Appendix - Dataset belonging to ‘The transparency of creoles’

Index
Abbreviations 2

1 Nubi and its source languages
1.1 Nubi 3
1.2 Bari 12
1.3 Sudanese Arabic 18

2 Pichi and its source languages
2.1 Pichi 26
2.2 English 34
2.3 Fongbe 40
2.4 Kikongo 49

3 Sri Lanka Malay and its source languages
3.1 Sri Lanka Malay 56
3.2 Ambonese Malay 64
3.3 Sinhala 70
3.4 Tamil 80

4 Diu Indo-Portuguese and its source languages
4.1 Diu Indo-Portuguese 90
4.2 European Portuguese 97
4.3 Gujarati 106
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3 = 1st, 2nd, 3rd person</td>
<td>EMPH = emphasis</td>
</tr>
<tr>
<td>I, II, - X = semantic classes</td>
<td>NOM = nominative</td>
</tr>
<tr>
<td>ABL = ablative</td>
<td>NPST = non-past</td>
</tr>
<tr>
<td>ABS = absolutive</td>
<td>OBJ = object</td>
</tr>
<tr>
<td>ACC = accusative</td>
<td>PASS = passive</td>
</tr>
<tr>
<td>ACT = active voice</td>
<td>PFV = perfective</td>
</tr>
<tr>
<td>ADVR = adverbializer</td>
<td>PL = plural,</td>
</tr>
<tr>
<td>AM = assertion marker</td>
<td>POSS = possessive</td>
</tr>
<tr>
<td>ANIM = animate</td>
<td>POT = potential</td>
</tr>
<tr>
<td>CAUS = causative</td>
<td>PROG = progressive</td>
</tr>
<tr>
<td>COMP = complementizer</td>
<td>PROX = proximal</td>
</tr>
<tr>
<td>CONJ = conjunctive</td>
<td>PRS = present</td>
</tr>
<tr>
<td>CONT = continuous</td>
<td>PST = past</td>
</tr>
<tr>
<td>COP = copula</td>
<td>PTCP = participle</td>
</tr>
<tr>
<td>DAT = dative</td>
<td>Q = question particle</td>
</tr>
<tr>
<td>DEF = definite</td>
<td>QUOT = quotative</td>
</tr>
<tr>
<td>DEM = demonstrative</td>
<td>RDP = reduplication</td>
</tr>
<tr>
<td>DIM = diminutive</td>
<td>REL = relativizer</td>
</tr>
<tr>
<td>DIST = distal</td>
<td>REM = remote past</td>
</tr>
<tr>
<td>DUAL = dualis</td>
<td>SBJV = subjunctive</td>
</tr>
<tr>
<td>F = feminine</td>
<td>SG = singular</td>
</tr>
<tr>
<td>FOC = focus</td>
<td>SIMIL = similitive</td>
</tr>
<tr>
<td>GEN = genitive</td>
<td>TR = transitive</td>
</tr>
<tr>
<td>FUT = future</td>
<td></td>
</tr>
</tbody>
</table>
1. **Nubi and its source languages**

1.1 **Nubi**

Nubi (also known as Kinubi) is a contact language spoken by approximately 26,000 people in Uganda and Kenya. My main source, Wellens (2003), describes the Ugandan variant (spoken by 15,000 people, mostly around Bombo.), but I also use data from the Kenyan varieties (spoken by 10,000 people around Kibera and Mombasa). These varieties are mutually intelligible. Nubi is very similar to Juba Arabic (the creole of South Sudan), and indeed the two languages are mutually intelligible as well (Owens 1990).

Nubi emerged between 1850 and 1890 (cf. Owens 1990), and was derived (according to Wellens) from a Central Arabic trade pidgin. This pidgin, which must have existed before 1820, was presumably based on Egyptian Arabic and Sudanese Arabic.

Owens (1991) argues that Sudanese Arabic is the most significant lexical source of Nubi. Bari, and to a lesser extent Mamvu, have been suggested as substrate languages. There may be other not yet identified languages contributing to Nubi.

**Redundancy**

1. **Apposition: Cross-reference**

There is no cross-reference in Nubi. As can be seen in example (1), there can be number marking on the predicate.

Wellens (2003:181)

(1) Anas kun kweisin.
    people.PL be-Ø good.PL
    ‘People are good.’
As there is no marking of person, but only of number, I do not consider this to be referential marking. I will hence not treat it as a case of cross-reference – the only referential marker is the independently expressed argument *anas*.

2. *Multiple expression of semantic information*

- There is no negative concord in Nubi. Clauses are negated by means of the markers *ma* or *mafi* (Wellens 2003:181). Negation of indefinites is expressed by means of an existential phrase, as in (2), so there is no negative concord there.

Wellens (2003:182)

(2) Mafi sokol al gu-rudu asede.

   EXIST.NEG thing REL PROG-answer now

   ‘There is no one who answers now.’

Wellens does give two sentences with two negative markers, of which I show one here as (3).

Wellens (2003:183)

(3) Dunia ta youminde ma je dunia ta zaman

   world of nowadays NEG like world of bygone.days

   ta waze tena de ma.

   of parent.PL my DEF NEG

   ‘The world of today is not like the world of the bygone days of our parents.’

The negative marker *ma* is used twice, while there is only one semantic negation. However, the use of two negative markers can in my eyes be explained by the fact that the phrase is so long: the negative marker is repeated to make sure the hearer remembers it. I have no strict evidence for this, but there are only two examples of double marking (both long sentences), against hundreds of simple negations. Hence, in my opinion, it is highly
unlikely that there is a grammatical construction negative concord in Nubi – I consider (3) to be a coincidental instance of a repeated marker.

- There is no plural concord in Nubi. Plural marking on nouns is optional in all cases. If a numeral is expressed, the noun never gets a plural marker, as shown in (4).

Luffin (2004:16)

(4) group taláta
    group 3
    ‘Three groups.’

- Nubi has no semantic classification system, hence double expression is not applicable here.

**Domain disintegration**

3. *Fusion*

- Person and number are expressed together in pronouns; this is cumulation.

- There is stem alternation in Nubi, mainly in the nominal domain. Nouns and adjectives are marked for plural not by particles or affixes, but by vowel alternations, as in example (5).

Wellens (2003:76)

(5) kebir kubar
    big.SG big.PL

Note that in Semitic languages (like Sudanese Arabic, one of the sources of Nubi), the stem in fact consists of three consonants, the so called roots or root consonants, combined with a CVCV template. In an Arabic language, the adjectival stem of example (5) would be
KBR, the template would be CVCVC. A change in the quality of the vowels would hence not affect the stem – it would still be KVBVR. However, in Nubi, this stem forming system is lost (Kihm 2011:58ff.). Word forms as in (5) are simply petrified copies from Arabic. Therefore, I analyze this as stem alternation.

4. *Discontinuity*

- There are no circumfixes in Nubi (Wellens 2003).
- There are no infixes in Nubi (Wellens 2003).
- I have not seen examples of argument raising in Nubi. The construction in (6) is a possible candidate for a raising construction, as the predicate ‘to seem’ often evokes it.

Luffin (2005:298)¹

(6) Ya gi-be#n rután de gu-ró wóduru.

  EMPH PROG-seem language DEM PROG-go disappear

  ‘It seems that this language will disappear.’

The argument *rután* ‘language’ is semantically speaking the first argument of the embedded clause. As it is unmarked, it is hard to say what its syntactic status is. If *rután* could be argued to function as the object of the main clause, we would have a case of raising. However, this is likely not the case. Usually, the subject precedes the object in Nubi (Luffin 2005:360), so that the only indication for syntactic function states that *rután* is a subject. In that case, this is not argument raising. There is no discrepancy between semantics and syntax.

- I have not found examples of extraposition in Nubi.

5. *Non-parallel alignment*

¹ The original example from Luffin is glossed and translated in French. The translation to English here is mine.
I found no examples of non-parallel alignment in Nubi.

**Form-based form**

6. **Agreement**

Nubi does not exhibit agreement. Number can be marked on nouns (Wellens 2003:230), and it appears on adjectives and demonstratives (cf. (7); Wellens 2003:76). As all plural marking is optional, it cannot be the result of an obligatory copying rule.

Wellens (2003:98)

(7) wele wadin dolde
   boy.PL other.PL DEM.PROX.PL
   ‘these other boys’

Agreement is also not exhibited in the verbal domain. As explained in the section on cross-reference, there can be number marking on the predicate, but this is again optional.

7. **Expletive elements**

There are no nominal expletives in Nubi, as shown in (8).

Khamis & Owens (2007:206)

(8) Ana aminu (ja gali) matara bi-waga asede.
   I believe (COMP) rain FUT-fall now
   ‘I believe that it will rain now.’

8. **Grammatical gender**
Nubi does not exhibit grammatical gender (Wellens 2003:61).

9. Sequence of tenses

There is no sequence of tense rule in Nubi. Wellens (2003:202) claims that there are constraints on complement clauses only after verbs of modality, but these only involve the subject. From that I gather that there are no constraints on TMA marking.

Another reason to think that there is no copying of tense to the embedded clause, is that tense marking is not obligatory – there are virtually no complement clauses that express tense at all. If tense marking is not obligatory and therefore hardly ever explicit, it is highly unlikely that there is obligatory copying of tense in Nubi. I will assume then that there is no *consecutio temporum* in Nubi.

10. Influence of complexity on word order

There can be influence of weight on morphosyntactic placement in Nubi. A Recipient argument is marked by the marker *na* and placed after the Undergoer, as in example (9). However, when the Recipient is pronominal, it can be placed before the Undergoer, as in (10) (Wellens 2003:180).

Wellens (2003:180)

(9) Bes, umon kutu sum na Mohamadi.
    well they put poison to M.
    ‘Well, they gave poison to Mohammed.’

(10) Mama de wedi n-ouo dawa fu akili.
    mother DEF give to-3.SG medication in food
    ‘The mother gave him medication in his food.’

That means that ordering can be influenced by morphosyntactic weight.
11. Influence of complexity on function marking

Nubi is a predominantly isolating language. It has some affixes and some suppletive marking (cf. the section above on fusion), but most functions are marked by means of independent markers like na (Recipient) and ma (negation). These markers can apply to units of different morphosyntactic nature and complexity. Hence, there is no influence of morphosyntactic information on function marking in Nubi.

12. Syntactic functions

There is at least a syntactic function subject in Nubi. Arguments are usually unmarked. Only arguments with the semantic role of Recipient are always introduced by the particle na (Wellens 2003:180), as in example (9), repeated here as (11).

Wellens (2003:180)
(11) Bes, umon kutu sum na Mohamadi.
    Well they put poison to M.
    ‘They gave poison to Mohammed.’

Semantic role is hence relevant for the expression of Recipients. Also, semantic information like animacy and definiteness is relevant for the expression of the verbal marker –u. Wellens (2003:100ff.) argues that this is a transitivity marker, but she acknowledges that there are counter examples. Regardless of one’s opinion on the exact status of –u, it is clear that it has something to do with the semantics of the denoted event. It is hence a reason to say that Nubi’s alignment system is semantically based.

However, there is a reason to postulate a syntactic function subject in Nubi, which is that Actors and Undergoers can be treated the same morphosyntactically.

Luffin (2005:358)
(12) Úwo bi-já.
    he FUT-come
    ‘He will come.’

Luffin (2005:359)

(13) Asker-íya bi-mútu.
    soldier-PL FUT-die
    ‘The soldiers will die.’

The argument in (12) is an Actor, while the argument in (13) is an Undergoer. Still, they look the same: they are both unmarked, have the same position and the predicates are identically marked as well. There is hence a neutralization of semantic roles and a syntactic function Subject.

13. (Morpho-)phonologically triggered alternations

There are alternations in Nubi. Firstly, vowels can be deleted in between two consonants of the same place of articulation or in word-final position (Wellens 2003:44) or when preceding another vowel, e.g. Recipient marker na + 3.sg pronoun ouo > nuou (Wellens 2003:180). A stressed vowel can fuse with another vowel (Wellens 2003:46). Furthermore, there is vowel harmony (Wellens 2003:47); vowels assimilate their height in the neighbourhood of other vowels, e.g. silu ‘take’ that may for some speakers become sulu in rapid speech. The verbal prefixes gi- and bi- then optionally assimilate their height too (bu-sulu), which shows that vowel assimilation can cross morpheme boundaries.

Consonants can acquire or lose voice when preceding voiced or voiceless consonants or pauses, e.g. bil ta balala ‘wetting of thread’ is pronounced /biːl̪̂ ta bɑlala/ (Wellens 2003:49). Nasal consonants adapt their articulation place to following consonants (Wellens 2003:49). Consonants are sometimes deleted when they stand in between two vowels, e.g. bahati ‘luck’ may become bati in rapid speech (Wellens 2003:50). While geminates are usually retained word-internally, there can be degemination at word
boundaries, e.g. *danab bagara* ‘tail cow > a cow’s tail’ is pronounced /danabɑɡara/ (Wellens 2003:51).

**References**


1.2  Bari

Bari is a Nilo-Saharan language of the Nilotic branch, spoken in South Sudan by 420,000 people (Lewis ed. 2009). Owens (1990) argues it is one of the source languages of Nubi. The Bari lived in Southern Sudan when from 1851 onwards, more and more trade camps were set up. Although established by Austrians and English colonialists, power was soon exerted by the Khartoum government. Hence, Sudanese Arabic was brought to the region and mixed with Bari (and potentially, other languages) to form the trade pidgin that formed the basis for Nubi.

Redundancy

1.  Apposition: Cross-reference

There is no cross-reference in Bari. Predicates can be marked for plural by means of reduplication (Spagnolo 1933:231), but this is optional. Otherwise, person and number are not marked grammatically on the predicate, as illustrated in (14) and (15).

Spagnolo (1933:226)

(14) Nye amët lbpëŋ.
    (s)he saw him/her
    ‘(S)he saw him/her.

(15) Se amët lbpëŋ-at.
    they saw him/her-pl
    ‘They saw them.’

---

2 Bari should not be confused with Barí, a South American language.
A lexical referential expression is optional as well, as long as the referent can be retrieved from a previous sentence (Spagnolo 1993:226, 236).

Spagnolo (1933:226)

(16) ḏɔ amɛt ’bulüt? E, amɛt.
    you saw hyena? yes saw
    ‘Did you see the hyena? Yes, I did.’

2. *Multiple expression of semantic information*

- I have not seen negative concord in Bari. Negation is expressed in simple clauses by means of the particle ti–tine, as in (17), or kɔ (Spagnolo 1933:120).

(17) Nαn tı dɛn.
    I NEG know
    ‘I don’t know.’

However, since I have not seen examples of negation of indefinites (where negative concord is likely to occur), I have no conclusive evidence on this matter.
- Plural concord does occur in Bari. If a numeral (>1) is used, the noun still has to be marked for plural, e.g. yidin musala ‘sheep.pl three’ (Spagnolo 1933:226).
- There is biological gender (i.e. semantic classification) in Bari, but no multiple marking of it. Nouns denoting animates trigger masculine or feminine forms of possessives, demonstratives (as in (18)18)) and adjectives (Spagnolo, 1933:19, 29).

Spagnolo (1933:29)

(18) lɔ ɲутu’ // na ɲутu’
    DEM.M person    DEM.F person
    ‘This man / this woman.’
Inanimates are masculine when they have male-like properties (active things or objects with long shape; Spagnolo, 1933:20) and feminine when they have female-like properties (flat or hollow shape, p.20). There are more semantic criteria, e.g. natural phenomena like trees and fruits are masculine (p. 21) and abstract nouns like countries and clans are feminine (p. 22). Since gender is not marked on the nouns themselves, there is no redundancy.

