

The Nominal Domain in Santome

Nélia Alexandre¹ & Tjerk Hagemeijer²

1. Introduction

In what follows we will discuss the properties of the nominal domain in Santome (ST)³ focusing on the elements that modify the interpretation of the noun and their respective interaction. It will be argued that in ST there is no substantial evidence for a Determiner Phrase (DP) (cf. Longobardi 1994, 2003). Rather, definiteness in this language is a compositional feature obtained derivationally. Furthermore, we argue that specific marker (SP) *se*, which behaves like a clitic, is the core element of the nominal domain, anchoring the identifiability of the noun. In section 2 we deal with the descriptive properties of the nominal domain in ST and in section 3 we will provide a data analysis of our findings. A structural outline of our proposal is to be found in section 4.

2. Data description

This section separately discusses the different lexical items that may be hosted by the noun. For the sake of clarity, we have split the elements that operate on the noun into a prenominal and a postnominal class.

2.1. The left-hand of the head noun

2.1.1. Indefinite article *ũa* ‘a, an’

This item corresponds to the indefinite article and numeral ‘one’.

¹ Universidade Autónoma de Lisboa.

² Universidade de Lisboa, funded by doctoral grant BD/3159/2000 of the Foundation for Science and Technology and the European Structural Fund within the IIIrd Community Support Framework, Portugal.

³ Santome is the Portuguese-based Creole spoken on the island of S. Tomé in the Gulf of Guinea and is better known in previous work as São-Tomense.

- (1) a. *Ūa ome/bwê...*
 ‘a/one man/cow...’

2.1.2. Plural marker *inen*

Inen marks plurality on the noun and also corresponds to the 3rd person pronoun, a common feature among Creole languages. Nouns that are exclusively modified by *inen* (i.e., without any other material modifying the noun), as in (2-3), are scarce in our corpus.

- (2) *Ise sa depositu ku kwa ku inen blanku ka fla nê.*
 This is deposit with thing that 3PL white ASP speak of-3SG
 ‘This is the deposit, and so forth, white people (=the white) talk about.’
- (3) *Inen bunzu, inen tslôkô, inen vwadô: yô pixi ku ngê na ka kume*
 PL whelk PL sea-devil PL flying-fish many fish that people NEG ASP eat
fa
 NEG
 ‘Whelks, sea-devils, flying fishes: there are many fish species people don’t eat.’

These two examples show that *inen* basically conveys a pluralizing reading. The absence of other functions associated to *inen*, for example definiteness or specificity, is particularly clear in (3) where we are dealing with an enumeration of species in general. The plural marker shows a preference for nouns with the feature [+human], which we consider a consequence of its sensitivity to a semantic principle that takes [+human] as being more individuated than items that are [-human].⁴ This preference is also expected if we consider that *inen* is the standard 3rd person plural pronoun and, as such, still retains part of its pronominal features. Interestingly, *inen* is very often found in direct speech addressing a number of hearers with the meaning ‘you’, as illustrated in (4-5).

⁴ A very common exception to this is the examples in (i). We consider these examples crystallized forms, as follows from the translation.

(i) *inen kwa se // tudu inen kwa se*
 3PL thing SPEC // all 3PL thing SPEC
 ‘these things, this // all this, everything’

- (4) *Jina solo, jina inen migu xê dai en.*
 since sun since PL friend leave here EMPH
 ‘Since sunrise, since you friends left.’
- (5) *Inen ke mu, sa pingada fan.*
 3PL house mine is gun EMPH
 ‘My friends (=people at my place), it’s a gun!’

The referred asymmetries with respect to plural marking on [+/-human] noun are confirmed by dislocation tests. Pluralized [-human] items are typically co-indexed with a 3rd person singular pronoun (cf. 6), although there seems to be some variation among speakers and syntactic environments, whereas [+human] dislocated constituents are always co-indexed with a 3rd person plural pronoun (cf. 7).

- (6) *Inen zanela se, bô {fis’e/fis’inen/*fisa}.*
 PL window SP 2SG {close-3SG/close-3PL/close}
 ‘The windows (in question), you closed them.’
- (7) *Inen ome se ala, Zon {bê inen/*bê’lê/*bê}.*
 PL man SP there Zon {see-3PL/see-3SG/see}
 ‘The men (in question) overthere, Zon saw them.’

