The Somali Language Lexical Project:
aims and methods

Annarita Puglielli

1.

The aim of this paper is to give an overall picture of the work on the Somali Language conducted since 1977 as a part of a Technical Cooperation Project financed by the Italian Ministry of Foreign Affairs in collaboration with the Somali National University in Mogadishu. Given the breadth of the project and the large number of people involved, both Italians and Somalis, it will be impossible to describe in detail each of the subject areas and all derived research. We will therefore concentrate on overall objectives and results, focusing on how this experience permitted using and verifying different research methods.

First let us briefly sketch the cultural setting in which our project was to operate. Somali has become the official language of the country since 1969 and has been transcribed officially since 1972.

What was the language situation before 1969? Although only an oral language, Somali had a very rich tradition. One indication is the relevance that verbal art had in Somali society (especially among the nomads). Oral poetry, in fact, still plays a very important role in Somali society; it represents a rich tradition with several poetic genres based on different metrical structures as well as on the subjects they deal with.

Thus, a common oral language was spoken all over the country; it coexisted with regional varieties and other dialects, as well as with minority languages. We may consider the variety of the North-central regions as the *common language*; it had in fact been used for radio broadcasting since 1945, to give but one example of its special status.

As for writing, Arabic was for many years the only written language used, largely for religious purposes. During the colonial period Somalia acquired two more written languages: English in the north and Italian in the centre and south. From the 1950s onward these languages became languages of instruction and administration, and they continued to be used as such even after independence up to 1972.

Attempts had been made to transcribe Somali between 1920-1950 using either Arabic characters or Latin or Osmanian[1]; it was, however, only in 1972 that, after a long debate, the Latin alphabet was officially adopted for transcription.

Somali successively became the official language of the state administration as well as the language of instruction. A literacy campaign was promoted all over the country. It gave Somalis at least a chance to regain their national and cultural identity, embodied in the use of a single native language at all levels of usage. This language had been used

primarily for everyday life communication up to 1972; suddenly it had to adapt to new situations and contexts of usage and cope with necessities determined by new, formal and differentiated communicative contexts alien to the traditional areas of Somali culture.

Given this background, which we can only briefly sketch here[2], we began the Somali language Project with the following aims:
a) the scientific description of Somali as a linguistic system;
b) extending knowledge of the usage of Somali within given socio-cultural and ethno-anthropological contexts.

It must be emphasized that, since our work began as part of a project of Technical Cooperation financed by the Italian Ministry of Foreign Affairs, one of our main aims, right from the beginning, was the training of Somali linguists who could continue research once the project had reached its end. Thus, scientific research was to be an essential instrument of training.

2.

The first objective we identified was, as stated above, a *scientific description* of Somali as a linguistic system (or systems, if one considers the different levels of possible linguistic description: i.e. phonology, morphology, syntax, lexicon, etc.).

Somali is one of the best described African languages, but although many scholars[3] had produced relevant studies on Somali language and culture, beginning in colonial times going into this century, there was still no comprehensive scientific treatment of the language. This was because different authors had worked on different varieties, either regional or ethnic, usually working with a single informant; moreover, recent developments in the field of linguistic research made it necessary to revise and complete the descriptions made.

As for the description of Somali as a linguistic system, it was decided that research be conducted simultaneously both on lexis and on grammar. Each area was to be circumscribed and studied independently using the more appropriate method of inquiry. We were, of course, mindful that there are connections between the two areas - extremely evident and extremely relevant ones. Ideally we would have liked to sequence our work in such a way that the results of research in one area (for example, morphological description) could be incorporated directly into our lexicographical work. Time did not allow us to do so. Moreover, the project called for the almost immediate application of research findings to such activities as the production of textbooks for schools. Thus, to borrow an image from computer science, we had to work within a parallel-processing architecture. Unfortunately, to continue with the image, our pipelining was makeshift; in particular, during the first phase of our work, interconnection between the different areas of research was rather limited. It was, however, sufficient to produce acceptable results, as we shall see.