**Domain disintegration**

3. **Fusion**

- There is cumulation of person and number in Bari, as these are expressed together in pronouns (Spagnolo, 1933:80). There is also cumulation in demonstratives: na ‘DEM.F’ (Spagnolo 1933:29).
- There is some stem alternation in Bari. Some nouns (±20) are marked for plural by means of suppletion, e.g. *kine* ‘sheep.SG’ > *yidin* ‘sheep.PL’ or stem mutation, e.g. *wöri* ‘son/daughter.SG’ > *wöti* ‘son/daughter.PL’ (Spagnolo, 1933:50).

4. **Discontinuity**

- There are no circumfixes in Bari.
- There are no infixes in Bari.
- I have not seen examples of argument raising in Bari.
- I have not seen instances of extraposition in Bari.

5. **Non-parallel alignment**

I have not seen examples of non-parallel alignment in Bari.
Form-based form

6. Agreement

There is agreement in Bari, as adjectives copy the feature +plural from the nouns they modify. This is illustrated in (19) and (20).

Spagnolo (1933:226)
(19) kadi duma
   house big
   ‘The big house.’

Yokwe (2001: 287)
(20) wate k ng’ a-jin
    woman.PL brave-PL
    ‘Brave women.’

7. Expletive elements

I have not found examples of weather predicates in Bari, nor of any sentences that provide information on the presence of nominal expletives.

8. Grammatical gender

As explained above in the section on multiple expression of semantic information, there is a semantic nominal classification system in Bari: animates can be masculine or feminine. Inanimates are assigned a gender on the basis of semantic properties as well. There are exceptions to the criteria (Spagnolo 1933:23), but the system is semantic in nature. Hence, Bari does not have grammatical gender.
9. *Sequence of tenses*

Tense marking is done in Bari by means of independent particles. These particles are not obligatory and therefore it is hard to find examples of explicit tense marking in embedded clauses. Sentences like (21) do not give conclusive evidence.

Spagnolo (1933:251)

(21) Kïŋósur a lô yîŋ ‘bôra’ dô nýar nûŋ.

sister COP NEG understand well you want what

‘Your sister did not understand properly what you wanted (lit., want).’

However, as tense marking is not obligatory and hardly ever explicit, it is highly unlikely that there is obligatory copying of tense in Bari. I will assume there is no *consecutio temporum* in Bari.

10. *Influence of complexity on word order*

I have not seen examples of heavy phrases moving right or light phrases moving left in Bari. The grammars do not mention anything about a possible influence of complexity.

11. *Influence of complexity on function marking*

Complexity does not determine function marking in Bari. Bari is strongly isolating: function marking is done by means of independent particles which can scope over simple words or entire clauses. The nature or complexity of the marked unit is not relevant.

12. *Syntactic functions*

I do not have enough evidence to judge whether Bari has syntactic functions. My impression from the data is that Bari has neutral alignment – arguments are unmarked for
pragmatic or semantic and therefore have syntactic functions. However, I have too little information to be sure.

Spagnolo (1933:108) claims that there is a passive construction - verbs could be passivized by means of a suffix, e.g. köt ‘to clear away’ > kötö ‘to be cleared away’. It is not clear whether this involves some kind of promotion of the Undergoer argument and whether it would be possible to add a by-phrase. If that would be the case, this would be a genuine passive and a case of morphosyntactic alignment. However, there is not enough evidence available to know whether this is really the case.

13. (Morpho-)phonologically triggered alternations

There are phonological alternation rules in Bari. Firstly, there is vowel alternation: the plural suffix for nominal stems ending in a consonant is –VN, where the vowel is influenced by the last vowel of the singular stem, e.g. kurit > kurit-ön ‘giraffe-PL’ but dak > dak-an ‘pipe-PL’ (Spagnolo, 1933:34). For nominal stems ending in a vowel, there are several possible plural suffixes. The choice depends on the stem-final vowel (p.41).

The final –n of nan ‘1.sg’ is elided before a verb stem beginning with y: nan yeye > na yeye ‘I am thinking’ (Spagnolo 1933:106). I do not know whether this alternation extends to other contexts.

References

1.3 Sudanese Arabic

Sudanese Arabic (henceforth SA, also known as Khartoum Arabic) is the Arabic variety spoken by approximately 15 million people in Sudan (Lewis 2009). Sudan in fact knows many varieties of Arabic, but the one described here is the dominant one originating in Khartoum (the capital and the financial and administrative centre of Sudan). It is the spoken language of Sudan – the written language is Modern Standard Arabic.

As Owens (1990) argues, one of the source languages of Nubi was a Sudanese variety of Arabic. Khartoum was the administrative centre of Sudan in the 19th century already, which makes it likely that Khartoum Arabic was the dominant variety and has influenced the emergence of Nubi (Caroline Roset, personal communication).

Sudanese Arabic is also the lexifier of Juba Arabic, the creole of South Sudan, a language very similar to Nubi (Owens 1990).

Redundancy

1. Apposition: Cross-reference

I do not know whether cross-reference is possible in SA. Person, number and gender are expressed by obligatory prefixes and suffixes on the predicate in Sudanese Arabic (Bergman 2002:22). It is not necessary to express arguments independently as well, but it might be allowed – I have not seen examples of it.

Independent pronouns appear to be used mostly in combination with the active participle, as in (22) (Bergman 2002:42). As there is no person marking on the predicate here, this is no cross-reference.

---

3 I am very grateful to Caroline Roset for helping me with the choice for an Arabic variety, finding literature, and providing additional data. All remaining errors are of course my own.
Bergman (2002:42)

(22) Anaa jaaya min tayriiban xamsa sanaw-aat.
    I come.ACT.PTCP from about 5 year-INANIM.PL
    ‘I came about five years ago.’

2. Multiple expression of semantic information

- There is no negative concord in SA. Sentences are negated by means of the particle *maa* (cf. example [23]), or *mif~muʃ* with non-verbal predicates (Bergman 2002:59).

Bergman (2002:59)

(23) Daaʃ akluu maa zariif.
    DEM.M.SG appearance NEG nice
    ‘That (one’s) appearance is not nice.’

These particles can appear only once in a sentences. Indefinites are negated by the same particles, as in (24).

Caroline Roset (personal communication)

(24) maa bərif hadd
    NEG know someone
    ‘I don’t know anyone.’

- SA exhibits plural concord. Nouns are marked for plural after the numerals 3-10, e.g. *xamsat agsaam* ‘five sections’, with some exceptions like *talaata yoom* ‘three day’ (Bergman 2002:53). There can also be dualis concord, e.g. *yom-een itneen* ‘day-DUAL two’, but this is optional (Bergman 2002:19, 53). With all numbers over 10, nouns are singular.

- All SA nouns are classified as either masculine or feminine (Bergman 2002: 18). This is a semantic classification: for singular nouns, male animates and inanimates get masculine
agreement, and female animates get feminine agreement. For plural nouns, human beings get masculine inflection and inanimates get feminine inflection. Nouns denoting multiple female animates can get masculine plural inflection or feminine plural inflection – in general, men use the former, and women the latter (Bergman 2002:52). Since the nominal classification is fully predictable by number, animacy and sex, this is a semantic classification system.

The semantic class of nouns is expressed on the noun itself and on adjectives, demonstratives, etc. (cf. section below on agreement). Hence, there is multiple expression of semantic class in SA.

Domain disintegration

3. Fusion


- There is stem alternation in Sudanese Arabic. Lexical stems in Arabic dialects are special, as they consist of three consonants (called roots or radicals), e.g. KBR ‘big’, together with a CV-pattern (e.g. CaCaC). Inflection and derivation often involve changes in this CV-pattern, e.g. KaTaBa ‘write.PFV’, yaKTiB ‘write.IPV’ (Bergman 2002:26). Such changes are all cases of stem alternation.

Apart from this kind of ‘regular’ stem alternation, SA also shows some irregular stem alternating inflection patterns. E.g. the verb wagaf ‘to stop’ is inflected wagif ‘stop/3.M.SG.PERF’, ya-giif ‘3.M.SG.IPFV-stop’ (Bergman 2002:27). The first radical is dropped in this context, hence not only the CV-pattern is affected, but also the stem’s

4. \textit{Discontinuity}

- I have not encountered infixes in Bergman (2002) or Trimingham (1946), while there are lists of circumfixes in both of them. I will therefore assume that SA does not have infixes.
- Sudanese Arabic exhibits argument raising. The subject of the embedded clause in (25) can become the subject of the main clause, cf. (26).

Caroline Roset (personal communication)

(25) \textit{ya-bdu annuMuhammad mariid}§

\hspace{1cm} 3.SG-seem that M. ill

‘It seems that Mohamed is ill.’

(26) Muhammad \textit{ya-bdu mariid}§

\hspace{1cm} M. 3.SG-seem ill

‘Mohamed seems ill.’

- Extraposition is not allowed in Sudanese Arabic. The integrity of a phrase is always maintained, as shown in examples (27) and (28). The relative clause \textit{al dayman ya-d\text{"i}hak} ‘who always laughs’ cannot be moved out away from the head of its phrase to the end of the sentence, despite its heaviness.

Caroline Roset (personal communication)

(27) \textit{[shuuf-ta ar-raajil al dayman ya-d\text{"i}hak] imbaarih}

\hspace{1cm} saw-1.SG DEF-man REL always 3.SG-laugh yesterday

‘I saw the man who always laughs yesterday.’
(28) *shuuf-ta ar-raajil imbaarih al dayman ya-d’hak
   saw-1.SG DEF-man yesterday REL always 3.SG-laugh
   ‘I saw the man yesterday, who always laughs.

5. Non-parallel alignment

I have not seen examples of non-parallel alignment in SA.

Form-based form

6. Agreement

There is agreement in SA: an attributive adjective or participle copies definiteness, number and gender from the noun it modifies, e.g. in (29)-(31) (Bergman 2002:51).

Bergman (2002:51)
(29) al-bitt as-samh-a
    DEF-girl DEF-pretty-F.SG
    ‘The pretty girl.’

(30) beet jadiid
    house new
    ‘A new house.’

(31) farabiyy-a jadiid-a
    car new-F.SG
    ‘A new car.’
7. *Expletive elements*

I have not seen examples of weather predicates in SA. However, there is an existential construction, shown in (32), suggesting that there is no expletive pronoun.

Trimingham (1946:58)

(32) Fih fingan?

\[\text{EXIST} \text{ cup}\]

‘Is there a cup?’

8. *Grammatical gender*

There is no grammatical gender in SA. There is natural gender, as discussed in 2, but no other nominal classification.

9. *Sequence of tenses*

There is no tense copying in Sudanese Arabic, which is shown in example (33).

Persson, Persson & Hussein (1979:171)

(33) Gaal, ḥa-yamʃí al ubeeyid bukra.

\[\text{say.PST.3.SG} \text{ FUT-go DEF Obeid tomorrow}\]

‘He said that he will go to El Obeid tomorrow.’

The past tense of the main verb is not copied to the embedded verb, which carries a future tense.

10. *Influence of complexity on word order*
The complexity of constituents does not bear an influence on morphosyntactic placement in Sudanese. Dickins (2009, 2010) studies sentence structure in SA and finds that this is largely determined by referentiality. Since his detailed analysis does not mention any influence of heaviness, I will assume that this factor does not contribute to placement.

11. Influence of complexity on function marking

There are some markers in SA (e.g. *maa* ‘emph’) that apply to units of all degrees of complexity. However, function marking is in most cases done by affixes or morphological patterns. There are for instance person, number, gender and tense markers that only attach to (simple) verbs and are hence sensitive to the complexity of their hosts. An example of morphological pattern marking is aspect marking, e.g. the roots KTB⁴ ‘write’ are perfective with the pattern CaCaC: *kataba*, and imperfective with CCiC: *ktib* (Bergman 2002:26). These patterns can only apply to words, never to more complex units. This means that function marking in SA is strongly influenced by the complexity of host units.

12. Syntactic functions

There is no evidence for a syntactic function Subject in Sudanese. Bergman (2002:34) shows that a passive-like verb can be derived from an active one by means of vowel change. Bergman does not give examples where the passive verb is used together with a by-phrase. Several Sudanese informants of Caroline Roset (personal communication) state that expressing the Agent in such a way in a passive sentence is not possible, cf. example (34). The same is true for other Arabic vernaculars in the region.

Caroline Roset (personal communication)

---

⁴ This is a stem from the first verb conjugation. The patterns are different for other verb conjugations.
(34) Ahmed itgattal *[min Muhammad]
A. kill.PASS.PST.3.SG by M.

‘Ahmed was killed (*by Muhammad).’

This construction is hence not a genuine passive. There is no reason to assume a syntactic function Subject in SA.

I have not seen examples of dative shift in Sudanese, but neither Bergman nor any other expert states that it is impossible. There is hence no conclusive evidence on the matter of a syntactic function Object in SA.

13. (Morpho-)phonologically triggered alternations

There are several alternation processes in SA. Firstly, there is final devoicing, e.g. shurab ‘folksong’ is pronounced [ʃuraːp] (Bergman 2002:2, 13). Long vowels are shortened in word-final position (p. 7).

Unstressed vowels are often deleted when they precede a single consonant (p.13). E.g. madras-at-uu ‘school-F.SG-3.SG.POSS’ becomes madrastuu.

There is place assimilation of consonants preceding other consonants, e.g. min been-um ‘from among-them’ is pronounced [mim beenum] (p.13). A frequent assimilation of this kind is the assimilation of the definite article prefix al-. The /l/ is assimilated when the first consonant of the host is an apical consonant, a dorsal consonant or a /ʃ/, e.g. as-salata ‘the salad’, aj-jaami‘a ‘the university’ (p.15).

References


2 Pichi and its source languages

2.1 Pichi

Pichi is spoken on the island of Bioko, off the coast of Equatorial Guinea. Yakpo (2009) calls Pichi ‘an offshoot’ of Krio, while others (e.g. Lewis (ed.) 2009) do not consider Pichi a separate language but a dialect of Krio. The difference between Krio and Pichi is that since the arrival in 1858 of Spanish colonialists, Pichi has been influenced by Spanish.

Krio is itself an English-based creole, spoken in Sierra Leone. Its origins are still heavily debated. This is not the place to go into the discussion too deeply, but let me outline some of the issues. There is obviously influence of (Black) English, which was spoken by the so-called Nova Scotian settlers: former slaves who lived in North America and then Nova Scotia, before being transferred to Sierra Leone. Another contribution must have been made by Trelawny Maroons from Jamaica, but it is unclear what language they spoke. What we do know is that English-based Jamaican and Suriname creoles show great similarities. This could be the result of substrate influence of African languages, as for instance argued by Kouwenberg (2004). She finds evidence for substrate influence of Bantu languages as well as for some transfer from Gbe languages to Surinam and Jamaican Creoles. Smith (2002) argues that at least the Bantu language Kikongo must have influenced Krio. As for the Gbe cluster, Smith points out that Ewe-Fon is a likely source of Krio, but practical considerations have made me include Fongbe in the sample.
Redundancy

1. Apposition: Cross-reference

Cross-reference does not exist in Pichi. There is no obligatory grammatical referential marking. There are only so called object clitics (Yakpo 2009: 180), i.e. clitics that resemble independent pronouns and can be attached to the predicate replacing an independent object. These clitics never appear together with an independently expressed object argument.

2. Multiple expression of semantic information

- There is negative concord in Pichi. As illustrated in (35), a clause is negated by using a negation marker. One marker suffices in a simple clause.

