

Annarita Puglielli and Francesca Bitocchi

QAAMUUSKA AF SOOMAALIGA
THE GREAT SOMALI MOOLINGUAL
DICTIONARY

1. Background

Now that work on the first edition of the *Qaamuuska Af Soomaaliga* (Dictionary of the Somali Language) is in its final stages, it is worth reporting on what has been done so far by examining the various phases of the project.

Lexicographical work was one of the central aims of our research from the very beginning, being part of the systematic description of Somali that we had embarked upon. Indeed, the adoption of Somali (officially transcribed in 1972) as the language of instruction showed that there was a need for the language to cover fields that had hitherto not been part of normal contexts of use.

When we commenced our work, the only existing monolingual dictionary was Yaasiin Cusmaan Keenadiid's dictionary, a seminal work for Somali lexicography,¹ which was published in 1976 and comprised approximately 3,500 words.

Over the years our project has evolved through several stages, moving towards the production of a 'comprehensive' monolingual dictionary, but, at the same time, also towards the creation of a lexical database that could be easily consulted and updated. This research is of particular importance since these data were the outcome of the first systematic and large-scale work on the language. In the future it is expected that the lexical databases can be expanded and specific research undertaken to include different areas that at the moment are scarcely present for reasons which are more than evident.

Let us now go rapidly through the different phases of our lexical project in creating the *Qaamuuska Af Soomaaliga* (QAS).

I. Our lexical work began with the production of the Somali-Italian Dictionary. The reasons for this choice were, as we have stated on previous occasions, the following:

- a) the definitions for the lexical entries were implicit rather than explicit. If we had started with the monolingual dictionary we would have probably had to face the need to coin new terms even before we had developed enough competence in Somali to be able to be of help to the native speakers involved in the work and who were at the same time being trained on the job;
- b) Somali had been officially transcribed only five years before our work started; this meant there was a certain amount of instability in the writing system;
- c) there was the need for Italian and Somali researchers to become acquainted with each other on the job in order to build up a solid and positive working relationship.

We started with the collection of all the relevant existing lexicographical sources. These were systematically checked with our informants and Somali experts, and then transcribed according to the official orthography. We then proceeded with the scanning of texts that had been published since the application of the official writing system in order to find words and different usages that were not recorded in our sources.² The result of this work was the *Dizionario Somalo Italiano* (approximately 30,000 lemmas) published in 1985.

II. The completion of the bilingual dictionary required the subsequent compilation of the Italian-Somali Dictionary which was produced mainly in Italy with the assistance of bilingual Somali experts. In this case the input language had to be Italian and the selection of entries was made on the basis of criteria such as frequency etc.

The Somali Italian dictionary was then reversed (through

computer elaboration), thus producing correspondences between Italian and Somali items with the Italian word as input. Evidently, this was only the initial input for the writing of the lexical entries since the final entry was the result of the work and skills of the various people involved in the compilation of the dictionary. The Italian-Somali dictionary was published in 1998.

III. Work on the QAS started as soon as the *Dizionario Somalo Italiano* was completed and, in this case too, the work involved several researchers and was split into various activities. These could be considered sub-projects due to the fact that different choices had to be performed given the different semantic fields that were being covered. The next paragraph will describe the aims and methods adopted for the compilation of the monolingual dictionary.

2. Phases

The work to be carried out was, on the one hand, to provide definitions for all the lexical entries in the bilingual dictionary, and on the other, to expand the lexical database towards an objective of 50,000 words. The work was carried out in both directions right from the start, and while one group of experts was writing the definitions for lemmas present in our database, different methods for collecting new words were determined.

We will briefly describe the various methods adopted for different lexical areas.

