Subject Marking in Somali

JACQUELINE LECARME

CNRS - L.L.A.OR. (Langues et Langage de l'Afrique Orientale) Valbonne - Sophia Antipolis, France

Introduction

The subject-object opposition characteristic of other Cushitic languages with case marking is particularly well defined in Somali, especially in the Northern dialect. In this apparently very complex system, Nominative Case Marking (tonal and segmental markings) interferes with a number of other grammatical factors, mostly concerned with the syntactic organization of the sentence and the pragmatic functions syntactically marked in Somali, topic and focus.

The main purpose of the present paper is to define, though informally, some general principles we assume to be present in the Nominative Case Assignment and Case Marking mechanism. Starting from data that are especially well described in the literature, the analysis proper centers on *waa* and *má* structures, then on *baa* and *waxa* structures. It is supposed that the conclusions can be extended to all other structures involving nominative case marking, and it is hoped that they will provide a better understanding of Somali syntax.

Generalities about Data

To sum up well-known data, Somali as other East Cushitic languages has an unmarked objective (or absolutive) case and a marked nominative case. The manifestation of case marking on the subject essentially consists in lowering of morphological tone (Hyman 1981),[1] this morphological tone being predicted on the basis of gender and declension class (Andrzejewski 1964, 1979). In addition, several feminine nouns, as *doqón* in (1), suffix *-i* (non derived feminine nouns with a high (H) tone on the last syllable, those belonging to Andrzejewski's fourth declension or Hyman's class D1). Lowering of tone affects the definitives which suffix the noun, but the definite article *-ku/-tu* is exempt from accent reduction (2):

(1) *Doqoni sokeeyé má ahá*
    [+nom]
    Fools are not relatives /Andrzejewski (1964: 56)

(2) *Soomalídu béen wày sheegtaa*
    [+nom]
    Somalis tell lies /Andrzejewski (1956: 112).

---

[1] More specificaly, according to this author, in syntactic reduction of morphological accents which cannot receive any tonal specification.

As a general rule, a subject *out of focus* is marked nominative ([+nom]). Lowering of tone suffices to establish the subject/object differentiation as shown in (3) and (4):

(3) a- *Nin wuu yimid*
      [+nom]
      A man came
   b- *Nín waan arkay*
      I saw a man

(4) a- *Nínkii libàax má dilín*
      [+nom]
      The man did not kill a lion
   b- *Nínkíi libaax má dilín*
            [+nom]
      A lion did not kill the man /Andrzejewski (1964: 21)

There is no nominative agreement, between a subject and its attribute for example, as *sokeeyé* in (1). If the NP is a complex NP, a coordinate structure for example as in (5), *subject marking* applies only once on the rightmost element in the NP. If the NP is a relative clause, the subject antecedent is never marked [+nom], and subject marking applies at the end of the relative clause, what seems to be a marked property of Somali among other Cushitic languages. Example (6) shows tone alternation between a relative clause in subject position out of focus and a relative clause in focus position using the same paradigm of the verb. Infinitives are treated as Hyman's D1 feminine nouns and show a *-i* suffix as in (7):

(5) a- *Àar iyo góol má qabteen?*
      Did they catch a male lion and a lioness?
   b- *Àar iyo gooli má qabteen?*
            [+nom]
      Did a male lion and a lioness catch him? /Andrzejewski (1975: 142)

(6) [*Nín* [*béen yaqaan*]] *waa* [*nín* [*fál yaqáan*]]
   NP S   [+nom]   NP   S
     A man who knows how to lie knows magic /Andrzejewski (1964: 62)

(7) *Nín [aan wáayo barani] waxbá má barán*
              [+nom]
      A man who has not learnt « why » has not learnt anything /Andrzejewski (1956: 117)

In fact, we find many alterations in the mechanism of subject marking, some of them being due to dialectal variations: as already observed by Antinucci and Puglielli (1984), subject marking applies less regularly in the Coastal dialect than in the Northern dialect. Saeed (1984: 205) observes that there is a tendency for long NPs not to be subject marked. Yet sentential *in*-complements when functioning as subjects are never marked for case, contrary to relative clause in the same position, as it can be observed from negative examples as (9):