Yakpo (2009:256)

(35) À no bin fit tek mòtô.
    I NEG PST can take car
    ‘I haven’t been able to take the car.’

With indefinites, it is obligatory to negate both the clause as a whole and the indefinite. This is shown in (36).

Yakpo (2009:255)

(36) Dën no dè gi no natin.
    they NEG IPFV give NEG nothing
    ‘They don’t give anything.’ (lit.: They not give not nothing.)

Yakpo (2009:257ff.) also provides sentences where the marker no appears twice. However, in these sentences, the double expression has an emphatic effect and hence it adds meaning. In example (36) above, the second negation has no pragmatic or semantic effect.
There is plural concord in Pichi. Yakpo (2009:173) shows that there are Pichi numerals that are used up to 7. For higher amounts, Spanish numerals are employed. The Pichi numerals are used with Pichi nouns, which cannot be inflected for number (Yakpo 2009:153). The Spanish numerals however, take Spanish plural nouns: *quinze años* ’15 years’ (Yakpo 2009:545).

There is no semantic classification in Pichi, hence no multiple marking of semantic class. The feature does not apply.

**Domain disintegration**

3. **Fusion**

- There is cumulation of person and number in pronouns in Pichi. All other grammatical categories are expressed in separate words – I have not seen any portmanteau morphemes.
- I have not seen examples of stem alternation in Pichi.

4. **Discontinuity**

- There are no circumfixes in Pichi. All affixes in Pichi are suffixes (cf. Yakpo 2009:123, 125).
- There are no infixes in Pichi. All affixes in Pichi are suffixes (cf. Yakpo 2009:123, 125).
- Yakpo (2009) claims that argument raising is allowed in Pichi. For instance in (37), *Boye* is the subject of the embedded clause, while the main clause has an expletive subject. The predicate *fiba* requires a complement clause with an overt subject, so (38) is ungrammatical. Example (39), however, where *Boye* is supposedly raised and an extra pronoun is introduced, is grammatical.

Yakpo (2009:416)
(37) È fiba se Bòyé get móní.
     it seem QUOT B. get money
     ‘It seems that Boye has money.’

(38) *Bòyé fiba get móní.
     B. seem get money
     ‘Boye seems to get money.’

(39) Bòyé fiba se è get móní.
     B. seem QUOT 3.SG get money
     ‘Boye seems to have money.’

However, in my opinion, this is not really raising. An argument is raised if it is taken out of the embedded clause where it semantically ‘belongs’, and becomes an argument of the main clause. Consequently, the complement clause becomes discontinuous. In the case of (39), the embedded clause still has an explicit argument – there is no evidence that Bòyé is ‘taken out’. There is no discontinuous complement clause. I will hence assume that there is no argument raising in Pichi.

- I found no examples of extraposition in Pichi.

5. Non-parallel alignment

I found no examples of non-parallel alignment in Pichi.

Form-based form

6. Agreement
There is no agreement in the nominal domain in Pichi. Features like number and definiteness are not expressed on nouns, nor on units modifying them (Yakpo 2009:153). In the verbal domain, no features are copied from arguments to the predicate (cf. section above on cross-reference).

7. Expletive elements

Pichi has expletive elements. In fact, the language has a list of verbs that take expletive elements as arguments (Yakpo 2009:412), for instance weather predicates (fɔl ‘to rain’), copula elements (fibɔ ‘to seem’) and evaluative verbs (gud ‘to be good’). These verbs occur with a syntactic argument that has no semantic counterpart, as for instance in (40).

Yakpo (2009:436)

(40) ... è want fɔl.
     it want rain
     ‘It’s going to rain.’

However, weather predicates are usually expressed without an expletive, that is, with a semantically full argument as in (41) and (42).

Yakpo (2009:436)

(41) Dì de dak.
     DEF weather dark
     ‘It is dark.’

(42) ... ren gò fɔl.
     rain POT rain
     ‘It might rain.’
The semantically full argument in (42) shows that the predicate ṭɔl does not in fact trigger an empty pronoun in (40). In fact, the è in (40) could also be analysed as an anaphoric pronoun, referring to a semantic argument like ren, that is present in the context. I therefore consider Pichi not to have expletives.

8. *Grammatical gender*

There is no grammatical gender or other nominal classification system in Pichi (Yakpo 2009:153).

9. *Sequence of tenses*

Yakpo (2009:209) states that there is no consecutio temporum in Pichi – tense in an embedded clause is determined on semantic basis and not automatically copied from the main clause.

10. *Influence of complexity on word order*

In Pichi, the possessor is expressed pre-nominally if it is light (cf. [43]), but post-nominally when it is morphosyntactically heavy (example [44]).

Yakpo (2009:180)

(43) ìn bɔyfrɛn
    her boyfriend
    ‘Her boyfriend.’

Yakpo (2009:171)

(44) pikín fɔ mì àntí
    child of my aunt
    ‘My aunt’s child.’
Weight hence influences morphosyntactic placement in Pichi.

11. Influence of complexity on function marking

As Pichi is an isolating language, there are only few bound morphemes. Most functions are marked by means of independent particles. These particles scope over units of different complexity. Complexity hence does not influence function marking in Pichi.

12. Syntactic functions

Pichi has no syntactic functions, but exhibits neutral alignment. Only personal pronouns get function marking: first person Actor arguments are for instance expressed by means of à, first person Undergoers and Recipients by means of mí. An Undergoer can be expressed by means of the ‘object’ clitic =àn, but other arguments cannot (Yakpo 2009:390-392). Units other than pronouns get no marking at all, which means that arguments with different pragmatic and semantic roles look the same morphosyntactically. Hence, Pichi has syntactic functions.

13. (Morpho-)phonologically triggered alternations

Assimilation processes occur in Pichi. Voiced stops are devoiced before a pause (Yakpo 2009:56). In bisyllabic verbs ending in –i and with a high tone on their first syllable, the –i often gets nasalized, e.g. [háŋgrĩ] ‘hungry’, or an –n is added [háŋgrín] (id.). Furthermore, the final nasal of several units (e.g. dɛn ‘they’) assimilates to the place of assimilation of the following consonant, e.g. dɛn bɔkũ ‘they many, they’re much’ is pronounced [dɛmbɔkũ] (Yakpo 2009:57).

Finally, there may be insertion of glides in between vowels (Yakpo 2009:59) and some consonant deletion to avoid complex consonant clusters (Yakpo 2009:59).


References


2.2 English

English is an Indo-European language of the Germanic branch. It has more than 58 million native speakers in the UK, 214 million in the USA (Lewis ed. 2009), and many more speaking it as an L2 over the whole world.

Redundancy

1. Apposition: Cross-reference

English does not exhibit cross-reference. Lexical reference is obligatory, i.e. a pronominal or NP argument cannot be left out. There is grammatical reference marking on the verb for 3rd person singular, which is obligatory as well (cf. section below on agreement).

---

5 I would like to thank Norval Smith and Evelien Keizer for providing and correcting English examples.
2. **Multiple expression of semantic information**

- Present-day Standard English does not have negative concord. However, in many non-standard varieties it does appear. Furthermore, Kallel (2007) argues that negative concord was common in Old English, but was lost in Late Middle English and Early Modern English, that is, in the English of 1450-1600. Krio, on which Pichi is based, emerged probably even before 1700. At that time, English still did have negative concord.
- English exhibits plural concord, as plural marking is obligatory after a $>1$ numeral: *five elephant-s*.
- English does not have a semantic classification system of nouns, hence this feature is not applicable.

**Domain disintegration**

3. **Fusion**

- There is cumulation of person and number in pronouns.
- There is stem alternation in English verbs. Past tense can be marked by means of a suffix *–ed*. The verbal stems that take *–ed* are called weak verbs. This set also includes verbs that undergo vowel change (hence stem alternation) after adding the suffix, e.g. *keep + -ed > kept*. There is also a subset of English verbs that do not get *–ed*, but are marked for tense by other stem change, e.g. present tense *sing*, past tense *sang*, past participle *sung*.

  There are also some cases of vowel change in stems of nouns inflected for plural, e.g. singular *woman*, plural *women*.

4. **Discontinuity**

- English has no circumfixes.
- Idiosyncratically infixation of interjections is possible in English, as for instance in *abo-*
  bloody-lutely.
- English allows for argument raising, as illustrated in example (45).

(45)  a. Marilyn believed that he was honest.
      b. Marilyn believed him to be honest.

In a, the object is the complement clause *he was honest*. The subject of that clause, *he*, is
the object of the main clause in b.
- Extraposition is also allowed in English, as illustrated in example (46).

(46) The boy cried yesterday, who never shows his emotions.

5. Non-parallel alignment

English has clitics that attach to hosts they semantically do not belong to. For example, in
(47), the modal verb attaches to the subject rather than to the verb.

(47) You’d think.

Semantically, ‘*d is part of TMA marking and as such belongs to the verbal domain (where
TMA is usually expressed in English). But phonologically, it attaches to the subject. The
alignment at the semantic and phonological levels is hence in this case not parallel.

Form-based form

6. Agreement
There is agreement in English. At the phrasal level, the number of nouns is copied to
demonstratives, e.g. that flower, those flowers. At the clausal level, there is agreement
between the subject argument and the verb for the third person singular. As both the
independent lexical argument and the verb marker –s are obligatory, this is agreement and
not cross-reference.

7. Expletive elements, dummies

English uses an expletive with some weather predicates, e.g. it in it is raining.

8. Grammatical gender

English has no nominal classification system.

9. Sequence of tenses

English has a sequence of tenses rule: past tense is copied obligatorily from the main clause
to the embedded clause, cf. example (48).

(48) Stacey told her brother that he had to go.

Stacey must have used a present tense in her original utterance. In (48), however, this tense
is adapted to the tense of the main clause.

10. Influence of complexity on word order

There is influence of complexity on word order in English, as heavy phrases can be moved
to the right of the sentence and light ones to the left – sometimes even obligatorily. In (49),
word order is neutral. There are two options that do not have a different meaning. In (50),
the object argument is a pronoun, i.e. a light unit. The order from (50b) is now ungrammatical – the light phrase cannot stand in sentence-final position.

Evelien Keizer (personal communication)

(49)  a. I handed the documents over.
     b. I handed over the documents.

(50)  a. I handed them over.
     b. *I handed over them.

(51)  I handed over THEM.

Example (51) shows that weight is decisive here. The morphosyntactic lightness of the pronoun is overruled by phonological weight: the unit is heavy because of the emphatic stress. Therefore, the unit can stand in sentence-final position.

11. Influence of complexity on function marking

English has affixes (e.g. plural –s, 3rd person sG –s) as well as clitics (e.g. possessive =s and reduced negation form =n’t; [Zwicky & Pullum 1983:502-13]). The former are in the majority, which is a reason to say that English favours head-marking (non-transparent) over phrase-marking (transparent).

12. Syntactic functions

English exhibits syntactic functions. Consider examples (52) and (53).

(52)  Thora breaks the vases.

(53)  The vases break.
In (52), the Actor *Thora* triggers agreement on the verb. In (53), the Undergoer argument *the vases* triggers the agreement. The semantic distinction between Actor and Undergoer is hence neutralized, there is a Subject function.

13. *(Morpho-)phonologically triggered alternations*

There are phonological alternation rules in English. Firstly, there can be place assimilation of alveolar consonants when preceding a labial consonant (for instance *cat burglar* is sometimes pronounced /cat burglar/) and when preceding a velar (e.g. *common good* is pronounced /commoŋ good/).

Secondly, according to an anonymous contributor on a phonology blog, English exhibits degemination in fossilized derivational morphology, but not so much in productive derivation. For instance the combination of *mean* and –*ness* is pronounced /miːnəs/, but the fossilized *innumerable* (from *in-* and *numerable*) is pronounced with one /n/.

**References**


---

6 [http://www.phon.ucl.ac.uk/home/wells/blog0610b.htm](http://www.phon.ucl.ac.uk/home/wells/blog0610b.htm)
2.3  Fongbe

Fongbe (or: Fon) is a Niger-Kordofanian language. It is part of the Gbe language cluster. It is spoken in the south of Benin and Togo by approximately 1.5 million speakers (Lefebvre & Brousseau 2002:1, henceforth L&B).

One of the sources of Krio (and hence of Pichi) is the language spoken by Trelawny Maroons. There is no consensus on the exact nature of their language (cf. the analysis of Pichi), but Smith (2002) argues that at least Ewe-Fon must have contributed to it. I have chosen to study another language from the Gbe cluster, Fongbe, for the practical reason that a more complete grammar was available for this language.

Redundancy

1. Apposition: Cross-reference

There is no grammatical (but only lexical) marking of person in Fongbe. There is hence no cross-reference.

2. Multiple expression of semantic information

- Fongbe does not have negative concord. Negation of a simple clause is done by means of negation markers, e.g. mà in (54) (L&B 2002:120).

Lefebvre & Brousseau (2002:120)

(54)  Kôkù mà wá.
     K.  NEG  arrive
     ‘Koku has not arrived.’

7 I want to thank Enoch Aboh for sharing his ideas and data on Fongbe.
Negation of indefinites obligatorily involves one of these negation markers and a negative indefinite, as in (55).

Lefebvre & Brousseau (2002:371)

(55) Kòkú mà mò mètì.
    K. NEG see nobody
    ‘Koku saw no one.’
    “Koku not saw nobody”

However, this negative indefinite cannot occur on its own, cf. (56).

(56) *Mètí wá.
    nobody come
    ‘Nobody came.’

This would be negative concord if both negating elements would be semantic negators (cf. Zeijlstra 2007). However, as the indefinite cannot negate on its own, it has no semantic value. Therefore, I do not consider this to be negative concord.

In fact, it is possible to analyse the construction above as agreement: the negative marker triggers negation on the indefinite. However, in this study, I will restrict myself to subject-verb agreement and head-modifier agreement in NPs, and leave ‘negation agreement’ for further research.

- There is plural concord in Fongbe, as shown in (57).

Lefebvre & Brousseau (2002:54)

(57) ɖìɖè Kòkú tòn wè lè
    sketch K. GEN two PL
    ‘Koku’s two sketches.’
- There is no semantic classification in Fongbe. This feature is hence not applicable.

**Domain disintegration**

3. *Fusion*

- There is cumulation of person and number in personal pronouns, e.g. *nyè* ‘1.SG’, *mí* ‘1.PL’ (L&B 2002:61). I did not find any portmanteau morphemes – pronouns are the only case of cumulation.
- There is stem alternation in Fongbe as a result of vowel harmony, e.g. *gómi* > *gómù* ‘eraser’ (L&B 2002:25). This, however, only occurs in loanwords. Fongbe itself has no stem alternation.

4. *Discontinuity*

- There are no circumfixes in Fongbe (L&B 2002:187).
- There are no infixes in Fongbe (L&B 2002:187).
- L&B (2002:277) state that argument raising exists in Fongbe. There are verbs that take an expletive subject, as in (58). An Undergoer argument, or even a Locative, can then also be raised to function as the subject, as in (59) and (60) respectively.

Lefebvre & Brousseau (2002:278)

(58) ]
\[
É \text{ hwè zè } dʒ \text{ núsúnù } \text{ ́ } \text{ mè.}
\]
\[
\text{It lack salt at soup \\DEF} \\
\text{in}
\]
\[
\text{‘It lacks salt in the soup.’}
\]

(59) ]
\[
Zè \text{ hwè } dʒ \text{ núsúnù } \text{ ́ } \text{ mè.}
\]
\[
\text{salt lack at soup \\DEF} \\
\text{in}
\]
\[
\text{‘Salt is lacking in the soup.’}
\]
Actually, I do not believe this is argument raising. In Fongbe, argument functions are in most cases unmarked (cf. the section below on syntactic functions), that is, only word order and semantic inference can tell us which argument functions as the subject. Since word order is actually very flexible in Fongbe (cf. section below on the influence of complexity on word order), it cannot be seen as a strong indicator of argument relations. Therefore, it is unclear to me whether in an example like (59) and (60), ‘salt’ is really a syntactic subject argument of the clause. There is no real evidence for such an analysis and therefore I will not consider this raising. Fongbe, then, does not exhibit argument raising.