2.1.3. Quantifiers and numerals

As far as we know, the following quantifiers modify the noun in ST: *yô* ‘many’, *maxi montxi* ‘many’, *tudu* ‘all’, *kwakwali* ‘any’, *ũa dôsu* ‘some’, *ôtlô* ‘other’, *nyũa* ‘no N’, *kada* ‘every’, *pôkô* ‘few’. They occur typically in the leftmost position within the nominal domain and are mutually exclusive (cf. 8). Our corpus shows that especially *tudu* ‘all’ is commonly followed by plural marker *inen*, a solution that also exist for *yô* ‘many’, *maxi montxi* ‘many’ and *pôkô* ‘few’ (cf. 9).

- (8) {*yô / maxi / montxi / tudu / kwakwali / ũa dôsu / ôtlô / nyũa / kada / pôkô*} *bisu ...*
 ‘{All /many/some ...} bird(s) ...’
- (9) {*tudu / yô / maxi montxi / pôkô*} *inen ngé*
 {all, many, many, few} PL person

These facts about quantifiers suggest that they sit in the topmost specifier position in the nominal domain. In section 4 we will argue that, in the absence of other evidence, this is the specifier of the Number Phrase (Spec,NbP).

Numerals, on the other hand, may occur in different structural positions:

- (10) a. *Inen dōsu mosu se...*
 PL two boy SP
 ‘The two boys...’
 b. *Dōsu inen mosu se...*
 Two of the boys...’
 c. *Inen mosu se dōsu...*
 ‘These/those two boys...’

The a. example corresponds to the numeral’s canonical position within the nominal domain. The b. and c. examples correspond to specific partitive-focus readings. In section 4 we will argue that numerals are base-generated in different positions, but the b. reading favors the intuition that in this position *dōsu* ‘two’ also occurs in the quantifier slot. This explains why its co-occurrence with other quantifiers is precluded.

2.1.4. Augmentation/Diminution

Operations of augmentation/diminution take place to the immediate left of the noun they modify.

- (11) a. *ũa mina ke* ‘A small house.’
 a child house
 c. *memen vapô ope* ‘An enormous foot.’
 big boat foot

These modifying items are apparently derived from nouns entertaining a compounding relation with the head noun (contrasting with adjectives, which occur to the right of the noun).

This hypothesis is supported by the occurrence of specific marker *se*, a clitic that in these cases can only attach to the right of the head noun (cf. section 2.2.1).

- (12) *Memen (*se) vapô (*se) ope *(se).*
 ‘The very big foot in question.’

2.2. The right-hand of the head noun

2.2.1. Specific marker *se* ‘the, this, that, these’

This marker establishes deictic/anaphoric proximity⁵ by referring textually, situationally or inferentially to specific objects in the world that represent shared knowledge of speaker and hearer. *Se* always occurs to the immediate right of the noun and requires at least one of the following hosts:

i) Pronouns⁶

- (13) ... *punda non se na tê zêté doxi fa.*
 because 1PL SP NEG have olive-oil sweet NEG
 ‘...because we (in question) don’t have sweet olive oil.’

ii) Nouns

- (14) *N mêsê pa men jê mina awa se da anzu se* (Common)
 1SG want for mom fetch little water SP give baby SP
 ‘I want you (affective way to address a woman) to bring a bit of that water (in question) for the baby (in question).’
- (15) *Fernanji se di Mate Ngola.* (Proper)
 Ferdinand SP of Mateus Angolar
 ‘Ferdinand of Mateus Angolar.’
- (16) *Kuma non lentla setembru se en sa non sa ni tempu suba za.* (Time)
 as 1PL enter September SP EMPH be 1PL be in time rain already
 ‘Since it’s September, we are already in the rainy season.’

⁵ Santome exhibits two other, less commonly used anaphoric/deictic markers, *xi* ‘that (less proximate)’ and *sala* ‘that (distant)’. The nominalization of *se*, *xi* and *sala* yields respectively *ise/isaki* ‘this (here)’, *ixi* ‘that’ and *isala* ‘that’.

