385). The last speakers were trilingual in Natchez, Creek, and Cherokee. The Cherokee also incorporated speakers of other languages into their communities via capture and intermarriage, and ‘multilingualism was common and valued’ (Fogelson 2004: 337).

The languages of the Southeast are typologically more similar to each other than those of California. Most at the core are mildly synthetic, with basic SOV order and postpositions. Most contain pronominal affixes on verbs with agent/patient alignment. Among the linguistic traits listed in the literature for Southeastern languages are labial fricatives, voiceless laterals, retroflex sibilants, sound symbolism, independent possessive pronouns, alienable/inalienable possession, agent/patient pronominal affixes, pronominal duals and plurals, inclusive/exclusive first person, diminutive noun suffixes, verb reduplication for distribution or repetition, locative/directional verbal prefixes, tense/aspect verbal suffixes, pervasive postural distinctions, auxiliary constructions, and quinary counting systems (Sherzer 1973, Campbell 1997, Martin 2004).

1.5.4 The Southwest

The Southwest (Map 6) is a strong culture area with some shared linguistic traits. Bereznak (1995) provides detailed discussion of these. Five genetic units are represented: Tanoan languages, Keresan languages, Hopi (Uto-Aztecan), Zuni (an isolate) and Apachean Athabaskan languages, particularly Navajo. The Pueblo communities have been neighbors for a long time, but the Apacheans are more recent arrivals. Population density was high, and groups were linked by trade relations, particularly in prehistoric times. There was later active trade between the Pueblos and the Navajo, in which Navajo was used as a trade language, spoken by many Arizona Tewas, Hopis, and Zunis (Kroskrity 1982, Ford 1983: 720). Intermarriage among Pueblo groups (Hopi, Zuni, Keresan, Tanoan, shaded in Map 6) and between them and Navajos was common, as was bilingualism (Parsons 1939 cited in Bereznak 1995: 77). Bereznak notes that ‘marriage partners from other villages introduced new songs, dances, and societies... Ceremonial dancers and ritualists commonly participated in ceremonies in other villages.’ (1995: 60)

Bereznak identifies 28 traits shared among various Southwestern languages, but considers just four to be reasonable areal indicators: glottalized consonants, tones, final devoicing, and pronominal duals. Not surprisingly, more features are shared among Pueblo languages than with the more recent Apachean arrivals. These include distinctive voiced stops, an s/š opposition, labio-velars, alienable/inalienable possessive noun prefixes, plural and locative noun suffixes, tense/aspect verbal suffixes, noun incorporation, a three-way demonstrative system, classificatory verbs, distinctions in the kinship system, and some actual morphemes, including the evidential ʔas, directional marker -mi, and passive -ti. Many shared traits extend beyond the Pueblo area westward
to Yuman languages, north to Great Basin languages, and eastward to Tonkawa and Caddo (Bereznak 1995: 159, 166).

Map 6. The Southwest Culture Area: Ortiz (1979: ix)

Sherzer explains the fact that this strong culture area is not matched by an equally robust linguistic area in terms of attitudes.

In other areas with dense populations (Northwest Coast, Plateau, California), many linguistic traits were found to have continuous distributions cutting across genetic linguistic boundaries. In the Southwest, traits tend to be more randomly distributed, suggesting that little mutual linguistic influencing has occurred. The explanation of this situation may be found in a sociolinguistic factor about which we rarely have data -- attitude toward language. The Southwest is one area for which many observers have reported attitudes towards one's own language and that of others, perhaps because [786] these attitudes are often quite explicit. Southwest Indians are very conservative with respect to language, taking pride in their own language and often refusing to learn that of others. When they do learn other languages, they seem consciously to avoid allowing alien linguistic traits to penetrate their own linguistic system. (Sherzer 1973: 785-786)
2 Evaluating shared linguistic traits

Sherzer (1973, 1976) provides lists of shared traits for all of the culture areas. As he realizes, not all contribute equally to areal strength.

2.1 Feature strength

An important consideration is the cross-linguistic frequency of the trait. The SOV order shared by Southeastern languages, for example, could be the result of contact: word-order patterns spread easily. At the same time, it is characteristic of a majority of the languages of the world, so it could be due to internal development. A second consideration is the relationship among traits. Many Southeastern languages contain postpositions. But postpositions often develop out of SOV structures, so the two traits are not necessarily independent.

Conspicuous absence of a trait is sometimes cited as an areal feature. The significance of the absence also depends on its cross-linguistic frequency and its source. Among the features cited by Sherzer (1973: 768) as characteristic of the Northwest Coast and the adjacent Plateau are ‘lack of one-stop-series languages’. But many languages in the world have multiple stop series. Sherzer also reports that languages of the Subarctic lack nominal reduplication, nominal case, and instrumental markers in verbs (1973: 765). Those of the Plains lack uvular $q$, labiovelar $k\text{ʷ}$, and nasals other than $m$ and $n$. These traits do distinguish the languages from those in other areas, but all are inherited from their respective parent languages. In a sense they are a testament to the lack of areal effects. Still, in some cases gaps are significant. A group of neighboring languages along the Northwest Coast lack $n$: Nitinaht and Makah (Wakashan); Quileute (Chimakuan); and Twana and marginally Island Comox and Upriver Halkomelem (Salishan). This gap can be traced to shifts of original nasals to prenasalized stops which ultimately lost their nasalization ($*n > ^{\text{d}}d > d$), habits of pronunciation transferred across genetic lines (Thompson 1972). Related languages outside the area still contain the nasals.

2.2 Levels of structure

Different domains of language are known to show different propensities for replication in contact situations. A commonly-cited scale is in (3).

\[(3) \text{ Borrowability}^2\]

---

2 This scale is based on the relative openness of levels of language. The most open is most prone to code copying (borrowing). However, because open classes show a high degree of consciousness for speakers they may well have the lowest degree of code copying if the speakers of the receiving community wish to distance themselves from the donor community (see Epps and Michael, this volume) or if the former occupies a lower position than the latter within a society (see, e.g., Hickey 2012) – Raymond Hickey,
Levels most affected

- Vocabulary (loanwords, phrases)
- Sounds (present in loanwords)
- Speech habits (general pronunciation, suprasegmentals [stress, intonation])
- Sentence structure, word-order
- Grammar (morphology: inflections)

Levels least affected

Interestingly, the strongest linguistic areas of North America do not show this pattern at all. Loanwords are strikingly rare, while structural parallelisms are often extensive. The distributions reflect both social circumstances and cognition.

The ranking in (3) corresponds to decreasing degrees of consciousness and control on the part of speakers. Vocabulary is highly accessible to consciousness: speakers can easily incorporate foreign lexical items into their own speech if they wish, even with little mastery of the donor language. They can also consciously avoid doing so. Morphological structure, by contrast, is routinized, generally below the level of consciousness and more difficult to manipulate intentionally.

The speech of second-language speakers can reflect the hierarchy in reverse. Learners usually focus on using new vocabulary, but they may retain speech habits (a foreign accent) and syntactic patterns from their first language. They may also carry over more subtle features, such as the relative frequency of certain stylistic choices (passivization, topicalization) or attention to particular semantic distinctions (location, direction, manner of motion, shape).

But shift is not the whole story. Even more balanced bilinguals making a conscious effort to speak a particular language tend to be more conscious of vocabulary. Where their two languages contain comparable constructions, they might transfer the relative frequencies of semantic and stylistic choices. The longer intimate contact persists, the more deeply it can shape language structure, as such choices become entrenched and routinized. In what follows, some of the mechanisms behind this shaping are described and illustrated.

3 Lexicon


3 This is true of the lexicon as an open class. A similar influence on grammar but not vocabulary can be seen in Old English from Brythonic (forms of Celtic in Britain), see the discussion in Hickey (2012).
languages, though Martin (1994) discusses some loanwords among Southeastern languages, including ‘money’, ‘black person’, ‘oyster’, ‘interpreter’, ‘peach’, ‘buzzard’, ‘turtle’, and ‘cedar’. Kimball observes that ‘Natchez seems to have been extraordinarily resistant to borrowing from European languages and from other American Indian languages. Most of the borrowings are personal and tribal names.’ (2005: 436–437). He notes loans ‘hackberry’ and ‘vulture’ from Chitimacha, ‘corn’ and ‘wild goose’, from Tunica, and ‘Creek Indian’ and ‘Alabama Indian’ from Creek. For the Southwest, Kroskrity (1982) notes the general dearth of loanwords, though Bereznak (1994) mentions borrowed ceremonial vocabulary among the Pueblos in the Southwest, not surprising given that ceremonies, with songs and prayers, were diffused.

4 Phonology

Sounds and sound patterns are usually ranked immediately after vocabulary in borrowability, because they often ride into a language on loanwords, though often the sounds which appear in borrowed words already are incipiently present in the borrowing languages. If lexical borrowing is rare, we might expect fewer phonological areal traits as well.

4.1 Sounds

Sounds do show areality in North America. Overall, consonant inventories are more elaborate in the West than the East. Ejectives are common in the West, from the Western Subarctic, through the Northwest Coast and Plateau, California, the Southwest, and into the Plains, but they are generally absent in the East. The Algic language Yurok, for example, spoken in northern California, has ejectives, while its Algonquian relatives to the east do not. In some cases, new sounds can be seen to have entered a language in loanwords. Lake Miwok, spoken in the Clear Lake area of California, contains plain, aspirated, and ejective stops, while related languages, spoken elsewhere, contain just a plain series. These did enter the language via loanwords then spread to native forms (Callaghan 1964: 47, 1987, 1991: 52). This is not the only possible route for transfer, however. Two areal traits often cited for the Southeast are a voiceless labial fricative and voiceless lateral. Each tells a different story.

Labial fricatives are present in all Muskogean languages, in the isolates Atakapa, Tunica, and Chitimacha, in the Siouan Ofo and Biloxi (marginally), in Yuchi, and in Timucua, but they are rare elsewhere in North America. One is reconstructed for Proto-Muskogean (Booker 2005: 254), where it was bilabial, though now in many languages it has become a labio-dental /f/. In the isolates, labial fricatives occur in only a few loanwords, such as Tunica káfi ‘coffee’ (< French café) (Haas 1941: 18) and Chitimacha doctrine ‘(Lake) D’Autre-Rive’ (< French) (Swadesh 1939: 19.30), reflecting the expected pattern. In the Siouan language Ofo, however, /f/ is pervasive, even in basic vocabulary. Robert Rankin (p.c.) reports that Ofo /f/ is the result of a regular sound change: Proto-Siouan *s > Ofo fh. Compare, for example, Biloxi isu, Ofo ifha ‘tooth’; Biloxi su, Ofo ifhu ‘seed’; Biloxi sa, Ofo afha ‘white’; Biloxi sindi, Ofo fxite ‘tail’ (Dorsey and Swanton 1912). Cognates in other branches of the family show s or š: Lakhota sište ‘tail’. The /f/ was
apparently not brought in via loanwords, but might have been heard as a prestigious variant of s by Ofo speakers, who then replicated it in their own language.

The voiceless lateral fricative l appears in most Muskogean languages and is reconstructed for Proto-Muskogean (Booker 2005:252). Voiceless laterals also occur in the neighboring Cherokee (but not its Northern Iroquoian relatives), as well as in the isolates Natchez, Tunica, and Atakapa. Scancarelli (2005: 360) explains their source in Cherokee. Cherokee resonants n, l, y, and w are voiceless adjacent to h (as throughout Iroquoian). There is also metathesis of lh to hl. hl sequences are pronounced as voiceless lateral fricatives. The effects of both processes can be seen in (4).

The voiceless lateral developed within the language by a familiar process, assimilation of voicelessness adjacent to h. Phonetic spirantization to l may have been stimulated by contact. Mid-sixteenth century explorers encountered Muskogean speakers in the Little Tennessee River area, but by the eighteenth century this was Cherokee territory (Fogelson 2004:338). Many Cherokee town names are earlier Muskogean forms (Booker, Hudson, and Rankin 1992: 432). In the eighteenth and early nineteenth centuries Cherokee populations expanded into northern Georgia and northeastern Alabama, ‘replacing, displacing, or assimilating previous populations’ (Fogelson 2004: 338). Muskogean speakers shifting to Cherokee could easily have substituted their native voiceless fricative lateral l for the Cherokee lateral that was already voiceless, a minor pronunciation variant. Loanwords were not necessarily involved.

For Natchez, Kimball posits a series of voiceless resonants /M/, /N/, /L/, /W/, and /Y/, but he also notes that syllable-final, morpheme-final resonants are automatically devoiced before a consonant: ʔeL-pa:-taN-ni-l-q ‘May you two look at me!’ (2005: 399). M, W, and Y were later lenited to h. The Natchez voiceless L is actually not fricative, however. For Tunica, Haas (1941: 18) explains that n, l, and r are automatically devoiced before voiceless consonants (except ʔ) and phrase-finally. She is careful to note that the Tunica voiceless lateral is also not fricative.

The voiceless lateral cited as an areal trait of the Southeast is thus more interesting than it first appears. It has varying phonetic status across the languages, a fricative in some, a simple voiceless liquid in others. It has varying phonemic status as well, a distinctive consonant in some, a predictable variant in others. Importantly, it did not enter the languages via loanwords. The devoicing of laterals adjacent to voiceless consonants could have occurred independently in each language, or it could have been stimulated by contact, as bilinguals transferred automatic habits of pronunciation from one of their languages to the other. Voiceless laterals are actually widespread in North America, also occurring in the Northwest, California, and the eastern Subarctic. Identification of the mechanisms by which they may have been transferred adds weight to their value as an areal feature.

4.2 Sound patterns
Another areal trait sometimes cited for the Southeast is fricative symbolism. Forms vary only in the point of articulation, with increasing backness corresponding to increasing intensity. Rankin (1987) cites numerous examples like those in (5) from the 1915 Byington dictionary of Choctaw, a Muskogean language with /ʃ/s/ʃ/ʃ/ alternations. Forms to the right are Rankin’s hypothesized roots, in which S represents the ablauting fricative. (Some of these include additional derivation.) Rankin notes that the process was productive.


<table>
<thead>
<tr>
<th>Choctaw</th>
<th>Rankin’s Hypothesized Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>fopa</td>
<td>murmuru</td>
</tr>
<tr>
<td>chopa</td>
<td>roar, as water</td>
</tr>
<tr>
<td>hompa</td>
<td>whoop, bang</td>
</tr>
<tr>
<td>shinka</td>
<td>to neigh, whinny</td>
</tr>
<tr>
<td>shinkachi</td>
<td>whiz like musket ball, tingle</td>
</tr>
<tr>
<td>chinka</td>
<td>to squeal</td>
</tr>
<tr>
<td>hinha</td>
<td>to groan</td>
</tr>
<tr>
<td>ak</td>
<td>noise made among dry leaves</td>
</tr>
<tr>
<td>chasha</td>
<td>to rattle</td>
</tr>
<tr>
<td>haṭa</td>
<td>stamp, tread</td>
</tr>
<tr>
<td>shana</td>
<td>to turn, twist</td>
</tr>
<tr>
<td>chanaha</td>
<td>round, coiled</td>
</tr>
<tr>
<td>Hana nukichi</td>
<td>be dizzy, cause dizziness</td>
</tr>
</tbody>
</table>

Fricative symbolism occurs in other Muskogean languages as well, such as Koasati wasiḥlin ‘to itch’, waciˈplin ‘to feel a stabbing pain’ (Rankin from Kimball p.c.) and Creek fiːˈpliː ‘unraveled’, siːˈpliː ‘torn, ripped’ (Rankin from Booker p.c.).


(6) Dakota: Boas and Deloria (1941: 16-17) cited in (Rankin 1987)

<table>
<thead>
<tr>
<th>Dakota</th>
<th>Rankin’s Hypothesized Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>zi</td>
<td>yellow</td>
</tr>
<tr>
<td>źi</td>
<td>brown</td>
</tr>
<tr>
<td>ţi</td>
<td>dark brown</td>
</tr>
<tr>
<td>sota</td>
<td>clear</td>
</tr>
<tr>
<td>šota</td>
<td>muddy, smoky</td>
</tr>
<tr>
<td>xota</td>
<td>grey</td>
</tr>
<tr>
<td>sleča</td>
<td>sliced, of bread</td>
</tr>
<tr>
<td>šleča</td>
<td>split, of logs</td>
</tr>
<tr>
<td>xleča</td>
<td>rent, of fabrics</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Winnebago</th>
<th>Rankin’s Hypothesized Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>-sox</td>
<td>‘frying sound’</td>
</tr>
<tr>
<td>-šox</td>
<td>‘bubbling sound’</td>
</tr>
<tr>
<td>-xox</td>
<td>‘breaking sound’</td>
</tr>
<tr>
<td>-zap</td>
<td>‘tear roughly’</td>
</tr>
<tr>
<td>-žap</td>
<td>‘peel’</td>
</tr>
<tr>
<td>-riš</td>
<td>‘bend wide’</td>
</tr>
<tr>
<td>-rixis</td>
<td>‘coil’</td>
</tr>
<tr>
<td>-γap</td>
<td>‘remove layer’</td>
</tr>
</tbody>
</table>

(8) Kansa: Rankin (1987)

<table>
<thead>
<tr>
<th>Kansa</th>
<th>Rankin’s Hypothesized Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>zi</td>
<td>yellow</td>
</tr>
<tr>
<td>źi</td>
<td>orange</td>
</tr>
<tr>
<td>ţi</td>
<td>brown</td>
</tr>
<tr>
<td>sabe</td>
<td>black</td>
</tr>
<tr>
<td>šabe</td>
<td>dark</td>
</tr>
<tr>
<td>leže</td>
<td>striped</td>
</tr>
<tr>
<td>leže</td>
<td>spotted</td>
</tr>
</tbody>
</table>
It can be reconstructed for Proto-Siouan and was inherited in Quapaw, from where it apparently spread to Muskogean languages. It was not transferred via loanwords: the actual fricatives involved are not the same, nor are the lexical items. The items in which it appears in the various Muskogean languages are not necessarily cognate. The pattern indicates that bilinguals can exploit patterns from one of their languages, here Siouan, in creating new forms in the other, without necessarily copying actual forms.

Another kind of sound symbolism occurs over a wide area of the West. Diminutive sound symbolism was noted early by Sapir for Wishram Chinook (1911, 1929), by Haas for northwestern California (1970), and by Langdon for Yuman (1971), and surveyed more widely by Nichols (1971). Sapir cites such examples as *i-č’iau* ‘snake’, *i-c’iau* ‘small snake’. Nichols characterizes the process as ‘the alteration, in point or manner of articulation, of consonants in verb or noun roots – expressing the diminutive category and, by extension, an attitude of endearment, affection, pity, or the like’ (1971: 826). In some languages the shifts are productive, in others unproductive but lexically well preserved, and in still others vestigial.

Nichols distinguishes three basic types of sound alternation: (i) strengthening, (ii) tonality, and (iii) apical resonant shifts. Only some types occur in each language. Her strengthening involves shifts from lenis to fortis (*s* > *ts*; *l* > *l̃*), continuant to non-continuant (*θ* > *č*, *s* > *č*, *w* > *b*), and glottalization (*C* > *C*’). The processes may affect all points of articulation or only some. Her tonality involves such shifts as *š* > *s*, *x* > *š*, and *q* > *k*. Dental, alveolar, and palato-alveolar consonants alternate symbolically, as do velar, postvelar, and palato-velar, but there is no interaction between the two groups. Her apical resonant shifts include *l* > *r*, *r* > *n*, and *l* > *n*.

Glottalization is a northern, typically Salish form of shift, concentrated particularly in the state of Washington and spread at least to neighboring Wishram. Shifts among dental resonants are centered farther south, spreading from Mexico to Oregon and central Idaho; the northernmost example, the Sahaptin and Nez Perce shift of *n* > *l*, is reversed in comparison to the more southern shifts, and its peripheral location may be partly responsible for its anomaly. No overlap in glottalizing and dental resonant shifts has been found. Tonality, however, is used in shifts throughout the area, either alone or together with any other shift. (Nichols 1971: 840).

The distribution of the phenomenon is clearly areal, crossing genetic boundaries. Nichols suggests three possible origins for it. One is dialect borrowing, which could result in doublets within a language. The alternants could then be given symbolic value (Jacobsen 1969: 150-151, Aoki 1962: 173). This could underlie Sahaptian *n* > *l* shifts. A second source could be morphophonemic alternations of consonants, especially when the conditioning factor has been lost. This could underlie Karok *r* > *n* shifts (Bright 1957: 39-40). A third source could be vowel alternations which in turn affected adjacent consonants. Rigsby and Silverstein (1969: 56) point to Sahaptian *k~q* and *s~c* alternations, which could have resulted from diminutive affixes, which triggered vowel harmony and then consonant shifts. Such events would have occurred in just a few languages, with subsequent extension of the patterns and spread through contact, as suggested by the geographical distribution of shift types. Again we find an unusual phonological pattern with a strong areal distribution not transferred via vocabulary. Bilingual speakers apparently replicated a pattern in one of their languages for expressive purposes in the other.
5. Basic replication of grammar

The strongest linguistic areas of North America show shared grammatical distinctions, categories, and structures, but the shared patterns are not generally attached to similar substance. In some cases, copying could occur relatively quickly.

5.1. Assembly from native components: Inclusive/exclusive

Many languages in the world distinguish inclusive first person (‘you and I’) from exclusive (‘he/she/it/they and I’). Jacobsen (1980) discusses the areal distribution of the trait and various paths of development. It is sometimes a family trait, but it often appears in just some members of a family; it occurs in Choctaw alone in the Muskogean family in the Southeast (Haas 1969:5), in Yuki alone in Yuki-Wappo in California, and in Shuswap alone in Salishan, and in Kwak’wala (Northern Wakashan) but not Nuuchahnulth (Southern Wakashan) on the Northwest Coast. It also goes beyond family boundaries. It is present in a continuous area including all languages of the Great Basin (Washo, the Numic branch of Uto-Aztecan) and adjacent California (Achumawi, Wintu, Miwok, Yokuts, Yuki), but it is absent from California languages outside that area, west and south of Yokuts, and from the Southwest. It is not found in languages in central California or southern Oregon It recurs, however, on the Oregon Coast (Coos, Siuslawan, and Alsea), and in Chinookan along the Columbia River and contiguous Sahaptin, but not the related Nez Perce.

Jacobsen points out that the distinction is easily diffused, because it is not bound to the syntactic structure of a language in the way that, for example, case might be. He catalogues various structural means exploited to replicate it. A new exclusive form might be built on a first person plus third person marker, as in Siuslaw, narrowing the original first person plural to an inclusive.

(9) Siuslaw: Frachtenberg (1922: 468)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>DU</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(EXCL)</td>
<td>-aʷxun</td>
<td>-nxan</td>
</tr>
<tr>
<td></td>
<td>(INCL)</td>
<td>-n</td>
<td>-ns</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-nx</td>
<td>-ts</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>-aʷx</td>
<td>-nx</td>
</tr>
</tbody>
</table>

A new inclusive form may be based on a second person marker, as in Yokuts, narrowing the original first person plural to an exclusive.

(10) Yawelmani Yokuts pronouns: Newman (1944: 231-232)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>DU</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(EXCL)</td>
<td>na’</td>
<td>na’ak’</td>
</tr>
<tr>
<td>1</td>
<td>(INCL)</td>
<td>mak’</td>
<td>may</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>ma’</td>
<td>ma’ak’</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>ama’</td>
<td>amak’</td>
</tr>
</tbody>
</table>

Other strategies can be found as well.
The inclusive/exclusive distinction is not uncommon cross-linguistically, but it is often the result of contact, as bilinguals replicate a distinction using native resources.

5.2. Stimulated reanalysis: Agent/patient systems

Alignment patterns have been hypothesized to be stable over time (Nichols 1992: 181). Yet even the rarer patterns show areal distributions. Nichols found that in her sample of 172 languages, 65% showed nominative/accusative or neutral systems, 19% ergative systems, 14% agent/patient or stative/active systems, and just 3% hierarchical systems (1992: 187). These last two patterns show strong areal distributions in North America.

Agent/patient systems distinguish two (or three) core arguments. Grammatical Agents are prototypically those who volitionally instigate and control events and states. Grammatical Patients are those who are not in control but are significantly affected by the situation. Examples of such a system can be seen in Creek, a Muskogean language of the Southeast.


\[
\begin{align*}
1SG \text{ AGENT} & \quad -éy- & 1SG \text{ PATIENT (a)}ca- \\
& a:l-éy-s & ca-tköl:s & \text{‘I’m going about’} \\
& homp-éy-s & ca-nocki:l:s & \text{‘I’m eating’} \\
& ta:sk-éy-s & ca-capákk'i:s & \text{‘I’m jumping’}
\end{align*}
\]

This is not an active/stative system: the distinction between events and states is not the determining factor. Creek grammatical Agents occur with voluntary, controlled states, and grammatical Patients with uncontrolled events.

(12) Creek (Muskogean): Martin (2004: 140, 139)

\[
\begin{align*}
1SG \text{ AGENT} & \quad -éy- & 1SG \text{ PATIENT (a)}ca- \\
léyk-éy-s & \text{‘I’m sitting’} & ca-latéyks & \text{‘I fell’}
\end{align*}
\]

A few verbs allow either an Agent or Patient, depending on volitionality: hosí:l-éy-s ‘I (AGENT) am urinating’, ca-hosí:lís ‘I (PATIENT) am urinating (unable to control it)’.

Both Agent and Patient forms occur in transitives. Third persons are unmarked.

(13) Creek (Muskogean): Martin (2004: 140)

a. \(\bar{C}i-na:fk-éy-s\)

\[
2SG.PATIENT-hit-1SG.AGENT-INDIC \\
\text{‘I (AGENT) am hitting you (PATIENT).’}
\]
b. \(\bar{O}:wa-n \quad ca-yd:ci:-s.\)

\[
\text{water-ACC} \quad 1SG.PATIENT-want-DUR-INDIC \\
\text{‘I (PATIENT) want water.’}
\]
c. \(\bar{O}:wa-n \quad ca-hos:i:i-t\)

\[
\text{water-ACC} \quad 1SG.PATIENT-want-SPONT-SS \quad \text{be-INDIC} \\
\text{‘I (PATIENT) forgot water.’}
\]

The system is thus neither ‘split intransitive’ (i.e. limited to intransitives) nor ergative.

Such systems are widespread across North America. They occur throughout the Southeast, in the Northeast, across the Plains, in the Southwest, in California, and along
the Northwest Coast. Some are ancient. In the Southeast, they can be reconstructed for Proto-Muskogean, the ancestor of Creek. In the Northeast and the Plains they occur throughout three families, Siouan, Caddoan, and Iroquoian, where they are all inherited. In the Southwest, they can be reconstructed for Proto-Kiowa-Tanoan. But the systems extend beyond these families. In the Southeast, they occur not only throughout Muskogean, the Siouan languages Quapaw, Ofo, and Biloxi, the Iroquoian language Cherokee, and the Caddoan language Caddo, but also in isolates Chitimacha, Natchez, Tunica, Atakapa, and Tonkawa, the western neighbor of Caddo. In Northern California they occur not only in all of the Pomoan languages, but also the adjacent Yuki, though not its relative Wappo. On the Northern Northwest Coast they appear in the isolate Haida and its neighbor Tingit, but not in its Eyak and Athabaskan relatives (Mithun 2008). The shapes of the markers in neighboring languages are not the same, however.