- Enoch Aboh (personal communication) indicates that extraposition is not easily allowed in Fongbe. However, he does give example (61) from Gungbe, which is similar to Fongbe in this respect.

Enoch Aboh (personal communication)

(61) Un do xo na Suru do vi eton na wa egbe
    I speak word to S. that child his will come today
    ‘I told Suru that his child will return today.’

The object xo and the clause do vi eton na wa egbe arguably ‘belong together’, but the clause was extraposed (possibly because of its weight). Apparently, extraposition is possible in Fongbe.

5. Non-parallel alignment

Lefebvre & Brousseau (2002:22) state: ‘Tonal domains correspond quite straightforwardly to morphological or syntactic domains.’ Apparently, alignment at the phonological and
morphosyntactic levels runs parallel. I have not seen examples of non-parallel alignment at the morphosyntax-semantics interface.

**Form-based form**

6. *Agreement*

There is no agreement in the nominal domain; no features are copied from nouns to nominal modifiers. This is illustrated in example (62).

Lefebvre & Brousseau (2002:51)

(62) ɖìqè ɖàgbè Kòkú tòn  screenWidth 1  le.

    sketch good K. GEN DEF PL

‘Koku’s good sketches.’

There is also no agreement in the verbal domain; no features are copied from arguments to predicates.

7. *Expletive elements, dummies*

In weather predicates, no expletive is used, cf. (63).

Lefebvre & Brousseau (2002:245)

(63) Jì jà.

    rain  fall

‘It is raining.’

As explained above in the section on cross-reference, L&B (2002:67) do postulate expletive pronouns, but I disagree with their analysis.
8. *Grammatical gender*

There is no nominal classification in Fongbe (L&B 2002:37).

9. *Sequence of tenses*

In Fongbe, temporal reference is expressed by means of phonologically independent markers (L&B 2002:85). These markers are not obligatory; usually, the interpretation of time is based on semantic properties of the predicate. It is therefore hard to find examples of embedded clauses with explicit tense marking – I have not seen any (only examples like (64), where temporal reference is implicit) and cannot prove that there is no sequence of tenses rule.

Lefebvre & Brousseau (2002:115)

(64) Kôkù ṃì dì Bàyí wá.
    K. believe COMP B. come
    ‘Koku believed that Bayi came.’

However, as tense marking is not obligatory and therefore hardly ever explicit, it is highly unlikely that there is obligatory copying of tense. Even though I cannot prove that the rule is not there, I will assume here that Fongbe has no *consecutio temporum*.

10. *Influence of complexity on word order*

A lot is possible when it comes to word order in Fongbe – there are, for instance, many constructions that involve left movement of a constituent (e.g. L&B 2002:306ff., 335ff.). All cases of dislocation, however, are motivated pragmatically. I have not found evidence that complexity is a factor here.
11. Influence of complexity on function marking

Fongbe is a strongly isolating language. Most function marking is done by means of independent particles that are blind to the complexity of the marked unit.

12. Syntactic functions

Fongbe has neutral alignment, i.e., no pragmatic, semantic or morphosyntactic roles are marked explicitly. For instance, the Actor in example (65) and the Undergoer in (66) are unmarked for their role.

Lefebvre & Brousseau (2002:241)

(65)  Kôkú  lăn.
  K. jump
  ‘Koku jumped.’

Lefebvre & Brousseau (2002:242)

(66)  Kôkú  kú.
  K. die
  ‘Koku died.’

In (67), the Actor and Undergoer arguments of the transitive predicate are unmarked as well.

Lefebvre & Brousseau (2002:247)

(67)  Kôkú  xò  Òsíbá.
  K. hit A.
  Koku hits Asiba.
Ditransitive clauses involve no marking either, cf. (68). Word order is also not informative as to the semantic role of arguments.

Lefebvre & Brousseau (2002:254)

(68) Kôkû ná Àsíbá àson / àson Àsíbá.
     K. give A. crab / crab A.
     ‘Koku gave Asiba (some) crab.’

All this means that pragmatic and semantic roles of arguments are not expressed morphosyntactically. Therefore, we should speak of syntactic functions in Fongbe.

13. (Morpho-)phonologically triggered alternations

There is tone spreading in Fongbe, e.g. àsá ‘thigh’ + mé ‘inside’ > àsá-mé (L&B 2002:22).

Another assimilation process in Fongbe is affrication, even though it is a superficial phenomenon that only occurs in rapid speech (L&B 2002:25). It involves the affrication of t and d preceding i, sometimes including palatalization, e.g. tì ‘squeeze’ is pronounced [či].

Fongbe has vowel harmony, but only in loanwords and in the diminutive suffix –i: àčú-ví ‘rat-DIM’ is pronounced [àčúvú] (L&B 2002:25).

Sonorants, as well as /b/ and /ɖ/ (which in many languages behave like sonorants, N. Smith, personal communication), are nasalized in Fongbe when they precede nasal vowels, e.g. /bɔ/ > /mɔ/ ‘to see’ (L&B 2002:27).

References


2.4 Kikongo

Kikongo (or Kongo, Koongo) is a Bantu language (Niger-Congo) spoken by 3 million people in Congo (Lewis 2009).

Kikongo is claimed by Smith (2002) to have influenced a pidgin spoken throughout the Caribbean in the 17th century. Trelawny Maroons arguably spoke this pidgin when they were transported to Nova Scotia and later to Sierra Leone, where Krio originated.

Redundancy

1. Apposition: Cross-reference

There is cross-reference in Kikongo. Predicates are marked for semantic class and number by prefixes. Apart from that, an argument can be expressed independently, as in (69). Since it is not obligatory to use an independent argument, this is cross-reference.

Dereau (1955:39)

(69) N-\text{zo} \quad \text{âme ya-\text{mbote}} \quad i-na.
   \text{II.SG-house} \quad \text{my} \quad \text{II.SG-good} \quad \text{II.SG-be}
   \text{‘My house is good.’}

Independent pronouns are usually not expressed in Kikongo, but it is possible to express them, to add emphasis, cf. (70) (Dereau 1955:29). This makes one of the referential units redundant.

Dereau (1955:29)

(70) Wa-\text{mbote kwâme} \quad \text{ngi-na}.
   \text{I.SG-good} \quad \text{EMPH} \quad \text{I.SG-be}
   \text{‘Me, I am good.’}
2. *Multiple expression of semantic information*

- I do not have enough information to establish whether there is negative concord in Kikongo. Sentences can be negated by means of the particles *ka* … *ko* that attach around the sentence (Dereau 1955:29, cf. Section 4). Since they cannot appear alone, I see this as one negative marker that cannot occur twice in a sentence. Since I have no information on the negation of indefinites, I do not know whether there is negative concord in that case.

- Plural concord exists in Kikongo. All numerals larger than 1 modify nouns that have plural prefixes (Dereau 1955:32), as shown in (71).

Dereau (1955:168)

(71) bâ-ntu kûmi ye bole
   1.PL-person ten with 1.TWO
   ‘12 people’

- As is common in Bantu languages, Kikongo has an extensive semantic classification system. Dereau (1955:17ff.) distinguishes 10 classes. Class assignment is semantically based – class 1 contains humans, class 2 contains humans with authority (e.g. ‘chief’), family members and natural things like ‘sun’, and certain higher order animal species. Class 7 contains abstract nouns that have no plural, class 9 contains diminutives, etc.

   Every noun is marked for its class (and number) by means of a prefix. Adjectives (p.19), demonstratives (p.94), numerals (p.32), possessive pronouns (p.35) and relative pronouns (p.90) are also marked for the semantic class of the noun they replace or modify, as shown for adjectives in examples (72) and (73) and for relative pronouns in (74). Semantic class can hence be expressed numerous times per phrase, clause or sentence in Kikongo.

Dereau (1955:23)
(72) mü-ntu wa-mbote
I.SG-person I.SG-good
‘A good person.’

(73) bâ-ntu ba-mbote
I.PL-person I.PL-good
‘Good persons.’

Dereau (1955:91)

(74) mü-ntu u yâ-mona
I.SG-person REL.I.SG I.SG-see.PST
‘The person that I saw.’

Domain disintegration

3. Fusion

- There is cumulation of semantic class and number in Kikongo, as these are expressed together in the prefixes described in the section above on multiple expression of semantic information.
- There is stem alternation in Kikongo. Past tense is usually marked by means of verbal suffixes, e.g. dy-a ‘eat-PRS’, dî-la ‘eat-PST’ dî-didi ‘eat-REM’. There are a few irregular verbs, e.g. kwênda ‘go.PRS’, êle ‘go.PST’, wêle ‘go.REM’ (Dereau 1955:61).

4. Discontinuity

- There is a ‘circumfix’ in Kikongo. I put the word between quotation marks as this does not involve a bound morpheme, hence not an affix in my definition – the unit under consideration attaches around clauses, rather than words. It should, however, be mentioned
here, as it is a discontinuous marker. Sentential negation can be marked by means of *ke* ... *ko* or *ka* ... *ko*, where the two particles stand around the entire sentence (cf. examples [75] and (76)). The separate particles cannot occur alone (Dereau 1955:29).

Dereau (1955:29)

(75) ba-mbote tu-na.
   1.PL-good 1.PL-be
   ‘We are good.’

(76) ka tu-na ba-mbote ko.
   NEG 1.PL-be 1.PL-good NEG
   ‘We are not good.’

- There are no infixes in Kikongo. Dereau (1955:78) mentions object marking infixes (e.g. *wã-*tu-vûluza ‘1.SG-1.PL.OBJ-save, he saved us’), but I consider these to be prefixes, as they attach to the left end of verbal stems. Dereau’s naive analysis is probably based on the fact that semantic class prefixes attaches to the left of the object markers, thus seemingly putting the object markers ‘in the middle’ of the word.

- I have not seen examples of argument raising in Kikongo.

- I have not seen examples of extraposition in Kikongo.

5. Non-parallel alignment

I have not seen examples of non-parallel alignment in Kikongo.

Form-based form

6. Agreement
There is no agreement in Kikongo. The semantic class and number of nouns is expressed on other units as well, as we have seen in the previous section on multiple expression of semantic information. But I consider agreement to be a purely morphosyntactic operation, while the class prefixes are semantic. Therefore, I do not consider this to be agreement.

7. *Expletive elements*

I have not seen examples of weather predicates in Kikongo, hence I do not know whether Kikongo exhibits expletive elements.

8. *Grammatical gender*

There is no grammatical noun classification in Kikongo, only semantic as described in the previous section on multiple expression of semantic information.

9. *Sequence of tenses*

I have not seen examples of embedded clauses in Kikongo that could help me to find out whether there is a tense copying rule.

10. *Influence of complexity on word order*

I have not seen examples of an influence of complexity on morphosyntactic placement.

11. *Influence of complexity on function marking*

Kikongo exhibits some markers that scope over sentences or clauses, like the negation marker discussed in the previous section on discontinuity. However, most function marking is done by means of affixes, e.g. the class prefixes on units of all kinds. Even though these markers look alike in different contexts, they have specialized forms for different
morphosyntactic units. So the class I plural prefix is \textit{ba-} on adjectives, \textit{tu-} on the verb ‘to be’, \textit{mi} when it is a relativizer, etc. Therefore, I do not analyze this as a clitic: there is influence of morphosyntactic class on function markers.

\textbf{12. Syntactic functions}

I do not have enough information on Kikongo alignment to be able to say whether there are syntactic functions. What I know is that there is no case marking and no obligatory voice marking on the predicate. Furthermore, the semantic class of the Actor argument of a transitive clause is marked on the predicate. I did not see examples of intransitive clauses with an Undergoer, so I do not know whether the Undergoer in that situation behaves like an Actor.

Dereau (1955:149) claims there is a passive construction in Kikongo. The verb is passivized by means of the suffix \textit{–wa}, e.g. \textit{kun-a} ‘plant-INF’, \textit{kun-wa} ‘plant-PASS, be planted’. However, since Dereau does not give any examples or any information on the properties of this construction, I do not know whether it is a genuine passive. In fact, the only thing that Dereau (1955:217) states is that ‘\textit{le kikôngo restraint le plus possible la voix passive’}, Kikongo is as reluctant as possible to use the passive voice (translation mine). I will therefore ignore it here.

\textbf{13. (Morpho-)phonologically triggered alternations}

There are several types of assimilation in Kikongo. For instance the person prefix \textit{n-} can adapt to the first consonant of its host: \textit{n- + bwêne} = \textit{m-bwêne} ‘1.SG-meet’ (Dereau 1955:14). In some cases, the stem-initial consonant is adapted, not the prefix, e.g. \textit{n- + lândidi} = \textit{n-dândidi} ‘1.SG-follow’ (p.14). Another possibility is the insertion of a consonant, e.g. \textit{n- + wâwasa} = \textit{n-gwâwase} ‘1.SG-reconcile’ (p.14).

If the first syllable of a stem contains a long vowel, and suffixes are added to the stem so that it becomes more than three syllables long, that vowel is shortened, e.g. \textit{lûmba} > \textit{lumbasana} (p.17).
The causative suffix is –isidi or –esele with monosyllabic verb stems, but –isi or –ese with polysyllabic verb stems (Dereau 1955:60).

References


3. Sri Lanka Malay and its source languages

3.1 Sri Lanka Malay

Sri Lanka Malay (henceforth SLM, also referred to as *Orang Java* or *Kirinda Java*, Nordhoff 2009:5) emerged in the 17th century. It is spoken on Sri Lanka by 46,000 people (Lewis 2009).

Sri Lanka was dominated by Dutch colonialists since their arrival in 1656 (Nordhoff 2011a). The Dutch transferred soldiers from all over the region, speaking a range of languages, to Sri Lanka. As a *lingua franca*, Eastern Trade Malay was used – a contact language that forms the basis of varieties of Malay that are nowadays spoken in East Indonesia (Paauw 2009). The Malay-speaking community mingled with the native Sri Lankan people, who spoke Tamil and Sinhala. Around 1800, the influence of Dutch diminished as the English took over control.
Redundancy

1. Apposition: Cross-reference

There is no cross-reference in SLM, since there is no grammatical person marking.\(^8\)

2. Multiple expression of semantic information

- There is no negative concord in SLM. Different types of clauses can be negated by means of different particles (Nordhoff 2009:671). One particle is always sufficient. Also in negating indefinites, no negative concord occurs. This is shown in (77).

Nordhoff (2009:673)

(77) See pukaran=hatthu=pon thama=gijja.

\[
\begin{align*}
\text{I} & \text{ work=INDEF=any} & \text{NEG.NPST=make} \\
\text{‘I don’t do any work.’}
\end{align*}
\]

- There is plural concord in SLM. For instance in (78), the noun \textit{mlaayu} has to be marked for plural (by \textit{pada}), despite the numeral that already indicates plurality.

Nordhoff (2009:243)

(78) Kandi=ka hathu thiga-pulu riibu=kee mlaayu pada arà-duuduk.

\[
\begin{align*}
\text{Kandy=} & \text{LOC INDEF three-ty thousand=SIMIL Malay PL NPST-exist.ANIM} \\
\text{‘There are 30,000 Malays in Kandy.’}
\end{align*}
\]

- There is no semantic classification in SLM, so no double expression of it.