(9) a- *Ninkii [aan dhimani] wuu soo noqday*
             [+nom]
      The man who did not die came back

   b- [*Iinaanuu* $\left\{ \begin{array}{c} dhiman \\ *dhimani \end{array} \right\}$] *waa hubàal*
       S   [+nom]

It is certain that he did not die

In any case, the mechanism in its whole appears to obey extremely logic and rigorous principles, as illustrated in (10), where the most embedded relative clause *sácu daaqáy*, in spite of the fact that it contains a nominative subject and a verb in a fully inflected paradigm, is marked *objective* (H tone) according to the grammatical function of its antecedent *béerta*, while the less embedded relative clause is marked [+nom] as aspected:

> (10) *Márkaasàa nínkíi* [*béerta* [*sácu daaqáy*] *lahaa*] *suldáankíi ú sóo*
> *ashtakooday* [+nom]
>> Then the man who owned the garden which the cow had grazed complained to the Sultan /Andrzejewski (1964: 43)

Thus we can expect that those principles we want to precise and make generalizations about are simple ones.

The Notion of Subject

To define the general principle which is the basis for [+nom] Case Assignment, it is first necessary to ask how to define the notion of subject. In earlier generative works as in traditional grammar, it has been assumed that grammatical functions (as *subject of*) which serve as a basis for Case Assignment, are determined in terms of syntactic configuration: the subject is thus defined by its syntactic position in a sentence.[2] But there are languages where grammatical functions are not represented configurationally (Somali, which is usually called a « free word order language », is a possible candidate) and for a sake of generalization, Chomsky (1981, 1982) developed a theory of Case where the notion of *government* plays a central role: subjects are nominative when they agree with the matrix verb - more precisely, with its inflections. The tensed verb contains a complex of features called agreement (AGR), which is the governing element that assigns Case on subject NPs. Thus it could be said that in Somali, the so-called objective (or absolute) unmarked case is assigned under government both by transitive verbs and prepositional particles (among other possible « case assigners »), and that the [+nom] case is assigned under government by the agreement features contained in the tensed verb.

If the verb is untensed, « infinitive », the subject cannot be normally marked for case. What is called « infinitive » in Somali, however, has quite different properties. Infinitives, which occur only in negative contexts (past negative contexts in main sentences), do have [+nom] subjects, while being untensed forms. In negative examples (11-12), tensed forms in present context are opposed to untensed forms in past context:

> (11) *Maroodígu takárta kú joogtá má arkó*
> [+nom]
>> The elephant does not see the camel fly which sits on him ... /Andrzejewski (1964: 77)

> (12) *Anigu ma arag shilkii dayaaradda*
> [+nom]
>> I didn't see the plane crash /Saeed (1984: 173)

---

[2] In standard generative-transformational grammar, the notion of subject is defined purely syntactically, with respect to a hierchical position in a tree structure.

These facts suggest that the neutralization of the AGR features contained in the verb is a marked property of Somali (among other Cushitic languages) and at an earlier stage of the derivation the Somali infinitives do contain the AGR features responsible for [+nom] Case assignment. What actually may be compared to infinitive in most other languages is the so-called « restrictive paradigm »,[3] which shares the same property of not having an overt subject in surface structure.

It is briefly argued in the present paper that in Somali main sentences, [+nom] Case is marked on the *topic* NP [4] (not the *subject* NP), and that in dependent clauses or parts of main sentences that are not in focus position, Case Marking is on the *subject* and can be assigned in terms of syntactic configuration. This is obviously due to the fact that in main sentences focus markers govern word order and may define grammatical functions, whereas in dependent clauses, which contain no grammatical marker of focus, grammatical relations depend on syntactic organization and fix order of elements as in configurational languages.