In the Southeast, the isolate Chitimacha distinguishes Agent and Patient pronominal suffixes on verbs. Pronominal suffixes exist only for first person singular and plural; second and third persons are identified by independent pronouns where necessary. The grammatical Agent suffixes (1SG -ik(i)) appear with intransitives (ʔušt’iš-ik ‘I am eating’, nušmišu-k ‘I shall work’) and transitives (k’et-ik ‘I beat (him)’, ʔam-ik ‘I see (him)’). Grammatical Patient suffixes (1SG -ki) similarly identify the single arguments of intransitives (nuc-p-ki-ću:š ‘if I die’, t’at’iwa-ki-kí ‘I felt cold’) as well as the goals/recipients of transitives (k’et-ki ‘(He) beat me’). Swadesh states that Patient forms occur with verbs denoting both events and states, with such meanings as ‘die’, ‘forget’, ‘make a hoarse sound in the throat’, ‘get sprained’, ‘shiver’, ‘fall asleep, sleep’, ‘become wearied’, ‘be tired’, ‘be afraid’, ‘be greedy’, ‘suffer pain’, ‘feel itchy’, ‘be pleased’, ‘want’. This is thus not an active/stative system. Some verbs occur with either Agents or Patients (Swadesh 1939: 94,119; Hieber ms.)

The isolate Atakapa also shows agent/patient pronominal affixes on verbs. Grammatical Agent suffixes appear with intransitives (híš-o ‘I plant’) and transitives (peni-o ‘I have healed him’). They appear with events (pál-o ‘I break it’) and states (hatpé-o ‘I am ready’). Patient prefixes similarly appear with intransitives (hí-láwet ‘I was burnt’) and transitives (hi-lóšat ‘(he) helped me’). They appear with both events (hímakaukit ‘I fell’) and states (hi-lak ‘I am strong’). (Gatschet and Swanton 1932). Second and third person singular agent affixes are unmarked, but independent pronouns are not infrequent.

The isolate Tonkawa, spoken immediately to the west of the Southeast area, also shows agent/patient patterning (Hoijer 1931). Arguments are identified by verbal prefixes and suffixes. Suffixes are used for agents. As in many such systems, third person singulars are unmarked.

(14)  Tonkawa: Hoijer (1931: 68, 72)

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Agents (Imm Present)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ge-</td>
<td>-c’</td>
</tr>
<tr>
<td>2</td>
<td>--</td>
<td>-n’ei</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>3PL</td>
<td>yagb-o’-c</td>
<td>‘I hit (him)’</td>
</tr>
<tr>
<td></td>
<td>ge-igab-o’</td>
<td>‘(he) hits me’</td>
</tr>
</tbody>
</table>
Prefixes are used for both goals/recipients of transitives, as above, and the non-volitional arguments of intransitives (-o’ = present declarative).

(15) Tonkawa: Hoijer (1931:70-71)

hedjin-o’-c  ‘I lie down’
ge:djin-o’  ‘I fall down, i.e. I lie down involuntarily, stumble and fall’
m’e:idj-o’-c  ‘I urinate’
ge-m’eidj-o’  ‘I urinate involuntarily’
ge-xadjlew-o’  ‘I am angry’
ge-die’abx-o’  ‘I have been punctured; I bleed to death’
ge-xamdj-o’  ‘I break my arm/legs’
ge-nc’ol-o’  ‘I have sores, blisters’


Kimball notes that verbs with Patient forms ‘refer to actions that are not controlled by their translation subjects. These verbs are uncommon, in contrast to the adjacent Muskogean languages.’ (2005: 439)

It is easy to see how agent/patient patterns could spread. Muskogean, Siouan, and Caddoan languages in the Southeast, as well as the adjacent Chitimacha, Atakapa, Tonkawa, and Natchez, share several other characteristics.

i  Intransitive and transitive verbs are not distinguished formally.
ii  Topical third persons are unmarked.
iii  Basic word order is predicate-final.

These features would facilitate reanalysis of a nominative/accusative system as an agent/patient system, or vice-versa.

(16) (SUBJECT) OBJECT TRANSITIVE VERB ‘It scared me (OBJECT)’

(it) me scared < > PATIENT INTRANSITIVE VERB ‘I (PATIENT) was scared’

Languages in other linguistic areas with agent/patient patterns show the same features: the Pomoan languages and adjacent Yuki of California, and Haida and the adjacent Tlingit in the Northwest. Yuki’s only relative, Wappo, shows a nominative/accusative system, as do Tlingit’s Eyak and Athabaskan relatives. We can catch a glimpse of reanalysis in the opposite direction in the speech of the last Wappo speaker. Her first language was (nominative/accusative) Wappo but she was bilingual in (agent/patient) Southern Pomo. When speaking Southern Pomo, she used the Agent pronouns as subjects and the Patient pronouns as objects.

This reanalysis is not the only mechanism by which agent/patient patterns can spread through contact. Still another Southeast isolate, Tunica, shows patterning that is
quite similar. Documentation of the language comes from three sources. Gatschet collected vocabulary and texts in 1886. Swanton collected additional material between 1907 and 1910 and compiled a grammatical sketch in 1921 based on all sources. Haas worked with the last fluent speaker between 1933 and 1939 and published a grammar (1941), grammatical sketch (1946), text collection (1950), and dictionary from all sources (1953).

In Tunica, core arguments are identified by pronominal affixes on the verb or auxiliary. Prefixes identify transitive goals and recipients. Suffixes (fused with aspect markers) identify transitive and intransitive agents.

(17) Tunica pronominal affixes: Haas (1941: 135)

(a) Tánisarahč, sáhkun, řúhtákanřákihč, támísara-hč sáku řúhk-táka-n-řáki-hč
the girl one 3M.SG-chase-CAUS-3F.SG.PFV-when she chased him
‘The girl chased one (of the puppies)’

(b) Tánisarahč, řákřámřěkěně, támísara-hč řáka-řámi-řáki-āni
DET-young.person-F.SG enter-disappear-3F.SG.PFV-HEARSAY
the girl she went in and disappeared, it is said
‘The girl had gone down (into the water) and disappeared.’


(18) Tunica pronominal prefix: Haas (1941: 60)

Tihči tīyaši.
tihči ū-yāši
she 3F.SG-be.angry
‘She is angry.’

The system looks much like the agent/patient systems in the area, but closer examination shows that it is not exactly the same. All arguments are overtly marked. Furthermore, the suffixes that identify transitive and intransitive agents of events also identify participants who are not in control, in verbs with meanings such as ‘fall’, ‘die’, ‘slip’, ‘yawn’, ‘snore’, ‘groan/moan in pain’, ‘mew/whimper/cry’, ‘cry/weep/bawl’, ‘vomit’, ‘topple over’, ‘break/snap in two (rope)’, ‘topple over’, and ‘start, jerk away as when startled’. The
crucial distinction is not control, but events versus states. This is an active/statative or
active/inactive system.

But there is another layer to the Tunica system, likely a more recent result of
close contact with the agent/patient systems in the area. There is a set of verbs termed by Haas
‘transimpersonal’. These are exactly the kinds of verbs that appear with Patient forms in
agent/patient systems: ‘bleed’, ‘wake up’, ‘cough’, ‘gasp for breath, pant (dog)’, ‘(eye) to
by a mosquito) or get a bump (from a bite)’, ‘get a cramp’, ‘be swollen’, ‘be sore’, ‘get a
‘have a fever’, ‘get hooked/caught/trapped/snared’, ‘have diarrhea’, and ‘itch’ (Haas
1953). Arguments of these verbs, not in control, are identified with the same prefixes as
the goals and recipients of transitives. These verbs also contain a pronominal suffix
‘referring to a nameless entity which cannot be expressed by a substantival referee’ (Haas
1941: 105). The same suffix is used in verbs Haas terms ‘impersonals’, such as šíhtun
*ya-ti-hč ‘when it became dark’ (dark become-3SG.SML-SUBORDINATE 1941: 58), and in
inchoatives based on statives, such as ḥu-yáhp-the ‘when he got hungry’ (M.SG-
be.hungry-3SG.SML-SUBORD 1941: 61). This suffix is also used for feminine singulars.

Some of the stems used in transimpersonal constructions also occur in intransitives:
le transimpersonal ‘lose in gambling’, intransitive ‘disappear’. Some also appear in
impersonal constructions: píra transimpersonal ‘become/turn into/revive/come to life/be
born’, impersonal ‘appear’. Some also appear in basic transitives: pála transimpersonal
‘get hooked/caught/trapped/snared’, transitive ‘catch/trap/snare’. But other stems appear
only in transimpersonals. Haas notes, for example, that ḥyú ‘wake up’ never occurs as a
transitive. This suggests that the transimpersonals were not simply basic transitives, but
were developing into a new construction. More recently, recente analysis was apparently
beginning to take them in a different direction. Haas noted that some of the
transimpersonal verbs recorded by Gatschet in 1886 were inflected as intransitives in the
1930’s by the last speaker: ‘There is evidence that stems denoting involuntary action
(e.g., ‘to breathe’, ‘to cough’) were formerly used as transimpersonals. The more usual
procedure now is to treat such stems as intransitives.’ (1941: 59).

(19) Tunica remodeling: Haas (1941: 59)

<table>
<thead>
<tr>
<th>Older transimpersonal</th>
<th>Newer Intransitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʔihkʔówikatí</td>
<td>ʔówikaní</td>
</tr>
<tr>
<td>ʔihk-ʔówi-katí</td>
<td>ʔówi-kaní</td>
</tr>
<tr>
<td>1SG.PAT-sweat-UNSPECIFIED.AGT.PFV</td>
<td>sweat-1SG.PFV</td>
</tr>
</tbody>
</table>

‘I am sweating’ or ‘I am sweating’

This last speaker, Sesostrie Youchigant, was born around 1870 and had not spoken the
language since 1915 or earlier. He also spoke French and English, so this later
remodeling could have been stimulated by contact with those languages. Heaton (ms)
points out that Sesostrie Youchigant was also apparently remodeling the shapes of
pronominal prefixes. Where all of the verbal prefixes recorded by Gatschet matched
those of alienable possessive prefixes on nouns, Sesostrie Youchigant used inalienable
possessive forms in stative and transimpersonal constructions.
In any case, Tunica provides an example of likely contact-induced replication of an agent/patient pattern through a second mechanism, with an overt unspecified/indefinite agent.

6 Gradual replication of grammar

We can see how bilinguals might replicate categories from native material (inclusive/exclusive) or reanalysis (nominative/accusative < > agent/patient). It is more difficult to imagine how speakers could replicate more abstract, tightly-integrated grammatical structure, that at the bottom of the hierarchy in (3). It is here that the length of contact can be key. Grammatical patterns need not be transferred in a single step.

6.1 Abstract grammatical structure

It is well known that word order spreads easily under contact. Mechanisms are clear. All languages allow some variation in order, though the frequency and pragmatic markedness of alternatives varies. Basic clause structure could be predicate-initial or predicate-final, for example, but constituents might appear before the nuclear clause in topicalization constructions, or after it in antitopic constructions. Under bilingualism, what may be copied is the relative frequency of alternatives, Johanson’s frequentional copying (2008). Over time, what was once a highly-marked order can become progressively less marked, even ultimately a basic order. Many Northwest languages show predicate-initial clause structure: all of those in the core Wakashan, Chimakuan, and Salishan families, the adjacent Tsimshianic and Chinookan families, and Alsea, Coosan, Siuslaw and Kutenai (strongly reflecting discourse status in some). Surrounding languages generally show predicate-final structure, more common cross-linguistically. The fact that predicate-initial order can be reconstructed for each family suggests that it spread early. It is also likely that it fostered further convergence.

6.1.1 Negation

Another feature often cited for Northwest languages involves negative constructions consisting of an initial negative predicate or auxiliary, followed by the negated clause. Describing Quileute (Chimakuan), Andrade says, ‘The negative morphemes wa or é: or the two in succession function as the main verb, and the action negated is expressed by a subordinate verb’ (1933: 268).

(20) Quileute (Chimakuan): Andrade (1933: 269)

\[
\begin{array}{ll}
\text{é:} & \text{wa-4-litš siyá’-a} \\
\text{NEG} & \text{NEG-INTENDED-2SG see-SUBORD}
\end{array}
\]

‘You do not intend to see it’.

For Coeur d’Alene (Salishan), Reichard comments, ‘The negative is almost certainly a verb, for it has many verbal characteristics. . . When used as an independent stem it
means ‘refuse’ . . . In the intransitive lutä- takes the possessive affixes with an s- prefix which may be nominal’ (1938:664-5). (y- another nominalizer.)

(21) Coeur d’Alene (Salishan): Reichard (1938: 665)
āmts ‘he shared it’
lutä-y’-āmts ‘he did not share it’
lutä-s-xʷāt’iš ‘he did not get up’

Similar constructions appear in neighboring unrelated languages.

(22) Nuuchahnulth (Wakashan): Nakayama (2003: 342)
 Wik’aƛ̕ wiinapuƛ̕ saayaa.
 wík’-aƛ̕ wí:náp-uƛ̕ sayaa
 not-FINITE stopping-MOM far
 ‘They didn’t stop for long distances.’

(23) Nisg̱a (Tsimshianic): Tarpent (1987: 357)
Niindii gipt
 ni:-nə-ti: kíp-t
 be.not-1SG.ERG-INTS eat.SG-3
 ‘I did not eat it.’

Complement constructions with an initial negative matrix are to be expected in predicate-initial languages, and could easily spread because of their fit within the existing systems. The propensity for spread is heightened by the fact that negative constructions typically cycle at a high rate, as speakers strive to restore the force of a crucial distinction which is especially susceptible to fading due to its frequency.

6.1.2 Lexical suffixes

The languages at the core of the Northwest Coast area share another notable construction involving suffixes. In many ways the suffixes resemble roots, with their large inventories, often numbering in the hundreds, and their often concrete meanings. Technically, they are suffixes, however, never serving on their own as the bases of words. They appear in both predicates and referring expressions.

(24) Coeur d’Alene (Salishan): Reichard (1938: 609)

a. Tsan-ts’ul-ts’ulxʷ-áxEn-ts
under-RDP-claw-arm-3>3
‘He clawed it under the arms.’

b. tšugw-tšugw-áxʷe-s
RDP-feather-arm-3SG.POSS
‘his wing feathers’
The meanings of the suffixes seem at first surprisingly concrete and specific for affixes, but they are often more diffuse and abstract. The Coeur d’Alene suffix -\textit{tsin}, for example is translated sometimes ‘mouth’, sometimes ‘eat’, ‘edge’, ‘shore’, etc.

(25) Coeur d’Alene (Salishan): Reichard (1938: 611)

a. \textit{gwiy’-tsin-il\textls}\-
\text{finish-mouth-3PL}

‘They finished eating.’

b. \textit{tšts-pänä”-yaR-tsí-stus-il\textls}\-
\text{hither-as.far.as-be.at.edge-mouth-CUST-3PL}

‘This way they brought it to shore.’


Similar structures can be found in unrelated neighbors.

(26) Quileute (Chimakuan): Andrade (1933: 279, 283)

\textit{Tsix hā’tšá:ƛowá’}.

tsx ha’tš-a-f-o-wa’

‘It was very good weather.’


The suffixes have free counterparts, but close scrutiny shows that the suffix is often more generic.

There are several words for the different types of canoes, but –qa may refer to any of them, as well as to a wagon or automobile. Also, there are free morphemes for bow and arrow, as well as for the modern gun . . . but all may be rendered by -pa. (Andrade 1933: 194)


Immediately to the north of this core area, the Tsimshianic languages also show similar suffixes often with relatively concrete meanings, but they are fewer in number and more transparently related to independent roots. Sm’algyax, for example, contains the suffixes –sk, -ks, -kwsa, -aks, -üks, and –iks which Dunn relates to the word aks ‘water’. The suffixes –g, –g, –xn, he relates to the word gan ‘tree/wood/stick’. The suffixes –g and –g he relates to gyet ‘man’. The suffixes –bn and –n he links to ban ‘belly’. And –mx, –mk, and –xi are possibly related to diilmx ‘respond’.

(27) Sm’algyax (Tsihsmianic): Dunn (1995: 36-37)

a. 
  batsgn
  batsk-gan
  arrive-tree
  ‘arrive in a canoe’

b. 
  yel-gan
  drill-stick
  ‘fire drill’

For all of the languages, the suffixes are necessarily translated with noun-like and verb-like English glosses, but the constructions do not specify a semantic or syntactic role.

The lexical suffix constructions of the Northwest are ancient, reconstructible for the various proto-languages. The suffix forms were not copied across family lines. But the constructions, unusual cross-linguistically, must have developed through contact, probably over an extended period. Rather than being grammaticalized gradually one at a time, they apparently originated in highly productive compounding comparable to the
noun-noun and verb-noun (noun incorporation) constructions of other languages. A very few of the suffixes show resemblances to roots in the same language or a related one like -aɫqixʷ ‘breath/smell’ and qixʷ ‘breathe/smell’ of Coeur d’Alene (Reichard 1938: 633). The initial –a of this suffix actually matches the linker that appears in compounds in the language. A source in compounding explains the large inventory of suffixes and their relatively concrete and specific meanings: constituents of compounds are drawn from the full inventory of lexical roots. Like non-heads of compounds, the lexical suffixes do not serve specific syntactic roles. The suffixes can serve the same range of functions as incorporated nouns (Mithun 1997). They are used to create lexemes for nameworthy concepts, such as the Kwak’wala da:xs ‘take aboard’, with suffix –xs ‘canoe’ (Boas 1947: 239). They are used to manipulate argument structure, as in wos wədaːw ‘have cold ears’, with suffix -ʔa ‘ear’, a use which explains the large inventories of suffixes evoking body parts. They are used to shape the flow of information, representing established or incidental entities unobtrusively. Over time, via regular processes of grammaticalization, the second constituents of compounds apparently eroded into suffixes, sometimes losing phonetic substance, sometimes gaining new or more abstract meanings as were extended in new formations, like the Kwak’wala –’sto ‘eye > door > round opening’. The grammaticalization of the compound constructions could have taken place individually in each language, but it was likely stimulated by ongoing contact.

This shared grammaticalization process is responsible for other areal features as well, such as numeral classifiers. These consist of lexical suffixes attached to number roots, as in Kwak’wala -okʷ ‘human beings’, me’ló:kʷ ‘two persons’; -xsa ‘flat’ qé:xsa ‘many (leaves)’, ’né:mxa ‘one (day)’ (Boas 1947: 240).

6.1.3 Means/manner and location/direction


(28) Atsugewi affixes: Talmy (1972: 432)
a. C’waswálmic’.
ʔ-w-ca-swal-mic’
3-FACTUAL—blowing limp.material.move-down onto ground
‘The clothes blew down from the clothesline.’

b. W’oswalic’ta.
ʔ-w-uh-swal-ic’t-a
3-FACTUAL—thrusting limp.material.move-into.liquid-EFFECTIVE
‘She threw the clothes into the laundry tub.’

The pattern crosses genetic lines between families and even the most ambitious proposed superstocks. The prefixes occur in the Palaihnihan, Pomoan, and Yuman families, and in Karuk, Yana, and Washo, all once hypothesized to be “Hokan”, but not in Shasta, Esselen, or Salinan also grouped as Hokan. They occur in the Maidun and Sahaptian families and in Klamath and Takelma, once grouped as “Penutian”, but not in Wintun, Utian, Yokuts, Coos, Siuslaw, or Alsea, also grouped as Penutian. They occur in Chumashan and Wappo-Yuki, each agreed to be a distinct family. They occur in the Numic branch of Uto-Aztecan, spoken throughout the Great Basin in eastern California, adjacent Oregon, Idaho, and into Utah and Wyoming, but not in Uto-Aztecan languages spoken further away.

Locative/directional suffixes show a similar but not identical distribution. They occur in the Palaihnihan and Pomoan families, in Karuk, Shasta, and Yana, all once labeled Hokan, but not in Yuman or Washo, also included in Hokan. They appear in the Maidun and Sahaptian families and Klamath, once grouped as Penutian, but they are not mentioned in descriptions of Wintun, Utian, Yokuts, Takelma, Coos, Siuslaw, or Alsea. They do not appear in Chumashan, Wappo-Yuki, or Uto-Aztecan.

The shapes of markers are not similar across family lines. Such tightly integrated, abstract morphological structures would not seem amenable to simple replication by bilinguals. Many of the markers are quite small, often just a consonant and sometimes just preglottalization or preaspiration. It is difficult to imagine an Atsugewi speaker deciding to replicate the ‘thrusting/ digging/sewing/leaning/propping’ prefix in (28b) above in another language. The affixes are derivational; speakers generally select full lexical items as they speak rather than assembling them online. Some languages at the edge of the area provide a clue to the mechanisms behind this areal distribution.

Both the means/manner and locative/directional constructions appear to be descended from compounds, as might be expected. We can still see the kinds of [NOUN-VERB]VERB and [VERB-VERB]VERB compounds that could give rise to means/manner prefixes at the eastern edge of the area, in the Numic languages. Tümpisa Shoshone still shows NOUN-VERB compounds in which the noun indicates an instrument, such as ki-kuttih ‘elbow-shoot’ = ‘jab with the elbow’ (Dayley 1989: 92). Tümpisa also contains a sizeable inventory of means/manner prefixes that Dayley traces to Proto-Uto-Aztecan noun and verb roots, among them ku- ‘with heat or fire’ < *kuh ‘fire’, kü- ‘with teeth or mouth’ < *küi ‘bite’, ma- ‘with the hand’ < *maa ‘hand’, mu- ‘with the nose’ < *mupi ‘nose’, ni- ‘with words, talking’ < *niya/niha ‘name’, pa- ‘involving water’ < *paa ‘water’, pi- ‘with the butt or behind’ < *pih ‘back’, sii- ‘from cold’ < *süp ‘cold’, sun- ‘with the mind, by feelings, sensing’ < *suuna ‘heart’ or *suuwah ‘notice, believe’, ta- ‘with the foot’ < *tannah ‘foot’, and tsa- ‘grasping in hand’ < Numic *tsa’i ‘grasp, hold’ (Dayley 1989: 95-96).

Examples of the kinds of compounds that could develop into locative/directional constructions can be seen at the northern edge of the area. Kathlamet Chinook of Oregon
contains a kind of [VERB-VERB] VERB compounding in which the second root indicates motion in a particular direction, such as -pa ‘motion out of’, -pq ‘motion into’, -pck ‘motion from open to cover, especially from water to shore or inland’, -ƛx ‘motion from cover to open, especially toward water’, -tį ‘motion, position down’, uulx ‘motion up’: ikuxuni-pck ‘she drifted ashore’ (-xuni-pck ‘drift-from.water.to.shore’). These roots also occur independently: txú-pck-a ‘Let us go inland’ (Hymes 1955: 218).

Confirmation of the compounding origins of these prefix and suffix constructions can be seen in what Jacobsen (1980b) and DeLancey (1996) term ‘bipartite stem’ constructions: some of the languages contain verb stems which now appear to consist only of a prefix and a suffix. Maidu, for example, contains a means/manner prefix wi- ‘with hands’ and a suffix -doj ‘upward’: wi-dók-doj ‘grasp with the hand and pull up’ (Shipley 1963: 187). It also contains stems like wi-dój ‘pull (something) up’. Such stems are likely descendants of lexicalized compounds.

While it is unlikely that bilinguals would replicate the small, abstract affixes seen in many of these languages directly, they could easily replicate compounding types, with initial means/manner roots or with final locative/directional roots. Over time, perhaps also stimulated by ongoing contact, frequently-recurring elements of such compounds could easily erode into affixes. Various stages of development can be seen across the area, with affixes of varying productivity, varying degrees of abstraction, and varying phonological size.

6.2 Alignment: Hierarchical systems

The alignment type Nichols found to be the rarest cross-linguistically is hierarchical, where ‘access to inflectional slots for subject and/or object is based on person, number, and/or animacy rather than (or no less than) on syntactic relations’ (1992: 66). This type appeared in only 3% of her sample. Hierarchical systems also show areal patterning in North America, and it is possible to discern mechanisms behind their geographical distribution (Mithun 2012b).

6.2.1 Northwest Coast

In Nuuchahnulth (Nootka), a Southern Wakashan language spoken on Vancouver Island, British Columbia, arguments are identified by first and second-person enclitics to predicates. Third person is unmarked. The first person singular is =s.


\(\text{wa}t\tilde{s}\text{i}a\tilde{\lambda}=s\)

‘I went home.’

\(\tilde{\eta}unaak=s-i:\$\)

‘I have a friend.’

\(\tilde{\eta}ith\tilde{s}\text{i}=s\)

‘I cried.’

\(\tilde{\eta}uyimt\tilde{c}k\tilde{w}=s\quad c’u\tilde{\eta}i\tilde{c}h\).

‘I was born in winter.’

\(\text{waa}\tilde{\eta}\tilde{\lambda}\tilde{a}t=s\ldots\)

‘He said to me.’

427
This pronominal enclitic is not simply a subject, object, absolutive, ergative, agent, or patient. This is a hierarchical system, in which only one argument may be specified. If there is a first or second person, this referent takes precedence over a third person. If there is both a first and second person, the agent takes precedence over the patient (imperatives allow both). The system is thus 1, 2 > 3, Agent > Patient.

The hierarchy is maintained through the use of a suffix –at. Much like a passive in many languages, it functions to ensure that the discourse topic is cast as a core argument (overt or not). Speaker George Louie was describing how to make a canoe. One begins by looking for a cedar in the forest.


ʔuuwaʔat  húmap,  
ʔu-wa-’at  húmapat  
it-find-SHIFT  cedar.tree  
‘You find a cedar tree

yaəp’ičuʔat.  
yac-p’ič-u-’at.  
step-at.base.of.pole-MOM-SHIFT  
and then step up to the base of it.’

Hiiistiʔat  n’an’aan’ičat,  ḥustʔas  takqiinuʔat,  
hiista-’iƛ-’at  n’an’a :n’ič-’at  ḥust-’as  tak-qi:nu-’at  
there-start.from-SHIFT  look-SHIFT  LOC-on.ground  facing-on.top.MOM-SHIFT  
‘You look it over from the ground to the top,...’

The -at suffix was used here to indicate that though the cedar was a semantic goal, it was the discourse topic at this point in the discussion. The same suffix is used whenever a third person acts on a first or second.


waaʔaƛat-s . . .  
wa :’aƛ-’at=s  
say-FINITE-SHIFT=1SG  
‘He said to me . . . ’ (‘it was said to me’)

This suffix cannot be used if a first or second person acts on a third.

The two other Southern Wakashan languages, Nitinaht (Ditidat) and Makah, spoken to the south, also show hierarchical systems, maintained by suffixes cognate with the Nuuchahnulth -at. Here the suffixes are also used pervasively to maintain topicality through discourse, as well as whenever a third person acts on a first or second. The hierarchical system has not penetrated quite as far as in Nuuchahnulth, however. When both first and second persons are present, both can be identified by clitics. The system is thus 1, 2 > 3.