---

\(^8\) However, Ansaldo & Nordhoff (2009: 350) notice that some speakers nowadays use person marking clitics on the predicate. Since this is still rare, I will not take it into account here, but it will be interesting to follow how this ‘incipient agreement’ (or in my terms cross-reference, since the clitic is optional) will develop.
Domain disintegration

3. Fusion

- There is cumulation of person and number in pronouns. Secondly, Nordhoff (2011b:106) argues that some allomorphs of case markers should be seen as portmanteau forms. The dative marker =nang is expressed as =dang after singular pronouns. The possessive marker =pe becomes =ppe after the same pronouns. The forms =nang and =ppe then, do not just express DAT and POSS, but also singularity. Furthermore, negative markers in SLM (Nordhoff 2009:671ff.) involve cumulation of negation and information on TMA, animacy and lexeme class.

- There is no stem alternation in SLM. There is one possible case. (In-)transitivity is usually not marked in SLM. There are however some rare verbs where, as Nordhoff states, transitivity is suppletive, e.g. thinggalam ‘sink.INTR’, cullop ‘sink.TR’ (Nordhoff 2009:484). However, I do not think we should see this as stem alternation – rather, I think we are dealing with two separate lexemes that happen to have an overlapping meaning but a different valency.

It is not plausible that there are other cases of stem alternation in SLM’s grammar, as Nordhoff (2011b) does not mention any.

4. Discontinuity

- There are no circumfixes in SLM nowadays. There used to be the nominalizers para-...-an and ka-...-an, but these are no longer productive (Nordhoff 2009:305).

- There are no infixes in SLM. I have not found any examples throughout the very thorough grammar of Nordhoff (2009). Furthermore, Nordhoff (2011b:101) states that there is no discontinuity in SLM.

- Argument raising is not found in SLM (Nordhoff 2011b:106).
- I have not seen examples of extraposition in SLM. Furthermore, Nordhoff (2011b:101) states that there are no discontinuous units in SLM. From this I gather that extraposition does not occur in the language.

5. **Non-parallel alignment**

Nordhoff (2011b:107) distinguishes various phonological units in SLM. Among others, there are Presuppositive Phrases (LH boundary tone) and Assertive Phrases (L boundary tone). Morphosyntactic NPs are always expressed as Presuppositive Phrases, while Assertive Phrases map onto predicates. Alignment at the phonological and morphosyntactic levels is hence parallel. I found no counterexamples to this parallel alignment – none on the morphosyntax/phonology interface, none on higher interfaces.

**Form-based form**

6. **Agreement**

Nordhoff (2009:495, 2011:106) states that there is no verbal agreement in SLM. This would of course not be possible, since there is no grammatical reference marking (cf. the previous section on cross-reference). In the nominal domain as well, no features are copied to other units (cf. Nordhoff 2009:415ff.).

7. **Expletive elements**

There are no expletives in SLM. As illustrated in (79), a semantically full argument is used in weather predicates.

SLM - Nordhoff (2009:504)
(79) Arà-uujiang.
    NPST-rain
    ‘It is raining.’

8. **Grammatical gender**

There is no nominal classification system in SLM (Nordhoff, 2011b:106).

9. **Sequence of tenses**

There is no sequence of tense rule in SLM (Nordhoff 2011b:105), which is proven by example (80), where the past tense of the main clause is not copied to the predicate in the embedded clause.

Nordhoff (2011b:105)

(80) Incayangsu-biilang duuva kaaka arà-duuduk katha.
    he PST-say two elder_brother NPST-exist.ANIM QUOT
    ‘He said that he had (lit. ‘has’) two elder brothers.’

10. **Influence of complexity on word order**

There can be influence of formal complexity on morphosyntactic placement in SLM. It is possible that a heavy constituent moves to sentence-final position (Nordhoff 2011b:107). For instance in (81), the heavy complement clause is moved to post-verbal position, while objects usually occur pre-verbally in SLM.

Nordhoff (2011b:107)
11. Influence of complexity on function marking

SLM has bound morphemes, of which the large majority fulfill the criteria for clitic-hood. For instance, Nordhoff (2011b:101) argues that case-marking enclitics such as =nang ‘DAT’ that attach to any kind of argument: nouns as in (82), pronouns, adjectives and clauses as in (83).

Nordhoff (2011b:101)

(82) Laayeng nigire=pe soojor pada=nang baae lakuvan=nang
different country=POSS European PL=DAT good price=DAT
anà-juuval.
PST-sell
‘(He) sold (it) to the Europeans from the other countries for a good price.’

Nordhoff (2011b:102)

(83) Suda buthul suuka nyaari siini su-dhaathang=nang.
thus correct like today here PST-come=DAT
‘So I very much liked that you came here today.’

As said before, most bound morphemes show this behaviour – there are only few bound morphemes (notably tense prefixes) that are more selective as to their hosts and should be considered affixes. Therefore, I conclude that there is no influence of complexity on function marking in SLM.
12. Syntactic functions

SLM has semantically based alignment, as semantic roles are expressed by means of case-marking clitics (Nordhoff 2011b:98ff.). This is illustrated in example (84). In (84a), the only argument of the intransitive verb is semantically an Actor. It is unmarked. In (84b), the argument of the intransitive verb (the corresponding transitive verb would be *cuppol*, hence it cannot be the case that this is a transitive verb with an unexpressed second argument) is semantically an Undergoer, and marked as such by =yang.

Nordhoff (2011b:100)
(84)  a. Dee su-thiidor baava=ka.
    he PST-sleep down=LOC
    ‘He slept downstairs.’
  b. Titanic kappal=yang su-thinggalam.
    Titanic ship=ACC PST-sink
    ‘The ship Titanic sank.’

The semantic role of the argument is relevant for case marking. Nordhoff (2009, 2011:99) argues that semantic role is never neutralized – there is no passive or other construction that would require proposing a syntactic function in SLM.

13. (Morpho-)phonological alternations

There are phonological assimilation rules in SLM. For instance, there is place assimilation of stem-final consonants, e.g. *dhaapan* ‘eight’ + *-blas* ‘-teen’ is pronounced as /dhlapamblas/ ‘eighteen’ (Nordhoff 2011b:108).

There can also be degemination in SLM, e.g. *baalek* ‘to return’ + *-king* ‘CAUS’ gives [baleking]. This, however, only happens when affixes are involved – degemination does not occur with enclitics (Nordhoff 2011b:108).
Nordhoff (2011b:109) furthermore mentions that nasalization of oral consonants adjacent to nasal consonants is most probably possible. He gives no examples.

References


3.2 Ambonese Malay

Ambonese Malay (also known as Ambon(g) Malay or *Malayu Ambong*) is the Malay variety spoken nowadays on Ambon and neighbouring islands in Indonesia by approximately 200,000 speakers (Lewis 2009). It is also spoken as a second language by around a million speakers throughout the central and south Moluccas (Paauw 2009:389).

Speakers consider Ambonese a variety of Indonesian, even though mutual intelligibility is ‘marginal’ (Paauw 2009:389).

There has been intensive trade contact with Portuguese in the 16th century, which is still visible in the Portuguese loan words that Ambonese Malay exhibits (Paauw 2009:391). Later on, there has been contact with Dutch and recently with local vernacular Ambon
languages (Paauw 2009:392). Collins (1980) discusses whether Ambonese should be seen as a creole language or not.

Sri Lanka Malay was formed on the basis of Tamil and Sinhala, which were spoken on Sri Lanka, in combination with a Malay variety that was used as a lingua franca between traders and soldiers. According to Nordhoff (2011), this lingua franca was Eastern (Indonesian) Trade Malay – a variety that is not spoken anymore, but must have been the origin of several contemporary Malay varieties of the Moluccas. These varieties are described and compared in Paauw (2009). I chose to study Ambong Malay here, as it is to my knowledge the only variety on which a good and recent grammar is available (i.e. Van Minde 1997) besides Paauw’s comparative grammar.

Redundancy

1. Apposition: Cross-reference

There is no cross-reference in Ambonese Malay. Person reference is marked by means of an independent nominal phrase, but not by any grammatical markers on the predicate or elsewhere (Paauw 2009:415). Even the independent argument can easily be left out when it is present in the context (Van Minde 1997:209).

2. Multiple expression of semantic information

- There is no negative concord in Ambon Malay. Simple clausal negation is expressed by means of a negating particle, most often *seng* (Paauw 2009:136, 401). The negation of indefinites does not require a second negator, as becomes clear in (85).

Van Minde (1997:160)
(85)  ... seng mo minta ap-apa
     NEG      want request   RDP-what
     ‘(He) doesn’t want to request for anything.’

- Ambonese Malay has no plural concord: if a noun is modified by a numeral or quantifier, it does not have to be marked for plural (Paauw 2009:405), e.g. in (86).

Paauw (2009:411)
(86)  parangpuang tuju
     woman    seven
     ‘seven women’

- There is no semantic classification of nouns in Ambon Malay; this feature does not apply.

**Domain disintegration**

3. **Fusion**

- In Ambonese Malay, cumulation occurs in personal pronouns, e.g. *beta* ‘1.SG’, *katong* ‘1.PL’. Person and number are expressed in one form and are inseparable.
- There is no stem alternation in Ambonese Malay. Apart from some changes resulting from assimilation processes (see the section below on [morpho-]phonological alternations), stems have an invariant form.

4. **Discontinuity**

- There are no circumfixes in Ambon Malay (Van Minde 1997:93).
- There are no infixes in Ambon Malay (Van Minde 1997:93).
- I have not seen examples of raising in Ambon Malay, so I do not know whether it is allowed.
- I have not seen examples of extraposition in Ambon Malay, so I do not know whether it is allowed.

5. *Non-parallel alignment*

I have not seen examples of non-parallel alignment in Ambon Malay, so I do not know whether it is allowed.

**Form-based form**

6. *Agreement*

Ambonese Malay does not exhibit agreement in the verbal or nominal domain; no features of units are copied to other units.

7. *Expletive elements*

There are no expletives in Ambon Malay. Weather predicates are zerovalent; they do not require an argument at all (Van Minde 1997:217). This is shown in example (87).

*Van Minde (1997:217)*

(87) Mo ujang.

\[
\text{FUT \ rain}
\]

‘It’s going to rain.

8. *Grammatical gender*
There is no classification of nouns on any formal basis in Ambon Malay.

9. **Sequence of tenses**

Tense is not obligatorily expressed in Ambonese Malay (Van Minde 1997:189). I have not seen examples of expression of tense in embedded clauses. However, as tense expression is not obligatory, it would be rather unexpected if it was obligatory to copy it to embedded clauses. Therefore, I will assume that there is no sequence of tenses rule in Ambon Malay.

10. **Influence of complexity on word order**

Ambonese Malay has a rigid SVO word order (Van Minde 1997: 281). Pragmatic factors can influence placement (cf. for example Van Minde 1997: 285 on left-dislocation), but I have not seen examples of influence of morphosyntactic or phonological weight.

11. **Influence of complexity on function marking**

Ambon Malay is an isolating language (Paauw 2009:397), which means that there are very few bound morphemes – Paauw (2009:413) claims that there is only one productive prefix, Van Minde (199793) distinguishes five. Function marking is hence in the vast majority of cases done by independent words that can take scope over simple and more complex units. This means that complexity does not influence function marking in Ambonese Malay.

12. **Syntactic functions**

On the basis of examples given below, I consider Ambon Malay to syntactic functions. Arguments are in most cases not marked for a (pragmatic or semantic) role, and predicates or not marked for valency or voice. In example (88) the Actor of an intransitive clause is zero-marked. In example (89), a ditransitive clause, both the Actor and the Undergoer are
unmarked and the predicate ‘prepare’ is formally the same as the intransitive predicate in (88). However, the Recipient in (89) is marked by a preposition.

Van Minde (1997:120)

(88) De lari cap’at.
    he run fast
    ‘He ran fast.’

Van Minde (1997:246)

(89) Ontua sadia ikang ni par kat’ong.
    she prepare fish DEM.PROX for us
    ‘She prepared this fish for us.’

13. (Morpho-)phonologically triggered alternations

There are several phonological alternations that are the result of adjacency of certain phonemes. The /h/ is often deleted between vowels, e.g. bah’asa ‘language’ can be pronounced [basa] (Van Minde 1997:47). There are also other cases of deletion of consonants when adjacency of consonants results in prohibited consonant clusters, or simply in rapid speech (Van Minde 1997:49, 57). E.g. d’ohod’oho ‘souvenir’ can be reduced to dod’oho in rapid speech.

There is also external sandhi in Ambonese, as consonants can also undergo assimilation of place or manner of articulation under the influence of following consonants across word boundaries. For example, seng b’isa ‘NEG possible’ is pronounced [sem b’isa] (Van Minde 1997:56). Vowels can undergo changes as well, e.g. d’ia seng s’ono ‘he NEG asleep’ > [de sen s’ono] (idem).

References


3.3 Sinhala

Sinhala or Sinhalese is an Indo-European language, from the Indo-Aryan, Sinhalese-Maldivian branch. It is the official language of Sri Lanka and spoken by more than 15 million people (Gair & Paolillo:1).

Redundancy

1. Apposition: Cross-reference

There is no cross-reference in Sinhalese. There are no grammatical referential markers (e.g. Gair & Paolillo 1997:23), only lexical ones.

2. Multiple expression of semantic information

- Sinhalese does not exhibit negative concord. Negation is expressed by means of particles at the right side of the predicate (Chandralal, 2010:12), cf. (90).

Chandralal (2010:135)

(90)  Ranjit bat ka-nn-e nææ.

R. rice eat-NPST-FOC NEG

‘Ranjit does not eat rice.

In sentences with indefinites, negative concord does not occur. When two negating forms are combined, they both have meaning and cancel each other out, as in (91).

Gair & Paolillo (1997:45)

I am thankful to James Gair and Sebastian Nordhoff for their help with the analysis of Sinhala.
(91) Mini-ha no-ka-nə dey-ak nææ.
man-SG NEG-eat-NPST.PTCP thing-INDEF NEG
‘There is nothing that that man doesn’t eat. (=He eats everything.)’

- There is plural concord in Sinhala: plural marking on the noun is obligatory with a >1 numeral.

Chandralal (2010:60)
(92) gas de-kə
tree.PL two-INANIM
‘two trees’

- There is semantic classification of nouns in Sinhala, as animates and inanimates behave different inflectionally. For example, the singular indefinite marker is –ek for animates, -ak for inanimates (Nitz & Nordhoff 2010:2). However, class is not marked on the noun itself – it is only apparent from affixes. Hence there is no double expression of semantic class.

**Domain disintegration**

3. *Fusion*

- There is cumulation of person and number in first and second person pronouns, e.g. *mama* ‘1.SG’, *api* ‘1.PL’. (Gair 1970:33). Third person pronouns are to some extent transparent, separating person and number, e.g. *mee-kə* ‘3.INANIM.PROX-SG’, *mee-wa* ‘3.INANIM.PROX-PL’ (Gair & Paolillo 1997:19).

   Another case of cumulation is found in nominal suffixes, which express number, definiteness and animacy in one form: –ā ‘SG.DEF’, –ek ‘SG.ANIM.INDEF’, -ak ‘SG.INANIM.INDEF’ (Nitz & Nordhoff 2010:3, 4).

  Active and middle voice is indicated by means of vowel change in the stem, e.g. *gotə-na-wa* ‘weave-NPST-IND’ > *gete-na-wa* ‘be_woven-NPST-IND’ (Chandralal 2010:69).