- a) Regarding the 'Humanities', we proceeded with the scanning of texts, articles and essays and any words that were not present in our corpus were added and defined. Other terms deriving from research into specific fields – for example ethno-musicology – were collected by the experts in the field and handed over to the dictionary team in order to have them included.
- b) For terms relating to all fields necessary for school text-

books – whose revision appeared to be an urgent need since books written by different authors used different terms – we worked alongside a Committee of the Curriculum Development Centre of the Ministry of Instruction since they were deemed responsible for the revision of the textbooks. For each subject – History, Geography, Mathematics, Physics, Biology, Chemistry – we chose a number of basic terms that we considered necessary (ranging from 150 to 300 for each area). They decided which words would be used in Somali for these terms and wrote the definitions for the dictionary. These same terms would then be used in the revision of the textbooks, the intention being uniformity of terminology.

- c) For terms relating to plants and animals, research started with the collection of bibliographical material produced by researchers and experts in the field and a bibliography of over 300 publications was compiled. The principle works were consulted and data relative to vernacular names as well as data concerning their classifications were put on a separate database. Clearly, the plan was also to check all the collected data through field work. There were obvious reasons for this: our materials had been collected at different times and by different authors in different places and therefore the corpus contained, for example, different names supposedly for the same animal or plant. The reasons for this ranged from the erroneous transcription of the name provided by informants to possible variations due to local differences.
- d) In order to collect data – both from a linguistic and an ethnological point of view – relating to traditional arts and crafts we prepared questionnaires to be used for field work in different regions of the country.

Predictably, this collection and/or checking of data through further field work was not possible as the civil war started in 1990. Therefore, some of the areas we intended to cover systematically and on scientifically solid grounds, will unfortunately be barely present in the current version of the QAS.

3. *The format*

The important thing to stress is that the present lexical corpus, whose origin and development we have described, and on which the first edition of the *Qaamuuska Af Soomaaliga* is based, is now a database of approximately 50,000 lemmas that can be consulted, expanded and questioned. This is the result of successive developments of the computational applications to our initial corpus.

From the initial computer format of the lexical entries in the Somali-Italian dictionary that we used for several computational elaborations³, we first expanded the lexical bases by adding additional entries through a system of semi-automatic comparison and at the same time decided to move to a new format – the Lexical Database – where we elaborated entries with all the relevant information. For each lexical entry we have: the grammatical category (both in Somali and in Italian), the inflection (plurals for nouns and the different paradigmatic forms for verbs), a semantic classification, the source, the Somali definition, the Italian translation and the English translation. The last two are not present for all the Somali words, while Italian equivalents are present for all entries deriving from the Somali-Italian Dictionary, but not for all the entries from other sources. Naturally the aim is to complete these at a future date.

Information relative to other aspects is also available, but will not be discussed here.

This is an example of a lexical entry:

GEEL

Lemma	1	geel
CtGrS	2	m. l. u.
CtGrI	3	n. m. coll. (con art. geela)
Fless	4	(-lal, mdh.)
CodFl	5	103+++36
CodSe	6	cum zoo
Var	7	
Fonte	8	DSI
Scol	9	N
TradI	10	Cammello.
Con.sin	11	Geelal g.:
Esprl	12	Più mandrie appartenenti a diverse persone.
Con.sin	13	Neef g. ah, tulus g. ah:
Esprl	14	Un cammello.
DefSo	15	Xoolaa la dhaqdo midka ugu weyn, lehna hal kurusama laba iyo cag ballaaran, harraadka u adkaysta, sidkiisuna yahay laba iyo toban bilood, oo looga faa'iideys; cad, caano, saan subag iyo gaadiidsi.
DefE	16	Camel.

This lexical database can be questioned for each individual entry or field.

It is therefore evident that the lexical data base is a flexible research instrument which will not only allow for our preliminary stage, i.e. the production of the monolingual dictionary, but also for future developments. These can be foreseen both in terms of the enlargement of the monolingual dictionary and in terms of a possible expansion of the Somali-Italian Dictionary.