⁶ Note that etymologically the 2nd person plural (*i)nanse~(i)nense* is likely to have its origin in *inen+se* ‘they+se’, i.e. ‘you (pl.)’, establishing proximity with respect to the point of view of the speaker (cf. also examples (4-5)).

- (17) *San se ê, fe mu favôlô axi an.* (Vocative)
 miss SP EMPH do me favor such EMPH
 ‘Oh miss, could you do me a favor...’
- (18) *Montxi Makaku, liba se dai ôbô.*⁷ (Prepositional)
 Mount Monkey on-top SP here jungle
 ‘Mount Monkey, over there up in the jungle.’
- (19) *Kengê se ku fe mu favôlô?* (Interrogative)
 who SP KU do me favor
 ‘Who (of the persons in question) does me a favor?’

iii) Numerals

- (20) *Non tlêxi se so.*
 1PL three SP only
 ‘Just the three of us.’

In the light of the available data, we postulate the hypothesis that *se* is a clitic requiring a nominal host.⁸ The following syntactic tests corroborate this claim:

a) Adjacency requirement:

- (21) *Inen mosu *(se) dôsu *(se).*
 PL boy (SP) two (SP)
 ‘The two boys in question’

Compare (21) to examples (10a-c), where we have shown the mobility of numerals. It follows that, despite their mobility, numerals are crucially not able to split up a sequence of noun-specific marker.

b) *Se does not license ellipsis of its nominal host:*

- (22) *Zon paga *(mwala) se.*
 Zon pay woman SP
 ‘Zon payed the woman in question.’

⁷ Some syntactic tests (e.g. intransitivity) show that several prepositions in ST, among which *liba* ‘on top of’, behave as nominals.

⁸ In a very few exceptional cases of temporal constructions like in the example below, *se* appears to modify verbs. Given ST’s tendency to derive nouns from verbs by null affixation, it might well be the case that there is some reanalysis going on, since the aspectual and temporal modification are not morphologically realized on the verb itself.

- (i) *Ola san ska pali se, san ka glita ...*
 when lady ASP give.birth SE lady ASP scream
 ‘While she was in labour, the woman screamed ...’

Whereas plural marker *inen* and indefinite marker *ũa* allow for gapping of the noun, *se* does not and should therefore be considered a more grammaticalized item.

c) The specific marker *cannot be stranded (23b) or fronted (23c)*

- (23) a. *Mwala se, Zon pag'e.*
 woman SP Zon pay-3SG
 'The woman, Zon payed her.'
 b. **Mwala, Zon {pag'e/paga} se.*
 woman Zon {pay-3SG/pay} SP
 c. **Se so Zon paga mwala [-].*
 SP FOC Zon pay woman

d) *Compounding*

Compounds form another piece of evidence. The reanalysis of two X^0 as a single one predicts correctly that it should not have internal structure. Hence, the clitic can only attach to the right of the newly coined word in the examples (24a-d).

- (24) a. *[[[boka] [sabi]] se]* 'The keyhole in question'
 Mouth key SP
 b. *[[[kwa] [kume]] se]* 'The food in question'
 Thing eat SP
 c. *[[[kota] [bega]] se]* 'The last-born in question'
 Cut belly SP
 d. *[[[tempu] [glavana]] se]* 'The dry season in question'
 Weather dry.season SP

The relevant contrast is with (25a-d), which may resemble compounds due to juxtaposition of nouns without any overt mediating case marker. The crucial difference with the examples in (24) is that the clitic, as expected, may modify the head noun, its complement, or even both if one would like to.

- (25) a. *[[opo] se] [(di) glavana]]* 'The dry season's dust'
 dust SP (of) dry.season
 b. *[[Fenanji] se] [(di) Mate Ngola]]* 'Ferdinand from Mateus Angolar'
 Ferdinand SP (of) Mateus Angolar
 c. *[[klonveson] se] [doxi-doxi]]* 'The very nice talking question'
 talk SP sweet.sweet
 d. *[[tudu ngê] [flêgêja] se]]* 'Everyone of the district in question'
 Every person district SP

2.2.2. Modifiers

There is nothing special to be said about modifiers, except for the fact that they are adjoined categories that always occur to the right of the head obeying the order possessive>adjective>relative clause, as in (26) below.