There is only a hint of the hierarchy in the Wakashan language immediately to the north of Nuuchahnulth, Kwak’wala. Here subjects are identified by enclitics attached to the first element of the clause, and objects by suffixes on the predicate. The language does not show a hierarchical system like those of its Southern Wakashan relatives, but there is a significant gap in the pronominal object paradigm: there are no suffixes for first
person objects. Periphrastic constructions are used in their place, built from a predicate ‘come’ for actions directed toward a first person, and ‘go’ for those directed toward a second or third. The other North Wakashan languages spoken still further to the north, Heiltsuk and Haisla, contain full sets of pronominal subject clitics and object suffixes, used in all combinations.

To the south of the Wakashan family is the Chimakuan language Quileute. Here, too, a hierarchical system can be seen, but it has not penetrated quite as deeply as in Wakashan, affecting only second persons: $2 > 3$. If a third person acts on a second (‘He saw you’), a passive suffix must be used, as in Wakashan (‘you were seen’), though the forms of the markers are completely different.

To the east of the Wakashan and Chimakuan families is the Salishan family. Some Salishan languages also show hierarchical systems, those closest geographically to Nuuchahnulth. To the east along the Central Coast, Northern Straits and Klallam privilege first and second persons over third, again maintaining the hierarchy via obligatory passivization: $1, 2 > 3$. Immediately to the north, Halkomelem and Squamish privilege just second persons over third: $2 > 3$. Salishan languages beyond this area show no restrictions.

The distribution of the hierarchical systems of the Northwest is clearly areal. Nuuchahnulth is at the core, privileging first and second persons over third and, among first and second, agents over patients. Immediately adjacent languages privilege just first and second over third. Languages immediately beyond these privilege just second over third. The hierarchies in all of them are maintained by obligatory passivization, though the markers differ across family lines.

It is easy to see how this areal distribution could come about. In languages with subjects various factors can enter into subject selection: person (first and second over third), humanness, animacy, givenness, identifiability, agency, etc., features associated with discourse topicality. A propensity to privilege first and second persons over third would be easily spread by bilinguals, becoming crystallized as a grammatical requirement to varying degrees. (Mithun 2007b).

6.2.2 Northern California

Another hierarchical area occurs further south, but it developed in slightly different ways. The Chimariko, Yana, Yurok, and Karuk languages of Northern California are genetically unrelated, but all exhibit hierarchical systems to some degree. The foundations of the pronominal systems differ from language to language, as do the forms of their markers. Chimariko shows basic agent/patient alignment, Yana and Yurok nominative/accusative alignment, and Karuk a mixture of the two.

In Chimariko, only one core argument is represented pronominally on any verb. First and second persons have priority over third; if both parties are speech act participants, the agent takes priority over the patient: $1, 2 > 3$; Agent $>$ Patient. The roles of first person arguments are distinguished by the shape of the pronominal affix: grammatical Agents have one form, and grammatical Patients another. Thus ‘I hit him’ = $1SG.AGENT$-hit, ‘He hit me’ = $1SG.PATIENT$-hit, and ‘You hit me’ = $2SG.AGENT$-hit. Independent pronouns can be added for clarification.

In Yana, core arguments are represented by pronominal suffixes on verbs. Third persons are unmarked: ‘I will eat it’ = ‘I will eat.’ If a third person agent is involved in a
transitive event, however, an element -wa appears in the pronominal suffix complex. Its source survives in modern Yana as a passive marker. The only way to say ‘he hit me’ is literally ‘I was hit’; ‘he hit you’ is literally ‘you were hit’. When both arguments are speech act participants (1/2, 2/1), the suffix includes -wa- plus a patient subject pronominal. The verb meaning ‘I love you’ is essentially ‘you are loved’. Such sentences are no longer necessarily interpreted as passives, however; they may contain overt lexical agents. (Other elements have since been added to the suffix complexes, so they are no longer identical with passive constructions.)

The Yurok language is spoken in the same area as Chimariko and Yana, but it is not related to either. It is Algic, remotely related to Algonquian. Yurok indicative verbs carry pronominal suffixes identifying their core arguments. The system appears to show a nominative/accusative basis. The same pronominal suffixes appear in intransitives (‘go slowly’, ‘be in pain’) and transitives (‘hear someone’). Some of the transitive pronominals are still transparent combinations of an object marker followed by a subject marker.

But not all core arguments are represented overtly in all combinations. Third person transitive patients are often not expressed at all. The pronominal inflection of ‘we hear them’, for example, contains no object marker for ‘them’: the form is the same as that for ‘we hear’. Third person transitive agents are also sometimes not overtly represented. The form for ‘he meets me’ contains no ‘he’. Such verbs do, however, show an extra element -y-. Its source is clear: a passive marker. Together these observations suggest a hierarchical system based on person priorities. The hierarchy has not penetrated the entire transitive paradigm, however. For some transitive combinations, there are choices between a marker overtly specifying both arguments and a passive formation. Taken together, the endings reflect a hierarchy 1PL > 2 > 3SG > 3PL. The strategies apparently used to achieve the priorities in Yurok are reminiscent of those in Yana: omission of some third person transitive patients and passivization in certain contexts. The Yurok pronominal suffixes, like those in Yana, do not constitute a regular, synchronic system, however. They reflect earlier priorities whose traces have become frozen in the pronominal strings.

Karuk is spoken directly to the east of Yurok. Core arguments are identified by pronominal prefixes on verbs. First person prefixes distinguish subjects and objects, though Bright (1957: 59) notes that one set of intransitive verb stems allows object forms, suggesting an incipient agent/patient pattern: ‘be hungry’, ‘be jealous’, ‘be afraid’, ‘bleed’, ‘be thin, lose weight’, ‘defecate’, ‘be cross-eyed’, ‘fall asleep’, ‘be tired’, ‘burn oneself’, ‘be hot’, ‘feel pain’, ‘thirst for’, ‘be sick’, ‘have good luck with’, ‘be unwilling, lazy, tired’, ‘be cross-eyed’, ‘be bald’, ‘be nervous, cranky, fretful’, ‘be late, be offended’, etc. But only one argument can be expressed within a Karuk verb. The choice of argument depends on person and number. Traces of a person hierarchy similar to those in Chimariko, Yana, and Yurok, can be perceived. First and second persons are always chosen over third: 1, 2 > 3. When a first or second person acts on a third, the third person is omitted, as in the other languages. Second person plurals have priority over all other participants: 2PL > 1 > 2SG > 3.

When a third person acts on a second person (3/2), the prefix refers to the second person, but a suffix -ap appears on the verb. There is no passive construction in modern Karuk (Macaulay 1992, 2000: 475), but the distribution of the -ap suffix suggests that it may have had a passive origin. The suffix appears in clauses describing transitive events in which the semantic agent would normally be viewed as less topical or lower on a
person hierarchy than the semantic patient, such as ‘A monster is going to eat you’ or ‘They do not like you’. Lexical agents in such constructions also carry a special postposition ʔi:n, labeled an agentive marker by Bright, suggestive of an earlier oblique marker for passive agents. The postposition occurs only in clauses describing semantically transitive events.

The four languages in this area of California thus share hierarchical alignment, but there is no shared substance, and the systems differ in detail and in entrenchment. As on the Northwest Coast, a propensity to privilege speech act participants over others was apparently spread across languages by bilinguals. What was transferred was not a grammatical structure, but rather a behavior pertaining to preferred choices among constructions already available in the languages. Over time the stylistic tendencies became crystallized in grammar, in different constructions with similar effects.

6.3 From posture to aspect

Evidence of the gradual development of grammatical patterns stimulated by contact can be seen in another cluster of constructions in the Southeast. A striking areal trait is the propensity of speakers to specify the posture or position of entities, typically ‘sitting’, ‘standing (vertical)’, ‘lying (horizontal)’, and often ‘moving’.

(32) Creek (Muskogean): Haas and Hill (ms 39, 26, 685), Martin (2011: 354)

Ci:pâ:na:t hî:ci t â:levkâti:s. ‘The boy was (sitting) there watching.’
Afânâ:kit sîho:katí:s. ‘They were (standing) looking around.’
Ísti hâmkit inôkkí:t wâ:kkâti:s. ‘A man was (lying) sick.’
Icînâpî:1 ahi:n payës o:mi:s. ‘I (go about) opposing you.’

(33) Quapaw (Siouan): Rankin (2005: 484, 459, 467, 458)

Koâsâkdábá tháhe má. ‘I was (standing) working for him.’
Áwîtwaâvé ažáâtta miêhe. ‘I will be (lying) watching you.’
Kôišóta ekízhì bdé ttânihë. ‘So I am going to go (moving) somewhere else.’

(34) Biloxi (Siouan): Dorsey and Swanton (1912: 28, 26, 275, 150)

K’hôni naxé nákí. ‘His mother was (sitting) listening.’
Skák’hônañi naxé nêdi. ‘The Opossum (stood) listening.’
Te ó mákí. ‘He was (lying) dead.’
Nkákítupe nkâdëdi. ‘I was (moving) carrying it on my shoulder.’

(35) Natchez (isolate): Mary Haas (p.c.)

O:ya sanaskuk kaseNeckik. ‘He was (sitting) eating persimmons.’
Waskupe: sewetik súcik. ‘The dog was (lying) dead.’
ŠínakaY suhtik. ‘He was (going) carrying it on his back.’

(36) Tunica (isolate): Haas (1946: 50, 51)

Sâku- Holocaust. ‘He was (sitting) eating.’
Látihč hër-ʔunáná-ní. ‘At night they kept watch (sitting).’
Léhpi-náuni-hk-ěni. ‘They two (alligators) were (lying) blocking it.’
ʔusólʔuwańi. ‘He was (going) creeping up on it.’

(37) Chitimacha (isolate): Swadesh (1946: 332, 322), texts A35d, A49b from Daniel Hieber (p.c.)

Waʔaš his kečmi:k’ hiʔin. ‘He was (sitting) waiting for the others.’
Him yaʔa ni wopmi:k’ ʔap ču:k’čiʔin ‘I’ve come (standing) to ask for your daughter’
Weʔ nekš kap nu:pkš pen. ‘That devil lay dead.’
Ku:kš mi:pk pentka, . . . ‘Although he lay drunk, [he wanted more].’

(38) Atakapa (isolate): Gatschet and Swanton (1932: 64, 123, 94)

Ké-ukámškinto. ‘I am (seated) paddling.’
Náu tá-uwalwálekit. ‘The feather is (standing) waving.’
Iškalí-núl-wilwilhiento I rock a child (lying down).’

In most of the languages, the posture is indicated by an auxiliary construction.

(39) Quapaw (Siouan): Rankin (2005: 459)

Wákiwébdábda théhe má. 1SG.AGT-BEN-INDEF-1SG.AGT-work 1SG.AGT-stand CONT 1SG.AGT-do.IMPF
‘I was (standing) working for him.’

(40) Biloxi (Siouan): Dorsey and Swanton (1912: 150)

Nkítxa nkákitupé nkadédi
nkí-txa nká-ki-tupe nká-de-di
1SG.PAT-alone 1SG.AGT-carry.on.shoulder 1SG.AGT-go-M.DEC
‘I carried it (moving) on my shoulder alone.’

(41) Creek (Muskogean): Martin (2004: 279)

Łaló pasát-i:pít akhoylí:pin.
łaló pasát-i:p-i-t ak-hoy-lí:p-i-n.
fish kill-SPONT-I-SAME.SUBJECT in.water-stand-SPONT-I-DIFF.SUBJ
‘He was standing there killing fish.’

(42) Chitimacha (isolate): Swadesh (1939: 238)

Ni:kí:k’ peken.
niːk-iːk’ pe-ke-n
sick-PARTICIPLE.SAME.SUBJECT lie-1SG-CONT
‘I am (lying) sick.’

These constructions are not exactly equivalent. All contain a lexical verb followed by an inflected auxiliary. In the Siouan languages, the lexical verb is also inflected, in
Muskogean it is not inflected but it ends in a same-subject linker, and in Chitimacha it is a same-subject participle. Swadesh notes that the Chitimacha neutral ‘sit’ auxiliary hit(h) had begun to contract: it ‘often loses its initial h and amalgamates with a preceding participle when the latter ends in -š (an optional element in participles), e.g. teyk’s hin > teykšin ‘he is seated’ (1939: 25.47). In Atakapa the positionals are generally loosely-attached verbal prefixes.

(43) Atakapa: Gatschet and Swanton (1932: 64)
Kéukámškinto.
ké-u-kam-š-kint-o
sit-HAB-paddle-be-CONT-1SG.AGT
‘I am (seated) paddling.’

In Tunica they are now an integral part of the verb word, which consists of a bare lexical root followed by the inflected positional.

(44) Tunica: Haas (1946: 50)
Látihč
hēr-ʔunaná-ni.
at.night watch-3M.DU-HEARSAY
‘At night they kept watch (sitting).’

The auxiliaries do not necessarily match such independent lexical verbs as ‘sit’, ‘stand’, ‘lie’ in form. Specification of position is not unknown elsewhere in the world. Many of the examples above have English counterparts (‘they sat waiting’). The pervasiveness of the postural auxiliaries across the Southeast suggests an areal trait, however. In a series of works, Rankin (1977, 1978, 2004, 2011) has demonstrated that it originated in the Siouan family. The positional roots ‘sit’, ‘stand (animate)’, ‘stand (inanimate)’, ‘lie’, and ‘move’ can be reconstructed for Proto-Siouan. In the modern Muskogean languages, positionals are used in similar ways, but they are not generally cognate across the family.

(45) Muskogean positional verbs: Rankin (1978)

<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th>DUAL</th>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘sit’</td>
<td>Choctaw-Chickasaw</td>
<td>binili</td>
<td>chiya</td>
</tr>
<tr>
<td></td>
<td>Creek-Seminole</td>
<td>talaya</td>
<td>taloha</td>
</tr>
<tr>
<td></td>
<td>Hitchiti-Mikasuki</td>
<td>čokooli-</td>
<td>wiik-</td>
</tr>
<tr>
<td></td>
<td>Alabama-Koasati</td>
<td>čokoka-</td>
<td>čikiika</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Choctaw-Chickasaw</th>
<th>hikiya</th>
<th>hiili</th>
<th>(hi)yoh-</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘stand’</td>
<td>Creek-Seminole</td>
<td>hoyílita</td>
<td>sihóókita</td>
<td>sapááklita</td>
</tr>
<tr>
<td></td>
<td>Hitchiti-Mikasuki</td>
<td>hačaali-</td>
<td>lokooka-</td>
<td>lokooka-</td>
</tr>
<tr>
<td></td>
<td>Alabama-Koasati</td>
<td>hačaali-</td>
<td>hikiili-</td>
<td>lokooli-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>‘lie’</th>
<th>Choctaw-Chickasaw</th>
<th>ittola</th>
<th>kaha</th>
<th>kah-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Creek-Seminole</td>
<td>wakkita</td>
<td>wakhokita</td>
<td>lomhita</td>
</tr>
</tbody>
</table>
The mismatches across the languages suggest that the distinction came into Muskogean after they had diverged. Positional distinctions were replicated from different native resources. Rankin provides detailed discussion of the formations. Choctaw and Chickasaw singular and plural binili and binohli ‘sit’ are derived from a noun denoting a type of abode: bina(h) ‘lodge, tent, camp’. The Hitchiti-Mikasuki čokoo(l)- and Alabama-Koasati čokoo(ka)- ‘dwell, sit’ came from the Proto-Muskogean *čokk- ‘house’. Choctaw and Chickasaw contain verbs derived from this noun, but they mean ‘enter’ and are not part of the positional paradigm. The Choctaw-Chickasaw dual čiya is the passive form of the verb celí ‘to bear, bring forth young’, so ‘to enter’ or ‘to exist’. Rankin notes that the use of verbs meaning ‘be born’ as auxiliaries is also characteristic of the Alabama-Koasati ‘sit’ positionals. The Creek cognates of these forms mean ‘enter’, and are not positionals.

The modern Choctaw ‘stand’ plurals hiyohli and hiyohmaáya originated as duals containing the dual marker *-oh-. The modern dual hiíli is based on the same root hi- ‘upright’ with the active suffix -li. The Koasati singular, dual, and plural positionals for ‘stand’ were formed from three different roots. All of the Hichiti-Mikasuki and Alabama-Koasati non-singular positionals are based on a Common Muskogean root *lokoo- ‘be in a group’. Cognates in Creek, Choctaw, and Chickasaw are not positional in meaning: Choctaw lakoli ‘to collect, flock, cluster, huddle’. Creek loyétv ‘to fill’ (as a container). Rankin reconstructs the ancestor of the Creek plural ‘stand’ as *(i)si-apaak-lita ‘to have or hold together in a group.’ Both the Choctaw inanimate ‘sit’ positionals and the Hitchiti-Mikasuki singular ‘lie’ are traced to a general locative verb ‘remain, be fixed’.

The various ‘lie’ positionals developed from verbs of falling, crawling, placing, and grouping. The Choctaw singular ittol- comes from the verb ittola ‘fall, remain, lie’. Creek has a cognate toletv ‘fall’. The Choctaw dual and plural kah- are based on the root kah- ‘lie, fall down’. Creek has cognates outside the auxiliary system, with the suppletive plural kááyita ‘to lay eggs, have young’. Rankin reconstructs the common ancestor as ‘put, place’. The Hitchiti-Mikasuki singular ‘lie’ positional talaakom is cognate with the Choctaw talaya, taloha ‘sit’, a more general locational. The Hitchiti dual and plural ‘lie’ positional sol(k)- is cognate with Creek sulkii ‘many, much, a herd’. The Alabama-Koasati ‘lie’ positional bal(a)- matches the regular Choctaw and Chickasaw verbs bala-li ‘crawl, creep’.

Positionals occur in other constructions as well. They appear in existential/presentative/locational constructions throughout the area: ‘There is a house (sitting) over there’, ‘There is a stream (lying) over there’, etc. They occur in possessive constructions. They also mark aspect. Already in Proto-Siouan, they had undergone grammaticalization to durative or continuous markers.

(46) Osage Dhegiha (Siouan): Quintero (2004: 312)

<table>
<thead>
<tr>
<th>Awáachi</th>
<th>aqhé.</th>
<th>awáachie</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-waachi</td>
<td>aqhé</td>
<td>wa-waachi-de</td>
</tr>
<tr>
<td>1SG.AGT-dance</td>
<td>1SG.CONT-moving</td>
<td>1SG.AGT-dance-DECL</td>
</tr>
<tr>
<td>‘I’m dancing.’</td>
<td></td>
<td>‘I danced’.</td>
</tr>
</tbody>
</table>
Quintero notes that in Osage, all positionals can serve as basic continuatives distinguishing position, but individual forms have taken on specialized implications: ‘moving’ auxiliaries imply that the action has been ongoing for some time, and the ‘standing’ continuative can convey imminence, being on the verge of undertaking an activity (2004: 315).

The positional aspect markers now co-occur with lexical positional verbs.

(47) Biloxi (Siouan): Einaudi (1976: 154)

a. *Xe nak†i.*
   sit sitting.\ CONT
   ‘She is sitting (sitting).’

b. *Tox maki.*
   lie lying.\ CONT
   ‘He was lying (lying)

c. *Sîhîx ne.*
   stand stand.\ CONT
   ‘It was standing (standing).’

Kaufman (2013:298) notes that in Biloxi, certain positionals were coming to be routinely associated with particular entities: the sun, forests, lakes, and villages sit; rain, hens and rivers lie; thunder moves.

Similar developments have occurred elsewhere in the Southeast to varying degrees. In some languages, all of the positionals serve as durative/continuative/progressive markers. In others, just one or two have become specialized in this function.

(48) Creek (Muskogean): Martin (2011: 304,304, 443)

a. *Im-iti†a:k-aha:nî-t apô:ki-t ó:-s.*
   DAT.APPL.get.ready-going-to-SAME.SUBJECT sit.PL-SAME.SUBJECT be-INDIC
   ‘They’re [sitting waiting] going to get ready.’

b. *Ahî:ci-t leyk-icki-n yéyc-al:i:-s*
   watch-SAME.SUBJECT sit.SG-2SG.AGT arrive.SG-FUT-DUR-INDIC
   ‘Keep [sit] looking, and they’ll come.’

c. *Ahkopan-î:t-t al-î”p-at-i:-s.*
   play-SPONT-SS go.about.SG-SPONTANEOUS-ANT-DUR-INDIC
   ‘He kept on playing.’

Martin notes that in their aspectual use, the forms based on ‘sit’ need not involve actual sitting. The aspectual auxiliary in (49) based on *apô:k* ‘sit pl.’ could be translated as ‘settled down’.

(49) Creek (Muskogean): Martin (2011: 304)

\begin{verbatim}
Opân-ka:t-t tàyi: hi”i-n apâ:nî-t
dance.GER-FOC be-DUR very-DIFF.SUBJ dance-SAME.SUBJ
apô:ki-t ômho:yi-n . . .
sit.PL-SAME.SUBJ be-DIFF.SUBJ
\end{verbatim}
‘And as they were really beginning to dance, [he sat down].’

The Natchez aspectual use of auxiliaries is less developed, but beginnings can be seen. The motion verb ‘be about’ could be spatial in meaning or temporal below.

(50) Natchez (isolate): Kimball (2005: 386)

\[ʔohoti:nuhc cu:tahaw pološaL  šupitine . . .\]
\[ʔohoti:nuh-c cu:tahaw polo-š-al-ki šu-piti-ne\]

wildcat-ERG rail split-HEARSAY-AUX-CONN HEARSAY-be-about-WHEN

‘When Wildcat was around splitting rails, so it is said, [Turkey arrived].’

The Tunica positionals ‘sit’ and ‘lie’ are also used as aspectual markers. With a basic verb they are stative or durative. With habitual aspect they function as progressives.

(51) Tunica (isolate): Haas (1941: 50)

\[kátaňň, . . . títihiʔé hopiʔurahč\]
\[kátaňň títihiʔé hopiʔ ūra-hče\]
where river get.out-M.SG.lie-SUBORD

‘where a river came out’

\[Táwišihč  híštaha wicíchʔaráni.\]
\[ta-wiší-hč  híštahaki wíčí-hk-ʔará-ní\]
ART-water-F.SG still rise-HAB-F.SG.lie-HEARSAY

‘The water was still rising.’

The Chitimacha ‘sit’ (neutral), ‘stand’ (vertical), and ‘lie’ (horizontal) positionals are also exploited for aspectual meanings.

(52) Chitimacha: Hieber (p.c.) from Swadesh texts A36d, A1b, A30d

a. ʔasi nanc’ip’u hikinakiš.
\[ʔasi nahe’ip’u hi-ki=mk=i=š\]
man small sit-1SG.PAT-LOC=NZR-SUBORD

‘When I was a boy’

b. k’ihkite hikin.
k’iht-ki-te hi-ki-n
want-1SG-PTCP sit-1SG.AGT-CONT

‘I am (sitting) wanting (it)’

c. Teyk’iš hiʔi.
\[tey-k’iš hi-ʔi\]
sit-PARTICIPLE.SAME sit-3SG

‘He just sat (sitting) there.’

d. Nat’ik’iš pekin.
\[nat’i-k’iš pe-kin\]
lie-PARTICIPLE.SAME lie-1SG

436
'I am (lying) lying down.'

The horizontal ‘lie’ auxiliary is often translated as an anterior.

(53) Chitimacha: Swadesh (1939: 234, 255, text A30e)

a. \(\text{We}^\nu \text{ po: sek'is tapši:k' c}i\)
\(\text{we}^\nu \text{ po: sek?is tapši-k' c}i\)
that plant among be.standing-PARTICIPIAL.SAME.SUBJECT stand
‘She had stood among those plants.’

b. \(\text{Wetk kas tuhyt:k' pe?anki}\)
\(\text{we-t-k ka tuhyte-}:ik' \text{ pe-}\text{=}nk=i\)
DEM-RFL-LOC back stoop-PARTICIPIAL lie-3SG=LOC-NZR
‘When he had stooped down...’

The co-occurrence of the auxiliary hi- ‘sit’ with the lexical verb tey ‘sit’, the auxiliary c\(i\) ‘stand’ with the verb tapši- ‘stand’, and pe- with nat’i- ‘lie’ confirm their aspectual function.

The Chitimacha positionals have become still further grammaticalized as elements of the verbal continuative aspect suffix, which consists of a continuative element -\(\bar{\text{i}}\text{s}\)- plus a positional.

(54) Chitimacha continuative: Swadesh (1939: 239)

\(\text{Kow-\(\bar{\text{i}}\text{s}\)-c}\bar{\text{i}}\bar{\text{t}}-i}\)
\(\text{call-CONT-standing-3SG}\)
‘He stood calling.’

Swadesh (1933) notes that the vertical (‘standing’) positional has developed a refined or respectful connotation, and the horizontal (‘lying’) a derogatory or disrespectful one.

The distribution of postural/positional constructions through the Southeast raises intriguing questions about the mechanisms behind the development of such areal traits. Postural distinctions originated in the Siouan languages and were well-established in existential/locational, possessive, auxiliary, aspectual, and demonstrative constructions in those languages in the Southeast. Speakers of other Southeastern languages apparently replicated the distinctions using native material. But without a philological record, the relative timing of developments can only be a matter of conjecture.

It is not uncommon cross-linguistically for postural verbs to develop into continuative/durative/progressive markers. Such developments have been noted for Ngambay, Diegueño, alyawarra, Imonda, Kxoe, Tatar, Tamil, Diola-Fogny, Mamvy, Nobii, Tibetan, and Kabyle, in addition to Siouan (Bblansitt 1975, Heine 1993, Bybee et al 1994, Kuteva 2001). Kuteva proposes a sit/stand/lie auxiliation chain, with examples from Swedish, Norwegian, Danish, Dutch, and Bulgarian (2001: 43-74). The postural verbs are first used with human subjects to specify the orientation of the human body in space (‘She is (sitting) over there’). Their meanings include inherent stative semantics, or temporal ‘unboundedness’. They can also appear in coordinate bi-clausal structures with co-referential subjects (‘She is sitting on the couch and writing a letter.’) Next they are extended to express the canonical spatial position of physical objects (‘The clothes are
Finally the inherent temporal unboundedness becomes a focal feature, yielding continuative/durative/progressive markers, (perhaps first with inanimates and then animates), and the bi-clausal structure is reanalyzed as monoclausal.

Each of these stages persists in the Southeastern languages. Any of them could have occurred in any of the languages in isolation. Given the geographical distribution, however, it is more likely that each was stimulated and/or facilitated by contact. Bilinguals may have begun by simply increasing the frequency of postural specification, with constructions that already existed in the languages. Specification via auxiliary constructions was not a big step, since auxiliary constructions already existed in the languages at various stages of development (cf Munro 1985 on Muskogean). The Atakapa positional prefixes may have been stimulated by contact with the neighboring Caddo, which contains stem-initial sit/stand/lie postural elements, cognate with other Caddoan languages out of the area (Chafe 2005: 432, David Rood p.c.).

The result is what Heine and Kuteva (2005: 5.2.1) term a ‘grammaticalization area’: ‘a group of geographically contiguous languages that have undergone the same grammaticalization process as a result of language contact.’

