

Annarita Puglielli

Somali language studies past and future lexicography in the foreground

In 1972 Somali became a national language with an official writing system: a shift then from oral to written language. At the same time it became the language of instruction. All this determined an important development in the research on Somali language and culture and if we look at bibliographies, the amount of research that has developed in the last forty years³⁹ is evident.

The new tasks the language had to perform showed immediately the need for 'development', necessary to cover the areas that did not belong to the previous functions the language was used to cover. Many of the terminologies needed to cover different subjects in the school curricula had to be selected and/or coined. The need for terminologies therefore brought to the attention the need for lexicographic work and dictionary writing but, as lexical research started, there was an increasing awareness that dictionary writing implies a deep knowledge of the whole system of the language being described.

In fact a dictionary is not just a simple collection of words; an adequate description of the lexicon of a language implies the knowledge of the whole system of the language in object. This is because a language is a coherent system where different levels of organization interact (from phonology to text/discourse), and because when we speak we do not use single words but sentences (i.e. combinations of them).

³⁹ The research on Somali language and culture has been promoted by institutions such as the Somali Academy of Arts and Sciences and the Somali National University in Mogadishu, by projects such as Studi Somali within the Technical Cooperation of the Italian Ministry of Foreign Affairs realized by Rome University: 'La Sapienza' and then 'Roma Tre', and of course many individual researchers both in Somalia and all over the world.

A lexical entry in a dictionary therefore must contain not only a description of the meaning of the word in semantic terms but also grammatical information (i.e. classification of words in terms of parts of speech as noun, verb etc.) and morphological information (conjugation for verbs, gender and number for nouns etc.), and finally phonological/phonetic information. This is necessary to understand how words must be used in 'context', that is when combined in sentences.

We shall try to explore two specific subjects with reference to dictionary writing for Somali:

- Words and their grammatical categorization with related morphological aspects
- The relationship between lexical aspects and syntax: i.e. the correlation between the lexicon of the language and its typological characteristics in a syntactic perspective.

Our reflections are based on our lexical research work started in 1978 which has given the following results:

- Dizionario Somalo Italiano (Agostini , Ciise M. Siyaad e Puglielli 1985)
- Dizionario Italiano Somalo (Puglielli 1998)
- Qaamuuska Af-Soomaaliga (Puglielli e Cabdalla Cumar Mansur 2012).

1. Words and grammatical categories

Words are made up of sound and corresponding meaning, and meaning is the union of semantic characteristics as well as functional properties. Consider the following examples:

- (1) *Magoolku wuu biqilay.*
'the bud sprouted up'
- (2) *Cawsku wuu magoolay.*
'the grass germinated'

In (1) *magool* is a noun and is the subject of the sentence, while in (2) *magool* is a verb and is the predicate. In other words the lexical entry *magool*, the meaning of which is 'blossom', can assume a different grammatical category and therefore have two different functions when inserted in a sentence.

We give one more example to illustrate this point:

Aadaan1 ml. 1. *Ogeysiin cod dheer leh oo mu'addinku ugu yeero muslimiinta waqtiyada salaadda.* 2. *Ci' dheer oo neef riyo ahi u sameeyo marka uu xanuunsado*
(1. the singing inviting to pray, 2. sheep groan.....)

Aadaan2 f.mgl 1. *Cod dheer ugu yeriid musliimiinta waqtiyada salaadda.* .2 *Kor u qaylin neef riyo ah xanuun dartiis.*

(1.call for prayer, 2. groan loudly used for sheep).

(3) *Aadaanka codkiisu waa dheeryahay.*

'The muezzin voice/ the prayer is loud'

(4) *Cali baa masaajidka ka aadaamay.*

'Cali called for prayer in the mosque'

Therefore the same word with one sound and one meaning, can assume a different function once it is inserted in 'context'; it is used as NOUN in (1) and (3) and it is used as VERB in (2) and (4).