Since the lexical project is the focus of this presentation, we will mention the grammatical description only to say that our aim was to produce a Reference Grammar of Somali, one that could successively be used: a) as a basis for further research in specific problematic areas, and b) as a basis for the production of school pedagogic grammars and teaching materials for Somali as a mother tongue.

By Reference Grammar we mean a *scientific* and *complete* description of the target language: *complete* means that all possible levels of description - i.e. phonology, morphology, syntax and pragmatics - should be covered and in as much detail as possible. By *scientific* we mean that the description should be systematic, coherent, and generalizing; in other words, through a careful collection and accurate analysis of data, we should be able to identify the underlying mechanisms regulating this language in general (data extraneous to our system should be perceived as peculiarities and exceptions and taken into consideration only in a second instance). The need to arrive at the underlying mechanisms appears even more evident if one considers that Somali belongs to the family of Afroasiatic languages, a group with little known structural characteristics. It takes more than routine data collection and analysis in order to obtain a full understanding of such phenomena as, for example, tonal accent. The role of tonal accent in the Somali phonological system has, in fact, been debated for many years. Another phenomenon that escapes superficial analysis are the focalization structures in syntax.

The production of a reference grammar was therefore seen as a necessary step for a better knowledge of the underlying structure and organization of the Somali language.

3.

A major sub-project of our study of Somali lexis was the compilation of a lexicon. Given the situation of the Somali language, as sketched earlier, this tool appeared most necessary. We identified the following kinds of lexicons to produce:
- A Somali-Italian Dictionary (SID)
- An Italian-Somali Dictionary (ISD)
- A Somali Monolingual Dictionary (SMD)
- A Somali School Dictionary (SSD)

We started in 1978 with the SID. Once this was published, in 1985, we moved on to proving the ground work for the other lexicons. This involved a) writing, in Somali, the definition for the lexical entries of the SID, thereby constituting the first systematic corpus of lexical data on the Somali language; b) expanding the lexical base in selected semantic fields; and c) writing the ISD. Only in 1989 did we start to work specifically on the SMD and the SSD. We will now illustrate each sub-project separately, illustrating the different techniques and methods and the different optimizing strategies we adopted.

3.1.

Before illustrating the methodology adopted in compiling the SID, let us explain why, in 1978, we decided to start with this dictionary.

In a bilingual dictionary the definition of entries is implicit, i.e. given in terms of equivalences. In other words, definitions are not made explicit, as in a monolingual dictionary where each entry is defined conceptually and/or by usage. But to start directly with an explicit definition of the Somali lemmas appeared too complex a task, since:

a) our linguistic starting point was such that defining certain entries might require creating new terms;

b) Somali was undergoing a great expansion as mentioned in our initial sketch, and, as a consequence, it was particularly unstable. The need to create western-style scientific terminology, completely alien to Somali culture, is one clear example. Such terms were created mainly by means of two different mechanisms: the semantic extension of already existing words (ex. *ukun = egg* becomes *cell*); lexical compositions and/or derivations consisting of calques of foreign terms, which at times required bending the structure of Somali[4].

c) regional variation and differences were still insufficiently described and this could generate sociocultural problems (for example, it would be risky to declare one citation form as the standard, in place of all others);

d) last but not least work on a bilingual dictionary could provide the occasion for an initial, valid exchange of knowledge between Somali informants (and aspiring linguists) and ourselves. At the same time, it could provide training, for our colleagues who were to oversee the writing of the monolingual dictionaries. For all these reasons, the task of writing a monolingual dictionary was delayed to a later stage, when the language would have reached for a greater degree of standardization.

The compilation of the SID required, first of all, collecting all the existing lexicographic sources, both bilingual and monolingual, compiled in the past, by various authors using various criteria: Reinisch (1902), Da Palermo (1915), Abraham (1964), Stepanchenko-Osman (1969), Panza (1974), Philibert (1976), Yaasiin (1976). These varied sources were made homogeneous by retranscribing each and every entry on the basis of the official system adopted for Somali, then comparing the various versions and thouroughly rechecking them with Somali informants.