7 Conclusion

North America, with its vast territory, large numbers of languages and genetic units, and diversity of environmental, social, cultural, and linguistic situations, provides an excellent arena for investigating the development of linguistic areas. The strongest areas have emerged out of long-term, intensive contact among small communities, in relatively densely-populated areas with long-standing intermarriage practices and multilingualism. Significantly, the traits that characterize these areas go against many expectations about the relative ease of borrowing: vocabulary > sounds > sound patterns > syntax > morphology. They generally show little lexical borrowing, but often extensive parallelisms in abstract structures that are deeply embedded in grammar. At least two kinds of factors might be responsible.

The first kind are social and cognitive. In many areas of North America, there was a marked absence of extensive code-switching: so far as is known, bilinguals worked hard to keep their languages apart, often for social reasons. In some cases language was viewed as a sign of identity, while in others it was simply socially appropriate to speak the language of the community one was in. In both cases, attention was paid to those aspects of language most accessible to consciousness, particularly vocabulary choice. Second-language speakers tend to focus on vocabulary as well. Less attention was paid to tightly-integrated grammatical constructions, particularly morphology, which tend to be less accessible to consciousness and less isomorphic across unrelated languages.

The second kind of factors involve the duration of contact. Some parallelisms can develop relatively rapidly. A bilingual may copy an inclusive/exclusive distinction in first person pronouns by simply combining existing first and second person forms to create a new inclusive form. Language learners may reanalyze a nominative/accusative system as agent/patient or vice-versa. Other parallelisms can develop more gradually, often in layers. The early spread of predicate-initial order can set the stage for certain kinds of complement constructions, like the negation on the Northwest Coast. The early spread of predicate-final order could facilitate the parallel development of auxiliary constructions. Grammatical parallelism can also arise simply from recurring patterns of expression.
Language communities differ not only in what speakers are required to specify, but in what they tend to express most often. Bilinguals might easily carry a propensity for elaboration of manner, location, or posture from one to another, exploiting native lexical resources. They might carry over relative frequencies of stylistic alternatives like subject selection without violating grammatical norms. The recurring choices can become routinized and undergo further integration into the grammar. Such grammaticalization processes could occur in any of the languages in isolation, but they can also be stimulated by ongoing contact, as speakers extend the uses of constructions in one language on the model of the other, like specification of position to continuative aspect in the Southeast.

There is much we will never know about the circumstances underlying the emergence of linguistic areas, but we continue to discover routes by which they can develop.

References


Heaton, Raina ms. Active/stative agreement in Tunica. University of Hawaii.


Kroskrity, Paul 1982. Language contact and linguistic diffusion: The Arizona Tewa speech community. Bilingualism and Language Contact: Spanish, English, and


Swadesh, Morris 1933. Chitimacha verbs of derogatory or abusive connotations with parallels from European languages. Language 9:192.
31 The areal linguistics of Amazonia

Patience Epps and Lev Michael

1 Introduction

Amazonia is one of the most linguistically diverse regions of the world, with a density of distinct genetic groupings – some fifty families and isolates – rivaled only by New Guinea. We know very little about the historical processes that have shaped the Amazonian linguistic picture, or gave rise to its plethora of languages. However, despite common assumptions, the Amazon basin provides ample evidence that the maintenance of diversity does not entail a lack of contact among groups speaking different languages. In many areas of the lowlands, contact among Amazonian language communities has been intense and long-term. Such contact situations have themselves profoundly shaped the linguistic profile of Amazonia and neighboring regions, giving rise to zones of typological similarity that cross-cut genetic linguistic differences.

Working out a precise account of how contact has influenced the languages of lowland South America presents a “vast and almost intractable” problem, as observed by Muysken (2012: 235). A significant aspect of this intractability lies in the paucity of descriptive and historical work that has been carried out on these languages. For many languages, it is already too late – they have been extinguished before they could be documented. However, the past few decades have seen an explosion of high-quality descriptive work on many surviving Amazonian languages, some of which are highly endangered, with new historical work building closely upon this foundation. We have thus entered an exciting new period of investigation into Amazonian language history, which is already yielding fresh insights into linguistic areality in the Amazon basin. In this chapter, we offer our assessment of the current state of the art in understanding linguistic areality in Amazonia.

Amazonia, which we define loosely here as the lowland region drained by the Amazon and Orinoco Rivers and extending to the northern and eastern litorals of the continent (cf. Dixon and Aikhenvald 1999: 4, Rodrigues 2000: 15), is bordered by the Andes mountains to the west, the Caribbean and Atlantic oceans in the north and east, and the drier regions of the Gran Chaco to the south. Most of this vast area is covered by tropical rain forest, with pockets of savannah on the margins. The majority of the linguistic diversity encountered within the Amazon region is concentrated on the western periphery, for reasons that are currently as mysterious as those behind the overall diversity itself. The major language families that do exist, most notably Arawak, Carib, Tupí, and Macro-Jê, are characterized by predominantly non-contiguous distributions, with their members interspersed by many other smaller families and isolates (see e.g. O’Connor and Muysken 2014, Campbell and Grondona 2012 for further discussion). While this linguistic patchwork renders the task of investigating language contact complex, it also makes it to some degree more accessible, since contact-induced change tends to be easier to identify among unrelated languages.

---

1 Epps gratefully acknowledges support from the National Science Foundation (HSD-0902114).
While our focus here is on the linguistic effects of contact, we emphasize that contact necessarily takes place among speakers, and it is the dynamics of these speakers’ interactions that produce the particular linguistic outcomes that we observe. The contact-motivated similarities that cross-cut language boundaries are evident not only in lexicon and grammar, but also in discourse and sociocultural practice more generally, and include the ways in which people tell stories, sing songs, prepare food, raise children, heal the sick, and so on. Speaker interactions are grounded in, and structured by, approaches to trade, intermarriage, and ritual and convivial practice. The extent of these interactions should not be underestimated; recent work has demonstrated that many areas of Pre-Colombian Amazonia hosted large population densities and relatively complex societies (e.g. Heckenberger and Neves 2009). Documentation of long-distance trade networks (e.g. Vidal 2000, Nordenskiöld 1922, cf. Hornborg 2005) and of migrations over large distances (e.g. Clastres 1995) likewise indicates that many native Amazonians had ample opportunity to interact with other language groups.

Our discussion is primarily concerned with the effects of contact among indigenous languages, and most of the patterns we consider are undoubtedly rooted in pre-Columbian social dynamics. However, the displacement and restructuring of many Amazonian societies in the centuries following the European invasion have introduced profound changes to those dynamics, and it is in many cases unclear to what extent contemporary socio-cultural practices represent continuity with pre-Columbian times, or how patterns of linguistic diffusion have been altered over the last five hundred years. Moreover, contact with European languages, particularly Spanish and Portuguese, as well as with European-mediated languages such as Quechua and Nheengatú (a Tupí-Guaraní language promoted by early Jesuit missionaries as a lingua franca), have profoundly affected many indigenous languages, in many instances culminating in language shift (see e.g. Muysken 2012). Interestingly, these contact scenarios are often characterized by a rather different mix of processes from those observed among indigenous Amazonian languages, with considerably more code-switching, lexical borrowing, and language shift – in keeping with the different sorts of social relations that pertain among these groups. We will not address these differences in detail here, but will focus on the observable outcomes of contact in indigenous Amazonian contexts, which have tended toward grammatical diffusion and language maintenance.

This chapter is organized as follows. We begin in Section 2 with a discussion of the principal localized contact zones that have been investigated within Amazonia (the Vaupés, the Upper Xingu, and other areas), where speakers of multiple languages live in close proximity and engage in frequent interaction. Section 3 provides a wider scope, with an assessment of evidence for larger areal diffusion zones within the lowlands. This section also addresses the possibility that the Amazon basin as a whole might represent a single large-scale diffusion zone, with substantive contrasts between it and other South American regions such as the Andes and the Southern Cone. Section 4 summarizes our current understanding of Amazonian areal linguistics and outlines directions for future research.

2 Localized diffusion within Amazonia

Most historically recent situations of regular contact among speakers of different indigenous languages are found in localized zones within the South American lowlands,
and it is in these contexts that areal diffusion is most easily identified. While all of these zones have been profoundly affected by the European-derived national society, they have maintained aspects of their traditional social structures and cultural practices, suggesting a degree of continuity with pre-Colombian dynamics. Many of these contemporary contact zones share some notable similarities: in particular, they are characterized by multiple groups speaking different languages, who maintain a relatively egalitarian relationship with respect to one another, and whose interaction is frequent, conventionalized, and profound – a context that clearly favors areal diffusion. Within these ‘regional systems’, as they are sometimes termed, groups tend to be characterized by a striking degree of cultural homogeneity on one hand, but on the other by a set of locally salient differences, such that they function rather like a set of interlocking cogs in a single machine (see Epps forthcoming). Map 1 illustrates the location of the principal regions discussed here.

Language plays a recurrent role as an emblem of difference in these zones, and local ideologies of language tend to strictly constrain the mixing of codes, even where frequent interaction among groups fosters intensive multilingualism (Hill 1996, Epps forthcoming). We widely encounter long-term language maintenance, limited code-switching, and low levels of lexical borrowing, often buttressed by explicit articulations of the quality and importance of linguistic difference. At the same time, particularly where individual multilingualism is high, there is considerable convergence among languages on a grammatical level. This combination of low lexical borrowing with substantial grammatical diffusion is in striking contrast with the outcomes of language contact in many other parts of the world, where lexical borrowing tends to precede and facilitate grammatical diffusion (e.g. Thomason and Kaufman 1988); nevertheless, it appears to be commonplace in Amazonia. This outcome is undoubtedly linked to speakers’ conscious efforts to avoid language mixing – of which they are most aware on the level of lexical forms – and their frequent exposure to multiple codes, which fosters the convergence of grammatical structures and categories below speakers’ “limits of awareness”, in Silverstein’s terms (1981; see also Aikhenvald 2001a and Mithun, this volume).

In many of these interactive zones, our assessment of the effects of contact on the languages themselves is facilitated by the presence of distinct language families, and by the possibility of comparison with related languages outside the region. These factors allow us to attribute similarities within the area to contact rather than to common inheritance, especially where the shared features are too numerous and/or too unusual to be easily explained as independent innovations (see Campbell et al. 1986, Epps et al. 2013). While this historically grounded approach yields relatively robust evidence of contact-induced change, cases where related languages are undocumented or do not occur outside the region require us to fall back on the identification of a set of features shared among languages within the area. Where these features cannot be shown to be significantly different from those that exist in the languages outside the area, the case for a regional contact zone is weakened (see Campbell and Grondona 2012, Muysken 2012); however, solid comparative analysis can provide evidence for areal diffusion regardless of whether we can define a precise contrast between a particular region and the neighboring areas.
Map 1. Lowland South American contact zones

2.1 The Vaupés region

The Vaupés region of the northwest Amazon has received the most in-depth attention of any South American contact zone. The area of the Vaupés river basin is home to dozens of languages belonging to four distinct families, Tukanoan, Arawak, Nadahup, and Kakua-Nukak (formerly lumped together with Nadahup to form the ‘Makú’ group).\(^2\) The Vaupés region is a particularly intensive contact zone within the larger Upper Rio Negro

\(^2\) The post-colonial arrival Nheengatú (Tupí-Guaraní) is also marginally represented in the region, but is not discussed here.
basin, which itself appears to be a region of less profound areal diffusion (see Aikhenvald 1999a, Epps and Stenzel 2013). While most of the existing work concerning language contact in the Vaupés has focused on particular languages or on particular linguistic features, the following discussion offers a short synthesis of our current understanding of diffusion across the region as a whole, and situates it within the wider lowland South American context.

The Vaupés has been described as a regional ‘system’, in which distinct groups function together, and as a culture area, in that these groups share many features in common (see e.g. Bruzzi 1977, Epps and Stenzel 2013). Similarities include various aspects of material culture, such as house construction, manioc-processing technology, and bodily adornment, and likewise many ritual practices. Discourse practices throughout the region are also strikingly similar, despite their delivery in different languages, with shared song traditions, stories, incantations, and conversational norms. On the other hand, certain salient differences underscore the systemic nature of the region, such that groups maintain an identity as distinct, interactive units. Among the most locally meaningful of these differences is language, which is emphasized in the context of linguistic exogamy – obligatory marriage across language groups – practiced by most of the East Tukanoan and some Arawak groups in the region. Another salient difference is subsistence pattern, with an opposition between the Nadahup and Kakua-Nukak ‘forest peoples’, who prioritize hunting and gathering, and the East Tukanoan and Arawak ‘river peoples’, who focus on fishing and farming. These subsistence distinctions are accompanied by social asymmetries (such that the river peoples are ‘ranked’ more highly than the forest peoples). Interaction among regional units is also promoted by trade specializations (for example, the Nadahup people provide meat and baskets; the East Tukanoan Tuyukas the canoes, etc.).

The systemic nature of the region engenders widespread multilingualism, coupled with language maintenance. East Tukanoan peoples tend to speak several of each other’s languages, as did Tariana (Arawak) speakers, facilitated by the linguistic exogamy system. The forest peoples in the region have traditionally spoken at least one river-Indian language; in their case, however, bi- or multilingualism has been almost always unidirectional, due to the local social imbalance. The active maintenance of distinct languages, despite intense multilingualism and even social asymmetry, has been viewed as an outcome of linguistic exogamy, which explicitly links marriageability to linguistic difference (Sorensen 1967, Jackson 1983, Aikhenvald 2001a, Stenzel 2005, inter alia). However, Epps (forthcoming) argues that the salience of the link between language and identity is probably as much a cause as an effect of linguistic exogamy, in light of broader trends across lowland South America (see below).

The Vaupés emphasis on language as an emblem of identity translates into tight constraints against the mixing of languages. Intrasentential code-switching is avoided and socially condemned (see e.g. Aikhenvald 2001a: 412, Chernela 2013: 213). East Tukanoan speakers explicitly characterize their awareness of linguistic differences, stating for example that some languages “flow slowly and smoothly, ‘like waves of water’” while others “‘sound like lightning’ … with sharp angles, stops, and starts” (Chernela 2013: 216-217). In keeping with these maintenance practices, the region’s

3 However, recent decades have seen Tariana and many other East Tukanoan speakers shifting to Tukano, which has gained a new, colonially mediated level of status (see Stenzel 2005, Aikhenvald 1999a: 387), and in some contexts to Portuguese.
languages have experienced remarkably low levels of lexical borrowing despite the intense, long-term contact among them. A systematic lexical study of ten different Vaupés languages reveals only 2-4% loans in basic vocabulary (Bowern et al. 2011, 2014; see also Epps 2009), with slightly higher levels in flora-fauna and material/ritual culture terms (around 10-12% in Nadahup and Kakua-Nukak languages; lower in Tukanoan and Arawak). On the other hand, the existing loans include various Wanderwörter (see Bowern et al. 2014, Epps 2012), which offer the impression that if a lexical item is to be borrowed, it is likely to travel widely, possibly because it loses an association with a particular language in the process and thus becomes ‘fair game’ (see also Muysken 2012: 252).

Despite the low levels of lexical borrowing, the effects of contact in the region’s languages have had significant effects on local lexicons, particularly involving the congruence of semantic categories and the calquing of lexical items across languages. Calquing is especially pervasive in place names and ethnonyms, binomial names for flora and fauna, and items of material and ritual culture (e.g. the name of a regional culture hero: Tukano o’ä-kó ‘Bone-Son’, Tariana yapi-riku-ri ‘One on the Bone’, Hup g’eg tēh ‘Bone Son’); see Aikhenvald (2002: 229), Epps (2013), Floyd (2013).

Grammatical structures and categories among the Vaupés languages have been profoundly affected by areal diffusion, giving rise to a significant degree of morpheme-to-morpheme and word-to-word intertranslatability (see Aikhenvald 2007a: 261). East Tukanoan languages have provided the model for many of these changes, but in some cases they too have converged to be more like their neighbors. The extent to which East Tukanoan languages have been shaped by contact with each other is not as easy to determine, but diffusion among these languages has also undoubtedly occurred (Gomez-Imbert 1993, Chacon 2013).

The heavy restructuring experienced by Tariana, the only Arawak language fully incorporated into the Vaupés linguistic exogamy system, has been explored in detail by Aikhenvald (1999a, 2001b, 2002, 2007a, inter alia). As Aikhenvald demonstrates through comparison with Baniwa, Tariana’s closest Arawak sister outside the Vaupés region, Tukanoan influence has led to changes throughout Tariana grammar, including a realignment of morphemes and constituent order; the development of case-marking on nominal arguments (topical non-subject and oblique, as well as the reduction of multiple locative markers to one catch-all marker); the elaboration and restructuring of the nominal classifier system; the augmentation of the evidential system from two categories to five; the development of verb compounding, switch-reference, new complementation strategies, discourse marking, etc. Tariana also exhibits significant convergence in its phonological inventory towards the typical Tukanoan profile, diverging from typical Arawak profiles (Chang and Michael 2014: 1-2). Examples (1-3) (from Aikhenvald 2007a: 245-246) illustrate the near-isomorphism between Tariana and Tukano, in contrast to Tariana’s sister Baniwa, as can be seen in the following examples.

(1)  Tariana (Arawak)

\[
\text{nes̄e paːma}\quad\text{di-na} \\
\text{then one+NUM.CL.ANIMATE.FEM 3sgfn-OBJ} \\
\text{du-yana-sita-pidana} \\
\text{3sgf-cook-ALREADY-REM.PAST.REP} \\
\text{‘She had reportedly cooked him already.’}
\]
The Nadahup languages have also undergone significant restructuring through Tukanoan influence, facilitated by the one-way bilingualism that predominates among many Nadahup speakers in the Vaupés. Hup appears to have been most profoundly affected (see Epps 2005, 2007a, 2008, *inter alia*), having developed prosodic nasalization, verb compounding, nominal classification, evidentials (as with Tariana, a five-way distinction built atop an earlier two-way one), future and recent/distant past tense distinctions, new strategies of number marking, and case-marking (including an object/non-subject marker that is sensitive to animacy and definiteness, and a catch-all locative), among many other features (compare example 4 to those above). Hup’s sister Yuhup, also spoken in the Vaupés, has likewise experienced significant Tukanoan influence (with diffusion-induced changes probably beginning in their common ancestor), while Dâw, located on the periphery of the Vaupés, has been less affected. Nadëb, their most distant sister, has been essentially unaffected by diffusion from Tukanoan (though it has almost certainly been influenced by Arawak), and its grammar is strikingly different from those of its sisters.

Kakua (Kakua-Nukak) likewise shows evidence of diffusion, probably also from Tukanoan languages. As can be seen in example (5) (from Bolaños 2012: 3), Kakua resembles its neighbors with respect to characteristics such as evidentiality, verb compounding, a recent/distant past tense distinction, non-subject case marking, and verb-final constituent order. Chang and Michael (2014) also find evidence of phonological convergence between Kakua and the Nadahup languages Hup and Yuhup.

Arawak languages have also exerted influence on some East Tukanoan languages. Evidence for diffusion in this direction includes the development of aspirated stops in Kotiria (Wanano), possessive proclitics in Kotiria, Kubeo, Tatuyo, and certain other
languages, and the use of shape classifiers with non-human animates in Kubeo and (to a limited extent) in Kotiria (Stenzel and Gomez-Imbert 2009, Gomez-Imbert 1996).

A number of further studies have undertaken in-depth comparative investigations of particular grammatical features across Vaupés languages, noting their many similarities; these have explored serial verb constructions, including grammaticalization of the verbs *come/go* as directionals (Ospina and Gomez-Imbert 2013, Aikhenvald 1999b); the expression of spatial relations (Stenzel 2013b, Epps and Neely 2014); possessive constructions, including the development of a comparable alienable/inalienable distinction and similar marking strategies (Stenzel 2013a); differential object marking strategies (Zuñiga 2007, Stenzel 2008); and nominal classification (Aikhenvald 2007b, Epps 2007b).

The Vaupés languages provide intriguing insights into the mechanisms by which areal diffusion takes place. In many instances, these mechanisms have involved the grammaticalization of native material to create new categories and structures. The process can involve the calquing of whole constructions, as can be seen for example in the close parallels in event packaging via serial verb constructions. Rather than resulting in simplification, the development of new material through contact-driven grammaticalization often produces an overall *increase* in grammatical complexity by adding to existing repertoires of categories and structures, as in the case of Tariana’s development of case-marking in addition to its earlier strategy of pronominal cross-referencing (see Aikhenvald 2003). Stenzel’s (2013a) discussion of possession-marking strategies among Vaupés languages highlights the fact that even while distinct languages may converge on a common model through processes of diffusion, the language-internal dynamics of change give rise to fine-grained differences, such that the outcomes are not isomorphic. Epps (2012) shows that despite the constraints against the borrowing of form in this region, perceived formal similarities nevertheless can and do play a role in facilitating structural adaptations, as in the case of the development of an evidential form *ni* from an existential verb in Tukanoan and Hup, and in Tariana from a functionally distinct marker * nhi* ‘anterior tense’ (Aikhenvald 2002: 123).

Particular examples of diffusion-driven change in Vaupés languages also highlight the role of discourse in motivating these processes (compare also Aikhenvald’s 2007a: 261 observation that “the more pragmatically motivated, the more diffusable”). The development of evidentials provides a good example: norms of conversation, storytelling, and other discourse forms, shared among speakers, foster an expectation that one’s source of information will be explicitly stated (and that if it is not, one’s reliability or responsibility may be in question). In Hup, this discursive expectation apparently led to an increase in the frequency of verb roots associated with information source appearing in verbal compounds encoding events, which in time grammaticalized into evidentials (see Epps 2005). Moreover, regional norms for narrating traditional stories may have given rise to a further change in Hup: In these stories, almost every clause is marked by the reported evidential =*mah*, followed directly by the distant past tense marker *j’am* (itself a fairly recent addition to Hup grammar), as can be seen in (6) – although in other discourse contexts the past markers only appear when the temporal information is contrastive or particularly emphasized. Moreover, in one dialect of Hup, the reported evidential appears fused together with the distant past marker in traditional stories to produce =*máam* (example 7). Given that Hup has almost no other portmanteau morphs, the parallel with the fused tense-evidential suffixes in Tukanoan languages is striking.
2.2 The Caquetá-Putumayo region

The Caquetá-Putumayo area of southern Colombia and northern Peru is another local diffusion zone, though it is not nearly as well studied as its northern neighbor the Vaupés. In this region, speakers of languages belonging to the Bora and Witoto families, the Arawak language Resígaro, and the isolate Andoke interact intensively through marriage, ritual contexts, etc.; refer to themselves together as the ‘People of the Center’; and are widely multilingual4 (see Echeverri 1997, Seifart et al. 2009, Londoño Sulkin 2012). Distinctive shared cultural practices include the ritual ingestion of powdered coca leaves and liquid tobacco, extensive song cycles, and particular styles of warfare, personal hygiene, etc.; at the same time, the groups have maintained distinct languages, origin stories, and certain other emblems of identity (Seifart 2011: 7-8, Whiffen 1915).

As in the Vaupés, the People of the Center share an “inhibition against lexical borrowing” (Seifart 2011: 88). Despite their close contact, Seifart (2011: 20) finds that only about 5% of Resígaro lexical stems have been borrowed from Bora (gauged via an extensive set of core and non-core vocabulary), of which many are flora-fauna terms; similarly low loan rates between Bora, Resígaro, and other languages of the region are identified in work by Bowern et al. (2011, 2014).5 However, Resígaro has borrowed a striking number of bound morphological forms, including whole sets of nominal classifiers (over half of Resígaro’s total), number markers, quantifiers, and other forms, totaling over 50 distinct items (Seifart 2011, 2012). Further diffusion has affected grammatical structures and categories in Resígaro, with or without the mediation of directly borrowed morphemes. These features include the development of an inclusive/exclusive distinction, second-position tense-aspect-mood clitics, the loss of object cross-referencing suffixes, and the restructuring of verbal morphology (Seifart 2011: 14, Aikhenvald 2001: 189). Resígaro phonology has also been affected, with the addition of new phonemes (/ɸ/, /dʒ/, /ʔ/), syllable structure restrictions, and a two-tone...

4 However, as in many of the other regions discussed in this paper, recent decades have seen significant language loss as speakers shift to Spanish (or other colonially mediated languages).

5 The figure of 24% loans in Resígaro given by Aikhenvald (2001b: 182; see also Eriksen and Danielsen 2014: 188) was probably erroneously inflated by the inclusion of borrowed classifier forms.
contrast (Aikhenvald 2001b, Seifart 2012, Chang and Michael 2014). Indeed, Chang and Michael (2014) show that Boran and Witotoan languages, Resígaro, and the isolate Andoke exhibit significant convergence in their phonological inventories, allowing us to pick out the People of the Center as a well-defined phonological area.

The People of the Center have been involved in longer-range processes of diffusion as well. They are in contact with the Arawak Yukuna to the north, which themselves are in close contact with the East Tukanoan Tanimuka (Retuarã; see Seifart 2007, Aikhenvald 2001b); the Caquetá-Putumayo groups share a number of characteristics with their northern neighbors in the Vaupés. These similarities are both cultural (e.g. large signal drums) and linguistic (e.g. the distributional and functional properties of nominal classifiers, nominative-accusative alignment, etc.; see Aikhenvald 2001b: 189, Seifart and Payne 2007).

Figure 1 models a subset of northwest Amazonian languages as a NeighborNet splitsgraph, with respect to 226 grammatical features (mostly morphosyntactic). The splitsgraph illustrates the extent to which areal diffusion has produced regional grammatical profiles among the Vaupés and Caquetá-Putumayo languages. The Vaupés cluster includes East Tukanoan, Hup, and Kakua, while the Arawak languages Tariana, Resígaro, and Yukuna pattern more closely with their non-Arawak neighbors than they do with their closest northern Arawak relatives. Also included are Kokama and Nheengatú (both Tupí-Guaraní) and Yagua (Peba-Yaguan); the association between these three languages is undoubtedly indicative of areal diffusion in the region where Yagua and Kokama are spoken, just south of the Caquetá-Putumayo area.

---

6 See [www.laits.utexas.edu/huntergatherer](http://www.laits.utexas.edu/huntergatherer).

7 We note that the tendency for grammars to change through areal diffusion while lexicons remain relatively conservative, as seen in these and other Amazonian languages, presents a challenge for the view that typological features will be in general be more likely to retain a deep-time genetic signature (e.g. Dunn et al. 2005, Sicoli and Holman 2014).
Figure 1. NeighborNet representation of grammatical structures in languages of the northwest Amazon divided into four groups

2.3 The Upper Xingu Region
Contact among the Upper Xingu peoples has been relatively well documented ethnographically, but its linguistic effects have only begun to be explored. This region is home to more than a dozen languages belonging to the Carib, Arawak, Jê, and Tupí families, as well as the isolate Trumai, although several of these groups are relative newcomers (arriving between the sixteenth and twentieth centuries in response to colonial pressures). Like the Vaupés and other Amazonian regional systems, the Xingu is an area of intense interaction among groups, particularly through ritual contexts and trade. The groups share a distinctive “cultural package” (Fausto et al. 2008: 137; see also Franchetto 2011), with particular rituals, hairstyles, house architecture, etc., to which newcomer groups have assimilated as they were absorbed into the Xingu system (see, e.g., Guirardello-Damien’s 2011 discussion of the Trumai).