This situation is very common in Somali therefore the question may rise: should we have in our dictionary a single entry i.e. just plain roots, without grammatical categorization? A root then once inserted in the context of a specific sentence will assume the category N(oun) or V(erb) according to its function and consequently all the other related morphological specifications. Nouns will co-occur with determiners and will have gender and number morphology, whereas verbs will incorporate tense/aspect morphology and person and number agreement with the subject according to the syntactic rules of the language.

This choice has not been adopted traditionally in Somali lexicography and has not been chosen in our dictionaries for several reasons: the most obvious being that at this stage it seemed that the double categorization and two different entries would

make the use of the dictionary more accessible for a larger public.

Therefore the entries for *aadaan* in our dictionary are two different entries:

Aadaan1 ml. 1. *Ogeysiin cod dheer leh oo mu'addinku ugu yeero muslimiinta waqtiyada salaadda.* 2 *Ci' dheer oo neef riyo ahi u sameeyo marka uu xanuunsado*

Aadaan2 f.mg1 1. *Cod dheer ugu yeriid musliimiinta waqtiyada salaadda.* 2. *Kor u qaylin neef riyo ah xanuun dartiis.*

Let us now consider the inventory of grammatical categories introduced in the Dictionary: Noun (*magac*), Verb (*fal*), preposition (*horyaale*), clitic pronoun (*dhimman*), particles (*qurub*), conjunction (*xiriiriye*), adverb (*falkaab*), focus (*diirad*). Given the limited time we have, we will concentrate only on MAJOR lexical categories i.e. **Nouns** and **Verbs**; as you can see **Adjectives** are not included. The choice is based on the analysis proposed by Andrzejewski (1969), according to which words like *waanagsan* 'good', *weyn* 'big', *yar* 'small', *dheer* 'tall', *fudud* 'light, easy' etc. because of their syntactic behaviour must be considered 'hybrid verbs' and assigned to the 4th conj where also other 'stative verbs' are assigned (for ex. *ah, leh, la*', and derived verbs like *furan, maqan* etc.).

This issue of the need of the category Adjective for Somali has been debated by several researchers⁴⁰, and experts have taken different positions on the matter, but I think that if we base our classification on an accurate analysis of the whole system of the language our choice must be maintained.

In fact what we try to capture in our description is that a noun modifier in Somali is necessarily a relative clause, not just an adjective:

(5) *Nin Ingriis ah baan la kulmay.*

'I met an English man'

(6) *Qof weyn waa abwaan.*

'an old person is wise'

40 Banti, G. (1988), Ajello, R. and A Puglielli (1988) and others.

(7) *Qof weynaa baa dhintay.*

'an old person died'

(8) *Rida cad baa luntay.*

'the white goat was lost'

(9) *Shalay baa ridii caddayd la qashay.*

'yesterday the white goat was slaughtered'

It is not the case to go into a detailed description of these examples, but notice that in (5) the N *nin* is modified by a heedless relative clause where *Ingriis ah* is the predicate with the verb in the reduced paradigm.⁴¹

In (6) *weyn* is what is left in the surface from the reduced paradigm of *ahaansho* present tense and in (7) the ending *-aa* attached to *weyn* is what is left of the past tense of the verb *to be*. In (8) and (9) *cad* is the predicate of the relative clause in the present tense of the reduced paradigm and *caddayd* is the same in the past tense.

These are the syntactic reasons for considering those lexical items that in Italian or English have been classified as adjectives, verbs of the 4th conj. in Somali; still it remains that from a semantic point of view they share the characteristics of expressing 'STATES or PROPERTIES' (just like adjectives), that are attributed to nouns.

We now turn our attention to the specifications introduced in our dictionaries for Nouns and Verbs.

1.1. Nouns

As for Nouns the choice to introduce in their entry morphological information such as gender and number was determined, once again, by the intention to make explicit features that are relevant for a correct construction in a sentence.