4. *Discontinuity*

- There are no circumfixes in Sinhalese (Chandralal, 2010:83).
- There are no infixes in Sinhalese (Chandralal, 2010:83).
- I have not seen examples of argument raising in Sinhala.
- Extraposition is possible in Sinhala, as illustrated in example (93). ‘Go’ and ‘to that place’ are separate morphosyntactically, while they ‘belong together’ semantically. The atypical word order in this sentence (Sinhala is usually SOV) is a result of pragmatic considerations (Gair, personal communication).

James Gair (personal communication)

(93) Ya-na-wa nan man kiya-nnan ee pættə-ţə.
    go-NPST-IND if I tell-FUT DEM place-DAT

  ‘I will tell (you) if I go to that place.’

5. *Non-parallel alignment*

No examples are found of non-parallel alignment in Sinhala.
Form-based form

6. Agreement

There is only very little agreement in Sinhala. In the nominal (Gair & Paolillo 1997:29) and in the verbal domain, features are in most cases not copied to other units. For the nominal domain, this is illustrated by examples (94) and (95): the number of the noun is not copied to the demonstrative. Example (96) shows that plural marking on demonstratives does occur, but not when the demonstrative functions as a modifier, that is, not as a result of agreement.

Gair (1970:39)
(94)  mee alut  pot-ə
     DEM new book-SG
     ‘This new book.’

Gair (personal communication)
(95)  mee alut  pot
     DEM new book.PL
     ‘These new books.’

(96)  Mee-wa alut  pot.
     DEM-PL new book.PL
     ‘These are new books.’

However, there is one case of agreement in Sinhala: numerals modifying a noun inflect for animacy, e.g. pot de-ka‘book.PL two-INANIM’, minissu de-nna ‘man.PL two-ANIM’ (Gair & Paolillo 1997:22).
7. *Expletive elements*

There are no expletives in Sinhala: in weather predicates (cf. [97]), no semantically empty argument is required.

Chandralal (2010:18)

(97) Wahi-nə-wa.
    rain-NPST-IND
    ‘It’s raining.’

8. *Grammatical gender*

As explained in the previous section on multiple expression of semantic information, there is some inflectional difference between animates and inanimates, but this is semantic, not grammatical classification.

However, according to Nordhoff & Nitz (2010:4) and Gair & Paolillo (1997:17), several subclasses of animate and inanimate nouns can be distinguished on the basis of morphophonological properties and concurrent inflectional behaviour. The difference lies in the way nouns are marked for singular (interestingly, bare nouns are plural in Sinhala, and a suffix is added to mark singular). The simple form of the singular suffix is -ə, but depending on the morphophonemic structure of the noun, this is realized as -yə, -və, ekə or –Ø. There is hence a relevant nominal classification based on morphophonemic, rather than semantic or pragmatic properties in Sinhala.

9. *Sequence of tenses*

There is no tense copying rule in Sinhala. As shown in (98), past tense in the main clause is not copied to the embedded clause.

Chandralal (2010:192)
(98) Labana sati-e kæmpæs waha-nə-wa kiəla upəkuləpəti
text week-LOC campus close-NPST-IND QUOT vice-chancellor
amaatyanshaya-ə waartaə kəla-a.
ministry-DAT report do.PST-IND
‘The vice-chancellor reported to the ministry that the university will be closed next week.’

10. Influence of complexity on word order

Chandralal (2010:189) states that there is no strict constraint on placements of (heavy) complement clauses. Usually, they appear in preverbal position, but they can move right for pragmatic reasons. James Gair (personal communication) also stresses the importance of information structure for word order and movement in Sinhala. Whether complexity can influence placement, is not clear.

11. Influence of complexity on function marking

There are quite some bound morphemes in Sinhala, which are in the large majority of cases affixes. They are highly selective as to their hosts (most bound morphemes attach to only one lexeme class) and always appear in the same position (Chandralal 2010:56ff.). There are also a few phrase-markers (Gair & Paolillo 1997:29), e.g. -mə ‘EMPH’ and the independent particle hemə ‘etcetera’. Also, there is a nominalizer -ekə that can attach to entire sentences (Gair & Paolillo 1997:52). However, since the majority of function markers is head-marking, I will assume that there is influence of complexity on function marking in Sinhala.

12. Syntactic functions

Sinhala predominantly shows semantic alignment, but there is reason to postulate at least a syntactic function Subject. The semantic nature of the predicate is marked by means of the
vowel in the verbal stem, e.g. kapə ‘cut.ACT’ kæpe ‘cut.MID’ kappa- ‘cut.CAUS’ (Gair 1970:34). The difference between the predicates here lies in the volitionality of the arguments: in the ‘active’ variant, there is an Actor volitionally performing an Action (possibly on an Undergoer, in a transitive predicate), in the ‘middle’ variant, there is only an Undergoer argument that undergoes an action unintentionally (Gair & Paolillo 1997:38ff.). The difference is especially clear in (99) and (100).

Gair & Paolillo (1997:32)

(99) Mini-ha duwə-na-wa.
    man-SG run.ACT-NPST-IND
    ‘The man is running.’

Gair & Paolillo (1997:33)

(100) Mini-ha-ʈə diwe-na-wa.
    man-SG-DAT run.MID-NPST-IND
    ‘The man is running (involuntarily).’

A second argument for semantic alignment, is that the Undergoer argument is marked for accusative case, also in intransitive clauses, e.g. in example (101). Semantic role is not neutralized, but relevant for case marking.

Chandralal (2010:121)

(101) Lame-a(-wə) waete-na-wa.
    child-SG(-ACC) fall.MID-NPST-IND
    ‘The child is falling.’

However, only animate arguments are (optionally) marked for their semantic role, by dative or accusative case marking (as in examples (100), (101) and (102); Chandralal 2010:45). Inanimates are always unmarked, even when they are Undergoer as in (103). This means that the semantic role of inanimates is neutralized in the case marking system of Sinhala.
Gair & Paolillo (1997:32)

(102) Mamə miniha(-wə) dækk-a.
        I man(-ACC) see.PST-IND
     ‘I saw the man.’

(103) Mamə ee potə(*-wə) kiyeww-a.
        I that book(*-ACC) read.PST-IND
     ‘I read that book.’

However, the semanticity of the alignment is still apparent from the predicate’s voice – there is no complete neutralization of the semantics. Therefore I do not see this as an instance of morphosyntactic alignment.

Some grammars (Chandralal 2010; Gair 1970) claim that there is a passive construction in Sinhala, cf. (104) and (105). If this is a genuine passive, it would be a reason to propose a syntactic function Subject for Sinhala.

Chandralal (2010:158)

(104) (Kolla atin) wæte hæde-na-wa.
        (boy by) fence make.MID-NPST-IND
     ‘The fence was made by the boy.’

Gair (1970:79)

(105) (Miniha atin) lamea-wə waive-wi.
        (man by) child-ACC fall.MID-POT
     ‘The child might be dropped (by the man).’

In these sentences, the Undergoer arguments (fence and child) grammatically behave like Actors in active sentences. The Actor arguments (boy and man) can be expressed by means of a by-phrase (atin). The construction is somewhat limited as only humans can function as
atin-arguments of a passive sentence. This shows that the construction is still semantically conditioned.

Gair & Paolillo (1997:38), however, claim that the construction outlined here is actually not a genuine passive construction. They say that these sentences are “more accurately designated as involutive”, that is, as what I call middle voice predicates. But in fact, Gair (1970:78-79) himself gives the decisive argument to label the construction a genuine passive: many middle voice predicates cannot get a atin-phrase. The predicates given here, as well as a few other, can get an atin-phrase. For this reason, I consider (104) and (105) genuine passives and treat them as morphosyntactic alignment.

In sum, Sinhala has a predominantly semantic alignment system, but shows one instance of morphosyntactic alignment, that makes it necessary to assume a syntactic function Subject.

13. (Morpho-)phonological alternations

Sinhala has several (morpho-)phonologically conditioned alternations. The affirmative marker has different allomorphs, the choice of which depends on the phonological form of the predicate. It appears as –i on the modal adjective puluan ‘possible’, as –y after vowels and as –uy after consonants (Gair 1970:42).

Stem-final nasals adapt their place of assimilation to a following consonant. This involves palatalization (e.g.: /koṭin/ ‘tigers’ + /-ṭa/ ‘DAT’ = /kotiṭa/; Chandralal 2010:35), velarization and labialization.

Disparate consonants can undergo assimilation when they form clusters. Such assimilation can be progressive, e.g. /ad/ ‘pull’ + /wə-nə-wa/ ‘CAUS-NPST-IND’ = /addənəwa/ ‘cause to pull’, or regressive, e.g. /puwak/ ‘aricanut’ + /gaha/ ‘tree’ = /puwaggaha/ ‘aricanut tree’ (Chandralal 2010:35).

Vowel collision leads to alternations as well (Chandralal 2010:91). For example, animate nominal stems, normally ending in a vowel, get suffix –o in plural. In that case, the stem-final vowel is deleted (e.g. waḍuwa > waḍuw-o, ‘carpenter.SG > carpenter-PL’; Chandralal 2010:45). In other cases of vowel collision, a glide is added between two
vowels to avoid deletion, e.g. amaaru ‘difficult’ + -ak ‘INDEF’ = amaaruwak ‘a difficulty’ (Gair & Paolillo 1997:8).

Finally, there are some instances of vowel harmony in Sinhala. For instance the feminine suffix –i can trigger this: munupuru ‘grandchild’ > minipiri-i ‘granddaughter-F’ (Chandralal 2010:83).

References


3.4 Tamil

Tamil is a Dravidian language spoken in India, mainly in the Tamil Nadu region and on Sri Lanka. It has between 50 and 70 million speakers (Andronov 2004). It is closely related to Malayalam and Irula (Steever, 2005: 31). There has been language contact with Portuguese, Telugu, Arabic, and Indo-Aryan languages (Schiffman 1999:11).

10 Different authors use different orthographies to represent spoken Tamil. To be consistent, I have tried to transliterate all examples to Lehmann’s (1989) orthography, and made some adaptations on advice of Mohamed Jaffar. I am very thankful to Mohamed Jaffar, Ian Smith, and Sebastian Nordhoff for their help in analyzing Tamil. All remaining errors are of course mine.
Redundancy

1. *Apposition: Cross-reference*

Tamil has cross-reference, as shown in (106). Finite verbs are obligatorily marked for tense, person and number (e.g. Steever 2005:61). Expressing the argument independently (by means of a pronoun or NP) is optional, as long as the referent is clear from preceding discourse (Lehmann 1989:173).

Asher (1982:53)

(106) (Raaju) va-nt-aaru.
    (R.) come-PST-3.SG.HON
    ‘Raju/he came.’

2. *Multiple expression of semantic information*

- Tamil does not exhibit negative concord. Negation is expressed by means of the particle *ille* (in copular sentences) or by a negative form of the verb (in regular verbal sentences; Asher 1982:78). Sometimes, a conjunctive negative suffix – *aama* is used, which can be combined with other negative forms. This does not lead to concord. For example, in example (107), there are two formal and two semantic negations.

Asher (1982:78)

(107) Avan veele ceyy-aama iru-kka maatṭa-aan.
    he work do-NEG.CONJ be-INF FUT.NEG-3.SG.M
    ‘He will not be not working.’ = ‘He is always working.’

Tamil does not have negative indefinites like ‘never’; to arrive at that meaning a simple negation is used, together with ‘always’ (Lehmann 1989:231).
- Tamil has plural concord with nouns of the rational class (humans and deities): if a numeral is used as a modifier of a rational noun, the noun still has to be marked for plurality (Schiffman 1999:27). This is illustrated in example Error! Reference source not found.

Schiffman (1999:27)
(108) reṇḍu manevīi-ṇkal=aa?
two wive-PL=Q
‘Two wives?’

- Tamil has a semantic classification system, but semantic class is not expressed redundantly. Singular nouns are divided over three genders: masculine, feminine and neuter (inanimate). Plural nouns are classified as human or non-human. This classification system is truly semantic (Andronov 2004:57, Asher 1982:136), as shown in examples (109) and (110). In (109), the child is marked on the predicate as being neuter. In (110), the predicate marking is feminine. This shows that the noun does not carry a (semantically empty) grammatical feature, but has natural gender.

Andronov (2004:58)
(109) Kulandai alu-d-adu.
child weep-PST-3.SG.N
‘The child wept.’

(110) Kulandai vishalam tan tayappanaar-iḍam varu-ŷir-aal.
child V. own father-LOC come-PRS-3.SG.F
‘The girl Vishalam comes near her father.’

Even though we see the expression of semantic class (i.e. biological gender) on verb markers, this is not a case of redundancy. As the noun itself is not marked for its gender,
and the predicate marker can appear only once in an utterance, there is no double expression.

**Domain disintegration**

3. **Fusion**

- There is cumulation of person, number, gender and social status in predicate markers and pronouns (Asher 1982:141).
- Verbal stems in Tamil can be divided over different classes according to their inflectional behaviour. Different authors propose different classifications, but I will use the classification by Schiffman (1991) here, as he provides most information. Schiffman’s (1999:51) class V has a paradigm that involves stem alternation. There are also verbal stems that do not fall into any of the classes, which have alternating consonants (Schiffman 1999:57). Then, there are irregular verbs (Lehmann 1989:65).

  There are also some verbal stems that undergo alternations depending on the voice they are used in, for instance *iṟaṅk* ‘descend.MID’ vs. *iṟakk* ‘descend.ACT’.

  Stem alternation occurs in the nominal domain as well. There is for instance a noun class where the oblique form of the stem is formed by means of doubling the consonant: *aaru* ‘river.NOM’, *aarru* ‘river.OBL’ (Lehmann 1989:14ff.).

4. **Discontinuity**

- There are no circumfixes in Tamil (Ian Smith, personal communication).
- There are no infixes in Tamil (Ian Smith, personal communication).
- According to Andronov (2004), Tamil has argument raising – he gives example (111). Andronov focuses on literary Tamil, but different informants indicate that the example is not ungrammatical or strange in spoken Tamil.
Andronov (2004:428)

(111) Ide-parri avaloo-ɖu peeci-v-ɖa veenɖum
     about-this her-with speak-FUT-3.SG.F want
     enru toonriy-adu avan-ukku.
     QUOT seem-SG.N 3.SG.M-DAT

‘It seemed to him that it was necessary to speak with her about it.’

However, it does not seem a genuine example of argument raising to me. It is not the case that an argument is taken from the complement clause and raised to be an argument of the main clause. Rather, the entire complement clause is in subject position. This analysis is supported by an example from Lehmann, given here as (112).

Lehmann (1989:175)

(112) Kumaar Raajaav-ai paarkk-a neer-nt-adu.
     K. R.-ACC see-INF happen-PST-3.SG.N

‘It happened that Kumar saw Raja.’

One might argue that this is an example of argument raising, where Kumar is promoted to subject of the main clause. The translation would have to be ‘Kumar happened to see Raja.’ This would be in line with the case marking (Kumar unmarked, Raja marked for accusative). But the predicate would in this analysis have to show masculine marking instead of neutral. This proves that the subject here is the entire complement, not Kumar.

I have not seen examples of genuine argument raising in Tamil, hence I do not know whether it is allowed.

- I have not seen examples of extraposition in Tamil, but according to an informant it is allowed. On the basis of the informant’s statement, I will assume it is possible.

5. Non-parallel alignment
I did not find examples of non-parallel alignment. None of my informants could come up with an example. I will therefore assume that alignment between pragmatics/semantics and morphosyntax/phonology is always parallel in Tamil.