As in the Vaupés, language plays an important role in the Xingu as a “basic diacritic” of ethnic identity and place within the regional system (Fausto et al. 2008: 141). Speakers provide explicit characterizations of linguistic differences, observing for example that the Carib groups speak ‘in the throat’ or ‘inwards’, while the Arawak peoples of the region speak ‘outwards’, ‘on the tip of the teeth’ (Fausto et al. 2008:143). These “rigorous and active processes of differentiation” are realized via a monolingual everyday ethos (Ball 2011: 93; see also Seki 2011: 85), in contrast to the Vaupés; yet multilingualism is extensive in ritual contexts, and the frequency of interaction has led to characterizations of the region as a “communications network” (Basso 1973: 5) or even a speech community (Ball 2011: 93). Little code-switching and relatively little lexical borrowing occur (though some loans exist; see Seki 1999, 2011).

Seki (1999, 2012) has characterized the Xingu as an “incipient” linguistic area, primarily on the basis of lexical similarities and common features of myth and ceremonial discourse. She identifies several grammatical features that have diffused within the region, and it is likely that further linguistic exploration will reveal more; these include the loss of a masculine-feminine gender distinction in the Arawak languages of the region (in cross-referencing and in independent pronouns), the development of a phoneme /ɨ/ in Arawak, a p > h shift in Carib and Tupí-Guaraní, and a change to CV syllable structure in Carib. Chang and Michael (2014) clarify the multilateral nature of phonological borrowing in the Xingu area, confirming the diffusion of /ɨ/ into Xinguan Arawak languages from their Carib or Tupí-Guaraní neighbors, and adding the diffusion of /ts/ into the Xinguan Carib languages from their Arawak neighbors and the diffusion of nasal vowels into the Xinguan Arawak and Carib languages from their Tupí-Guaraní neighbors. Lexical restructuring and calquing have also taken place, including the development of comparable systems of post-nominal elements meaning ‘big, supernatural, hyper’, ‘similar to’, ‘true/genuine’, and ‘bad/worthless/unsatisfying’, which are used productively to create new complex nouns. For example, in Yawalapiti úi ‘snake’ + kumã ‘hyper’ yields úi-tyumã [kumã] ‘snake-spirit’ (Viveiros de Castro 2002), and in Trumai fi ‘cigar, cigarette’ (tobacco)’ + yuraw ‘hyper’ (itself a loan from Tupí-Guaraní Kamayurá) yields fi yuraw ‘marijuana cigarette’ (considered abnormal, dangerous; see Guirardello-Damien 2011: 120). The diffusion among the Xingu languages appears to be generally multidirectional.

2.4 The Guaporé-Mamoré region
The Guaporé-Mamoré of southwest Brazil and northeast Bolivia is home to over fifty languages from a wide range of families (Arawak, Macro-Jê, Chapakuran, Tupí, Nambikwará, Pano, Takanan, and many isolates). Ethnographic documentation indicates that many of these groups have engaged in extensive interethnic contact, intermarriage, and exchange, yielding a regional culture area with commonalities in territorial subgroups, bodily adornment, mythological themes, etc. (Levi-Strauss 1948, Maldi 1991).

The investigation of contact among the Guaporé-Mamoré languages has not yet been extensive, but work by Crevels and van der Voort (2008; see also van der Voort 2005) indicates that diffusion has certainly taken place and that this is a rich area to explore. As in the other regions considered here, levels of lexical borrowing among these languages appear to be quite low (estimated at around 5% by van der Voort 2005: 395; see also Crevels and van der Voort 2008: 164); however, they share many structural similarities, including evidentials, an inclusive/exclusive distinction, a high incidence of prefixes, and a strong preference for verbal number (i.e. alteration of the verb via suppletion, reduplication, or affixation to express the number of the subject or object), accompanied by a general lack of nominal number (Crevels and van der Voort 2008: 167). Further similarities include verbal cross-reference systems with similar morpheme positions, and rich systems of directional morphemes, such as Kwaza’s ‘movement in a circle’, ‘into fire’, ‘behind the house’ (van der Voort 2005: 399; see also Guillaume forthcoming). Eriksen and Danielsen (2014: 175) also note areal influence in verbal morphology and personal pronominal distinctions in two Arawak languages of the region (Paunaka and Moxo).

A particularly interesting aspect of areal diffusion in the Guaporé-Mamoré involves the direct borrowing of morphological forms, particularly within nominal classifier systems (van der Voort 2005: 397, Crevels and van der Voort 2008: 167) – a striking parallel to the effects of contact in the Caquetá-Putumayo. Several of these forms are given in Table 1; none of the five languages listed are known to be related.

<table>
<thead>
<tr>
<th></th>
<th>‘bar k’</th>
<th>‘fruit’</th>
<th>‘bone’</th>
<th>‘tooth’</th>
<th>‘liquid’</th>
<th>‘round’</th>
<th>‘thorn’</th>
<th>‘porridge’</th>
<th>‘powder’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kwaza</td>
<td>-kalo</td>
<td>-ko</td>
<td>-su</td>
<td>-mäi</td>
<td>-mũ</td>
<td>-tẽ</td>
<td>-nĩ</td>
<td>-mẽ̃</td>
<td>-nũ̃</td>
</tr>
<tr>
<td>Kanoë</td>
<td></td>
<td>-ko</td>
<td>-mũ</td>
<td>-tæ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-nũ</td>
</tr>
<tr>
<td>Aikana</td>
<td></td>
<td>-zu</td>
<td>-mũ̃j</td>
<td>-mũ</td>
<td>-ðãw</td>
<td></td>
<td></td>
<td></td>
<td>-nũ</td>
</tr>
<tr>
<td>Arikapú</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-mĩ̃</td>
<td>-mrẽ̃</td>
<td></td>
<td></td>
<td>-nũ</td>
</tr>
<tr>
<td>Nambi-kwara</td>
<td>-kalo</td>
<td>-sũ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-nũ̃x̃</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As Muysken et al. (forthcoming) point out, the fact that so many of the Guaporé-Mamoré languages are small families or isolates located only within this region makes it difficult to test for contact by comparison with relatives outside. However, these authors apply quantitative measures to demonstrate that at least some features must have converged via diffusion in this region. Other features that are more widely attested may themselves be the outcome of more far-reaching contact networks that existed in the past, as discussed in Section 3 below.

2.5 Other regional diffusion zones in lowland South America

Several other localized diffusion zones exist in the Amazon basin and adjacent lowland regions, and still others probably remain undetected. One area that has had some documentation is the Southern Guiana region, where Carib, Arawak, and Salivan groups exhibit notable cultural continuity resulting from “constant interaction through marriage, trade, and migration” (Rivière 1984: 8; see also Arhem 1989, Carlin 2011: 226). Much like the other regions discussed here, linguistic distinctions are maintained as key markers of ethnic identity (Howard 2001: 341), and loanwords are relatively few (Carlin 2007, Bowern et al. 2011, 2014). Carlin (2007) discusses grammatical diffusion from Carib languages (principally Trio and Waiwai) into the Arawak language Mawayana, established in part via comparison with Mawayana’s sister Wapishana. Contact-induced changes in Mawayana include the addition of a first person exclusive distinction (via borrowing of a Waiwai pronominal form), the development of nominal tense marking, affective marking on nouns or verbs to express ‘pity’ or ‘recognition of unfortunate circumstance’, a frustrative marker on verbs (indicating that the action was carried out ‘in vain’), and a ‘similative’ or ‘as if’ marker on nominals (see examples 8-9, from Carlin 2007: 329). Carlin notes that all of these categories are obligatory in the local Carib languages.

(8) Mawayana (Arawak)

\[waata-ni\ r-ayăďyă\]
opossum-SIMIL 3s-transform,PAST

‘He changed into an opossum.’

(9) Trio (Carib)

\[kaikui-me\ tēmetē\]
jaguar-SIMIL he.transformed

‘He transformed into a jaguar.’

The Southern Guiana region may itself be part of a much larger contact zone that includes parts of the Orinoco and northern Amazon watersheds (Migliazza 1985: 20; cf. Campbell 2012: 306). Further possible diffusion zones include the area of the Tocantins and Mearim Rivers in northeastern Brazil, home to Jê and Tupí-Guaraní languages.
Finally, the region of the Gran Chaco, located just to the south of the Amazon basin, bears striking similarities to the Amazonian regional ‘systems’ discussed here: The speakers of its six distinct language families share many cultural similarities (Braunstein and Miller 1999: 9-11), shun code-switching and lexical borrowing (Campbell and Grondona 2010, 2012: 657; Vidal and Nercessian 2009: 1023), practice a form of linguistic exogamy, and identify strongly with a single language while understanding many (Campbell and Grondona 2010). The Chaco languages share many structural features, including animal classifier(s), active-inactive verb alignment, similar strategies of pronominal affixation, and rich demonstrative systems, many of which are undoubtedly due to diffusion (Campbell and Grondona 2012).

We will probably never know the extent to which the dynamics of these regional diffusion zones applied more generally in the Amazon basin, across both space and time. However, the wide distribution of regional ‘systems’ and their notable social and linguistic parallels suggest that these are not isolated cases. If Eriksen and Danielsen (2014: 163) are correct in their assertion of “a vast socio-religious and economic exchange system that affected the lives of all inhabitants of northern South America between 1000 BCE and 1000 CE,” then longer-range diffusion would undoubtedly have linked these local zones within wider networks. We turn to this question below.

3 Long-range language contact and macro-areality in Amazonia

The linguistic areas discussed in Section 2 generally arose through interactions between neighboring peoples in networks that span relatively small geographical regions in comparison to Amazonia as a whole. There is clear evidence, however, that Pre-Columbian Amazonian peoples were linked via large-scale social, ritual, and trade networks that spanned considerable areas of the continent, raising the possibility that Amazonia may exhibit linguistic areality on a similarly large scale. The fact that many cultural features and practices, such as perspectivism in mythology (Viveiros do Castro 1998, 2012: 48-51), the use of sacred flutes (Hornborg and Hill 2011: 17), and a variety of subsistence techniques (see e.g. Carneiro 2000 on manioc-processing technology), have diffused over large areas of Amazonia suggests that linguistic features may have diffused on a similar scale.

Although the full extent of pre-Columbian South American trade networks is impossible to recover at this point, ethnohistorical and archeological evidence clearly indicates that such networks spanned large areas of Amazonia, and linked Amazonia to adjacent regions, such as the Andes and Chaco. The earliest accounts of the colonial period provide ample evidence of vast trade networks criss-crossing the continent. Citing Oviedo y Valdés (1851-1855[1535]), for example, Nordenskiöld (1922: 7) describes Arawak traders then living near the mouth of the Amazon making 1500 kilometer trading journeys along the coast in groups of 50-60 canoes and 500-800 men. He similarly discusses evidence of trade between the Cariban peoples of the Guianas and the Cariban Carijona of the Putumayo, and further still to peoples on the Amazon proper (ibid: 149-150). Trade links between the Amazon and the Guianas are echoed by Fritz (Edmunson 1922) who indicates that the Omaguas, who occupied much of the Amazon between the Japurá and Napo, formed a node in a trade network linked to the Guianas, and traded with
peoples deeper into the Amazonian headwaters regions. Nordenskiöld presents evidence of similarly long-distance trade between the Guaraní peoples of the Paraná and Paraguay River Basins and the Andean Inca Empire (ibid: 7-10, 133-134). Trade of specific rare products is similarly known to have extended over thousands of kilometers, as in the case of salt – mined by the Arawak Ashéninkas in the Perené River basin in the Andean foothills, and compellingly argued by Rydén (1962: 652) to have reached the Tupinambás who had resettled on the Amazon at Tupinambarana, downriver of the mouth of the Río Negro. Eriksen and Danielsen (2014) cite ethnohistorical and archeological evidence to argue for the existence of an extensive Arawak-dominated trade network spanning much of western Amazonia, extending the arguments of Vidal (2000), who provides evidence for a major Arawak-dominated pre-Columbian trade network in northwestern Amazonia.

Long-distance trade of this sort survived well into the modern period. Lathrap (1973), for example, describes the varied networks in which the Shipibo communities, located on the Ucayali River, participated. In some of these networks, raw materials for ceramic production circulated up to 240 kms from the communities studied. Other networks were even more expansive and interethic, and included Yagua communities some thousand kilometres downriver, through which Shipibos obtained blowguns, as well as Tikuna communities, located a further five hundred kilometres along the Amazon proper, from which they obtained blowgun darts. Roth (1924) similarly describes a network spanning the eastern Orinoco basin and the Guianas, characterized by circulation of rare materials and high degrees of craft specialization (see also Dumont 1991). This network linked the Tupí-Guaraní communities on the Oyapock, producers of valued grinding stones; the river-craft-producing Warao at the mouth of the Orinoco; the Cariban Waiwai, located along the hills separating the Guianas from the Amazon watershed, and who specialized in a number of palm products; and the Cariban Makiritare, located in the central Orinoco basin, who produced hammocks, cassava graters, and ornamental products.

Whether far-flung networks of interaction like these have led to widespread grammatical borrowing and convergence, and to the emergence of linguistic macro-areas within Amazonia, has become a topic of increasing prominence in Amazonian linguistics, as has the question of whether Amazonia as a whole constitutes a linguistic area. One of the earliest macro-area proposals considered the region encompassing the Orinoco River Basin and the portion of the Amazon Basin containing the northern tributaries of the Río Negro (Migliazza 1985). Features that Migliazza attributed to this area included ergative alignment, OV order, lack of a passive construction, relative clauses formed by apposition and nominalization. Most Amazonianists would now recognize this list as including several features of broader Amazonian distribution, and indeed, the features that Derbyshire (1987: 311) tentatively proposed as defining an Amazonian linguistic area include most of these, and in addition, a preference for OS order, subject and object verb agreement, null free argument realization, head-modifier order, and complex morphology. Derbyshire and Payne (1990) subsequently added noun classifiers to this list of tentative features.

Perhaps the most promising macro-area proposals have been based on typological divisions bisecting the east-west axis of the continent. In one of the earliest such proposals, Doris Payne (1990: 5) suggested the existence of a Western Amazonian area consisting of the lowland areas to the west of the Andes characterized phonologically by complex stress and ‘pitch accent’ systems and morphosyntactically by
a strong tendency towards polysynthesis and complex verbal morphology, directional, locational, and positional morphemes, and a distinctive type of noun classification system. The validity of this proposal is somewhat difficult to evaluate because neither the precise limits of the area, nor the distribution of indicated features inside and outside the proposed area are given; but as discussed below, recent quantitatively based work lends support to an east-west areal split in South America.

As descriptions of Amazonian languages increase in both number and quality, work identifying macro-areas on the basis of relatively fine-grained linguistic phenomena will probably become more common. A promising example of this type is Guillaume and Rose’s (2010) suggestion that sociative causatives may be an areal feature of southwest Amazonia, with the distribution of sociative causatives outside this area attributed to the spread of Tupían languages from their southwestern homeland. A systematic examination of the distribution of such morphemes both within South America and beyond is an obvious target for future research.

The question of linguistic macro-areas within Amazonia naturally leads to an issue already raised in our discussion: whether Amazonia as a whole constitutes a linguistic area. Although this remains an open question, an emerging consensus points to Amazonia not forming a linguistic area *sensu strictu*. Dixon and Aikhenvald (1999: 7-10) is perhaps one of the best known and most explicit efforts to enumerate characteristics that define Amazonia as a linguistic area, and as complementary to an Andean linguistic area. The grammatical features proposed to be ‘shared by all (or most)’ Amazonian languages range from polysynthesis and head-marking, to TAM categories being expressed as optional suffixes, to adverbs and adpositions being incorporable into verbs. As Birchall (2014) observes, however, the empirical basis for the claimed areality of these features is unclear (as, indeed, they are for Derbyshire’s (1987), Migliazza’s (1985), and Payne’s (1990) proposals, discussed above), raising the need to move beyond impressionistic claims regarding areality to explicit quantitatively-grounded methods that make use of suitably organized, sufficiently large and dense, and ideally, publicly available datasets (Haspelmath 2004).

Steps towards databases of these types for South America include the South American Indigenous Language Structures database (Muysken et al. 2014), and the South American Phonological Inventory Database (SAPhon; Michael et al. 2012), and quantitatively sophisticated work based on these resources has recently begun to appear. Significantly, these works support not an Andean–Amazonian areality split, that leaves Amazonia as a clearly defined area, but a different west-east split – where the western area corresponds roughly to the Andes, Southern Cone, and Doris Payne’s Western Amazonian region, while the other large linguistic area to the east consists of the remainder of the continent (see also van Gijn et al., this volume).

One illustration of this east-west division can be seen in Birchall’s (2014) study of argument marking features in 74 South American languages, in which he tests for a statistically significant concentration in each of seven South American geographical macro-areas. Echoing Krasnoukhova’s (2012) qualitative results on noun-phrase features in South America, Birchall’s analysis suggests that the features he examined pattern similarly within an Eastern South American Linguistic Area (ESALA) on one hand, comprising northern and southern Amazonia and the Chaco-Planalto area, and within a Western South American Linguistic Area (WSALA) on the other, comprising the north and central Andes, western Amazonia, and the Southern Cone. For example, Birchall (ibid: 203) finds that ergative alignment, suggested by Dixon and Aikhenvald (1999) to
be a general Amazonian feature, has a statistically significant association only with the southern Amazon region and ESALA, and not with northern or western Amazonia. Similarly, clusivity distinctions, proposed by Adelaar (2008: 29) to be an Andean feature, do not emerge as particularly Andean, but once again, as WSALA feature (Birchall 2014: 205-206). In contrast, several features, such as the use of both indexation and case as argument marking strategies, and accusative alignment for NP and pronoun arguments, turn out to be not mainly Andean, as proposed by Dixon and Aikhenvald (1999), but WSALA characteristics, providing further evidence for an E/WSALA areality split rather than an Andean/Amazonian one.

These results are broadly congruent with quantitative computational work on phonological areality in South America, which finds that lowland languages exhibiting phonological similarities to Andean ones – for example, uvular and ejective consonants, palatal laterals, multiple liquids, and small vowel inventories – cluster near the central Andes and in the Southern Cone (Michael et al. 2014), forming a phonological area corresponding roughly to WSALA. Similarly, languages in the remainder of the continent, corresponding approximately to ESALA, exhibit larger vowel inventories that include mid and high central vowels, nasal vowels or supersegmental nasality, labial fricatives, and a glottal stop, among other features.

Despite the lack of support for a general Amazonian linguistic area evident in these results, there are nevertheless phenomena that are found in many Amazonian languages, although these are not pervasive enough to be diagnostic of a linguistic area in the usual sense of the term. In recognition of this fact, Aikhenvald (2012: 68-71) has more recently introduced the term ‘language region’, in contrast to the more rigorously defined notion of ‘linguistic area’, to characterize Amazonia as a whole. The precise explanation for recurrent but sporadically attested ‘Amazonian’ features, such as antipassives and complex classifier systems (ibid.: 70), is unclear, but there is suggestive evidence that in some cases, linguistic items and features have diffused along long-range networks of the kind discussed earlier in this section. For example, Epps (2014; see also Bowern et al. 2014) identifies several dozen Wanderwörter that indicate the widespread diffusion of terms associated with important animal and plant species and food items across much of northern Amazonia. Epps also observes a widespread tendency across Amazonia for numeral terms indicating ‘4’ (and occasionally ‘3’ and ‘5’) to be formed using ‘relational’ nouns or verbs, usually meaning ‘companion’ or ‘accompany’, a strategy that appears to be rare outside of Amazonia (Epps et al. 2012, Epps 2013). Epps suggests Tupí-Guaraní languages in particular may have played an important role in the diffusion, but observes that the ultimate source of these apparently diffused items remains an open question. Regardless, the fact that items like these have circulated widely but sporadically across Amazonia suggests that borrowing may be mediated by sparse networks that link relatively distant languages without directly affecting closer neighbors.

Another line of research addressing Amazonian areality seeks to identify areal patterns not in grammatical structure, but in discourse practices and language ideologies. Beier et al. (2002), for example, argue that particular discourse practices, such as the dialogical discourse genres, ritual wailing, and the pragmatically-motivated use of evidentials are common over large areas of Amazonia in a manner consonant with that of a language region. Bowern et al. (2011) similarly find that lexical borrowing is unusually low in Amazonia, in comparison with other global macro-regions (see also Section 2 above), a tendency that Epps (forthcoming) attributes to a widely diffused language ideology that discourages language mixing.
4 Conclusion

Amazonia offers important insights into the dynamics of language contact and the development of areal linguistic patterns. As this chapter has explored, striking similarities exist among regional contact zones throughout the Amazonian lowlands, in which intense interaction and a degree of cultural homogeneity tend to go hand in hand with high linguistic diversity. In these regions, language is afforded special salience as a marker of distinct social identities, and thus while individuals are often highly multilingual, there is a marked absence of phenomena that are commonly associated with multilingualism in other parts of the world – most notably code-switching and lexical borrowing. Similarly, linguae francae and clear examples of language shift (aside from those associated with the expansion of the European and Quechuan spheres), while common in many other regions of the world where multiple languages are represented, appear to be rare or even unattested in these Amazonian contact zones. At the same time, the diffusion of grammatical categories and structures across languages is ubiquitous in these regions, although for most of these contact zones these processes have only begun to be explored. In some cases, the grammatical borrowing even includes the transfer of bound morphology, most notably that associated with nominal classification. This prevalence of extensive grammatical diffusion coupled with restrained lexical borrowing is typologically significant, since current conceptions of language contact dynamics stress the importance of lexical mediation in the diffusion of grammatical material, and particularly of bound morphology.

On a larger scale, Amazonian linguistic areality provides some intriguing glimpses into the dynamics of human interaction in prehistory, when extensive trade routes would have criss-crossed the Amazon basin and linked it with other parts of the continent. The broad east-west division outlined above suggests that social networks linking the Andes, the western lowlands, and the Southern Cone on one hand, and the eastern and central lowlands on the other, would have been particularly active, whereas networks that set the Amazon basin apart from other regions may have been less significant or functioned in different ways.

The study of areality in Amazonia and South America more generally has entered an exciting new phase. The empirical basis for such studies has been greatly enriched by the recent surge in basic descriptive work, which in turn is feeding a number of databases that will facilitate systematic study. Likewise, new analytical and theoretical approaches to areality are being developed through the application of quantitative techniques and via holistic approaches that take geography, demographics, and culture into account. We look forward to exciting findings in the coming decades, and to the new insights these will provide into South American prehistory and the mechanisms and processes involved in language contact.

References


Guillaume, Antoine. Forthcoming. Sistemas complejos de movimiento asociado en las lenguas de las familias Tacana y Pano: perspectivas descriptiva, comparativa y...
tipológica [Complex systems of associated motion in the Takanan and Panoan families: Descriptive, comparative, and typological perspectives]. Amerindia. (Special edition, Estudios sincrónicos y diacrónicos sobre lenguas Pano y Takana: fonología, morfología y sintaxis [Synchronic and diachronic studies of Panoan and Takanan languages: Phonology, morphology, and syntax], ed. by A. Guillaume and P. M. Valenzuela).


Seifart, Frank, Doris Fagua, Jürg Gasché, and Juan Alvaro Echeverri (eds.). 2009. *A Multimedia Documentation of the Languages of the People of the Center.* Online publication of transcribed and translated Bora, Ocaina, Nonuya, Resigaro, and Witoto audio and video recordings with linguistic and ethnographic annotations and descriptions. Nijmegen: DOBES-MPI.


31 Linguistic areas, linguistic convergence, and river systems in South America

Rik van Gijn, Harald Hammarström, Simon van de Kerke, Olga Krasnoukhova, and Pieter Muysken

1 Introduction

The linguistic diversity of the South American continent is quite impressive. There are 576 indigenous languages attested (well enough documented to be classified family-wise), of which 404 are not yet extinct. There are 56 poorly attested languages (which nevertheless were arguably not the same as any of the 576), of which 3 are known to be not yet extinct (Nordhoff et al. 2013). Furthermore, there are 71 uncontacted living tribes, mainly in Brazil, which may speak anywhere between 0 and 71 languages different from the ones mentioned (Brackelaire and Azanha 2006).

Perhaps the most impressive aspect of the linguistic diversity of South American indigenous languages is their genealogical diversity. The languages fall into some 65 isolates and 44 families (Nordhoff et al. 2013, Campbell 2012a), based on a comparison of basic vocabulary. Nearly all of the families show little internal diversity, indicative of a shallow time-depth. Alongside this picture of genealogical diversity, many areal specialists have noted the widespread occurrence of a number of linguistic features across language families, suggesting either sustained contacts between groups of people, or remnants from times when the situation in South America may have been genealogically more homogeneous.

We only have historical records going back about 500 years, and the archeological record is scanty and fragmentary (Eriksen 2011). It follows that there are large tracts of South American prehistory that are unknown. Areal-typological analysis aims to address this gap by finding out which languages have been in contact and how, as evidenced by their typological features. Potentially, the location and intensity of language contact could answer questions about the (un)likelihood of deep genealogical relations, duration of interaction, and provide clues regarding the settlement of the continent. Intertwined with the goal of unraveling history is the modeling of different contact processes. We are not yet in a position to predict the linguistic outcome from a given contact scenario or vice versa (Muysken 2010). To achieve this, attested contact scenarios and attested typological feature distributions need to be confronted further.

However, there is another potentially significant set of variables that is so far largely unexplored in areal-typological linguistics, namely those related to physical geography. If physical geography mediates or even to some extent determines the location and intensity of contact in the region, as well as the dispersal of languages, specific historical events may play a less important role in accounting for the data. In the present paper, we take up one variable from physical geography commonly thought to play an important role in South America, namely waterways.

We begin this paper with a selective overview of previous studies in the field of areal linguistics in South America, discussing how the present study relates to them (section 2). Section 4 describes the Noun Phrase structure questionnaire used for this study, and sections 3 and 5 discuss the aspects of physical geography relevant to this study. In
section 6 we present the results of our survey, and in section 7 we discuss the implications of our results and suggest some further lines of inquiry.

2 A brief history of areal-linguistic studies in South America

2.1 Linguistic areas

Traditional areal studies in South America have their roots in diffusionism in cultural anthropology, as reflected, for instance, in the writings of the Swedish archaeologist and anthropologist Nordenskiöld (1877–1932). A linguistic example of this approach is the work of Tessmann. Tessmann (1930: 617-627) attempted to group the languages of the Peruvian Amazon on the basis of lexical criteria (using a 33-word list). He assumes that the five largest accepted families represent “pure tribes” and groups the remaining languages as “mixed”, depending how many lexical resemblances they show vis-à-vis words typical of the pure tribes. Areal influence is conceived here in terms of influence from large language families, as Tessmann (1930: 618) holds that “there is a considerable number of isolates” but “nothing is achieved by positing a family of its own for each isolate”.

A recurring theme in later work in the areal linguistics of South America is often the Andean / Amazonian divide. The dominant perception (also held in neighboring disciplines such as anthropology, ethnohistory, and archeology) is that there is a basic two-way split in the indigenous languages of South America, between Andean (surveyed by Adelaar with Muysken 2004) and non-Andean or Amazonian (surveyed by Dixon and Aikhenvald 1999). This split is reflected in ideological constructs, with concomitant mental images, of the Andean highland societies and the Amazonian river-side village cultures, of the llama and the canoe. Andean societies are also commonly assumed to be “more complex, with age-old sedentary habits and a highly diversified and technically well-developed architecture”, to cite Adelaar (2008: 23-33).