41 For a description of relative clause structure see Antinucci F. (1981), Puglielli A.e F. Antinucci, (1980), Puglielli A. e M. Frascarelli, 2005

If we consider, for example the noun *inan* we find two identical entries, one for the masculine noun (*inan*) and one for the feminine one (*inán*), since the difference in stress is not registered in writing. But we know that the information about gender is needed, given that in a sentence the verb must agree with its subject and verbs have different endings for third person masculine and third person feminine. Therefore we have:

(10) *inan* (masculine) - *inanku hilib buu cunay*
'the boy ate meat'

(11) *inan* (feminine) - *inantu hilib bay cuntay*
'the girl ate meat'

As for Number, we introduced the forms for the plural right from our first dictionary, for two reasons:

A) Rules for plural formation are rather complex in Somali with different morphological mechanisms at work such as addition of inflectional endings with related polarity effects (i.e. change in gender), change of stress, etc.⁴² Here are some examples:

(12)

Singular	plural
<i>Buug</i> m..	<i>buugag</i> m.
<i>War</i> m.	<i>warar</i> m.
<i>shabeel</i> m.	<i>shabeello</i> f.
<i>magac</i> m.	<i>magacyo</i> f.
<i>tuke</i> m.	<i>tukayaal</i> f.
<i>habar</i> f.	<i>habro</i> m.
<i>gacan</i> f.	<i>gacmo</i> m.
<i>mind</i> f.	<i>mindiy</i> m.
<i>qaaddo</i> f.	<i>qaaddooyin</i> m.

⁴² For a detailed description of plural formation rules see, Puglielli A. e Ciise Mohamed Siyaad (1984).

B) The nouns that can be pluralized are not necessarily the same in all languages. So, for example in Somali, the nouns related to plant names are subcategorized as collectives and are not used in the plural form. Examples are: *yaaq*, *tiintiin*, *qurac* etc. that are used only in the singular.

More information about subcategorization of nouns appeared as necessary information to be included, in particular for a subclass of words that we named **predicative nouns**. Their behaviour has been accurately described by Ajello (1984). Examples of these nouns are: *irtid* 'reckless', *ceyrin* 'unripe, uncooked', *fircoon* 'cruel', *teneg* 'stupid' etc.

But let us examine their use in context:

(13) *Nin irtid ah baa yimid.*
'a reckless man has come'

(14) *Cali waa irtid.*
'C. is reckless'

(15) ?? *Irtid waa yimid.*

(16) ? *Irtidkii waa yimid.*

In (13) *irtid* has the function of the predicate of a relative clause that modifies *nin*, and in (14) is the main predicate of the sentence. In (15), where that noun is used as subject, the sentence though possible is completely unclear (as signalled by the 2 question marks). The sentence becomes more acceptable in (16) where the presence of the anaphoric article *-kii* allows the identification of the referent of the noun; in fact the anaphoric article can be used only if the person is already present in the context of discourse i.e. has been already mentioned.

A consequence of the peculiar distribution of these nouns is their behaviour in terms of agreement. So in a sentence with a plural subject a noun like *macallin* used as predicate must be used in the plural, as shown in (17) and (18), that is ungrammatical:

- (17) *Cali iyo Cabdalla waa macallimiin.*(pl.)
 'Cali and Cabdalla are teachers'
 (18) * *Cali iyo Cabdalla waa macallin* (sing.)

While a predicative noun as *washensi* must be used in the singular

- (19) *Cali iyo Ahmed waa washensi.*(sing.)
 'Cali and Ahmed are rude'

We must conclude therefore that the subclassification we propose helps us to understand the different syntactic behaviour shown in (17-19) and therefore tells us how to use these words in context. As a consequence this information is useful – if not indispensable – in a dictionary.