Waa and *má* Structures

Consider the structures (13 to 16) where *waa* is assumed to be an assertive focus marker and *má* some negation of it, in which the Case-assignment mechanism appears to be essentially the same:

(13) *Ninku$_i$ wuu$_i$ shaqeynayaa*
$\qquad$ [+nom]
$\qquad$ The man is working /Saeed (1984: 46)

(14) *Warqadi ma iman maanta*
$\qquad$ [+nom]
$\qquad$ A letter did not arrive today /Saeed (1984: 111)

(15) *Way$_i$ dhaadhèeryihiin daaráhaasu$_i$*
$\qquad\qquad\qquad\qquad\qquad$ [+nom]
$\qquad$ Those houses are high /Andrzejewski (1956: 119)

(16) *Wax ma aqoon ninku*
$\qquad\qquad\qquad$ [+nom]
$\qquad$ That man was completely ignorant /Hassan (1974: 117)

It can be observed that the [+nom] Case can be assigned only to a NP which is external to the *waa* or *má* sentence. If a [+nom] NP appears at the right of the VP, as in (15), the subject position is in most cases filled by a weak pronoun with the same index. As (17) shows, any [+nom] NP in this position should produce an agrammatical structure:

(17) **Waa ninku shaqeynayaa*

$\qquad$ **Ma warqadi iman*

---

[3] On the origin of the « extensive » vs. « restrictive » distinction, see Andrzejewski (1956: 125). The Restrictive Paradigm is invariable, except for a feminine third singular and a first plural person feature specification.

[4] *Topic* NP is to be understood here in a purely syntactic sense, independent of pragmatic notions, i.e. a NP in topic position, base generated (left or right dislocated) outside of the normal structure of the sentence (cf. Lecarme, 1984).

The subject position in *waa* (or *má*) sentences can be filled only by an element which cannot be tone bearing, as impersonal pronoun *la* in (18):

(18) *Ínankíí yaráa na wàa la qabtay*
     And the young boy was captured /Andrzejewski (1964: 81)

These facts suggest that the NP that is assigned the [+nom] Case is in topical position, i.e. left or right dislocated out of the structure of the sentence, and inherits the [+nom] Case from the pronoun, if marked coreferential with it at an earlier stage of the derivation (call it D (deep)-structure).

There are, however, many examples such as (19), where there is only one NP in the sentence, the one that is assumed here to be in topical position, without any coreferential pronoun:

(19) *Ninku waa tegay*
     [+nom]
     The man went

This must be related to a special property of Somali (among other languages as Italian or Arabic, but not English or French), which allow pronominal subjects to remain unexpressed, and is called the « *null subject parameter* » or « *pro-drop parameter* » (cf. Chomsky 1981, 1982). (As we shall see later, Somali has a mixed system, permitting the subject to drop in certain syntactic configurations, not in others). This property is usually connected with the richness of inflexion, but as it appears from many languages, this is not always the case. In Somali, there is no overt subject in the so-called « independet » paradigm (20), where the subject is in some sense contained in the verbal inflexion, but even in focus constructions as (21 to 25) subjects can be dropped even if the verb is negative and thus appears without any overt inflexion:

(20) *Garánnay*
     We have understood /Andrzejewski (1956: 113)

(21) *Waa* [ ] *keenay*
          NP
     He brought it /Andrzejewski (1975: 158

(22) *Má* [ ] *keenay?*
          NP
     Did he bring it? /Andrzejewski (1975: 158)

(23) *Má* [ ] *keenín*
          NP
     He did not bring it /Andrzejewski (1975: 158)

(24) *Xamar ma* [ ] *tago*
                NP
     I do not go to Mogadiscio

(25) *Waa* [pro$_i$] *yimid* [*Cali$_i$*]
          NP   TOP   [+nom]
     Ali came

We may thus conclude from these examples that the subject must precede the verb in the structure, and, if we take as a general principle that any sentence must have a subject at every level of syntactic structure (Chomsky 1981: 198), that there is an empty pronominal (call it *pro*) in place of the subject as figured in (25) which is assigned [+nom] Case at D-Structure, and that the NP that is marked [+nom] at S (surface)-structure is in topic position (left or right dislocated) in (14, 16, 19, 25).