The Peruvian linguist Alfredo Torero attempted in his final magnum opus (2002: 511-544), *Idiomas de los Andes*, to characterize the languages of the Andes typologically (see also Van de Kerke and Muysken 2014). He contrasts the Andes with Mesoamerica, the Area Intermedia proposed by Constenla (see below), and a possible Amazonian area (the latter with a question mark). The nine languages/small families taken into consideration are: Aru (Aymaran), Cunza, Cholon, Huarpe, Quechua, Uru(ullqa), Mochica, Puquina, and Mapuche. Torero uses 40 internal differentiation features to classify the nine languages. Out of the resulting 360 data points, only 11 are question marks. On the basis of his feature classification, Torero distinguishes a number of sub-areas (sometimes limited to a single language): *Nuclear* (Quechua, Aru), *Altiplano* (Uru, Puquina, Cunza, Huarpe), *Mochica*, *Cholon*, and *Mapuche*.

The approach taken by Adelaar (2008) for the Andes is relatively hard to summarize, since it is not very systematic and at the same time very concise. Over 52 languages are taken into consideration, spanning the entire length of the continent, and including the slopes on both sides of the Andes. It is clear that Adelaar conceives of the Andes in very wide terms. On the basis of his survey of structural features, Adelaar concludes that (2008: 31) “there is still very little evidence that can be helpful for recognizing and delimiting linguistic typological areas, let alone, an Andean linguistic
area that would encompass the entire region.” In fact, the area can be best defined negatively (2008: 31), when contrasted with the rest of the continent:

Negative defining features of Andean languages (after Adelaar 2008)

- Suffixing
- Case marking
- No prosodic nasality
- No tone
- No complex vowel systems
- No nominal classifier systems
- No gender
- No stative-active systems
- No well-developed ergativity

The Amazonian linguistic area is arguably more controversial than the Andean area. Nevertheless, specialists have observed several linguistic traits that are spread over large areas, and across languages families. We summarize the different characterizations of the Amazonian features in the above list. Work has been done in this area by, among others, Derbyshire and Pullum (1986: 16-20), Doris Payne (1990), and Dixon and Aikhenvald (1999: 8-10). Those features that are mentioned by several authors are underlined; those features also found in the Andes are italicized.

Table 1. Features used to classify the languages of the Amazon

<table>
<thead>
<tr>
<th>WORD ORDER</th>
<th>Derbyshire and Pullum</th>
<th>Payne</th>
<th>Dixon and Aikhenvald</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OS languages, Noun-Adjective, Possessor-Possessed, Noun-Postposition</td>
<td>Verb initial orders</td>
<td>Possessor possessed</td>
</tr>
<tr>
<td>CO- AND SUBORDINATION</td>
<td>No coordinating conjunctions, juxtaposition (Remnants of) ergative case marking</td>
<td>Subordination often through nominalization</td>
<td></td>
</tr>
<tr>
<td>ALIGNMENT</td>
<td>Few oblique cases, many signs of ergative, purely nominative/accusative rare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MORPHOLOGY</td>
<td>High degree of polysynthesis</td>
<td>Polysynthesis, head marking, agglutination, many prefixes, but more suffixes, person marking towards the periphery of the verb, many TAME categories optional</td>
<td></td>
</tr>
<tr>
<td>CATEGORIES</td>
<td>Directional suffixes in the verb, nominal</td>
<td>Extensive semantic gender or noun class marking, person markers in the nominal paradigm identical to the verbal</td>
<td></td>
</tr>
</tbody>
</table>
The Costa Rican Chibcha specialist Constenla (1991) has argued for a linguistic area in between the Andes and the well-known Meso-American area, which we may call the Colombian-Central America area. He identifies almost twenty features as characteristic of the southern part of his sample, including parts of Ecuador and Peru. This part of his sample is assumed to be characterized by the features in Table 2.

Table 2. Features of the Colombian-Central America area in Constenla (1991)

<table>
<thead>
<tr>
<th>SEGMENTAL PHONOLOGY</th>
<th>SYNTAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absence of mid-high contrast in back vowels, absence of voiced/voiceless distinction in affricates, a voiceless alveolar affricate, palatal consonant subsystem, retroflex fricatives and affricates (not Quechua), mid-high contrast in front vowels (not Quechua), no glottalized sounds, no uvulars, voiceless labial fricative, roundedness opposition in back vowels (not Quechua)</td>
<td>Adjective Noun order, Numeral Noun order, question word in initial position, accusative case, genitive case, no nominal person marking, tense and aspect with prefixes (not Quechua)</td>
</tr>
</tbody>
</table>

Apart from these three major proposed areas of diffusion, proposals for smaller areas have also been made. Perhaps the best-known example of this is the Vaupés area (Aikhenvald 2002) in the border area between Columbia and Brazil, where Arawakan, Tucanoan, and – to a lesser extent – Nadahup languages share a complex contact network resulting in convergence between the languages. The Guaporé-Mamoré area between the Bolivian and Brazilian border (Crevels and Van der Voort 2008) is composed of two to three cultural areas. It is characterized by extreme genealogical diversity, but the languages share a number of structural traits; as argued by Crevels and Van der Voort these shared features are the result of contact. Seki (1999) argues that the upper Xingú area in central Brazil should be regarded as an incipient linguistic area on the basis of lexical borrowing, and a few shared structural traits. Further cases of linguistic areas are summarized in Campbell (2012b: 301-309). It is interesting from the perspective of this paper that in particular these smaller areas are associated with river systems.

2.2 Areal typology

Until the present century, all proposals for linguistic areas, although drawn from decades of expertise in certain areas, were argued on the basis of impressions and selectively chosen feature lists. Recent studies have effectively closed the methodological gap by defining objective procedures that find and test prototypical linguistic areas, given languages features and language locations as input. Muysken et al. (2015) provide a method for testing whether a proposed linguistic area actually has more homogeneity
than expected by chance, and a method for finding the area delimitation that has the strongest areal signal. Similarly, Michael et al. (2014) provide a slightly different method for testing whether a proposed area is distinguishable from its complement, and Chang and Michael (2014) define a method for inferring borrowing (as opposed to inheritance) and show that a series of borrowings between neighboring languages is tantamount to finding a linguistic area.

Another set of recent studies set out to objectively test various proposals concerning feature distributions, typically involving (and calling into question) the Andes-Amazon divide, using standard statistical techniques to explore systematically filled-in databases. Van Gijn’s survey (2014a) of the distribution of Andean and Amazonian features in the upper Amazon area shows that the transition from the Andean to the Amazonian area is gradual and complex, consistent with the intricate history of contact between the different ethnic groups of the area, and casting doubt on the reality of the Andean-Amazonian divide.

In the domain of verbal argument marking, Birchall (2014a, 2014b) examined the diverse array of verbal argument marking patterns encountered across the continent and tested for regional distributions of certain often-discussed features. Statistical tests showed that many proposed areal distinctions in the literature are in fact not significant, and that an east-west division was often more significant than the classic Andean-Amazonian division. For instance, an inclusive / exclusive distinction in verbal argument marking does have an east vs. west distinction. Contrary to the suggestion by Adelaar (2008: 31), Birchall (2014a: 205-206) finds that “the languages of the Andes show a lower distribution of clusivity in indexation than the rest of the continent” and that eastern South America shows “a statistically significant areal distribution” of clusivity in indexation.

In the nominal domain, Krasnoukhova (2012, 2014) has also shown that in Noun Phrase structure there is a split between languages spoken in the western part vs. the eastern part of the continent, and not between the Andes and the Amazon as has been traditionally assumed. While the western part corresponds to the Andean sphere, the eastern part includes languages spoken far beyond the Amazon region. Furthermore, in a case-study on semantic features encoded by demonstratives, Krasnoukhova (2014) has shown that the Chaco and the Southwest Amazon region stand out on the continent for encoding verbal categories with demonstratives (a case of the “Trait-sprawl areas” suggested by Campbell, this volume).

Other areal-typological studies suggest that linguistic features that are statistically overrepresented in South America as opposed to other parts of the world do not necessarily show clear systematic patterns of diffusion within the continent. Van Gijn (2014b) showed that nominalization as a subordination strategy is significantly more pervasive in South America than would be predicted on the basis of global patterns. The patterns found within South America are most consistent with a scenario of several smaller spreads, possibly promoted by a few language families with major extensions (e.g. Quechuan, Tupian, Cariban). In the domain of Tense/Aspect/Mood/Evidentiality (TAME) systems, Müller (2014) presented evidence that there is little systematicity in either the genealogical or the areal distribution of specific features, although there are several continent-wide general patterns, including highly frequent desiderative marking.

3 Exploring the role of geography: river systems
So far, physical geography has remained an understudied factor both in the study of linguistic areas and of areal typology. Yet, several authors (e.g. Nichols 1990, 1992, Nettle 1999, Dahl et al. 2012, Evans and Levinson 2009, Hammarström and Güldemann 2014) have surmised that physical geography plays an important role in shaping patterns of diversity around the globe. This is also the picture that emerges from the overview of the linguistic area studies given above.

The present study is intended as a first step towards integrating factors of physical geography into areal-linguistic studies. We are fully aware that a direct connection between geography and linguistic diversity is grossly oversimplifying the complex relation between humans, their environment, and their personal, social, and cultural answers to this ecology. It is therefore also not our goal to present a deterministic picture of human behavior. Rather, we want to try and isolate the possible contribution of geography to the patterns of linguistic diversity, in order to assess the facts unaccounted for and which should most probably be explained in terms of human choice and socio-cultural organization and innovation. Having thus acknowledged the relative simplicity of the model presented here, we must try and assess which aspects of geography are important to the rise and maintenance of ethno-linguistic diversity.

We conceive of the role of geography in shaping patterns of diversity as a rather indirect one. In broad terms, the physical surroundings potentially influence the migratory and contact-seeking or -avoiding behavior of groups of people, which in turn influences diversification patterns and creates new contact situations. The role of geography in these processes has two sides: there is a set of factors to do with the incentive to spread or move, or get into contact with one’s neighbors, and on the other hand a set of factors that either facilitate, or hamper or impede mobility, in terms of pathways and barriers.

3.1 Incentive factors

Incentive factors can be described by what Nettle (1999) calls ecological risk, defined as “the probability of a household facing a temporary shortfall, at whatever timescale, in food production” (p. 79). In a constant and favorable climate and a rich environment (such as a seashore with plenty of mollusks), people are less forced to look for strategies to cope with ecological risk. If the climate is less dependable and the environment less rich, peoples may resort to a variety of strategies. The most important one in terms of language diversity is ‘exchange’, since that creates ties between peoples that become so strong that people start to identify or at least interact with each other, thus facilitating the spread of linguistic features.

What are the ecological factors that play a role in the degree of ecological risk? Nettle (1996, 1999) uses climatic data (rainfall and temperature per month) to calculate the length of the growing season. Based on these figures, each month is either included or excluded as part of the growing season. Months for which the average daily temperature is above 6°C are included in the growing season if the average precipitation in millimeters exceeds twice the average temperature per month. The rationale behind this measure is: the longer the growing season, the lower the ecological risk and therefore the lower the incentive to seek contacts with other groups or to expand or move to other
territories. Therefore linguistic diversity is expected to be especially high in areas of low ecological risk.

The growing season is a proxy for the degree of ecological risk, and as such it could be refined further on the basis of e.g. soil fertility, risk factors - especially the risk of inundation - factors facilitating the presence and thriving of high-protein food (like larger animals and fish), like the presence of water and of dense forest. The notion of growing season is an important one, but should be interpreted also in terms of the availability of specific crops or gatherable plants.

3.2. Mobility factors

Another set of factors of potential influence on linguistic diversity are mobility factors. It is probably true that a people with an incentive to migrate will do so, no matter what the difficulties they encounter on the way (see e.g. Nettle 1996 for a critique of explanations for linguistic diversity on the basis of topographical isolation hypotheses). Considering the topographical context in terms of barriers or pathways of mobility is nevertheless useful, because it may still predict likely directions of spread, or perhaps highlight the fact that topographical barriers may raise the threshold for incentives to leave. Nichols (1990, 1992) observes that mountainous areas are in some cases linguistically diverse (Himalayas, Caucasus) but in others (e.g. the Andes versus the foothills and adjacent lowlands in South America) the diversity patterns are rather reverse. We will show that the Andes is much more homogeneous than neighboring areas.

In the absence of a fully-fledged theory of the interaction between topographical factors and human mobility our approach is sketchy and tentative, and focuses only on one part of the programmatic sketch outlined above: the mobility factor of waterways. Especially in the Amazon area, the many rivers provide possibilities for faster travel over relatively long distances, and increased carrying capacity. To the present day, boats are the main means of transportation for many of the indigenous peoples of South America.

In order to get a firmer grip on the role of physical geography in shaping patterns of diversity, we aim to answer the following three questions:

1. Are river system networks congruent with land distances as the crow flies? If not, do riverine distances mirror typological distances better than distances as the crow flies?
2. Do languages on the same river system converge on the same structural profile?
3. Are there specific features that spread easily along rivers?

To answer these questions, in the next two sections, we give a more detailed introduction to the two major components of our study, the Noun Phrase database (section 3) and the classification of geographical areas in South America (section 4).

4. The Noun Phrase (NP) database and earlier results

As mentioned, the structural features that we take into account in this paper relate to NP structure. Technically speaking, the basis for the analysis is a questionnaire database, developed by Krasnoukhova (2012) in order to outline the structure of the NP in South
American languages. The NP questionnaire consists of 6 general questions (e.g. constituent order at the clause level) and 50 questions which specifically relate to the NP. The latter group of questions is divided into 31 main questions and 19 dependent questions. The areas of the NP which are explored in the questionnaire include those listed below.

Areas of the Nominal Phrase examined in research questionnaire

NP structure:
- Constituent order within the NP. Such modifier categories as demonstratives, numerals, and adjectives were all approached as semantic categories.
- Presence and realization of agreement within the NP.

Modifiers within the NP:
- Articles, demonstratives;
- Adjectives, grammatical status of adjectives;
- Numerals, grammatical status of numerals;

NP related issues:
- Grammatical expression and conditions on the realization of number within the NP;
- Noun categorization devices, such as classifiers, and gender and noun class systems.
- Attributive possessive constructions. The parameters under investigation include: head vs. dependent marking of possession and the presence and formal realization of (in)alienability.
- Spatial deixis, with a focus on semantic features that can be encoded by adnominal demonstratives.
- Grammatical marking of temporal distinctions within the NP.

This questionnaire was filled in by Krasnoukhova and Van de Kerke for 97 languages using descriptive materials and whenever possible using information provided by specialists working on a specific language. The language sample used for this paper includes representatives of 30 language families and 16 isolates. When constructing the sample we paid attention to the areal distribution of languages in order to ensure an adequate geographic coverage. Appendix 1 below shows the languages included in our sample with their genealogic affiliation.

In what follows, we give a summary of observations based on Krasnoukhova (2012, 2014) with respect to the areal distribution of NP features. Krasnoukhova observed that there is evidence in the NP domain for a split between languages spoken in the western vs. the eastern part of the continent, and not so much the split between the Andes and the Amazon. The western group consists of languages spoken along the western part of the continent and roughly corresponds with the Andean sphere, while the eastern group includes languages spoken in the rest of the continent, and thus is not limited to the Amazon region. Namely, the eastern group includes languages spoken far beyond the Amazon region and includes, for instance, the eastern and southern part of Brazil, the Chaco, and the Southern Cone. What are the linguistic features that suggest the west vs. east split in the NP domain?

The most robust features of the languages spoken in the ‘western’ part are the following:
(1) a. word order within the NP: pre-head position for all modifiers (demonstratives, lexical possessors, numerals, property words);
   b. property words are morphologically nominal;
   c. absence of gender in NP;
   d. absence of gender distinction in personal pronouns;
   e. absence of classifiers;
   f. absence of temporal distinction within the NP;
   g. absence of a class of inalienable nouns.

The most robust features of the languages spoken in the ‘eastern’ group can be condensed to the following list:

(2) a. word order within the NP: pre-head position for demonstratives, lexical possessors and numerals, and posthead position for property words;
   b. small or no distinct adjective class;
   c. property words are verbal;
   d. presence of gender in the NP;
   e. presence of classifiers, often of the multifunctional type;
   f. presence of temporal distinction within the NP (a less robust feature);
   g. presence of a class of inalienable nouns (a very robust feature).

Just to give a few concrete illustrations, for instance, hardly any language of the Andean sphere encodes property words through verbs, whereas languages in which property concepts are encoded by verbs are predominant in the Northwest Amazon and the Southwest Amazon regions; they are also found in the Chaco (e.g. Tapieté and Wichí), the eastern and southern part of Brazil (e.g. Timbira and Bororo), and in the Southern Cone (Tehuelche).

As for NP constituents, templates in which all modifiers tend to occur on one side of the noun are all found along the western edge of the continent, e.g. in Aymara, Huallaga Quechua, Imbabura Quechua, Leko, Mapuche, Tsafiki and Yaneshá’. In Miraña, a language of the Northwest Amazon region, all modifiers also tend to occur pre-head, therefore it constitutes an exception in this observation. Conversely, a template in which some modifiers (mainly, demonstratives, possessors and numerals) always precede the head noun and some modifiers (mainly, property words) always follow it, is found predominantly in languages outside the Andean sphere: e.g. Warao, Ninam, Dâw, Hup, Puinave, Urarina, Matsés, Yaminahua, Jarawara, Baure, Movima, Itonama, Mekens, Gavião, Wari’, Karo, Kanoè, Mamaindê, Sabanè, Wichí, Pilagá, Chamacoco, Bororo, Kamaiurá, Trumai, and Timbira. There are a few exceptions here: the latter template is also found in three languages spoken in the northwestern part of Colombia, viz. Ika, Nasa Yuwe, and Northern Embera.

To give one more example, a great majority of languages spoken along the western edge of the continent lack a class of inalienable nouns, whereas languages in the rest of the continent predominantly have such a class. Specifically, the following languages do not have a class of inalienable nouns: Quechua and Aymara, spoken in the Andes; Mapuche, spoken in the Southern Andes; Tsafiki, Awa Pit, Nasa Yuwe and Northern Embera, spoken in the western part of Ecuador and Colombia. Inalienable nouns are also absent in Aguaruna, in the northern Peruvian foothills, and Shipibo-Konibo and Urarina,
spoken in the Western Amazon (Peru). Exceptions from the observations are Warao (spoken in western Guiana and northeastern Venezuela) and Kwaza and Sabanê (spoken in the Southwest Amazon region). All other languages in the sample have a class of inalienable nouns.

The data on the NP also show that the following features cannot be treated as characteristic of any particular larger area (western vs eastern part, specifically Andes vs Amazon), as they are found in the languages across the continent (see also the discussion in Campbell 2012b: 301-304):

(3) a. number distinction/marking in personal pronouns;
   b. number distinction/marking in the NP;
   c. inclusive/exclusive distinction in free personal pronouns.

Such feature as ‘locus of possession marking’ was carefully examined for areality too, since the head-marking pattern has been proposed in the literature among the features of the ‘Amazonian linguistic area’ and the double-marking pattern was proposed among the features of the ‘Andean linguistic area’ (Dixon and Aikhenvald 1999: 8, 10). Our NP data showed no evidence for this areal division. It is correct to say that the double-marking pattern is found predominantly in the languages spoken in the Andes (specifically, among the Aymaran and many of the Quechuan variants, and in Aguaruna), but this is not the only possession strategy found in the languages spoken in the Andes. For instance, Chipaya is dependent-marking, as well as Imbabura Quechua. Likewise, there is no good evidence to generalize that Amazonian languages are predominantly head-marking for possession: both a dependent- and head-marking pattern are equally common in the Amazonian languages in the sample. What was found was a much smaller areal clustering: namely, dependent-marking languages are more concentrated in the Western Amazon region, while head-marking languages are more present in the Bolivian lowlands and in the Chaco. In addition, head-marking languages are found along the western coast of South America (e.g. Tsaiki, Ika) and in the Southern Cone (e.g. Tehuelche) (see Krasnoukhova 2012 for more details).

5 The major drainage basins of South America

The importance of rivers and river systems for the population dynamics in South America cannot be overestimated (Dunne and Mertes 2007). It is no coincidence that many proposals for linguistic areas of the continent are associated with river systems: the Vaupés-Içana (Aikhenvald 2002), the Guaporé-Mamoré (Crevels and Van der Voort 2008) and the Xingú (Seki 1999). River systems, and the ecologies associated with them, to a considerable extent determine the possibilities for gathering and cultivating food, cultivating plants, and they facilitate mobility, and with it contact between peoples. The effect of the ecological circumstances can be expected to be even greater for locally organized, rural economies (Nettle 1999), as is the case for most indigenous cultures of South America.

The connection between linguistic areas and river systems, mentioned above, raises the question of whether there is a more general interaction between river systems/ecologies and the degree of linguistic convergence. What we set out to do in this chapter, therefore, is to define a number of ecological zones mainly based on river
systems, and look for patterns of structural diversity and convergence within each of these areas.

5.1 The South American physical space

The elements responsible for creating the ecological “hardware”, i.e. the physical landscape and the basic ecological conditions of an area are tectonism (the shaping of the Earth’s crust) and climate (Orme 2007). In turn, these two constrain flora and fauna, and to a certain extent, human activity. The South American continent is delimited in the northwest by the border of Colombia with Panama, and on all other sides by seas and oceans, the Caribbean Sea to the north, the South Pacific Ocean to the west, and the Atlantic Ocean to the east. Broadly speaking, the physical geography of South America is dominated by three highland areas: the ancient, Precambrian (prior to 541 million years ago) Guyanan and Brazilian highlands, and the much younger Andean mountain range, which only assumed its present shape in the Cretaceous period (145-66 million years ago). The rivers that spring from these highland areas flow through the vast lowland areas, forming some of the largest river systems in the world. These elements form the basis that underlies the different ecological systems of the continent.

In order to achieve a more local perspective (assuming that contact in local ecosystems or river systems becomes even more manifest), we divide the continent into several smaller regions. First, we distinguish between the very different ecosystems of the Pacific, the Andes and the Amazon. Second, we make further distinctions within the Amazon and Andean area, based on more local river systems and other factors.

5.2 Ecological zones

The Andean mountain range is usually divided into a southern, central, and northern part. However, there are different ways to determine the borders of these areas (Torero 2002: 13): to the south the Bolivian altiplano may or may not be counted as part of the Southern Andes, and to the north, the Ecuadorian Andes are either included in or excluded from the Northern Andes. In line with our continental north-south divide, and with the secondary north-south divide in Western Amazonia, we put the border between the central and Northern Andes in Northern Peru/Southern Colombia, where the Andean mountain range takes the shape of three parallel Cordilleras. To the south, the border is determined by the very broad Bolivian Andean range, where the rivers flowing north join the Amazon whereas those flowing south join the Paraguay River, and eventually the River Plate. Because of low language diversity to the south of the Gran Chaco, the Southern Andes and Southern cone are considered together as a single area. The Andean areas are strictly speaking not river-defined areas, but they do form ecologically coherent areas, and they cannot easily be grouped with any of the river-based areas.

We also regard the Pacific coast as a separate area. Few Pacific coastal languages have survived, causing the focus of our Pacific area for the purposes of this paper to be on the northern Peruvian/Ecuadorian coast, where a relatively low passage of the Andes between Northern Peru and Southern Ecuador made trade between the coast and the Amazon possible (Adelaar with Muysken 2004: 6).
In the north-east, we separate the Orinoco basin in Venezuela from the adjacent Guyanas and Northern Brazil, and south of this area we recognize smaller river systems, which are defined in Table 3 below. The Río de la Plata basin (in particular the western part, the Gran Chaco in Northern Argentina, South Bolivia, and Paraguay) and the Southern Cone plus Southern Andes are also separated. In the list that follows, we briefly outline the smaller areas and indicate which of the sample languages are spoken in each region. The numbers in the first column of the table refer to the numbers on the accompanying map (Map 1).

Table 3. Regions, locations and sample languages

<table>
<thead>
<tr>
<th>Region</th>
<th>Delimitation/description</th>
<th>Sample languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Northern Andes</td>
<td>Columbian Andes</td>
<td>Arhuaco, Embera, Páez</td>
</tr>
<tr>
<td>2 Orinoco</td>
<td>Venezuelan Orinoco basin</td>
<td>Nhengatú, Ninam, Panare, Puinave</td>
</tr>
<tr>
<td>3 Guyanas</td>
<td>French and British Guyana, Surinam, Northern Brazil.</td>
<td>Emérillon, Hixkaryana, Papiamento, Trio, Warao</td>
</tr>
<tr>
<td>4 Pacific</td>
<td>Pacific coast: practically the Ecuadorian Pacific.</td>
<td>Awa-Cuaquier, Colorado, Mochica</td>
</tr>
<tr>
<td>5 Napo-Marañón</td>
<td>Ecuadorian and North Peruvian eastern slopes and lowlands</td>
<td>Bora, Aguaruna, Kokama, Muniche, Shuar, Urarina, Cofán, Iquito, Arabela</td>
</tr>
<tr>
<td>6 Vaupés</td>
<td>The Columbian-Brazilian border.</td>
<td>Hup, Cubeo, Dâw, Desano, Tariana</td>
</tr>
<tr>
<td>7 Solimões</td>
<td>Western Brazil between the Juruá and Purus Rivers, and up to the Ucayali in the west.</td>
<td>Jarawara, Matsés, Yaminahua</td>
</tr>
<tr>
<td>8 Tapajos-Madeira</td>
<td>River basins in Southern Amazonia directly east of the Madeira-Guaporé line.</td>
<td>Apurinã, Bororo</td>
</tr>
<tr>
<td>9 Xingú-Tocantins</td>
<td>Easternmost river basin in Southern Amazonia</td>
<td>Kamaíurá, Krikati-Timbira, Trumai</td>
</tr>
<tr>
<td>10 Central Andes</td>
<td>Ecuadorian, Peruvian and Bolivian Andes</td>
<td>Aymara, Cusco Quechua, Huallaga Quechua, Uru, Chipaya</td>
</tr>
<tr>
<td>11 Huallaga-Ucayali</td>
<td>Central-South Peruvian lowlands</td>
<td>Shipibo-Konibo, Yaneshá, Cholon</td>
</tr>
<tr>
<td>12 Beni-Guaporé</td>
<td>Area between the Beni and Guaporé Rivers in Bolivia and Rondónia</td>
<td>Baure, Cavineña, Gavião, Itonama, Kanoë, Karitiana, Karo, Kwazá, Lakondê, Leko, Mamainde, Moseten, Movima, Sabanë, Mekens, Warí, Yurakaré Wichí, Chamacoco, Mocoví, Pilagá, Tapiete Tehuelche, Mapudungun, Selknam/Ona, Güñúña Küne / Puelche</td>
</tr>
<tr>
<td>13 Río de la Plata basin</td>
<td>Northern Argentina, Paraguay</td>
<td></td>
</tr>
<tr>
<td>14 Southern Cone</td>
<td>Central and Southern Argentina and adjacent areas in Chile</td>
<td></td>
</tr>
</tbody>
</table>
Map 1. Smaller regions considered in the present study

Mapping the NP questionnaire database discussed in section 3 onto the geographical zones discussed in this section, we can now move to a discussion of the results.
6 Results

Given the geographical information about the various river systems, and the data sets on NP features for the languages in our sample, we are now in a position to try and answer a number of questions. As mentioned above, and repeated here for convenience, we address the following issues:

1. Are river system networks congruent with land distances as the crow flies? If not, do riverine distances mirror typological distances better than distances as the crow flies?
2. Do languages on the same river system converge on the same structural profile?
3. Are there specific features that spread easily along rivers?