1.2. Verbs

The subclassification of verbs is also based on their syntactic structure, i.e. the number and type of argument Noun Phrases (NP) they require in a sentence on the basis of their meaning. Verbs are traditionally subclassified in intransitive and transitive depending on their construction with one or two NP argument. Here are some examples:

- (20)
 Vint. – 1 argument: *dhoof* 'leave', *soco* 'walk', *qosol* 'laugh'...
Cali waa dhoofay.
 'Cali travelled'
- (21)
 Vtr. – 2 arguments: *cun* 'eat', *akhri* 'read', *qor* 'write'...
Cali cambe wuu cunay.
 'Cali ate a mango'
Caasha warqadda way qortay.
 'Caasha wrote the letter'

But there are verbs in Somali that can occur with 3 NP arguments without the addition of any extra element (i.e. a preposition). Those are the di-transitive verbs (*labagudbe*) such as:

- (22)
 1° cnj.: *Duldhig* 'put something over something else'
tus 'show something or someone to someone'
 2° cnj.: *Dhansii* 'make someone drink something'
fahansii 'make someone understand something'
 3° cnj.: *Daaqso* 'make something eat something (1°obj.animals, 2° obj. Grass)
 4° cnj. ---- there are no di-transitive verbs that belong to this group, and this is not surprising since, as we know, the verbs of the 4th conjugation are STATIVE verbs and there are few transitive ones. It does not seem to be possible from a semantic point of view for a verb of this class to be di-transitive.

A last observation on verbs in Somali. Once we understand the interaction between the different levels of organization i.e. the interconnections between meaning, derivational morphology and inflection we are not surprised by the following quantitative data taken from our database:

conjugation	I	II	III	IV
V intr.	2506	574	1178	1993
V tr.	1535	2343	934	42
V di-tr.	11	80	7	--

The number of intransitive and transitive verbs across first and second conjugation is symmetrical. This can be explained in terms of the derivational morphology mechanisms of Somali verbs: by adding to an intransitive root of a 1st conjugation verb a suffix *-i/-is* we derive a transitive verb of the second conjugation:

(23)

Vroot intr. I + *is/in* = Vtr II

as in.

buux in *koobku waa buuxay* 'the glass was full'*buuxi* in *Cali koobka waa buuxiyay* 'Cali filled the glass'.

An adequate description of morphological mechanisms is therefore a prerequisite for lexical work.

2. Relationship between lexical aspects and syntax

In describing sub-categorization of verbs we have somehow started to explore the relation between lexical meaning, semantic structure and morphology (in particular derivational morphology).

The last aspect on which we would like to spend a few words now is how the typological characteristics of the language – Somali in our case - are relevant for lexicography: this is because of the relationship between syntactic structure of sentences and semantic structure of verbs. In fact, as we know a minimal structure for a sentence is made up of a predicate (V) and the arguments (NPs) required by the semantic grid of the verb.

The syntactic structure of a sentence in Somali is characterized by the presence of a Verbal Complex (VC) that is the verb plus a clitic cluster in which the NP arguments related to the verb are realized as clitic pronouns and disposed with other elements such as prepositions in a templatic structure (Puglielli 1981):

Impersonal	one	Preps	Obj cl	deictics	some ad-	Verb
Subj. cl. <i>la</i>	Obj.cl.	(1/2)	(poss.pro.)	<i>soo/sii</i> etc.	verbial	

Table 1

Now let us consider the following sentences:

- (24) *Axmed baa guri-gii nooga* [(= Ø + na + u + Ø + ka)⁴³]
 Axmed fm house-an ocl3s-ocl1pl +for+ocl3s+from
soo qaaday buugga
 take.pst book -the
 'Axmed took the book from home for us'

- (25) *Xaawo baa igu [(i+ku)] kaa aamintay*
 Xaawo fm ocl1sg-to you (lit.: your) entrust.pst
 'Xaawo entrusted me to you'

In (24) and (25) the clitic pronouns present in the Verbal Complex (VC) are the elements that tell us the different function of the NPs (direct object, object of a preposition) while the full NPs are in a dislocated position to the left and /or right of the verb. In these sentences all NP are resumed by a clitic pronoun.