Baa and *waxa* Structures

Consider now structures with *baa* and *waxa*, which are assumed here to be grammatical pre- and post-verbal focus markers (often contrastive as for pragmatic function, but there is not necessarily close correlation). The general principle holds that a *subject out of focus* is marked [+nom]. The hypothesis that the [+nom] Case is marked on the topic, not the subject NP, also holds in a number of cases, those illustrated by examples (26-28 and 29-31), where the left-dislocated NP in topic position is obligatorily linked to a clitic pronoun of the same index, which fills the subject position:

(26) [*Nínkani$_i$*] *awr buu$_i$ keenay*
     TOP [+nom]
    This man brought a he-camel /Andrzejewski (1975: 163)

(27) *Awr buu$_i$ keenay* [*nínkani$_i$*]
                 TOP [+nom]
    This man brought a he-camel /Andrzejewski (1975: 163)

(28) *Awr buu$_i$* [*ninkani$_i$*] *keenay*
           TOP [+nom]
    This man brought a he-camel /Andrzejewski (1975: 164)

(29) [*Duqsigu$_i$*] *wuxuu$_i$ ka mid yahay ...*
     TOP [+nom]
    The fly is one of ... /Saeed (1984: 207)

(30) [*Wasirka warshadaha oo shir furay*]$_i$ *wuxuu$_i$ yiri ...*
    TOP   [+nom]
    Il ministro dell'Industria che ha aperto la' riunione ha detto ... /Gebert (1981: 76)

(31) *Wuxuu$_i$ yidhi* [*turjubaanku$_i$*] « ... »
               TOP [+nom]
    The interpreter said « ... »

There are examples, however, that are never found in *waa* or *má* structures, where the subject position can be filled by a [+nom] NP alone (without any pronoun):

(32) *Aad baa roobku u dhacay*
             [+nom]
    Much rain fell

(33) *Súbixíi bàa ilaaládíi yóo gaajaysani degmádíi ká luudday*
                          [+nom]
    In the morning the scouts trudged hungry out of the village /
                                    Andrzejewski (1960: 104)

Such examples seem to be less common in *waxa* structures than in *baa* structures. Yet (34b) is to be considered as grammatical as (34a):

(34) a- [*Naagtu$_i$*] *waxay$_i$ keentay lacagtii*
      TOP [+nom]
    b- *Waxa naagtu keentay lacagtii*
              [+nom]

In those examples, it cannot be said that the NP marked [+nom] is topic. [+nom] Case Marking obey another principle we must define.

It can be observed that in the part of the sentence that is not in focus position (on the right of *baa*) the subject position *must* be occupied by an overt NP, and

that the phenomenon we called *pro-drop* in *waa* and *má* structure is not possible (35):

(35) **awr baa [   ] keenay*
　　　　　　　　NP
　　　He brought a he-camel

In structures where the *subject* is focalized by *baa*, which is shown in ex (36), the empty element we noted [*e*] cannot be *pro*, since the paradigm of the verb is restrictive. In fact, not only is *pro-drop* impossible in these structures, but, as (37-38) show, the order of elements SOV is strictly configurational and cannot be modified:

(36) *Cáli bàa [e] yimí*
　　　　　　　NP
　　　Ali came

(37) *Márkaasàa*
$$\left\{ \begin{array}{c} \textit{libáaxii nínkíi} \\ [+\text{nom}] \\ \textit{*nínkíi libáaxii} \\ [+\text{nom}] \end{array} \right\}$$
dilay

　　　Then the lion killed the man /Andrzejewski (1964: 77)

(38) *Góor dhów bàa*
$$\left\{ \begin{array}{cc} \textit{náagtii} & \textit{awrkíi} \\ [+\text{nom}] & \\ \textit{* awrkíi} & \textit{náagtii} \end{array} \right\}$$
keentay
　　　　　　　　　　　　　　　　　　　[+nom]

　　　Soon afterwards the wife brought the he-camel /Andrzejewski (1964: 76)

This situation is exactly the same in dependant clauses, *in*-complements and relative clauses in examples (39-41), where neither *pro-drop*, nor free inversion of the elements is possible:

(39) *Warqaddii [ay$_i$ noo soo dirtay [Shammado Liibaan]$_i$]*
　　　　　　　　　　　　　　　　TOP
　　　*Warqaddii noo soo dirtay Shammado Liibaan
　　　The letter sent us by Shammado Liibaan/Hassan (1974:118)

(40) *Wáxa muuqatá* [
$$\left\{ \begin{array}{c} \textit{ínay yimaaddeen} \\ \textit{* in yimaaddeen} \end{array} \right\}$$
]
　　　　　　　　S
　　　It appears that they have come

(41) *Kuma ayay jecelyihiin*
$$\left\{ \begin{array}{c} \textit{[in Cali iyada guursanayo]?} \\ \text{S} \quad\quad [+\text{nom}] \\ \textit{* in iyada Cali guursanayo} \\ [+\text{nom}] \end{array} \right\}$$

　　　Who do they prefer that Ali marries? /Saeed (1984: 146)

The fixed order of the elements SOV that may be supposed to be the basic order in Somali shows up only in sentences or parts of sentences that escape to Focus government. In those structures, the [+nom] Case is assigned configurationally, as in languages where grammatical functions are determined in terms of syntactic configuration and specific order of the elements.