Several methodological remarks are in order here. First, the added geographical information by itself does not help decide whether a certain feature distribution could reflect deep time geographical relationships or later contact. A river may be a conduit for exchange relations between unrelated groups as well as a path along which a population spread. Second, not all measures can be applied globally. It does not make sense to calculate river distance between two points which are not directly connected by rivers.

Our data on rivers in South America come from the freely accessible country-by-country river information of DIVA-GIS\(^1\). The granularity of these data is too fine for the purposes of this paper, as it includes what appears to be every small creek on the continent. We selected the 600 largest shapes resulting in what is impressionistically a granularity suitable for the purposes in the present paper. This yields the river system shown in Map 2, shown at a resolution of 1/8 of a degree per point. Languages are represented with their centrepoint coordinates from Nordhoff et al. (2013).

We define river distances only for pairs of languages which are in the same drainage area, i.e., languages that are actually connected by a river system. The river distance between two languages A and B is defined by the distance of A to the nearest point on a river plus the distance of B to the nearest point on a river plus the shortest river way path between those two points. Map 2 shows a line from every language location to its nearest point on a river (based on the river data explained above). As an example, in Map 3 the river path between the languages Hixkaryana and Nhengatu is indicated. The distance between Hixkaryana as the crow flies is 917 km whereas the river distance, following the river path, is 1642 km.

\(^1\) http://www.diva-gis.org/gdata accessed the 24 November 2013.
Map 2. River data and language locations. Lines from every language location to their nearest point on a river
Map 3. The river path between the languages Hixkaryana and Nhengatu
6.1 River vs bird distances

Before relating any linguistic data to water distances, we must ask if there are any essential differences between riverine distances and distances as the crow flies. There are some logical possibilities. One possibility is that river distance for all language pairs merely prolongs the bird distance by a certain percentage. In this case, the landscape of contact effects would be the same as if there were no rivers, and the only reason to confront linguistic data and river distances would be to estimate the size of the effect, not the locus. On the other hand, it may be that some language pairs differ very much from other pairs in terms of water distance. That is, some pairs of languages connected by a straight river path have a river distance which is very close in kilometres to their bird distance, while other pairs, that require a lengthy river-detour, have a river-distance far higher than their bird distance. In this case it would be possible to test for riverine contact as a factor different from contact by proximity in general. If so, the linguistic distances should correspond better to the riverine distances than the bird distances.

Figure 1 shows all 510 pairs of languages for which water distances are defined, plotted against their bird distance. The water distance for most language pairs can indeed be found by adding a percentage to their bird distance, though not for all. The river distances between Cuzco Quechua and Quechuan languages further north requires a very long downward route via the Madre de Dios, up the Madeira, and back up the Amazon and the Ucayali, totaling around 5000 km. Similarly, there are other pairs which require river travel down the Napo in Ecuador only to turn around at the Japurá and go as far back up, totaling around 2000 km.
Figure 1. Bird vs water distance

For the linguistic distance between a pair of languages, we use the Hamming distance which is simply the percentage of features which do not have the same value in the two languages (cf. Hammarström and O’Connor 2013). For example, if 80 features are defined for the pair of languages and they agree in value for 20 of them, then the distance is 60/80=0.75. The correlation between bird distance and linguistic (NP) distance is shown in Figure 2. The correlation between water distance and linguistic (NP) distance is shown in Figure 3.
Figure 2. Bird vs NP distance

$r = -0.255826$
$y = 0.415 + 9e^{-0.05x}$
The size of the correlation for bird and NP distance is $r=–0.26$, while for water and NP distance it is slightly less strong ($–0.23$). So there is little overall improvement in using water distances to predict NP distance. We also did a more specific test, looking only at the regions where the river systems were thought to be more significant, i.e., Huallaga-Ucayali, Orinoco, Guaporé-Mamoré, Vaupés, Napo-Marañon, Tapajos-Madeira, Xingu-Tocantins and Solimões, thereby excluding pairs such as those involving the highly unnatural river route between Cuzco Quechua and other Quechuan languages. In those cases, there were plausible land routes, as indeed created during the Inca empire.

Doing this drastically changes both bird/water distance correlations with NP distance to the values $r=0.11$ and $r=0.12$ respectively. So again there was no appreciable
difference of water vis-a-vis bird distance. We can thus answer the first question. Riverine distances are very different from distances as the crow flies for many language pairs, but the linguistic differences are not better predicted by the riverine distances.

6.2 Riverine areas

While there is no simple relationship involving water distance and NP distance in our data, there may be a more indirect effect of river systems. Languages on the same river system may be hypothesized to interact along both the main waterways as well as cut across headwaters. Given sufficiently long time, these languages may then influence each other structurally. The hypothesis is that they influence each other more than any set of geographically adjacent languages which do not share a river system. We thus test if the languages belonging to the same riverine area are more similar to each other than 1) random collections of the same size (Sig Random), and 2) geographically coherent sets of languages of the same size (Sig Geo), and 3) random collections of languages of the same size from different families (FamI). To generate a random geographically coherent set of languages, we start from a random language, and then continue picking the geographically nearest language, until the desired size has been reached. The difference between riverine areas and geographically coherent areas is merely that the riverine areas are geographically coherent sets of languages necessarily associated with a river system. To generate a random set of languages from different families, we first select a random set of families of the desired number, and then select a random member from each. Table 4 shows the average distance of the languages within each area and the significance vis-a-vis random sets of languages, random geographically coherent sets of languages, and random sets of languages from different families, of the same size as the area in question. The significance number counts the number of random sets with a higher average distance. A number of 950 or higher can be equated with a conventional significance level of 0.05.

Table 4. Convergence of the NP-profile in different areas. Avg dist is the average distance for languages in the same area. Sig Random indicates how many (out of 1000) random sets of languages of the corresponding size have a lower average similarity. Sig Geo indicates how many (out of 1000) random sets of geographically coherent languages of the corresponding size have a lower average similarity. Sig FamI indicates how many (out of 1000) random sets (of the corresponding size) of languages from different families have a lower average similarity.

2 While we were concerned with differences between water and bird distance, the reader may have noted that the correlation in the first case was negative, i.e., that longer distance predicts a smaller distance! This unintuitive result is, however, due to the many pairs of languages belonging to the shallow but geographically widespread Quechua family. Quechuan languages are much more similar to other Quechuan languages far away than to many neighbouring non-Quechuan languages and, in general, much more similar to each other than a random pair of languages from the rest of the sample. In the second case, the Quechuan languages were not included, and the geographic vs linguistic correlation then showed the intuitively expected positive polarity.
Few, if any, riverine areas show significant convergence. The only one to clearly do so is the well-known Central Andes area, where the similarity is known to be due to genealogical inheritance and intense language contact, neither involving rivers in any significant way. Also the North Andes and Napo-Marañon areas owe their high significance numbers to the fact that they encompass subsets of very close Quechua varieties. Interestingly, the set of riverine areas tested includes prospective and established linguistic areas like the Guaporé-Mamoré and the Vaupés, whose NP convergence are well within the bounds of similarity expected for random sets of languages. Thus we can answer the second question: merely being in the same river system is not sufficient for convergence of the entire NP profile.

6.3 Riverine features

While entire typological profiles do not appear to converge in languages of the same river system, a weaker effect of rivers may be operational. Certain unstable features, not entire profiles, may spread easily along rivers. The hypothesis is that this concerns only the most unstable features, because if it concerned every feature the entire profiles would have converged. To get an idea of which features are likely to be inherently stable versus unstable, we take over the stability measure calculated in Muysken et al. (2014). This essentially measures the probability that a feature value remains the same in the transition from a proto-language to a modern language within a language family.

Suppose we define a “riverine” feature as one which tends to have the same value in each riverine area, i.e., homogenous within a riverine area. For example, if 2/3 languages in an area have the same value in a certain riverine area, then the homogeneity in that area is 0.67. The overall homogeneity is the average it has across all riverine areas. (For this test, we have used only the strictly river defined areas, excluding the Central Andes and Southern Cone areas.) We may also test for significance of homogeneity by comparing simulated random and geographically coherent sets of the same size (similar to the above). The riverine-ness of each feature is shown below (with the number of languages in the full sample for which the feature is defined, the overall homogeneity and significance of this).
Table 5 shows figures for how riverine a feature is along with its stability value. Although there are a number of NP features which achieve high significance individually, notably many that concern possession, the numbers have not been controlled for multiple testing. There are about equally many that achieve a correspondingly low significance, suggesting that there are random fluctuations in both directions. Also, more importantly, there is no correlation between low stability and riverine-ness ($r \approx 0.01$).

We are thus inclined to also answer the third question in the negative: we find little evidence for specific features whose spread would be facilitated by rivers.

Table 5. The columns of the table consist, from left to right, of (i) every NP feature, (ii) the number of languages for which it is defined, (iii) its average homogeneity in riverine areas, (iv) the significance of this homogeneity with respect to randomized mirrors of the riverine areas, (v) the significance of this homogeneity with respect to geographically coherent randomized mirrors of the riverine areas, and (vi) the feature’s stability value.

<table>
<thead>
<tr>
<th>Feature</th>
<th># lgs</th>
<th>Hom.</th>
<th>Sig Rando m</th>
<th>Sig Geo</th>
<th>Stabili ty</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP406: Do possessive constructions with optionally possessed nouns differ from those with obligatorily possessed nouns? (= is alienable and inalienable possession formally distinguished?)</td>
<td>61</td>
<td>0.683</td>
<td>0.999</td>
<td>1.000</td>
<td>0.872</td>
</tr>
<tr>
<td>NP405: Are there nouns denoting obligatorily possessed items?</td>
<td>94</td>
<td>0.809</td>
<td>0.999</td>
<td>1.000</td>
<td>0.901</td>
</tr>
<tr>
<td>NP408: Does the language have a separate form for third person possessive reflexive? Cf. example</td>
<td>34</td>
<td>0.824</td>
<td>0.996</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>NP160: What is the most frequent order of adjective and degree word?</td>
<td>31</td>
<td>0.807</td>
<td>0.994</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>NP13: In what way are indefinite pronouns realized?</td>
<td>30</td>
<td>0.807</td>
<td>0.994</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>NP409: Can pronominal possessors (pronoun, affix, clitic) be related to main tense and verbal person markers (pronoun, affix, clitic)</td>
<td>25</td>
<td>0.794</td>
<td>0.993</td>
<td>0.996</td>
<td>0.825</td>
</tr>
<tr>
<td>NP150: What is the most frequent order of relative clause and noun?</td>
<td>82</td>
<td>0.709</td>
<td>0.989</td>
<td>0.997</td>
<td>0.789</td>
</tr>
<tr>
<td>NP404: In possessive constructions with a pronominal possessor, is the POSSESSED noun usually marked? (state most frequent construction; NB: if marking depends on alienability, state marking of possessor in ALIENABLE constructions)</td>
<td>95</td>
<td>0.632</td>
<td>0.962</td>
<td>0.921</td>
<td>0.837</td>
</tr>
<tr>
<td>NP740: Is there a morphologically indicated collective marker?</td>
<td>81</td>
<td>0.691</td>
<td>0.947</td>
<td>0.962</td>
<td>0.869</td>
</tr>
<tr>
<td>NP305: How many distance contrasts do adnominal demonstratives encode?</td>
<td>91</td>
<td>0.668</td>
<td>0.900</td>
<td>0.951</td>
<td>0.757</td>
</tr>
<tr>
<td>NP304: What system of demonstratives is present in the language? (distance-oriented system vs. person-oriented)</td>
<td>80</td>
<td>0.725</td>
<td>0.891</td>
<td>0.920</td>
<td>0.820</td>
</tr>
<tr>
<td>Question</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Can adnominal demonstratives encode altitude?</td>
<td>92</td>
<td>0.974</td>
<td>0.889</td>
<td>0.872</td>
<td>0.968</td>
</tr>
<tr>
<td>Which word class do (the majority of) words denoting SPEED belong to?</td>
<td>66</td>
<td>0.550</td>
<td>0.882</td>
<td>0.958</td>
<td>0.792</td>
</tr>
<tr>
<td>Which word class do (the majority of) words denoting HUMAN PROPENSITY belong to?</td>
<td>67</td>
<td>0.587</td>
<td>0.866</td>
<td>0.882</td>
<td>0.861</td>
</tr>
<tr>
<td>In possessive constructions with a nominal possessor, is the POSSESSED noun usually marked? (state most frequent construction; NB: if marking depends on alienability, state marking of possessor in ALIENABLE constructions)</td>
<td>96</td>
<td>0.686</td>
<td>0.855</td>
<td>0.819</td>
<td>0.806</td>
</tr>
<tr>
<td>In possessive constructions with a nominal possessor, is the POSSESSOR usually marked? (state most frequent construction; NB: if marking depends on alienability, state marking of possessor in ALIENABLE constructions)</td>
<td>96</td>
<td>0.700</td>
<td>0.851</td>
<td>0.964</td>
<td>0.896</td>
</tr>
<tr>
<td>Can adnominal demonstratives encode gender?</td>
<td>93</td>
<td>0.837</td>
<td>0.830</td>
<td>0.826</td>
<td>0.975</td>
</tr>
<tr>
<td>Is there agreement between demonstrative/ determiner and the noun in the NP?</td>
<td>88</td>
<td>0.727</td>
<td>0.823</td>
<td>0.897</td>
<td>0.896</td>
</tr>
<tr>
<td>How many gender distinctions are realized within the NP?</td>
<td>94</td>
<td>0.853</td>
<td>0.822</td>
<td>0.926</td>
<td>0.922</td>
</tr>
<tr>
<td>Can adnominal demonstratives encode number?</td>
<td>92</td>
<td>0.737</td>
<td>0.809</td>
<td>0.561</td>
<td>0.850</td>
</tr>
<tr>
<td>What is the most frequent order of property word (adjective) and noun?</td>
<td>95</td>
<td>0.626</td>
<td>0.797</td>
<td>0.626</td>
<td>0.771</td>
</tr>
<tr>
<td>Can adnominal demonstratives encode movement?</td>
<td>93</td>
<td>0.928</td>
<td>0.782</td>
<td>0.854</td>
<td>0.963</td>
</tr>
<tr>
<td>Is nominalization the main strategy for complement clause formation?</td>
<td>60</td>
<td>0.605</td>
<td>0.774</td>
<td>0.772</td>
<td>0.900</td>
</tr>
<tr>
<td>What is the morphological composition of adverbial demonstratives (as compared to adnominal demonstratives)?</td>
<td>86</td>
<td>0.645</td>
<td>0.713</td>
<td>0.748</td>
<td>0.785</td>
</tr>
<tr>
<td>How many distance contrasts do adverbial demonstratives encode?</td>
<td>77</td>
<td>0.603</td>
<td>0.688</td>
<td>0.815</td>
<td>0.849</td>
</tr>
<tr>
<td>Which word class do (the majority of) words denoting PHYSICAL PROPERTY belong to?</td>
<td>69</td>
<td>0.574</td>
<td>0.685</td>
<td>0.677</td>
<td>0.882</td>
</tr>
<tr>
<td>Can adnominal demonstratives encode position of the referent in space or posture?</td>
<td>93</td>
<td>0.896</td>
<td>0.684</td>
<td>0.673</td>
<td>0.963</td>
</tr>
<tr>
<td>What kind of adposition type is most dominant?</td>
<td>87</td>
<td>0.684</td>
<td>0.666</td>
<td>0.721</td>
<td>0.861</td>
</tr>
<tr>
<td>Which word class do (the majority of) words denoting COLOR belong to?</td>
<td>70</td>
<td>0.574</td>
<td>0.647</td>
<td>0.686</td>
<td>0.864</td>
</tr>
<tr>
<td>Is nominalization the main strategy for adverbial clause formation?</td>
<td>61</td>
<td>0.562</td>
<td>0.622</td>
<td>0.676</td>
<td>0.925</td>
</tr>
<tr>
<td>What is the most frequent order of possessor and possessed?</td>
<td>97</td>
<td>0.897</td>
<td>0.622</td>
<td>0.525</td>
<td>0.941</td>
</tr>
<tr>
<td>Do demonstratives receive different inflectional features when used pronominally or adnominally?</td>
<td>80</td>
<td>0.642</td>
<td>0.616</td>
<td>0.858</td>
<td>0.828</td>
</tr>
<tr>
<td>NP</td>
<td>Question</td>
<td>Value</td>
<td>Ref.</td>
<td>Q20</td>
<td>Q30</td>
</tr>
<tr>
<td>----</td>
<td>----------------------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>NP313</td>
<td>Can adnominal demonstratives encode physical properties of the referent?</td>
<td>92</td>
<td>0.879</td>
<td>0.609</td>
<td>0.585</td>
</tr>
<tr>
<td>NP720</td>
<td>Do nouns have a morphologically marked dual?</td>
<td>87</td>
<td>0.834</td>
<td>0.601</td>
<td>0.473</td>
</tr>
<tr>
<td>NP820</td>
<td>Are there any (grammaticalized) sex-markers? (realized on the noun itself)</td>
<td>90</td>
<td>0.726</td>
<td>0.573</td>
<td>0.505</td>
</tr>
<tr>
<td>NP730</td>
<td>Do nouns have a morphologically marked paucal form?</td>
<td>89</td>
<td>0.899</td>
<td>0.570</td>
<td>0.443</td>
</tr>
<tr>
<td>NP2</td>
<td>What is the pragmatically unmarked word order for a transitive clause?</td>
<td>94</td>
<td>0.632</td>
<td>0.544</td>
<td>0.561</td>
</tr>
<tr>
<td>NP314</td>
<td>Do adnominal demonstrative roots require further derivation?</td>
<td>90</td>
<td>0.868</td>
<td>0.531</td>
<td>0.377</td>
</tr>
<tr>
<td>NP403</td>
<td>In possessive constructions with a pronominal possessor, is the POSSESSOR usually marked? (state most frequent construction; NB: if marking depends on alienability, state marking of possessor in ALIENABLE constructions)</td>
<td>96</td>
<td>0.635</td>
<td>0.523</td>
<td>0.726</td>
</tr>
<tr>
<td>NP310</td>
<td>Can adnominal demonstratives encode animacy?</td>
<td>93</td>
<td>0.836</td>
<td>0.481</td>
<td>0.275</td>
</tr>
<tr>
<td>NP9</td>
<td>Is there a number distinction in third person pronouns?</td>
<td>96</td>
<td>0.800</td>
<td>0.479</td>
<td>0.311</td>
</tr>
<tr>
<td>NP301</td>
<td>Are there indefinite articles in use?</td>
<td>86</td>
<td>0.655</td>
<td>0.466</td>
<td>0.585</td>
</tr>
<tr>
<td>NP7</td>
<td>Is there an inclusive/exclusive distinction in personal pronouns?</td>
<td>96</td>
<td>0.685</td>
<td>0.436</td>
<td>0.182</td>
</tr>
<tr>
<td>NP407</td>
<td>Is a word/morpheme meaning 'pet' (or similar) required in possessive constructions involving nouns which denote (domesticated) animals, and/or food?</td>
<td>83</td>
<td>0.717</td>
<td>0.368</td>
<td>0.384</td>
</tr>
<tr>
<td>NP303</td>
<td>Are third person pronouns and demonstratives related?</td>
<td>91</td>
<td>0.597</td>
<td>0.334</td>
<td>0.373</td>
</tr>
<tr>
<td>NP830</td>
<td>Are there classifiers?</td>
<td>93</td>
<td>0.690</td>
<td>0.329</td>
<td>0.172</td>
</tr>
<tr>
<td>NP503</td>
<td>Which word class do (the majority of) words denoting AGE belong to?</td>
<td>68</td>
<td>0.524</td>
<td>0.328</td>
<td>0.224</td>
</tr>
<tr>
<td>NP4</td>
<td>Is there a copula for predicate nouns? (only present tense clauses are taken into account)</td>
<td>79</td>
<td>0.590</td>
<td>0.327</td>
<td>0.548</td>
</tr>
<tr>
<td>NP1020</td>
<td>Can temporal or aspectual distinctions with a propositional scope be marked on the head noun?</td>
<td>84</td>
<td>0.818</td>
<td>0.308</td>
<td>0.412</td>
</tr>
<tr>
<td>NP501</td>
<td>Is there a class of synchronically underived adjectival elements (lexemes)?</td>
<td>96</td>
<td>0.650</td>
<td>0.302</td>
<td>0.297</td>
</tr>
<tr>
<td>NP10</td>
<td>Is nominalization the main strategy for attributive (relative) clause formation?</td>
<td>70</td>
<td>0.579</td>
<td>0.259</td>
<td>0.338</td>
</tr>
<tr>
<td>NP1</td>
<td>What is the pragmatically unmarked word order for an intransitive clause?</td>
<td>95</td>
<td>0.711</td>
<td>0.197</td>
<td>0.286</td>
</tr>
<tr>
<td>NP5</td>
<td>Is there a copula for predicate locative phrases? (only present tense clauses are taken into account)</td>
<td>61</td>
<td>0.526</td>
<td>0.173</td>
<td>0.313</td>
</tr>
<tr>
<td>NP6</td>
<td>Is there a copula for predicate adjectives? (only present tense clauses are taken into account)</td>
<td>76</td>
<td>0.560</td>
<td>0.161</td>
<td>0.356</td>
</tr>
</tbody>
</table>
Discussion and conclusion

The results of our study suggest that river systems do not contribute significantly to patterns of diffusion of structural NP features in South America. This is a surprising result given the fact that waterways have played such an important role in the mobility of the indigenous peoples of the continent. A few reasons for this discrepancy can be put forward, leading to a number of suggestions for further research.

Firstly, it may be the case that structural features of the noun phrase are not very prone to diffusion through contact, perhaps due to the fact that overt NPs in South-American languages with their highly synthetic verbal morphology are not altogether common. This would for instance also explain why some of the established or proposed linguistic areas associated with particular river systems do not show any significant degree of convergence either. On the other hand, Krasnoukhova (2012) does show that there are some large-scale areal patterns in South America for some of the NP-related features in her study. One suggestion for further research may nevertheless be to look at other domains of grammar.

The picture regarding word borrowing, for instance, is also quite complex. Bowern et al. (2011) report that word borrowing is not frequent among the Amazonian languages. Altogether, there is no evidence that suggests the contrary. Greenberg (1987) mentions some lexical items that have a surprisingly wide distribution, but without a systematic pattern. Muysken (2012: 252) draws attention to specific culture items that have a very wide distribution, such as the words for ‘chicken’ or ‘hundred’. It would be good to test
the distribution of specific lexical items in terms of river systems. Similarly, the spread of phonological features should be studied keeping river systems in mind. The same holds, of course, for other morphosyntactic features.

Secondly, it may be the case that river systems are simply less relevant to human migratory and contact patterns than is often assumed. It is probably too early to say whether or not this conclusion holds up, especially given the fact that many of the linguistic areas discussed in Campbell (2012b) for South America are connected to some river system. Furthermore, it may be a matter of finding the right scale. Even the smaller areas in our study are still rather large, and perhaps a more fine-grained analysis is required to find areal patterns of a more local nature. Finally, as mentioned in section 2.3 above, waterways are only one element of a potentially much more encompassing geographical approach to linguistic diversity. Embedding the role of river systems into such a larger programmatic approach may also yield more refined results.

Thirdly, the relation between river systems and specific structural features may simply be too indirect to yield significant results. After all, though physical geography may influence whether people decide to expand, or to contact other people, and also the directions they are most likely to travel to these ends, but there is still a large field of unknown variables that will influence the eventual patterns of structural linguistic diversity.

Fourthly, rivers can unite peoples on both sides if they are easily navigable and have good communication systems. On the other hand, very large rivers or those difficult to negotiate can separate peoples in which case we have the opposite effect. So size and navigability play an important role.

Fifthly, these unknowns lie in the field of human agency. Rivers exist ‘in the mind’ as well. People may have different strategies to reduce their ecological risk, which do not necessarily involve either migration or trade (like diversification or storage, cf. Nettle 1999: 80) and therefore not necessarily language contact. And even if they respond to disadvantageous ecological circumstances by mobility, this may lead to different types of encounters between peoples: they might for instance be trade-oriented or bellicose, intensive or shallow, short- or long-term, and policies with respect to contact may vary considerably from contact situation to contact situation (the Vaupés area as described by Aikhenvald 2002 is a case in point).

Lastly, as mentioned in the introduction, there is no obvious connection between contact scenarios and their linguistic outcomes to the extent that we can predict one on the basis of the other. Further research should ideally take into account as much as possible of the ethno-historical and sociological information available.

References


Adelaar, Willem. 2008. Towards a typological profile of the Andean languages. In: Alexander Lubotsky, Jos Schaeken and Jeroen Wiedenhof (eds.) Evidence and...


Campbell, Lyle. Why is it so hard to define a linguistic area?, this volume.