**Form-based form**

6. **Agreement**

There is no agreement in Tamil. There is a semantic noun classification system and nouns are marked for number, but such information is not copied to other units (e.g. Asher 1982:188). An example is given in (113).

Andronov (2004:153)

(113) periya kaŋ-kal
    large eye-PL
    ‘Large eyes.’

As explained previously, Tamil exhibits cross-reference but no agreement at the clausal level.

7. **Expletive elements**

Weather verbs are expressed in Tamil without using dummies, as illustrated in (114).

Asher (1982:53)

(114) Ma | -ai pey-r-adu.
    rain-ACC fall-PRS-3.SG.N
    ‘It’s raining.’
8. Grammatical gender

As described above, nouns are classified as masculine, feminine, or neuter when they are singular, and as human or non-human when they are plural. This classification is based on semantics (see examples (109) and [112]), so this is not grammatical gender.

However, Tamil does have a nominal classification system based on morphophonemic properties. Asher (1982:137) and Lehmann (1989:14) argue that there is classification of nominal stems on the basis of their behaviour when they get an oblique marker. For instance, there are nouns that end in –a in nominative case and get the oblique marker –att- (pa/a- ‘fruit.NOM’, pa/-att- ‘fruit-OBL’). Another class contains stems ending in –tu or –ru. To form an oblique stem, the final consonant is doubled: aaru ‘river.NOM’, aarru ‘river.OBL’ (Lehmann 1989:14ff.). As these noun classes or motivated by (morpho)phonemic information, and not by semantic or pragmatic considerations, this is non-transparent.

9. Sequence of tenses

- Tense copying is possible, but not obligatory in Tamil. The present tense from the original utterance is retained when the clause is embedded and the pronoun is shifted, as in (115)\textsuperscript{11}.

Lehmann (1989:374)

(115) Avan nii-nkal uur-ukku poo-kir-iir-kal enru con-n-aan.
   he 2-PL town-DAT go-PRS-2-PL COMP say-PST-3.SG.M
   ‘He said that you would go out of town.’

The same is true in a different type of embedded clauses that involves nominalization of the embedded predicate.

\textsuperscript{11} As Lehmann says that the second person in (115) cannot refer to the addressee, we must be dealing with real syntactic embedding and shifting of the original pronoun.
However, tense adaptation does exist: Asher (1982:49) says that the tenses of different clauses ‘tend to match’, as illustrated in example (116). This is true for ‘sentence-type’ as well as nominalized embedded clauses.

Asher (1982:20)

(116) Avan iNge va-nd-adu aaccariyamaa iru-nt-adu.

he here come-PST-NLZR surprising be-PST-3.SG.HON
va-r-ru-adu iru-ku-du
come-PRS-NLZR be-PRS-3.SG.HON

‘His coming here was/is surprising.’

10. Influence of complexity on word order

According to Asher (1982:99), Tamil is reluctant to move heavy phrases. It does occur, but ‘movement of heavy structures is relatively rare and never obligatory’. Lehmann (1989:176) shows that word order is relatively free in Tamil, but the verb is strictly sentence-final. Even though it is restricted, there is influence of weight on morphosyntactic placement in Tamil.

11. Influence of complexity on function marking

Asher (1982:195ff.) and Lehmann (1989:150) analyse some of Tamil’s bound morphemes as clitics. This concerns modal particles (e.g. /um ‘should’) and several other particles (e.g. the interrogative particle =aa, hearsay particle =aam, coordinator =um). I agree with this analysis because these particles all match the following criteria: they are phonologically dependent, their hosts can be of different categories and degrees of complexity and they can stand in different positions (e.g. the interrogative particle can attach to any constituent in the sentence). From the modal particles, there is also a strong form, which is another indication that we are dealing with clitics. Hence, there are phrase-markers in Tamil.
However, the majority of bound morphemes is head-marking in nature, e.g. all case-markers and reference markers on predicates. These units match the following criteria for affix-hood: they attach to hosts of one category (nouns and verbs respectively), they always occur in the same position and they do not attach to more complex units.

12. **Syntactic functions**

Tamil’s alignment system is partly semantically based, as the semantic role of arguments is relevant for the inflection of 60% of the verbs (Lehmann 1989:50). For these verbs, different stem forms\(^\text{12}\) are used for an intransitive clause with a volitional performer of the action than for a process involving a non-volitional Undergoer. Steever (2005:58) and Lehmann (1989) refer to the Undergoer-inflection as ‘affective voice’ and to the Actor-inflection as ‘effective voice’, but I will use middle and active voice respectively. An example of the difference is given in (117).

Lehmann (1989:51)
(117)  a. peru-ku-
    increase-MID-
    ‘increases’ (e.g. it increases)

b. peru-kku-
    increase-ACT-
    ‘increases’ (e.g. he increases s.th.)

However, for the other 40% of verbs, there is no semantic alignment. For instance, the Actor-argument ‘boy’ in (118) and the Undergoer-argument ‘door’ in (119) behave

\(^{12}\) Authors disagree on whether the voice marker is a part of the verbal stem, a separate suffix or a portmanteau morpheme together with the tense marker. I assume the first option here, in agreement with Lehmann (1989), for practical reasons: he provides the best glosses. This choice has no relevance for the outcome of the transparency analysis.
identically. There is no voice marking. Also, the fact that *door* functions as Undergoer in (119) is not expressed by means of case-marking.

Lehmann (1989:27)

(118) Paiyan caaviy-aal katav-ait tira-nt-aan.
    boy   key-INSTR door-ACC open-PST-3.SG.M
    ‘The boy opened the door with a key.’

(119) Katavu tira-nt-adu.
    door   open-PST-3.SG.N
    ‘The door opened.’

Here, semantic roles are neutralized and there is evidence for a syntactic role Subject.

13. *(Morpho-)phonologically triggered alternations*

If a suffix that starts with a vowel is added to a stem ending in a vowel, different processes might occur: a glide can be inserted between the vowels (Schiffman 1999:20) or the stem-final vowel is deleted (p.22), e.g. *naakku* > *naak-ile* ‘tongue-LOC’ (Asher 1982:238).

    Intervocalic −v− or −h− is often deleted, except when the −v− is the causative marker (Schiffman 1999:10).

    If a (nominal or verbal) stem ends in a rhotic, it is deleted before the past marker, which results in stem alternation. E.g. *okkaar− ‘sit’ > okkaa-nd-an ‘sit-PST-3.SG.M’* (Schiffman 1999:54). Sonorants at the end of a stem or suffix are also deleted before a suffix starting with a plosive (Schiffman 1999:22), e.g. *paar-u* ‘see-IMP’ *paa-tt-een* ‘see-PST-1.SG’ (Asher 1982:239). Other consonants assimilate their place of articulation to the following consonant, e.g. *coll-u* ‘say-IMP’ *con-n-een* ‘say-PST-1.SG’ (Asher 1982:235).

References
4. **Diu Indo-Portuguese and its source languages**

4.1 **Diu Indo-Portuguese**

DIP is a contact language spoken in India on the island Diu, off the coast of Gujarat state. It emerged in the 16th century, when Portuguese-speaking colonists took control over the island, where people spoke Gujarati. In the first centuries of DIP’s emergence, various other languages and varieties were present on Diu, as many visitors (slaves, soldiers, traders, etc.) came to the island who spoke different varieties of Portuguese, African, and Asian languages. Intermarriage between the people of Diu and Daman (a neighbouring Portuguese colony) territories created an even more multilingual and mixed population. Nowadays, there is language contact between DIP and Gujarati, Hindi and English.

**Redundancy**

1. **Apposition: Cross-reference**

There is no grammatical expression of reference in DIP (cf. [120]) – only lexical. Hence, there is no cross-reference.
DIP - Cardoso (2009:198)

(120) ɔm larg-o ped.
\hspace{1cm} \text{man release-PST fart}
\hspace{1cm} ‘The man let out a fart.’

2. *Multiple expression of semantic information*

- DIP shows negative concord with negative indefinites, as shown in (121). Clausal negation is expressed by the particle *nã*, which is still obligatory in combination with ‘nobody’ and other negative indefinites.\(^{13}\)

DIP - Cardoso (2009:211)

(121) ĩge nã apĩŋ-o pex
\hspace{1cm} \text{nobody not catch-PST fish}
\hspace{1cm} ‘Nobody caught fish.’

- Plural nouns in DIP are optionally marked by means of the particle *tud*, literally meaning ‘all’ (Cardoso 2009:174). With a numeral, the noun need not be marked for plural: *sey adiw* ‘six fox’ (Cardoso 2009:19). There is no plural concord in DIP.

- Biological gender, i.e. masculine or feminine, is marked in 3\(^{rd}\) person possessives in DIP (Cardoso 2009:125). However, since the gender is not marked on the noun, there is no double marking.

\(^{13}\) Cardoso (2009:304) notes that Gujarati does not have negative concord, but Portuguese does, making it likely that DIP inherited negative concord from the latter. However, in Portuguese, negative concord only appears when the negative quantifier precedes the clause – when ‘nobody’ stands to the right of the negating particle, that particle is optional. This is a difference with negative concord in DIP, which is obligatory regardless of word order.
Domain disintegration

3. Fusion

- There is cumulation in pronouns (Cardoso 2009:127) and in auxiliary stems. For instance, *te* ‘IPFV.NPST’ (p. 146) combines tense and aspect, and *a(d)* ‘IRR.NPST’ (p.148) cumulates mood and tense.
- Stem alternation exists in the verbal domain. The frequent verbs ‘to come’, ‘to give’ and ‘to go’ have suppletive stem paradigms, e.g. *vay* ‘go.NPST’, *foy* ‘go.PST’, *id* ‘go.PTCP’ (Cardoso 2009:111).

4. Discontinuity

- There are no circumfixes in DIP (Cardoso 2009:254ff).
- There are no infixes in DIP (Cardoso 2009:254ff).
- Cardoso (personal communication) has not seen any instances of argument raising in DIP. I will hence assume that it is not allowed.
- I have not seen examples of extraposition in DIP. Cardoso (2009) gives examples like (122), saying that a noun (*gate*) and the adjective that modifies it (*pikanin*) are separated morphosyntactically.

DIP - Cardoso (2009:177)

(122) Ali ikəl gate nə te apərəs-e pikənin?
    there DEM gate NEG IPFV.NPST show.INF small
    ‘Doesn’t that small gate show over there?’

However, in my opinion, *pikanin* is not necessarily an attributive adjective. A different analysis is possible (and more plausible, considering the similarity between *apərəs* and ‘appear’) where *pikanin* is an adjectival predicate, giving the translation: ‘Doesn’t that gate
appear small?’ In that analysis, there is no discontinuous constituent. Sentence (122) can hence not constitute evidence for discontinuity in DIP.

5. Non-parallel alignment

I found no examples of non-parallel alignment.

Form-based form

6. Agreement

There is no agreement in DIP. In the nominal domain, no features are copied from the noun to its modifiers, as can be seen in (123).

(123) ikòl fin ped
     DEM small stone
     ‘That small stone.’

In the verbal domain, there is no marking of argument features on the predicate.

7. Expletive elements

In weather verbs, no expletive is used, as shown in (124).

Cardoso (2009:260)

(124) ikòl an tiŋ muyt chuv.
     DEM year exist.PST much rain
     ‘That year there was a lot of rain.’ (Litt.: “That year much rain existed.”)
8. *Grammatical gender*

There is no nominal classification at all in DIP (Cardoso 2009:117).

9. *Sequence of tenses*

There is no tense copying in DIP, as illustrated in (125). The past tense in the main clause is not copied to the embedded clause.

Cardoso (2009:134)

(125) Yo sab-ŋ ki el ta fal-a Liza mem.  
1.SG know-PST COMP 3.SG IPFV.NPST say-INF L. EMPH  
‘I knew that he would say (lit.: is saying) “Liza” only.

10. *Influence of complexity on word order*

Word order is mostly determined by pragmatic and semantic factors (Cardoso 2009:212). There is, however, at least one effect of complexity. In DIP, possessors are usually expressed post-nominally. Only when the possessor is morphosyntactically light, it can be expressed pre-nominally, as in example (126).

Hugo Cardoso (personal communication)

(126) da tete kaz  
of T. house  
‘Tete’s house.’

11. *Influence of complexity on function marking*

DIP is a predominantly isolating language (Cardoso 2009:107), so there are hardly any bound morphemes. Pragmatic or semantic information is expressed by means of
independent particles scoping over simple and complex units of all categories. Complexity hence does not influence the way functions are marked in DIP.

12. Syntactic functions

DIP predominantly exhibits semantic alignment, but it does have at least a syntactic function Subject. Cardoso (2009:191ff.) demonstrates that inanimate arguments of intransitive clauses are always unmarked. In transitive clauses, inanimate Recipients and inanimate pronominal Undergoers obligatorily get dative case marking, while for nominal Undergoers dative marking is optional (cf. [127]).

Cardoso (2009:191)

(127) Aviaw  arəm-o pared.
   plane  hit-PST  wall
   ‘The plane hit the wall.’

Animate Undergoers in transitive clauses, however, always get dative marking, pronominal or not (cf. [128]).

Cardoso (2009:195)

(128) Leopard  foy  murd-e pə  lion.
   leopard  go-PST  bite-INF  DAT  lion
   ‘The leopard went and bit the lion.’

This shows that animacy and semantic role are relevant for alignment, hence, that alignment is generally semantically based.

However, in intransitive clauses, semantic roles are neutralized. Compare for instance the Actor in (129) and the Undergoer in (130). These behave the same – they are unmarked. The argument of (130) cannot get dative marking (H. Cardoso, personal communication), so animacy is no longer a distinguishing factor here.
Cardoso (2009:195)

(129) Virgilo bate-w pə kāw.
V. hit-PST DAT dog
‘Virgilio hit the dog.’

Cardoso (2009:198)

(130) Armando kai-w.
A. fall-PST
‘Armando fell.’

Since semantic role is neutralized here, we need to postulate at least a syntactic function Subject.

13. (Morpho-)phonologically triggered alternations

There is some assimilation in DIP: vowels that are adjacent to a velar nasal consonant are nasalized, though this is not obligatory (Cardoso 2009:93). This nasality spread is highly restricted and infrequent (Cardoso, personal communication).

Some speakers insert a palatal glide before mid-high and mid-low vowels, e.g. es ‘this’ may be realized as [jes] (Cardoso 2009:85), or a labial-velar approximant before a back vowel (p.88). This is strictly optional and individual behaviour and the reason for insertion is unclear – therefore I will not treat it as a (morpho-)phonologically triggered process here. The insertion of a glide in between two vowels (e.g. a el ‘to him’> [ajel]) is, however, a regular process and I will treat it as an alternation.

References

4.2 European Portuguese

Portuguese is an Indo-European (Romance) language. There are two major varieties: Peninsular or European Portuguese (spoken in Portugal by 10 million people, Lewis 2009) and Brazilian Portuguese. The decision which variety to study here was a difficult one, as both varieties are different from colonial Portuguese to a similar degree – as far as we know what the latter language looked like. I have chosen for European Portuguese as I am under the impression that morphosyntactically, that variety is closer to the colonist’s Portuguese (Kees Hengeveld, personal communication). Brazilian might very well be closer phonologically, but in this study, morphosyntax is more important than phonology.

Redundancy

1. Apposition: Cross-reference

Portuguese has cross-reference, shown in (131). The referential marking on the predicate is obligatory, but the independent lexical expression is optional.

(131) (Eu) chege-i.
   1.SG arrive-1.SG
   ‘I arrived.’