Before concluding, I would like to precise that in spite of the fact that some

correlation has been made on the basis of the [+ nom] Case Assignment mechanism, between sentences or parts of sentences that share the common property not to be in focus position, it does not imply that they are the same structures, or that the part of the sentence that is on the right of *baa* is in some sense a relative clause.

As often emphasized in the literature, there is a close relation, which may have a diachronic explanation, between *baa* or *waxa* structures, and relative clauses. But as it seems to me, relative clauses are structurally much more constrained than focus structures in the sense that they must contain an element which is bound to the antecedent (subject or object) of the relative clause. *Baa* on the contrary may focus almost any category (as an adverbial clause in (42)), which is not necessarily bound to any element in the following sentence, and which develops freely in non-focus position:

(42) *Márkíi dádkii kala wada seexseexdáy bàa afádìi* [*ínanka galabnimádíi waraabisay*] *debéddíi ú sóo baxday ôo is tidhi*, « ... »
    [+ *nom*]
    When the people all went to sleep, the woman who gave water to the boy in the evening came out into the open and said to herself « ... » /Andrzejewski (1964: 82)

I don't know whether the analysis proposed in (36) for *baa*-structures is the right one (it would be the subject of another paper), but here I assume that there is no lexical NP source for *waxa* and that *baa* and *waxa* are grammatical Focus markers until there is a better explanation.

Conclusion

From the analysis of [+ nom] Case Assignment and Case Marking, we have made the assumption that in some contexts the relevant notion is *topic*, while in other contexts the notion of *subject* is basic. One may thus conclude that the order of constituents is not free in Somali, that grammatical functions, where not constrained by grammatical focus or topic,[5] are determined in part by syntactic configuration, and that the two systems coexist in the syntactic structure of Somali.

References

Andrzejewski, B.W. 1956, « Accentual pattern in verbal forms in the Isaaq dialect of Somali », *Bulletin of the School of Oriental and African Studies (BSOAS)*, XVIII/I: 103-129.

Andrzejewski, B.W. 1964, *The declensions of Somali nouns*, London, School of Oriental and African Studies (S.O.A.S.).

Andrzejewski, B.W. 1975, « The role of indicator particles in Somali », *Afro Asiatic linguistics*, 1, 6: 1-69.

Andrzejewski, B.W. 1979, « The case system in Somali », London, S.O.A.S.

Antinucci, F. and A. Puglielli 1984, « Relative clause construction in Somali: a comparison between the northern and the coastal dialects », in J. Bynon (ed.), *Current progress in Afro-Asiatic linguistics*, Amsterdam, John Benjamins.

Chomsky, N. 1981, *Lectures on government and binding*, Dordrecht, Foris.

Chomsky, N. 1982, *Some concepts and consequences of the theory of government and binding*, Cambridge, Mass., MIT Press.

---

[5] « Grammatical » is opposed here to « pragmatic », though we assume that some elements that relate to pragmatics belong to this level of grammar called Logical Form (LF)-Representation.

Gebert, L. 1981, « Il sintagma nominale », in A. Puglielli (ed.), *Sintassi della lingua Somala, Studi Somali 2*, Roma, MAE.

Hassan Sheikh Mumin 1974, *Leopard among the women* (*Shabeel naagood*), Oxford, Oxford University Press.

Hyman, L.M. 1981, « Tonal accent in Somali », *Studies in African Linguistics*, 12, 2.

Lecarme, J. 1984, « On Somali complement constructions », in T. Labahn (ed.), *Proceedings of the 2nd International Congress of Somali Studies*, Vol. 1, Hamburg, Helmust Buske.

Saeed, J.I. 1984, *The syntax of Focus and Topic in Somali*, Hamburg, Helmust Buske.