(26) *vinpema mu doxi ku n bēbē.*

Palm-wine my sweet that I drink

‘My sweet palm wine that I drank.’

2.3. Bare nouns

In addition to the modifying lexical items discussed so far, a minimal account of the nominal domain in ST would be very incomplete without dealing with bare noun phrases (BNPs). These nouns, characterized by what we call a zero morpheme, are multifunctional and complex in the sense that they license a wide range of interpretations that are highly dependent upon syntactic (e.g. subject/object position), semantic (e.g. verbal aspect) and discursive anchoring (e.g. new/old information) which we cannot discuss in this venue for reasons of space. We will limit ourselves to providing some examples of their main uses.

In subject (cf. 27-28) and object (cf. 29) position, [+human], BNPs are more readily associated to singular readings, but this a mere tendency and not a rule (cf. discussion in section 2.1.2). In addition to the singular/plural readings, BNPs in object position can be definite singulars/plurals or indefinite plurals (cf. 29).

(27) *Piskadô ba ple.*

man go beach

‘The {fisherman/(fishermen)} went to the beach.’

(28) *Kabla ba matu.*

goat go bush

‘The {goats/goat} went into the bushes.’

(29) *Mwala se pya ome/kabla.*

woman SP look-at men/goat

‘The woman looked at {men, goats, the man/men, the goat/goats}.’

Furthermore, BNPs in ST license generic (cf. 30), existential (cf. 31) and characterizing (cf. 32) readings.

(30) *Maji vinpema ka fe ome mali.*

but palm-wine ASP do man bad

‘But palm wine isn’t good for {men/a man}.’

(31) *Mwala sa maxi montxi dōkê ome ni Santome.*

woman be more many than man in Santome

‘There are more women than men in Santome.’

(32) *N sa ome ku ka kume fluta muntu.*

1SG be man who ASP eat breadfruit a lot

‘I’m the type of man that eats a lot of breadfruit.’

3. Data analysis

3.1. Identifiability

Lambrecht (1994) distinguishes between (i) presupposed proposition: (some) shared knowledge between speaker and hearer, and (ii) asserted proposition: speaker’s representation at the time of utterance. The article system of Creole languages was already an important aspect of Bickerton’s (1981) Bioprogram. This author claimed that these languages exhibit a threefold determiner system based on the cognitive elements presupposed-specific (definite article), asserted-specific (indefinite article) and nonspecific (zero marker, i.e. BNPs).

Although this system may work for other languages, it does not for ST, since we argue that this language has no definite articles proper. Rather, definiteness is a compositional feature that obtains derivationally, whereas the semantically realized core consists of specificity (cf. following sections). Furthermore, zero markers in ST commonly do range into the specific domain. Example (33) exemplifies the relevance of discourse anchoring with respect to the determiner system.

(33) *Avia ũa sungê ku mina sun. Sun se sa ve ketekete. Mina se sa*
was a man with child man man SP be old IDEOPH child SP be

ai, sun ka sam’e...

here man ASP call-3SG

‘Once upon a time there was a man (formal) with his child. He was very old. He called his child who was close by.’

Under normal circumstances (here at the start of a folk story), the indefinite reading gives rise to a specific one that anaphorically recovers this information with *se*. As a consequence, the information is assumed to be familiar to the hearer/speaker and licenses a BNP with a specific and singular interpretation. Table 1 illustrates these findings.

Table 1. Information structure.

	NEW	OLD	
		ANCHORING	ANCHORED
Singular	Ũa {kabla/Ome}	{Kabla/Ome} se	{Kabla / Ome}
Plural	{Kabla / Ome}	Inen {kabla/ome} se	{Kabla / Ome}

It follows from the table that in ST BNPs are underspecified items with respect to identifiability.