2. Multiple expression of semantic information

---

14 Many thanks to Lachlan Mackenzie for filling out the questionnaire on Portuguese and to Joceli Stassi Sé for her help with glosses.
- There is negative concord between the negative head and negative indefinites in Portuguese. Cardoso (2009: 304) shows that this occurs only when the negated clauses (containing the indefinite) appear after the clausal negator, as in (132).

(132) O João não telefonou ninguém.
   DEF J. NEG phones no-one
   ‘João does not call anyone.’

- There is also plural concord in European Portuguese. As shown in (133), plural marking on the noun is obligatory after a >1 numeral.

(133) cinco elefant-es
   five elephant-PL
   ‘Five elephants.’

- There is a gender system in Portuguese: nouns are masculine or feminine. This is semantically based for animates, but arbitrary for inanimates. Therefore, I will consider this grammatical gender and not semantic classification. There can hence not be double expression of semantic class – this feature does not apply.

**Domain disintegration**

3. *Fusion*

- There is cumulation in Portuguese. Person and number are cumulated in pronouns. There are also portmanteau verb markers, e.g. *lhos* in (134), which expresses person, number, gender and directional information.
(134) dei-lhos
    gave.1.SG-3.M>3.PL
    gave.1.SG-3.PL>3.PL
    gave.1.SG-3.PL>3.M/F
    ‘I gave it(M) to them / I gave them to them / I gave them to him/her.’

- There is stem alternation in Portuguese. Notably, there are suppletive verb paradigms, e.g. the verb *ser* ‘to be’ has the forms *s-ão* ‘they are’, *e-ram* ‘they were (IPFV)’, *fo-ram* ‘they were (PFV)’, etc.

4. Discontinuity

- It has been claimed (e.g. by Van der Leeuw 1997) that Portuguese has mesoclisis, e.g. the unit *=lo* in (135). This is a highly controversial point of view – most authors think there is no such thing as mesoclisis and that forms like *=lo* should be seen as suffixes.

(135) Comprá=lo-ei.
    buy=3.SG.M-FUT.1.SG
    ‘I will buy it.’

In my view, this unit cannot be a clitic since it does not meet my main criterion: it cannot attach to hosts of different lexeme class and different degrees of complexity, but only to verbs. Phonologically speaking, there may be reasons to call this a clitic, but from a morphosyntactic point of view I consider this to be a suffix.

There is another infix candidate: the diminutive marker *–inh*, for instance *Carlos > Carl-inh-os*. However, I will consider this unit to be a suffix as well, as the only morphemes that are allowed after *–inh* are gender suffixes. *–inh* attaches to the end of the stem, and other inflection follows – I see no reason to analyze this as an infix.

- There is circumfixation in Portuguese, viz. the derivational circumfix *en-X-ec* verbalizes adjectives, e.g. *frac-* ‘weak’ > *en-fraqu-ec-er* ‘VBLZ-weak-VBLZ-INF, to weaken’.
- Argument raising exists in Portuguese, as shown in example (136). The subject of the embedded clause (o João) is raised to be the subject of the main clause.

(136) O João pareceu estar doente.
    DEF J. seems be ill
    ‘João seems to be ill.

- Extraposition is possible in Portuguese. In example (137), the NP is split up in two parts as the relative clause is moved right, leaving behind its head.

(137) Vi [aquele gajo] ontem, [que vestia um manto roxo].
    see.PST.1.SG that guy yesterday who wear INDEF cloak red
    ‘I saw that guy yesterday, who was wearing a red cloak.’

5. *Non-parallel alignment*

Parallel alignment can be violated in European Portuguese. An example is shown in (138). Phonologically, *se* and *não* are taken together to form one Phonological Word. Semantically, however, the units do not combine.

(138) Se não chegar ...
    if NEG arrive-FUT.SBJV.3.SG
    /snāw ʃgar/
    ‘If (s)he does not arrive…’

**Form-based form**

6. *Agreement*
There is agreement in the nominal domain in Portuguese, as number and gender are copied from nouns to adjectives, demonstratives and relativizers (when preceded by prepositions) – all shown in (139).

(139) dez motiv-os pel-os quais ainda est-ás solteiro
ten reason-PL.M for-DEF.PL.M REL.PL still be-PRS.2.SG bachelor
‘ten reasons why you are still a bachelor’

There is no agreement in the verbal domain. Person and number of the Actor argument are always marked on the verb, but as the expression of an independent argument is optional, this is cross-reference and not agreement.

7. *Expletive elements*

There is no expletive pronoun in Portuguese. For instance in (140), it is not necessary to use an explicit argument at all – no semantically empty subject needs to be inserted.

Hutchinson & Lloyd (2003:149)

(140) Está a chover.
be.3.SG CONT rain
‘It is raining.’

8. *Grammatical gender*

As discussed previously, nouns are masculine or feminine in Portuguese, and trigger agreement accordingly. For animates, gender is semantically based (even though there are exceptions), but the masculine/feminine gender of inanimates is not semantic. In many cases, nouns ending in –o are masculine, nouns in –a are feminine (form-based classification), but there are again exceptions (cf. Hutchinson & Lloyd 2003:9ff.). Gender in Portuguese is hence non-transparent.
9. Sequence of tenses

Portuguese has a sequence of tenses rule, as can be concluded from comparing (141) and (142). The present tense form in the reported speech clause in (141) is shifted to a past tense in (142). This past tense is copied from the past tense in the main clause.

(141) Di-sse: “Estou cansado.”
    say-PST-3.SG be-1.SG tired
    ‘He said: “I’m tired.”’

(142) Di-sse que estava cansado.
    say-PST-3.SG that be-PST.IPFV.3.SG tired
    ‘He said that he was tired.’

10. Influence of complexity on word order

Movement of phrases is possible in Portuguese. It is for instance allowed to move a heavy subject to post-verbal position. However, in thetic statements the subject is obligatorily in post-verbal position, so that in this case it is not clear whether movement occurs because of this pragmatic consideration or because of complexity.

In any case, it is possible to move a relative clause to the right (as in example [137] above). This does not have a pragmatic or semantic effect, so I will assume here that placement is influenced by the weight of the relative clause.

11. Influence of complexity on function marking

Portuguese has many bound morphemes, e.g. the gender and number marking suffixes that appear in example (139). If an NP is marked for number and/or gender, this is indicated on the head noun and its dependents, rather than marking the phrase as a whole. I have not
seen any cases of phrase-marking in Portuguese. There is hence influence of complexity on function marking.

12. Syntactic functions

Portuguese has a semantic function Subject – pragmatic and semantic roles can be neutralized. The intransitive clause in (143) has an Actor argument, while the intransitive clause in (144) has an Undergoer. Yet, they have the same form.

(143) Ele corre.
    He run.3.sg
    ‘He runs.’

(144) Ele cai.
    he fall.3.sg
    ‘He falls.’

The Actor in (145), ela, and the Undergoer in (146), a, are not neutralized: they have a different form. However, the Undergoer in (145) and the Actor in (146) have the same form, o Paulo. For pronouns in transitive clauses, there is no neutralization of semantic role, but for nouns, there is.

Hutchinson & Lloyd (2003:45)
(145) Ela viu o Paulo.
    She see.PST.3.SG DEF P.
    ‘She saw Paulo.’

Constructed example
(146) O Paulo a viu.

Paulo saw her.

Since pragmatic and semantic information can be ignored in alignment, there is reason to postulate at least a syntactic function Subject.

13. (Morpho-)phonologically triggered alternations

There are some alternations in Portuguese. For instance, there can be insertion of glides, e.g. in (147b).

(147) a. Come! b. Come-o!

/ko’m/ /ko’mju/

eat.IMP eat.IMP-it

‘Eat!’ ‘Eat it!’

A possible analysis here is that there is an underlying stem-final vowel that is elided in (147). When the suffix is added in (147), the glide is inserted in between the vowels, giving ko’meju. Then, the stem-final vowel is elided.

A stem-final consonant can be deleted when an object suffix is added, which itself gets an extra l, e.g. seduz ‘seduct’ + -a ‘her’ > sedu-la.

Furthermore, voiceless consonants can get voicing when adjacent to a voiced consonant or vowel, e.g. dentes /dəᵗʃ/ ‘teeth’ > dentes brancos /dəᵗʃ браkuʃ/ ‘white teeth’ and dentes amarelos /də’tʃ ɐmɐ’reluʃ/ ‘yellow teeth’.

References

4.3 Gujarati

Gujarati is an Indo-Aryan language. It is spoken by 50 million people, of which the majority lives in India in Gujarat state. There is language contact with Marwari, Hindi, Marathi, and nowadays with English (Doctor 2004:1).

Redundancy

1. Apposition: Cross-reference

There is cross-reference in Gujarati. Verbs obligatorily inflect for tense, person and number (Doctor 2004:38). Arguments are also expressed independently, but not obligatorily: Gujarati is a pro-drop language (Grosz & Patel 2006). This means that if an argument is expressed independently (as in [148]), we are dealing with cross-reference.

Doctor (2004:41)

(148) Hū pəɖ-y-o ch-ū.
   1.SG fall-PST.PTCP-1.SG be-PRS.1.SG
   ‘I have fallen.’

Note that both the past participle and the auxiliary are marked for person and number – the double verbal marking is redundant as well.
2. **Multiple expression of semantic information**

- Gujarati does not exhibit negative concord (Cardoso 2009:304).
- Gujarati does not show plural concord, e.g. *tren kalak* ‘three hour’ (Cardona 1965:67).
- There is no semantic classification of nouns in Gujarati, hence there cannot be double expression of it. This feature does not apply.

**Domain disintegration**

3. **Fusion**

- There is cumulation of person, number and case in pronouns, e.g. *hū* ‘I.NOM’, *mane* ‘I.ERG’ (Doctor 2004:31). Person, number and tense are cumulated in referential markers on verbs (cf. example [148]).
- Stem alternation exists in the nominal domain. There are plural suffixes, but also some irregular plurals, e.g. *manes* ‘man.M.PL’ (Doctor 2004:27). Furthermore, there are nominal stems that alternate under the influence of case marking. For instance the stem *chokro* ‘boy’ becomes *chokra* when marked for case: *chokra*-*e* ‘boy-ERG’ (Doctor 2004:25).

  There is also stem alternation in the verbal domain. The base forms of the past participle and remote past participle are usually formed by adding respectively *–y* or *–el* to the verbal stem (Doctor 2004:38). Several verbs, however, show irregular behaviour, e.g. *pēs* ‘to get in’, *bēs* ‘to sit’ and *nas* ‘to flee’ get the past participle base forms *pēth*, *bēth* and *nath* (Doctor 2004:48).

  Causative marking triggers stem change as well. While most verbs get the causative suffix *–aw*, certain verbs undergo stem changes (Doctor 2004:51), e.g. a change of vowel (*mār* ‘to die’ > *mar* ‘to kill’), a change of a consonant (*phaṭ* > *phad*, ‘to tear’ > to make s.o. tear’), or both (*chut* > *chod*, ‘to loosen’ > to make s.o. loosen’).
4. *Discontinuity*

- There are no circumfixes in Gujarati (Doctor 2004:35).
- There are no infixes in Gujarati (Doctor 2004:35).
- I have not seen examples of argument raising in Gujarati.
- I have not seen examples of extraposition in Gujarati.

5. *Non-parallel alignment*

I have not seen examples of non-parallel alignment.

**Form-based form**

6. *Agreement*

There is agreement in Gujarati, as a subset of Gujarati adjectives inflects for the gender and number of the noun they modify (Doctor 2004:26), as illustrated in (149) - (152). The same is true for possessive pronouns (idem: 32) and demonstratives (idem: 33).

Doctor (2004:26)

(149) sar-o chokr-o
    good-M.SG child-M.SG
    ‘A good boy.’

Doctor (2004:27)

(150) sar-a manes
    good-M.PL man-M.PL
    ‘Good men.’
There is no agreement in the verbal domain. Verbs obligatorily inflect for tense, person and number (Doctor 2004:38), but expression of arguments lexically is optional (Doctor 2004:79). This hence involves cross-reference.

7. Expletive elements

There are no nominal expletives in Gujarati. For instance, in example (153), there is no semantically empty pronoun.

Doctor (2004:54)

(153) Peṭmā duːkh-e ch-e.
    stomach-LOC pain-PRS.3.SG be-PRS.3.SG
    ‘My stomach hurts.’ (Lit.: ‘It is paining in the stomach.’)

I have not seen examples of weather predicates in Gujarati, but I expect that no expletives will be used in that context.

8. Grammatical gender

There are three genders in Gujarati: masculine, feminine and neuter. Male animates have masculine gender and female animates have feminine gender, but many inanimates can
have masculine or feminine gender as well. The gender of a noun is hence not always predictable on the basis of its semantics – this is grammatical gender (Doctor 2004:23-24).

9. **Sequence of tenses**

There is no tense copying in Gujarati, cf. example (154). The past tense of the main verb is not copied to the embedded clause.

Doctor (2004:84)

(154) Məne lag-yū ke e av-še.

I.ERG feel-PST.3.SG COMP he.NOM come-FUT.3.SG

‘I felt that he will come.’

10. **Influence of complexity on word order**

I have not seen examples of influence of complexity of units on morphosyntactic placement in Gujarati.

11. **Influence of complexity on function marking**

Function marking in Gujarati is performed by means of bound morphemes. These bound morphemes typically appear in one position and on one or two lexeme categories – I consider them affixes. I have not seen examples of clitics – apparently Gujarati strongly favours head-marking over phrase-marking. There is influence of complexity on function marking.

12. **Syntactic functions**
Gujarati shows a complicated alignment system that involves syntactic functions. In the present tense, Gujarati has an accusative alignment pattern. In the past tense, the argument of an intransitive clause (example (155)) and the Undergoer of a transitive clause in the past tense (example (156)) are marked for absolutive case, while the Actor argument of a transitive clause is marked by ergative suffix -e. Furthermore, the Undergoer-argument of past tense transitive clauses determines gender marking on the verb, as shown in (156).

Doctor (2004:79)

(155) Chokr-o av-yo.
    boy-ABS come-PST.3.SG
    ‘The boy came.’

(156) Chokri-e laḍu khadh-o.
    girl-ERG sweet.ABS.N eat-PST.3.SG.N
    ‘The girl ate the sweet.’

Semantic role is ignored here (the Actor-argument in (155) and the Undergoer (156) are treated identically), hence there is at least a syntactic function Subject.

13. (Morpho-)phonologically triggered alternations

There are alternations in Gujarati. Some verbs undergo vowel alternations when they are marked with passive suffix –a, e.g. ap > ap-a ‘give > give-PASS’ (Doctor 2004:53).

---

15 My analysis differs here from Doctor (2004), who states that Gujarati has nominative, oblique and vocative case. He calls the construction in (11) an objective. However, basing myself on the examples given here, I will for now consider Gujarati to have an alignment split conditioned by tense – in the present tense, Gujarati exhibits accusative alignment, in the past tense an ergative system. The situation is in fact more complicated, as discussed by Patel (2007), but for reasons of space and time I will not go into this any further and present this simplified account. This simplification has no consequences for the transparency analysis.
A schwa in a stem can be deleted before a suffix starting with a vowel, but not when the suffix is the plural marker -o, e.g. \(a\ddot{a}s\) ‘laziness’+ -\(\ddot{u}\) ‘ADJ’ = \(a\ddot{l}s\ddot{u}\) ‘lazy’ (Mistry 1997:661).

Furthermore, an /a/ is reduced to a schwa before a suffix starting with a, e.g. \(kap\) ‘to cut’ + -a ‘pass’ = \(k\ddot{a}pa\) (Mistry 1997:662).

References