This provided us with a corpus of about 20,000 words. Each one constituted an entry, which was then completed with its Italian translation. "Il risultato dell'operazione è stato largamente soddisfacente: il disporre di più testimonianze ci ha permesso di verificare la validità di molte parole in varie epoche e di integrare diverse accezioni semantiche" (Agostini 1981:143)[5].

The first lexical basis thus obtained, was then enlarged through the systematic scanning of texts written in Somali and published between 1973 and 1981. The corpus consisting of books, several periodicals, the law bulletin and a certain number of issues of the daily newspaper *Xiddigta Oktoobar*. Some particularly problematic domains - as, for example, scientific treaties - were left aside for the moment.

The result of this pains taking work, conducted by Somali scanners, was incorporated in what constituted the lexical base of the SID.

The SID was published in 1985 and its entries constitute the first extended lexical corpus of the Somali language. As such they immediately became the source of all successive lexicographic work, and were used: a) as input for computer elaboration of the data contained in the corpus; b) as an initial lexical base for the compilation of the Somali Monolingual Dictionary (SMD).

A few words should be spent on explaining the computational elaboration carried out in collaboration with the Institute of Computational Linguistics of the Italian National Research Council (C.N.R.) in Pisa. The aim of the elaboration was not only to make the available data more widely accessible, but also furnishing a computational basis for

further research in other fields (phonology, morphology, etc.). The computations produced include:

- an inverse dictionary, i.e. Somali lemmas organized in alphabetical order starting from the end of the word. This is an extremely useful tool for linguistic research (e.g., morphology);
- lists of words grouped according to grammatical categories;
- diagrams and trigrams for the study of phonological constraints;
- an initial Italian-Somali lexicon obtained by automatically inverting those items for which there was an Italian equivalent.

   - a query system implemented for PC's, using the Italian equivalences as the search field; this means that in its electronic form, the SID could be used immediately as an Italian-Somali dictionary. Upon entering an Italian word the PC displays the following information: how many Somali words translate the query word, and how many times the Italian word occurs in definitions, descriptions or expressions (when there is no one to one correspondence between Somali and Italian, paraphrases are used to translate from one language to the other). The PC will then display the list of the Somali words involving the requested Italian word, and each single record may be consulted.

3.2.

Compilation of the SMD, considered the main objective of the lexical project, began in 1985. The lexical entries of the SID served as the basis for it. Definitions for each entry were written by a group of Somalis once the entry format had been decided. At the same time, the enlargement of the lexical base was carried out in several different semantic areas. Different methods were adopted for different lexical areas, in order to optimize procedures and, at the same time, to verify the validity of the methods used. We will now describe the different lexical research sub-projects.

3.2.1.

One of the lexical areas identified as requiring enlargement was that of traditional oral culture. In other words, we sought to include varieties of oral discourse not yet present in our corpus. We thus proceeded to:

a) collect data by means of ad hoc questionnaires on traditional arts and crafts, then used in extensive field testing; this allowed us to collect not only lexical data but also ethno-anthropological data of wider interest;

b) create of a corpus of oral data, obtained by means of recordings of conversations on specific themes of traditional culture conducted by the elders of the National Academy of Arts and Sciences. These were the elders who had already been called upon to recognize words, obtained from pre-existing lexicons, that were unknown to our regular and younger informants.

The questionnaires, regarding different types of activities (house building, pottery making, etc.) and crafts (blacksmith, etc.) were never used, however. Data collection in the field throughout Somalia became impracticable after 1988/89 for reasons widely known.

As for the elders' conversations, we have a collection of a certain number of them on traditional themes such as marriage, religion, traditional rights, etc.
The tapes have been transcribed, but we have not scanned them yet in order to extract new words, attribute definitions to them and add them to the existing corpus.