3.2. Lexicalized functional items

The data have shown that ST has three nuclear lexicalized functional items and a non-lexical item in the nominal domain, namely *ũa*, *inen*, *se* and the zero morpheme, each with the following semantic core feature:

Table 2. Core semantic features of lexicalized functional items

	Plural	Specific
ũa	-	
Inen	+	
Se		+
Zero morpheme	α	α

Hence, the items with lexical realization have a single specified feature corresponding to its basic use (singular marker, plural marker, specific marker). BNP ‘exhibit’ a zero morpheme

with no lexical content we assume to correspond to unspecified features. The derivation will assign a positive or negative value to this zero morpheme.

Clusters of these items occur frequently and assemble these features to derive the interpretation.

Table 3. Clusters of lexicalized functional items

	Plural	Specific
Ūa + N + se	-	+
Inen + N + se	+	+
Zero morpheme	α	α

The following examples illustrate the findings of table 2 and 3. The relevant features of table 2 are transferred to the cluster. BNP remain of course unspecified and have to pick up their features in the course of derivation.

- (34) *A pali ūa kabalu blanku d'e. D'e ūa kabalu se*
 IMP prepare a horse white give-3SG give-3SG a horse SP

blanku.

white

‘They prepared a white horse for him. They gave him a white horse.’

- (35) *Kum'ê sa inen ome se ku ka tëndê kwa mwala ka fla,*
 as-3SG be PL man SP who ASP listen thing wife ASP say

ê na xê fa.

3SG NEG go-out NEG

‘Because he is one of these men that listens to his wife, he didn’t go out.’

Examples (34-35) nicely show the relevance of discourse in assigning a specific meaning to the noun. The anaphoric function of *se* clearly comes about.

4. Tendencies and structural representation

4.1. Tendencies

The data above lead us to the following findings with respect to the nominal domain in ST:

- (i) The left periphery (pre-nominal) in ST is essentially restricted to ‘quantity-denoting’ modifiers (numerals, augmentation, diminution, quantifiers and number marking elements).
- (ii) The right periphery (post-nominal) is used for specificity operations on the noun (cf. section 2.2). The arguments that support this hypothesis are (a) the presence of *se* and of other modifying constituents (e.g. possessives, adjectives and relative clauses) and (b) the specific focus reading of numerals in the post-nominal domain.
- (iii) Although it may yield syntactic and semantic nuances, the semantic feature of ‘animacy’ doesn’t seem to be a core feature of the nominal domain in the same sense as specificity and number. (cf. section 2.1.2).
- (iv) BNPs show unspecified features. The wide range of interpretations of BNPs seems to be linked to the early stages of creolization. Especially if we assume an initial *pidgin* (pre-creole) stage in which nouns must have been the essence of identifiability (a less functional stage), it is not surprising that BNP still exhibit a whole range of specific and generic readings which were only restricted by more functional material when the *pidgin* developed into a full-blown language. The number of BNPs in early Sranan, for instance, was much higher than it is nowadays (cf. Bruyn 1994, 1995).

4.2. Directionality of the DP and its structural representation

Taking into account considerations (i-iv) of the section above, we propose a tree structure that subsumes the following two fundamental aspects:

- (a) *se* as a clitic, which is the lexical representation of ‘specificity’, a central functional category (SpP) to this system; and

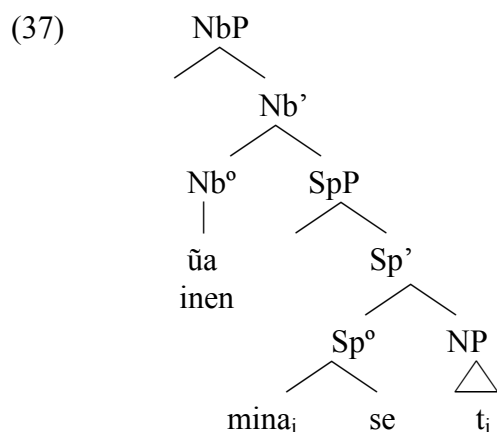
- (b) the importance of ‘number’, since the data have shown that there is no evidence for a Determiner node and that the structure of the nominal domain in this language only reaches as far as the Number node.

In the light of the available data, we argue that the DP structure in Santome is right-branching based on the following evidence:

- (i) absence of postpositions;
- (ii) relative clauses, adjectives and possessives occur to the right of the noun;
- (iii) compounds are head-initial;
- (iv) economy principles of representation.