But let us consider now the following sentence:

- (26) *Annag[oo guri-ga fadhina] bay (*na) yimadeen*
 we-prt house-det stay foc.cls3.3pl (1pl) come.prs-3pl
 'They came while we were at home'
 [lit.: we, who were at home, they came]

In (26) the NP *annagoo* cannot be resumed in the VC (as shown by the * in front of the clitic pronoun *na* that cannot be present in that sentence) and this means that the NP *annaga* does not belong to the core structure of the sentence and is rather an additional element (an adjunct or adverbial).

Therefore not every dislocated NP is resumed by a clitic pronoun in the VC; in fact only those that are part of the semantic grid of the verb are. Therefore we can understand the semantic grid of a verb by putting it in a sentence and observing which of the NPs present in the sentence is resumed by a clitic pronoun. All this can be relevant for a good and complete definition of verbs in a dictionary where the description of the meaning of a verb must contain both its lexical content and the number and type of arguments that that lexical content involve in the construction of the meaning of a sentence.

43 Notice that the object clitic third person is morphologically zero, i.e. not realized (see Puglielli 1981).

Another observation relevant for the description of verb argument structure, is that the morphological mechanisms of derivation of verbs in Somali determine new verbs with new meanings and related changes in argument structure. Here is an example:

-- the verb *buux* 'to be full' is a 1st conjugation intransitive verb, therefore a one argument verb with a 'Stative' meaning:

(26) *koobku waa buuxaa* 'the glass is full'

-- the verb *buuxi* (*buux* + *-is*) 'to fill' is a 2nd conjugation transitive verb, with two arguments and its meaning is 'Causative':

(27) *Cali koobka buu buuxiyay* 'Cali filled the glass'

-- the verb *buuxso* (*buux+is+an*) 'to fill something for himself' is a 3rd conjugation transitive verb, with two arguments, and its meaning is 'Causative + Autobenefactive'

(28) *Cali koobka buu buuxsaday* 'Cali filled the glass for himself'

-- Finally if we add the suffix *-am* to *buuxi* we obtain a verb with passive meaning where the agent cannot be expressed:

(29) *Koobku waa buuxsamay* (*buux+is+am+ay*)
glass-det-Nom DECL be full-CAUS-STAT-PST3SG.M
'The glass was filled up'

In fact a sentence like

(30) **koobku waa ka buuxsamay Cali*

cannot be interpreted as the glass was filled up *by Cali*; it could only be accepted, eventually, with the interpretation *for Cali*.

Once again we conclude that an adequate analysis of the derivational mechanisms of the language will be of great help for an adequate description of the lexical items to be introduced in a dictionary and their meanings.

3. The future

The great impulse that research on Somali language and culture has had in the last 40 years has certainly produced a lot of knowledge about it, but there is still a lot to discover. Thinking about Somali language there are still plenty of phenomena that need to be thoroughly described at all levels of linguistic description. We will mention only some in order to illustrate what we mean.

Starting from the description of the sound system and rules in Somali, we should know more about vowel harmony rules, an area still not described, as well as all aspects of intonational patterns and their relation to different pragmatic functions, and we could certainly add more topics.

Moving to other aspects, more relevant for lexicography, one topic that certainly needs further investigation is that of Verbal Nouns.