3.2.2.

The problem of the lexicon for the natural sciences was approached in a completely different manner.

By *natural sciences* we mean in particular the fields of botany and zoology. We obtained an initial corpus of data in these areas based on the existing specialized literature in each scientific discipline. A bibliography of about 300 titles was carefully established; the main sources were identified and all available data was extracted from these volumes. This resulted in a corpus of about 20,000 entries with their scientific classification (two thirds of which are plant names).

This preliminary data must still be processed: linguistically spelling must be normalized since most of the sources were written by foreign scientists before the official writing system for Somali was adopted. This means that the same word often has more than one spelling, or that different names exist for what appears to be the same object (plant or animal). The data clearly needs to be checked with experts in these fields, both Italian and Somali. This should allow us to eliminate possible redundancies and to assess possible dialectal differences.

Field work to double check at least some of the data, was also planned. Lexical inquiry cards were prepared for animals so as to assure a certain degree of certainty in collecting data: for example, the cards include a photo of the animal in order to avoid the kind of misunderstanding generated by verbal descriptions.

Once this research is completed, it will furnish not only a broader lexical base for the SMD, but also a collection of data of extreme interest for ethnoscience studies, as well as for zoologists and botanists, interested in that particular geographical area.

3.2.3.

The second, all-encompassing domain included, under the heading of *scientific lexicon*, the following disciplines : mathematics, physics, chemistry, biology, geography, history and medicine. It was decided to deal with these areas, during the initial phase, only as far as needed for the compilation of the School Dictionary (SSD). In the years that followed the adoption of Somali as the language of instruction, the Curriculum Development Centre of the Ministry of Education undertook to write textbooks for all subjects. This meant that authors had to cope with terminological problems and, in fact, they produced lists of terms for different subject areas. However, a cross check of a

number of textbooks showed that books written by different authors at different times did not use identical terms for the same referents. Thus, it became clear that this lexical domain required reordering. We therefore planned the following strategy. First of all, a certain number of terms, considered basic, were selected on the basis of criteria internal to individual scientific disciplines. These terms were handed over to the experts of the Curriculum Development Centre who were asked to give, for each of them, a Somali equivalent. They were provided with a revision of the existing lexical lists. Experts then wrote the Somali definition for each term. The entries thus compiled went to make up the SSD. They were also to be used in the routine revision of current textbooks, planned in 1990 but not yet undertaken.

3.2.4.

In the field of social sciences (law, economy, etc.), there exists a large corpus of printed texts in Somali. Given this situation we thought it might be useful to set up a computerized scanning system which would also automatically check for the correspondences with the existing corpus of the SMD. The Somali experts would therefore be required to write only the definition in Somali for each new term encountered.

Once the automatic system for scanning is completed, it will be possible to use it for any written source. The system requires:

a) a corpus consisting of texts in machine readable form easily obtainable by the use of a scanner;

b) an automatic parser for Somali, capable of lemmatizing the words found in the texts;

c) a relational retrieval system, capable of comparing the data obtained through the lemmatization process and the SMD data base.

The most interesting part of this work, from the point of view of linguistic research, concerns the preparation of the morphological analyser for Somali. The analyser enables the computer, by means of a set of rules and procedures, to produce all the inflected and/or agglutinated forms and to recognize them when they are given as input. Since Somali has a rich system of inflections both for verbs and nouns, defining the rules in terms that can be translated into algorithms, has been a rather complex task. Another aspect that has made our work rather difficult is that Somali has a certain number of elements that occur together rather than independently. Thus the definite article occurs after the noun but attached to it, and the focus particle fuses with the noun that is focused. The result is a series of rather complex forms, i.e. ones made up of several lexical elements. In spite of these difficulties, our work on the morphological analyser is proceeding as scheduled: linguistic analysis and rule explicitation have in fact been completed and there is already a first computer implementation of the morphological analyser, capable of handling both verbs and nouns. We are now proceeding to verify systematically all the forms the analyser produces in order to eliminate possible mistakes. Given the complexity of the system, mistakes are, indeed, probable. Then the system will be compared with the entire SMD data base, and only then will we be able to use it for lemmatizing texts.