4. *The present state and future prospects*

We are now completing the final revision of the QAS and the intention is to make it available for consultation on the Internet. Our objective is to encourage Somali speakers to participate in a forum with their observations, comments and so

on. Even if only a small part of the information we obtain is included in the first edition of the dictionary, it is worth obtaining people's reactions and, potentially, their suggestions to add to our database. Even when the first edition of the *Qaamuuska Af Soomaaliga* is published, we certainly will not consider that our work is complete.

In fact, we hope that our work can proceed towards expanding the QAS, drawing upon data which are to a certain extent already available since they can be derived from various existing sources. These are:

- Words and expressions that can be derived from cross-checking the Somali definitions in the Italian-Somali Dictionary with the entries in our lexical database;
- Additions from the scanning of literary texts (novels etc.);
- A corpus of oral language (90 hrs.) obtained from recording conversation among the elders of the Academy of Arts and Sciences in Moqdishu in which themes relative to traditional culture are discussed;
- Consulting internet sites in Somali which will certainly uncover neologisms as well as additional meanings for words already present in the database.

Evidently, the QAS will be the initial result of a lexical project that we hope will continue and be continued by Somali and Italian researchers.

In concluding this report, we would like to mention all the people who since 1985 have contributed in some way to this project. There have been many people involved, with different roles, at various periods and for various lengths of time; some of them are still alive in our memory, even after they have disappeared, or died, in the civil war.

The members of the Scientific Committee:

Annarita Puglielli, Ahmed Abdullahi Ahmed, Abdalla Omar Mansur, Ciise Maxamed Siyaad, Ahmed Cartan Xaange, Maryan Farax Warsame, Dahabo Farax Xasan, Cabdi Daahir Afey.

Contributors:

Ahmed Muuse Kayd, Abdullahi Mahamed Shirwac, Maxamed Cilmi Toxow, Cumar Mahamed Siyaad, Ibraahim Nuur Faqaburaale, Xasan Ibraahim Macallin, Cabdi Daahir Afey, Cabdullahi Xasan Rooble, Ahmed Cabdikarim Xirsi, Amuun Maxamed Cali, Maxamed Soofe Xasan, Daahir Cali Cumar, Saciid Cusmaan Keenadiid, Aw Jaamac Cumar Ciise, Cali Axmed 'Mudiir'.

NOTES

¹ Other bilingual dictionaries had been published: Reinisch (1902), Da Palermo (1915), Abraham (1964), Stepancenko-Osman (1969), Panza (1974), Philibert (1976).

² A more detailed description of the method adopted in the *Dizionario Somalo Italiano* (1985) can be found in the Introduction to the dictionary itself.

³ In collaboration with the Istituto di Linguistica computazionale del CNR di Pisa, we produced several models that could be used for different research purposes: an inverse dictionary, lists of words ordered by digrams, trigrams, lists of words according to grammatical category etc.

REFERENCES

- Agostini, F., Puglielli, A., Ciise Moxamed Siyaad (1985), *Dizionario Somalo Italiano*, Roma: Gangemi.
- Abraham, R.C. (1964), *Somali English Dictionary and English Somali Dictionary*, London: University of London Press.
- Da Palermo, G.M. (1915), *Dizionario della lingua somala*, Asmara: Tipografia Franceseana.
- Panza, B. (1974), *Af Soomaali*, Firenze: Le Monnier.
- Philibert, Ch. (1976), *Petit lexique somali-français*, Paris: Librairie C. Klincksieck.
- Reinisch, L. (1902), *Somali-Deutsch Wörterbuch*, Wien: Alfred Molder.
- Stepancenko, D.I., Moxamed Xadzi Osman (1969), *Kratkij Somali-Russkij I Russko-Somali slovar*, Moscva: Sov. encik.
- Yaasiin Cusmaan Keenadiid (1976), *Qaamuuska Af Soomaaliga*, Muqdisho/Firenze: Madbacadda Qaranka: Madbacadda Qaranka.