It is very systematic for each verb to find two corresponding verbal nouns, one with the same form of the verb characterized by masculine gender, and one derived with a derivational suffix - different for different conjugations - characterized by feminine gender. Here is one example for each conjugation:

	Verbal	verbal noun 1 masc.	verbal noun 2 fem.
1st conj.	dhadhan 'have taste'	dhadhan(ka)	dhadhamid(da)
2 nd conj.	dhadhami 'taste sth.'	dhadhamis(ka)	dhadhamin(ta)
3 rd conj.	dhadhansad 'taste sth.'	dhadhansad(ka)	dhadhansasho(ta) ¹
4 th conj.	weyn 'big'	weyn(ka) ²	weynaan(ta)

From a theoretical point of view this system appears to be redundant with two nominal forms for each verb, since apparently there seems to be no semantic relevant difference between the two forms of verbal nouns. If this is so, we should look for other differences; we should try to identify contexts where one can occur and the other cannot. This cannot be easily obtained by asking native speakers, but it could probably be obtained by the systematic and extensive scanning of texts, and this is certainly a topic for future research.

Coming now more specifically to lexicography let us make some consideration about future development. We have now a very good basis, considering the monolingual dictionaries recently produced, and we should start from there if we want to make progress rapidly with the intent not only to increase the number of words defined, but also to work towards a better coverage of all semantic areas. In other words we should start planning actions focused on specific topics and/or lexical areas to enlarge their coverage.

What we should work for, from now on, is a better connection between groups of researchers working on Somali lexicography, and modern technologies make it possible. Different groups could plan and develop their work in a complementary way and this would result in a more rapid development for Somali lexicography.

References

- Abdalla Omar Mansur e A. Puglielli, 1999, Barashada Naxwaha Af Soomaaliga, Londra, Haan Publisher.
- Agostini F., Ciise M. Siyaad e A. Puglielli (a cura di), 1985, Dizionario Somalo Italiano, La Casa del Libro, Roma.
- Ajello R.,1984, 'Substantives as predicates in Somali' in T. Labahn (ed.) Proceedings of the II International Congress of Somali Studies ,vol.1,Hamburg, Helmut Buske Verlag.
- Ajello R e A. Puglielli 1988, "More on 'hybrid verbs' in Somali", in M. Bachhaus-Gerst and F. Serzisko (eds.), *Cushitic-Omotic : Papers from the International Symposium on Cushitic and Omotic Languages*, Hamburg, Helmut Buske Verlag.
- Andrzejewski B.W., 1969, 'Some observation on hybrid verbs in Somali', *African Language Studies*, 10:47-89.
- Banti G.,1988, "Adjectives' in East Cuschitic' in M. Bachhaus-Gerst and F. Serzisko (eds.), *Cushitic-Omotic : Papers from the International Symposium on Cushitic and Omotic Languages*, Hamburg, Helmut Buske Verlag,.
- Puglielli A.e F. Antinucci, 1980 "The Syntax of Indicator Particles in Somali.

Part one: Relative Clause Construction", in *Afroasiatic Linguistics*, vol.VII, n.3.

- Puglielli A., 1981,'La struttura della frase semplice' in A.Puglielli (a cura di) *Sintassi della lingua somala, Studi Somali2*, Roma, MAE.
- Puglielli A., 1984."La derivazione nominale in Somalo", in A. Puglielli (a cura di) *Aspetti morfologici, lessicali e della focalizzazione, Studi Somali 5*, Roma, MAE.
- Puglielli A..e Ciise Mohamed Siyaad, 1984 'La flessione del nome' in Puglielli A. (a cura di) *Aspetti morfologici, lessicali e della focalizzazione, Studi Somali 5*, Roma, MAE.
- Puglielli A. (a cura di), 1998 Dizionario Italiano Somalo, Roma, Carocci, (II edizione 2011) Puglielli A. e M.Frascarelli, 2005 "A comparative analysis of restrictive and appositive relative clauses in Cushitic languages", in L. Brugè, G. Giusti, N. Munaro, W. Schweikert, and G. Turano (eds.), *Contribution to the IGG XXX*, 279-303, Venezia, Cafoscarina.
- Puglielli A. e Cabdalla Cumar Mansur (a cura di), 2012, Qaamuuska Af-Soomaaliga, Roma, Roma TrE-Press,. <http://hdl.handle.net/2307/720>.