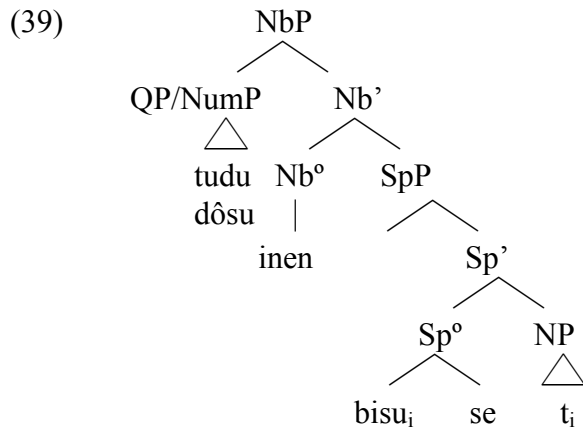
The data description now leads us to the tree structures that are outlined below.

- (36) a. *Ũa mina se* ‘a child in question’
 b. *Inen mina se* ‘the children in question’



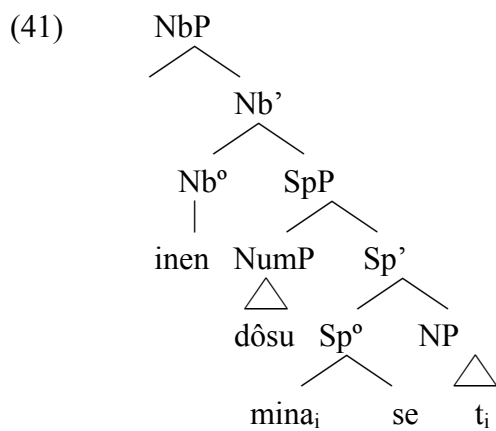
Based upon (36a-b), the noun *mina* in (37) moves through left-adjunction to the head of the Specific Phrase (Sp°) to check its feature. *Ũa* and *inen* are the referred lexical functional items that occupy the head of Number (Nb°), in accordance with their indefinite and plural nature.

- (38) a. *Tudu inen bisu se* ‘all the birds in question’
 b. *Dôsu inen bisu se* ‘two of the birds in question’



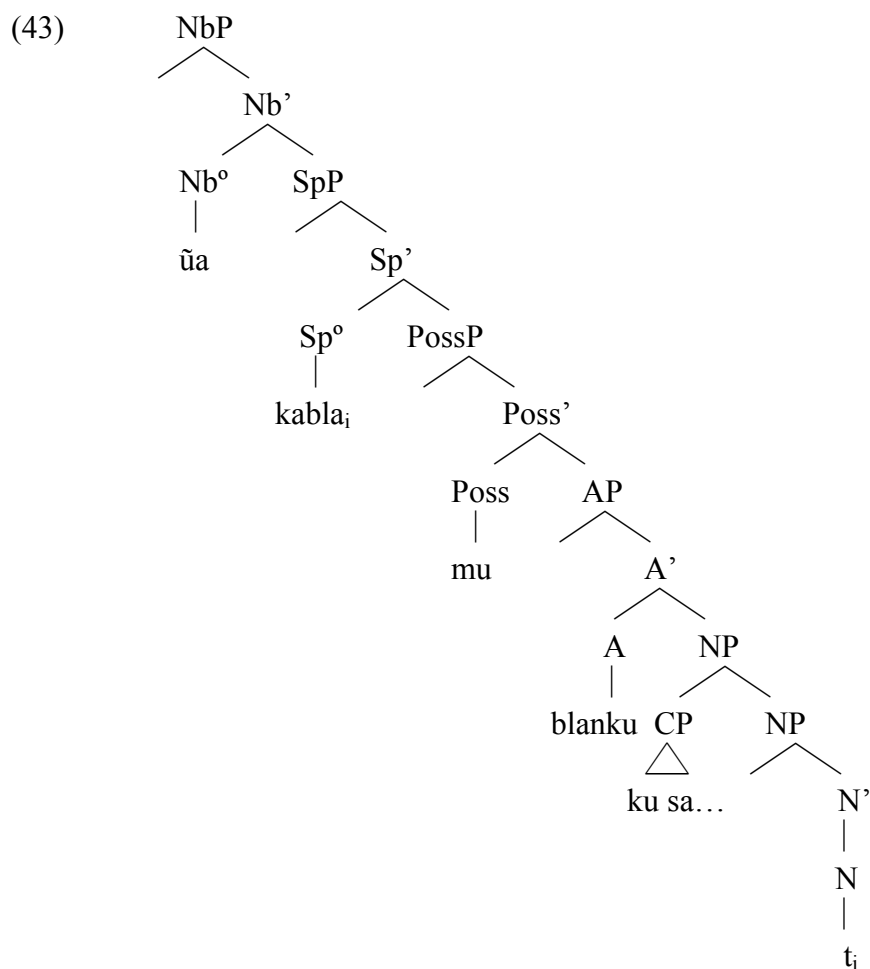
In (39), representing examples (38a-b), we propose that the Quantifier Phrase (QP) *tudu* is base-generated in the specifier of the NbP (Spec,NbP), an A'-position, where it has scope over the full sentence. When numerals precede *inen* and trigger the (specific) partitive-focus reading (cf. section 2.1.3.), the Numeral Phrase (NumP) has to be base-generated in Spec,NbP to scope over the whole sentence, as in (10b). Note that numerals are ‘quantity denoting elements’ and thus it is a natural assumption to relate them to a [number] feature.