**Appendix: The Language Sample**

<table>
<thead>
<tr>
<th>Language</th>
<th>Genealogical affiliation</th>
<th>ISO-code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arhuaco</td>
<td>arh</td>
<td>Chibchan</td>
</tr>
<tr>
<td>Inga</td>
<td>inb</td>
<td>Quechuan</td>
</tr>
<tr>
<td>Páez</td>
<td>pbb</td>
<td>Paez</td>
</tr>
<tr>
<td>Chamacoco</td>
<td>ceg</td>
<td>Zamucoan</td>
</tr>
<tr>
<td>Ayoreo</td>
<td>ayo</td>
<td>Zamucoan</td>
</tr>
<tr>
<td>Kadiwéu</td>
<td>kbc</td>
<td>Guaicuruan</td>
</tr>
<tr>
<td>Wichí Lhamtés Nocten</td>
<td>mtp</td>
<td>Matacoan</td>
</tr>
<tr>
<td>Santiago del Estero Quichua</td>
<td>qus</td>
<td>Quechuan</td>
</tr>
<tr>
<td>Maca</td>
<td>mca</td>
<td>Matacoan</td>
</tr>
<tr>
<td>Tapieté</td>
<td>tpj</td>
<td>Tupian</td>
</tr>
<tr>
<td>Pilagá</td>
<td>plg</td>
<td>Guaicuruan</td>
</tr>
<tr>
<td>Toba</td>
<td>tob</td>
<td>Guaicuruan</td>
</tr>
<tr>
<td>Mocoví</td>
<td>moc</td>
<td>Guaicuruan</td>
</tr>
<tr>
<td>Emerillon</td>
<td>eme</td>
<td>Tupian</td>
</tr>
<tr>
<td>Hixkaryána</td>
<td>hix</td>
<td>Cariban</td>
</tr>
<tr>
<td>Trió</td>
<td>tri</td>
<td>Cariban</td>
</tr>
<tr>
<td>Warao</td>
<td>wba</td>
<td>Warao</td>
</tr>
<tr>
<td>Tehuelche</td>
<td>teh</td>
<td>Chonan</td>
</tr>
<tr>
<td>Puelche</td>
<td>pue</td>
<td>Puelche</td>
</tr>
</tbody>
</table>

Andes (North)

La Plata

Matacoan

La Plata

Southern Cone
<table>
<thead>
<tr>
<th>Ona</th>
<th>ona</th>
<th>Chonan</th>
<th>Southern Cone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapudungun</td>
<td>arn</td>
<td>Araucanian</td>
<td>Southern Cone</td>
</tr>
<tr>
<td>Shipibo-Conibo</td>
<td>shp</td>
<td>Panoan</td>
<td>Huallaga-Ucayali</td>
</tr>
<tr>
<td>Matsés</td>
<td>mcf</td>
<td>Panoan</td>
<td>Huallaga-Ucayali</td>
</tr>
<tr>
<td>Yaneshá'</td>
<td>ame</td>
<td>Arawakan</td>
<td>Huallaga-Ucayali</td>
</tr>
<tr>
<td>Cholón</td>
<td>cht</td>
<td>Hibito-Cholon</td>
<td>Huallaga-Ucayali</td>
</tr>
<tr>
<td>Yaminhua</td>
<td>yaa</td>
<td>Panoan</td>
<td>Huallaga-Ucayali</td>
</tr>
<tr>
<td>San Martín Quechua</td>
<td>qvs</td>
<td>Quechuan</td>
<td>Huallaga-Ucayali</td>
</tr>
<tr>
<td>Uru</td>
<td>ure</td>
<td>Uru-Chipaya</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>North Junín Quechua</td>
<td>qvn</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Cusco Quechua</td>
<td>quz</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Imbabura Highland Quechua</td>
<td>qvi</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Huaylas Ancash Quechua</td>
<td>qwh</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Ayacucho Quechua</td>
<td>quy</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Huangascal-Topara-Yauyos Quechua</td>
<td>qux</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Chipaya</td>
<td>cap</td>
<td>Uru-Chipaya</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Callawalla</td>
<td>caw</td>
<td>Mixed Language</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Pacaraos Quechua</td>
<td>qvp</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Cajamarca Quechua</td>
<td>ayc</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Southern Aymara</td>
<td>qvb</td>
<td>Quechuan</td>
<td>Aymaran</td>
</tr>
<tr>
<td>Huallaga Huánuco Quechua</td>
<td>qub</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Jauja Wanca Quechua</td>
<td>qxw</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>South Bolivian Quechua</td>
<td>quh</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Jaqaru</td>
<td>jqr</td>
<td>Aymaran</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Salasaca Highland Quechua</td>
<td>qx1</td>
<td>Quechuan</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Central Aymara</td>
<td>ayr</td>
<td>Aymaran</td>
<td>Andes (Central)</td>
</tr>
<tr>
<td>Awa-Cuaiquer</td>
<td>kwi</td>
<td>Barbacoan</td>
<td>Pacific</td>
</tr>
<tr>
<td>Mochica</td>
<td>omc</td>
<td>Mochica</td>
<td>Pacific</td>
</tr>
<tr>
<td>Northern Emberá</td>
<td>emp</td>
<td>Chocoan</td>
<td>Pacific</td>
</tr>
<tr>
<td>Colorado</td>
<td>cof</td>
<td>Barbacoan</td>
<td>Pacific</td>
</tr>
<tr>
<td>Ninam</td>
<td>shb</td>
<td>Yanomamic</td>
<td>Orinoco</td>
</tr>
<tr>
<td>Puinave</td>
<td>pui</td>
<td>Puinave</td>
<td>Orinoco</td>
</tr>
<tr>
<td>Panare</td>
<td>pbh</td>
<td>Cariban</td>
<td>Orinoco</td>
</tr>
<tr>
<td>Itonama</td>
<td>ito</td>
<td>Itonama</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Karo (Brazil)</td>
<td>arr</td>
<td>Tupian</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Leco</td>
<td>lec</td>
<td>Leko</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Movima</td>
<td>mzp</td>
<td>Movima</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Gavião Do Jiparaná</td>
<td>gvo</td>
<td>Tupian</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Wari’</td>
<td>pav</td>
<td>Chapacuran</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Kanoë</td>
<td>kxo</td>
<td>Kanoe</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Mosetén-Chimané</td>
<td>cas</td>
<td>Moseten-Chimane</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Lakondê</td>
<td>lkd</td>
<td>Nambiquaran</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Cavineña</td>
<td>cav</td>
<td>Tacanan</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Sabanê</td>
<td>sae</td>
<td>Nambiquaran</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Sakirabíá</td>
<td>skf</td>
<td>Tupian</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Language</td>
<td>Code</td>
<td>Family</td>
<td>River-Region</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td>Ese Ejja</td>
<td>ese</td>
<td>Tacanan</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Reyesano</td>
<td>rey</td>
<td>Tacanan</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Yuracaré</td>
<td>yuz</td>
<td>Yurakare</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Baure</td>
<td>brg</td>
<td>Arawakan</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Mamaindé</td>
<td>wmd</td>
<td>Nambiquaran</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Yuqui</td>
<td>yuq</td>
<td>Tupian</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Karitiâna</td>
<td>ktn</td>
<td>Tupian</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Kwaza</td>
<td>xwa</td>
<td>Kwaza</td>
<td>Guaporé-Mamoré</td>
</tr>
<tr>
<td>Dáw</td>
<td>kwa</td>
<td>Nadahup</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Nhengatu</td>
<td>yrl</td>
<td>Tupian</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Desano</td>
<td>des</td>
<td>Tucanoan</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Tariana</td>
<td>tae</td>
<td>Arawakan</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Hupdê</td>
<td>jup</td>
<td>Nadahup</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Cubeo</td>
<td>cub</td>
<td>Tucanoan</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Bora</td>
<td>boa</td>
<td>Boran</td>
<td>Vaupés</td>
</tr>
<tr>
<td>Southern Pastaza Quechua</td>
<td>qup</td>
<td>Quechuan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Urarina</td>
<td>ura</td>
<td>Urarina</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Tena Lowland Quichua</td>
<td>quw</td>
<td>Quechuan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Iquito</td>
<td>iqu</td>
<td>Zaparoan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Muniche</td>
<td>myr</td>
<td>Muniche</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Shuar</td>
<td>jiv</td>
<td>Jivaroan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Siona-Tetete</td>
<td>snn</td>
<td>Tucanoan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Cocama-Cocamilla</td>
<td>cod</td>
<td>Tupian</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Arabela</td>
<td>arl</td>
<td>Zaparoan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Aguaruna</td>
<td>agr</td>
<td>Jivaroan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Cofán</td>
<td>con</td>
<td>Cofan</td>
<td>Napo-Marañon</td>
</tr>
<tr>
<td>Bororo</td>
<td>bor</td>
<td>Bororoan</td>
<td>Tapajos-Madeira</td>
</tr>
<tr>
<td>Kamayurá</td>
<td>kay</td>
<td>Tupian</td>
<td>Xingu-Tocantins</td>
</tr>
<tr>
<td>Canela</td>
<td>ram</td>
<td>Nuclear-Macro-Je</td>
<td>Xingu-Tocantins</td>
</tr>
<tr>
<td>Trumai</td>
<td>tpy</td>
<td>Trumai</td>
<td>Xingu-Tocantins</td>
</tr>
<tr>
<td>Jamamadí</td>
<td>jaa</td>
<td>Arawan</td>
<td>Solimoes</td>
</tr>
<tr>
<td>Apuriniâ</td>
<td>apu</td>
<td>Arawakan</td>
<td>Solimoes</td>
</tr>
</tbody>
</table>
Index (Part 2)

Africa, 2
Africanisms, 5
  advanced tongue root vowel (ATR) harmony, 8
  affirmative answers to negative questions, 10
  anastasis, 9
  clefting constructions, 9
  conceptualisation strategies, 10
  diphthongs, 7
  European-based pidgins and creoles, 6
  genetic origin, 6
  ideophones, 8
  implosives, 8
  Khoisan languages, 7
  labiodental flaps, 7
  lack of case inflections, 8
  lack of consonant clusters, 7
  logophoric marking, 9
  nominal polysemy, 11
  ‘foot’ and ‘leg’, 11
  ‘hand’ and ‘arm’, 11
  ‘meat’ and ‘animal’, 11
  open syllables, 7
  reduplication, 8
  SVO word order, 8
  various proposals, 6
  word classes, 8
  word-initial prenasalised consonants, 8
  areal comparison with rest of world, 2
  areal groupings within Africa, 2
  assessment as linguistic area, 12
  grammaticalisation areas, 13
  summary, 14
distribution of typological properties in world regions, 12
Ethiopian Language Area, 4
geographically defined language groupings, 2
Kalahari basin, 5
Khoisan phylum (Greenberg), 3
Macro-Sudanic Belt, 5
Niger-Congo, 2
Nilo-Saharan phylum (Greenberg), 3
phonological zones
  Center, 3
  East Africa, 3
  North Africa, 3
  Rift Valley, 4
  South Africa, 3
  Sudanic belt, 3
  Rift Valley Convergence Area, 3
  Sudanic languages, 5
  Tanzanian Rift Valley Area, 4, 36
  verb-initial syntax, 3
  West Africa, 2
  word order, 3
  VSO-languages, 3
Australia, 274
  languages
    history, 274
    scholarly standpoints, 278
    lack of phonological differentiation, 275
    macro level patterns, 275
    non-identical lexical correspondences, 276
    Pama-Nyungan and non-Pama-Nyungan, 278
Pilbara
  colonial settlement, 282
Pilbara languages, 279
  alignment shift, 289
  assessment, 289
  classification, 280
  common features
    shared inheritance, 283
    expected shared replacements, 291
    genetic relationship and language contact, 291
    innovation of accusative alignment, 287
    innovations in pronouns and demonstratives, 285
    lexical differences, 282
    multilingual contact, 282
    paradigm reflexives of demonstratives, 286
    pronoun paradigms, 283
    shared inheritance, 284
    reconstructed demonstrative stems, 285
    relative salience, 289
    role of social factors, 290
    sound change, 281
structural isomorphism in Lemerig and Koro, 290
switch-reference, 288
verbal systems, 281
Punctuated Equilibrium model, 277

China, 201
Altaic, 202
areal diffusion in East and South-East Asia, 207
Austroasiatic, 204
Austronesian, 204
features of South-East Asian languages, 207
Gansu-Qinghai, 201
Guangxi, 201
Hmong-Mien, 204
Hunan, 201
Language Atlas of China, 205
linguistic area(s)
assessment, 220
linguistic areas, 209
Central Transitional zone, 210
Far Southern, 210
Northern, 210
Southeastern, 210
Southwestern, 210
linguistic demography, 201
Sinitic languages, 205
analytic structure, 206
geographical spread, 206
grammaticalisation pathways, 209
Northern-Southern division, 207
feature tendencies, 208
Hashimoto’s dichotomy, 208
numbers of speakers, 205
preference for monosyllabicity, 206
word order typology, 209
Sino-Tibetan, 202
status of Standard Mandarin (pǔtōnghuà), 205
studies of language contact, 210
gender affixes in languages in Nanning, Guangxi, 214
Guangxi, Southern China, 213
verb-recipient-theme, 214
Hunan, Central China, 215
Hmong languages, 216
Southwestern Mandarin
guest language (kèhuà), 216
Xianghua, 216
archaic features, 217
differential object-marking, 218
polarity comparative, 218

Northwestern China, 210
case markers, 212
SOV word order, 211
Southern China
Tai substratum in Yue, 213
subordination markers in languages in Nanning, Guangxi, 215
Tai-Kadai, 203
Ethnologue, 216, 297
Jharkhand, 111
areal features
accusative alignment, 115
lack of grammatical gender, 115
numeral classifiers, 115
eastern-central South Asia areality ‘from’ and ‘to’, 118
‘start’ vs. ‘keep on’, 117
alienable vs. inalienable possession, 117
anticipatory categories, 119
assessment, 128
convergence patterns, 129
genitive, 121
lexicon, 116
new dual category, 121
eastern-central South Asia as a convergence area, 116
eastern-central South Asia in the larger context, 114
South Asian features, 115
Indo-Aryan and Munda convergence, 114
languages of area, 111
Indo-Aryan, 112
Munda, 113
Sadri existential/locative copula, 128
Sadri identificational copula, 128
Santali existential/locative copula, 126, 127
South Asian convergence area, 130
Kalahari Basin, 66
‘Khoisan’ families, 66
typological profile, 68, 69
discussion of areality, 81
feature summary, 79
features shared by other local languages, 82
lexicon, 71
borrowing patterns, 71
isoglosses, 74
polysemy, 73
morphosyntax, 74
associative plurals, 79
clause-second elements, 77
multipurpose oblique, 77
multi-verb construction(s), 75
simple attributive numerals, etc., 78
verb serialisation, 74
phonology
isoglosses, 70
previous research, 68
Mainland South-East Asia, 225
core area, 227
geographical delimitation, 225
greater area, 226
human movement in history, 226
languages, 229
assessment, 238
descriptive studies, 232
distribution, 229
diversity, 230
historical-comparative linguistics, 234
methodologies, 233
morphosyntax-semantics system
features, 231
nominal classification, 236
overview of scholarship, 231
perceptions, 235
sesquisyllables, 236
sociolinguistics, 234
sound system features, 230
status as linguistic area, 237
tone phonetics and phonology, 237
types and prototypes, 235
linguistic history, 227
tone, 249
assessment, 265
Austroasiatic, 257
Austronesian, 258
Burmese, 254
characteristics of language families, 257
contact-induced tonogenesis, 262
contrasting tones and phonation types, 256
definition of tone language, 249
domain of contrast, 256
Hmong-Mien, 260
laryngeal features, 254
phonation types, 251
pitch realisation, 251
register systems, 252
register systems
phonetic correlates, 253
Sino-Tibetan, 259
Tai-Kadai, 259
tone as an areal feature, 263
tonogenesis, 260
uncommon paths, 261
typing, 249
variation in systems, 250
word shape in tonogenesis, 264
trickle-effect colonisation, 228
Melanesia
Eastern, 348
‘aberrant’ Oceanic languages, 351
‘pidginisation’ hypothesis, 352
assessment, 366
Australian, 367
contact-induced change, 351
cultural/linguistic areas of the Pacific, 350
Eastern Solomon Islands, 354
indeterminate, 366
influence of non-Oceanic languages, 352
Melanesian languages, 351
New Caledonia, 358
Australian substrate, 359
ergative, 359
possible source, 359
standard account of prehistory, 359
Papua New Guinea and the Western Solomon Islands, 353
Papuan, 366
Polynesian outliers in Melanesia, 364
post-linguistic areas, 365
pot-and-handle loanwords, 362
Proto-Oceanic, 353
quinary numeral systems, 361
rapid internal change, 352
serial verb constructions, 360
The Reefs Santa Cruz group, 354
toponyms, 362
Vanuatu, 355
‘linguistic Melanesianisation’ theory, 358
key features, 355
lack of /p/ in phoneme, 356
lexical replacement, 358
pre-Lapita (Oceanic-speaking) population, 355
areality
lexical and structural elements, 377
areal-typological phonological features, 387
arguments for a sprachbund, 378, 379
effects of borrowing, 379
historical reasons, 380
assessment, 383
common phonological developments, 380
C_initial and C-final, 381
lack of contrastive voicing, 381
rhotacism and assimilation, 381
common structural features, 378
contact languages, 375
contact-induced change, 375
loans in Micronesian languages
Chamorro, 384
Chuukese, 385
Palauan, 384
Puluwatense, 385
Woleaian, 385
Yapese, 384
marked shared features, 378
Micronesian area, 374
partial relexification, 386
Proto-Malayo-Polynesian, 377
Proto-Malayo-Polynesian items, 391
rise of sprachbund, 382
structural parallels, 389
transfer in interlingual contact, 391
undiffused features, 391
metatypy, 21, 139, 305, 317, 327, 391
New Guinea, 297
languages
‘mixed’ (Oceanic/Papuan) languages, 304
‘southeast interior’ and ‘northwest coastal’, 307
areal studies, 306
areality, 311
investigation and results, 313
morphosyntactic ordering, 312
semantic encoding, 312
Austronesian languages, 297
arrivals in the region, 298
background, 301
Bismarck Archipelago, 311
contact studies, 304
contact-induced change, 319
eyear settlement, 299
effects of contact, 305
exclusive/inclusive distinction in pronouns, 307
geographical extent, 297
géolinguistics subregions, 300
Papuan and Austronesian populations, 300
key variables
alignment of argument indices, 330
alignment of verbal argument indices, 331
basic tense/mood distinctions in verbal morphology, 331
clause sequences, 327
clause-chaining, 329
clause-final negators, 325
closivity and number in free pronouns and gender in pronominals, 334
inalienable and alienable possession constructions, 334
negator position in declarative realis clause, 325
noun/adjective order and noun/demonstrative order, 328
possessor index on nouns, 324
possessum/possessor order, 328
subject and object indices on verbs, 322, 323
switch-reference, 329
tense- and mood-prominence, 332
verb/object and adposition/NP orders, 321
label ‘Papuan’, 297
lineage symbols, 318
linguistic area status, 310
microgroups within the Trans New Guinea family, 308
Northern Arc, 310
constituent languages, 310
sVo as salient feature, 323
Papuan classification, 301
Papuan lineages
East Nusantara and mainland New Guinea, 304
northwest island Melanesia, 304
possible Papuan lineages, 302
question of lineage, 297
relatedness discussion, 335
Sepik-Ramu phylum, 303
shared inheritance or contact-induced change, 336
the New Guinea Region, 298
Trans New Guinea phylum, 301
scholarship, 303
typological clustering, 309
WALS variables as delimiters of Austronesian and West Papuan, 306
linguistic diversity, 297
Sunda and Sahul, 299
Niger-Congo, 42
areal patterns in family, 46
areal studies of phylum, 58
Atlantic group, 44
contact among groups and families, 52
Bantu – Ma’á contact, 53
Bantu – Nilotic contact, 53
Bantu-Khoisan contact, 52
Benue-Congo / Chadic contact, 52
impact of Mande languages, 53
SOV word order, 53
core area, 44
distribution of Bantu languages, 44
distribution of Niger-Congo languages, 44
genealogical overview, 42
geographic overview, 45
Gur, 44
key features
noun class systems, 47
SVO word order, 46
tone, 46
Kwa subgroup, 48
language ideologies
general cultural patterns, 57
Lower Fungom study, 56
local dynamics and areal patterns, 57
Macro-Sudan Belt, 47
ATR harmony, 48
labial-velar stops, 48
VO-Neg word order, 49
new urban varieties, 54
continuity with rural varieties, 55
Sheng, 55
Wolof, 54
peripheral languages, 50
Atlantic group, 51
similarities with Bantu, 51
Bantu, 50
canonical type, 50
conservative morphology, 50
dynamics of expansion, 50
interlacustrine languages, 51
time depth, 51
Proto–Niger-Congo, 47
reduced morphology in core area, 52
role of ecology in areality, 46
sociocultural dynamics and language ideologies, 54
spread zones vs. accretion zones, 49
status of Mande group, 45
Nilo-Saharan, 19
(split) ergativity, 20
Central Sudanic, 21
Central Sudanic-Ubangi contact zone, 22
discourse markers, 24
dyad order, 22
evaluation of areal features, 37
languages allowing for postverbal subjects, 32
marked nominative systems, 37
Nilotic and Surmic languages, 33
expansion zones, 34
Northeastern typological break with Eastern Sudanic, 31
post-Northeastern contact zone, 26
converbs, 28
converbs in Beria, 29
coverb plus light verb, 30
number-marking systems, 28
population movements and areal influence, 30
Pre-Leiterband populations, 31
river topography, 30
typological dichotomy, 27
post-positional particles, 26
Southeastern Eastern Sudanic languages, 34
subclassification, 20
subclassification of Eastern Sudanic, 31
Surmic languages
typological convergence with Omotic, 35
SVO word order, 23
the Nilotic borderland, 35
Datooga cluster, 36
Kalenjin contact with Bantu structural borrowing, 36
Luo contact with Bantu, 35
structural borrowing, 35
typological convergence, 37
verb position, 37
Ubangi languages, 23
North America
linguistic areas, 398
Northern California, 401
Northwest, 398
Plateau, 400
Southeast, 403
culture area, 405
Southwest, 404
native languages, 394
areal boundaries, 396
contact areas, 394
culture areas, 395
inheritance versus contact, 397
linguistic profiles, 395
nature of contact, 394
shared features
  assessment, 435
  borrowability, 406
  duration of contact, 435
  feature strength, 406
  grammar
    agent/patient systems, 413
    alignment, 424
Northern California, 426
Northwest Coast, 424
  from posture to aspect, 428
‘lie’ positionals, 431
’stan’ plurals, 431
auxiliaries, 430
postural to aspectual meaning, 434
  further verbal features, 415
  inclusive-exclusive distinction, 412
  lexical suffixes, 419
  locative/directional, 423
  means/manner and location/direction, 422
  negation, 418
  replication, 418
  transimpersonal constructions, 417
  verbal prefixes, 416
levels of structure, 406
lexicon, 407
phonology, 408
  fricative symbolism, 410
  glottalisation, 411
  sound patterns, 410
  sounds, 408
voiceless sonorants, 409
social and cognitive factors, 435
type of areal trait, 396
North-eastern Siberia
  case marking, 179
  case systems, 181
  assessment, 196
  change due to Russian contact, 191
  dative vs. allative contrast, 181, 188
  instrumental
    vintage, 184
  instrumental and comitative, 181, 182
  locative and participal verb form, 192
  pre-Russian contact situation, 193
  prolate or prosecutive case, 181, 186
  sociative (Koryak), 184
Evenki
  code-switching, 195
general SOV word order, 191
Northern Tungusic languages, 184
Proto-(Northern-)Samoyedic, 187
Proto-Chukotko-Kamchatkan, 183
self-reported ethnic identity, 194
self-reported language use, 194
Yukaghiric languages, 182
Siberia
  indigenous languages, 180
  contact, 180
South Africa, 90
  Afrikaans and contact, 93
    associative plurals, 95, 98
diminutives, 97, 99
    double negation, 94, 97
effects in relation to other languages, 97
    reduplication, 96, 98
    the tag nè, 97, 99
  Afrikaans borrowings in English, 92
  areal convergence evaluation, 107
  Bantu borrowings in English, 93
  Black English
    ante-deletion, 104
    salient features, 103
    traditional and crossover varieties, 103
code-switching and mixing, 105
  District Six, Cape Town, 105
  Xhosa in Cape Town, 106
English and Second Language Acquisition, 102
  English influence on Afrikaans, 101
  English influence on Bantu language syntax, 104
  influence of Khoesan on South African languages, 91
  Khoesan borrowings in Xhosa and Zulu, 92
  Khoesan languages, 91
  official languages, 90
  substrate influence vs. universals of second language acquisition, 106
weak convergence
  busy, 99
South America
  Amazonia, 443
    areality of discourse practices, 460
    areality of language ideologies, 460
    assessment, 461
    Caquetá-Putumayo region, 451
    contact, 443
    language families, 445
linguistic effects, 444
localised diffusion, 444
pre-Colombian dynamics, 445
South American lowlands, 444
contact zones, 446
linguistic areality, 461
linguistic status of entire area, 459
long-range language contact and
macro-areality, 457
NeighborNet representation of
grammatical structures, 453
pre-Columbian trade networks, 457
status of lexical borrowing, 460
The Guaporé-Mamoré region, 454
classifier forms, 456
contact-induced change, 455
The Upper Xingu Region, 453
typological bisecting the east-west axis
of continent, 458
Vaupés region, 446
Arawak languages, 449
code-switching, 447
contact-driven grammaticalisation,
450
language as emblem of identity, 447
multilingualism, 447
Nadahup languages, 449
restructuring, 448
Western Amazonian area, 458
polysynthesis and complex verbal
morphology, 459
Andean/Amazonian split, 460
areal patterns
noun phrase, 476
areal distribution of features, 478
modifiers within NP, 477
non-defining features, 479
NP related issues, 477
NP structure, 477
areal typology, 473
areal-linguistic studies, 471
Andean / Amazonian divide, 471
Area Intermedia, 471
classification features of Amazonian
languages, 472
classification features of the Colombian-
Central America area, 473
contact scenarios, 470
Eastern South American Linguistic Area,
459
genealogical diversity, 470
geography
convergence of noun phrase profiles,
490
ecological zones, 480
incentive factors, 475
major drainage basins, 479
mobility factors, 476
physical space, 480
regions, locations and sample
languages, 481
relative river distances, 483, 486
river systems, 475
riverine areas, 490
riverine features, 491, 492
discussion, 495
language sample, 499
role of rivers in language development,
483
smaller regions, 482
historical records, 470
negative defining features of Andean
languages, 472
phonological areality, 460
regional diffusion zones in lowland South
America, 456
river systems, 470
South American Indigenous Language
Structures Database, 459
South American Phonological Inventory
Database, 459
typological feature distributions, 470
Western South American Linguistic Area,
459
Sri Lanka and South India, 134
classification of Sri Lanka, 136
case systems, 138
common grammatical features, 137
Kaffir population, 136
South-South Asia (SSA), 136
Sri Lanka Malay, 136
Sri Lanka Portuguese, 136
status of Sinhala, 136
Vedda languages, 136
influence of Dravidian, 134
grammar, 135
phonology, 135
multilingualism and language contact,
138
typological considerations, 140
typology and ecology, 141
Transeurasia, 143
assessment, 170
delimiting areality, 166
deviations from the prototype, 166
diffused vs. inherited features, 168
isolated position of Japanese, 169
isomorphism, 170
maximal coherence in Mongolic and Tungusic, 169
prototypical Transeurasian, 169
recurring grammaticalisation, 169
reduction of features in Korean and Japanese, 169
family tree, 168
feature values, 164
grammaticalisation, 157
direct insubordination, 157
first person pronoun augmented with a collective-plural marker, 162
negative verb to verbal negator, 160
lexicon and semantics, 148
inclusive-exclusive first person plural pronouns, 149
non-verbal strategy with (extra-family) verbal borrowing, 148
partial emphatic reduplication of nominal property words, 151
property words can show switched encoding, 151
property words may be verbally or nominally encoded, 150
two-way proximal-distal distinction in demonstrative pronouns, 148
morphology, 152

absence of obligatory numeral classifiers, 152
agglutination, 152
mi-Ti opposition in first vs. second singular person pronouns, 153
secondary oblique stem of personal pronouns, 154
suffixal inflectional morphology, 152
overview, 163
phonology, 145
absence of complex tonal distinctions, 145
absence of initial consonant clusters, 147
absence of initial r-, 147
absence of initial velar nasal, 147
polysyllabic root structure, 145
presence of vowel harmony, 145
tongue root vowel harmony, 146
voice distinction for stops, 148
syntax, 154
ablative case form to encode predicative comparison, 156
extensive use of converbs, 155
genitive-noun / adjective-noun phrase order, 155
locative existential construction to encode predicative possession, 156
SOV word order, 154