- (40) *Inen dôsu mina se* ‘the two children in question’



Sentence (40) and its corresponding tree (41) exhibits the canonical word order found with numerals (cf. (10a) above). Although *dôsu* ‘two’ and numerals in general express number, we assume that the NumP is base-generated in Spec,SpP, where it focuses on and licenses the specific reading.⁹

- (42) *Ûa kabla mu blanku ku sa kinte*
 One goat mine white REL be garden
 ‘A white goat of mine that’s in the garden’



Sentence (42) shows that modifiers occur structurally to the right of the noun, where they are subcategorized by the functional category Sp°, which licenses their specific reading. The

⁹ The reanalysis of *ũa dôsu* ‘some’ as a quantifier supports this view.

surface word order is obtained by raising the noun *kabla* ‘goat’ to Sp^o where it checks its features.

5. Final remarks

We have tried to show that specificity is the core of ST nominal domain system and therefore directly linked to the NP. The nature of number and quantification require these domains to be in a higher position where they have scope over specificity. Additionally, the high degree of grammaticalization and functionality of *se*, and determiners in general, make a good point for its nuclear status with respect to the noun. In a certain respect, the nominal domain structure in Santome resembles this language’s verbal domain, where the aspect marker shows the highest degree of grammaticalization and functionality and behaves like a clitic hosted by the verb. Our findings for Santome suggest that, in absence of positive evidence, not all languages should receive a classical DP analysis, since the D(eterminer) feature may be obtained compositionally (in a similar fashion, the evidence from Santome arguably does not support the Split-I hypothesis in the verbal domain). This is why we believe that the structural representation we propose for ST still mirrors the primitives of the nominal domain that was expanded upon over time.

References

- Bickerton, Derek (1981). *Roots of Language*, Anna Arbor: Karoma.
- Bruyn, Adrienne (1994). “Noun Phrases”, in *Pidgins and Creoles: An Introduction*, pp. 259-269. Amsterdam/Philadelphia: John Benjamins.
- Bruyn, Adrienne (1995). *Grammaticalization in Creoles: the Development of Determiners and Relative Clauses in Sranan*, PhD, Amsterdam: IFOTT.
- Lambrecht, Knud (1994). *Information Structure and Sentence Form – Topic, Focus and the Mental Representation of Discourse Referents*, Cambridge: Cambridge.
- Longobardi, Giuseppe (1994). “Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form”, in *Linguistic Inquiry*, vol. 25, n° 4, pp. 609-665, Massachusetts: MIT.
- Longobardi, Giuseppe (2003). “The structure of DPs: some principles, parameters, and problems”, in BALTIN, Mark & COLLINS, Chris (Eds.), *The Handbook of Contemporary Syntactic Theory*, pp. 562-604, Mass.: Blackwell Publ.