3.2.5.

The last area on which work has begun is the poetry lexicon. Since poetry is a very important part of Somali culture and has a very rich tradition we considered it a very important source, one we ought to take into consideration immediately. We began by compiling concordances of published collections of poems of the most famous poets; at present entries are being written for those words that are either not present in the SMD or present but with a different meaning.

Once the scanning of the printed sources is completed it will be possible to proceed by collecting oral poems. This will furnish new lexical material and, at the same time, serve to maintain oral traditions that could otherwise be lost.

3.2.6.

To complete the picture of the sources used to enlarge the lexical base of the Somali monolingual dictionary, we must mention that new lemmas have come from the actual writing of the SMD entries. That is to say, compiling the entries sometimes required using words that did not appear in the ISD (the initial source for the SMD) and thus had to be included with their definitions, in the SMD itself.

More lemmas also came from research conducted in other fields: for example studies in the field of ethnomusicology gave us a small corpus of related words that were not present in the original corpus. We are, in fact, open to any other lexical source that becomes available, independently of our initial target corpus, as defined in our project specifications.

3.3.

The last two objectives of the lexical project need not be described in much detail since they are more obvious and less relevant from a methodological point of view.

As for the ISD, we started by selecting the Italian words that we thought should constitute the corpus of our dictionary. The basis of comparison with a certain number of already existing bilingual dictionaries (Italian - other language). We then verified that this corpus indeed included the approximately about 7,000 words of the basic vocabulary of Italian (De Mauro 1980). Then we proceeded with establishing the Somali equivalences utilizing, when possible, the SID query system described in 3.1. The final product will be a bilingual dictionary of about 25,000 words. We are now in the stage of the final revision.

3.4.

It was decided fairly recently to produce the Somali School Dictionary using a selection of lexical entries from the SMD. This selection has been made on the basis of explicit

criteria, to permit future development on a well-defined basis. The defining criteria aim at a user base of native Somali, secondary school students who are using a dictionary for the first time. In practical terms, this means a dictionary with a fairly limited number of words, including lemmas from the so-called scientific areas (see 3.2.3.) as well as from the SMD with the exclusion of all highly infrequent and highly predictable words. The latter include, for example, derived abstract nouns or agent nouns highly predictable on morphological grounds and semantically transparent. If, on the other hand, the meaning of a derived noun is not predictable from its base, then the noun has been included.

4.

In pursuing our project we have met various descriptive as well as theoretical problems in the field of lexicography. For some we have found satisfactory solutions, for others we are still searching. A couple of examples will suffice to give a general idea of the complexity of the task.

One of the thorniest problems related to the formal description of Somali has already been mentioned: on the citation form for verbs. Contrary to the solution proposed by Yaasiin Cismaan Keenadiid, who adopted the form of the verbal noun, we adopted the form of the imperative (2nd person singular) for the verbs of 1st, 2nd and 3rd conjugation, together with the invariable form of the present tense
of the restrictive paradigm. One might object against our solution that, in normal usage, not all verbs can be used in the imperative (this is the case of stative verbs). However, the use of derived nouns as the citation form for verbs seems an even less adequate solution, and this for two different reasons: first, there is not always a verb corresponding to a verbal noun; second, more than one verb may correspond to a nominal form. Our choice may therefore be justified as the most efficient and economic way to recover the lexical information.

Within the theory bound aspect, one of the more interesting issues was that of defining the basic lexical categories. This relates both to the morphological/syntactical description of Somali and to the categorization of the lexical entries in the dictionary. For example, we were lead to eliminate the category *adjective* in favour of a 4th conjugation of verbs, following Andrzejewski (1969), because of the morphological and syntactical behaviour of the lexical items making up this group.

In addition, we had to introduce a new category, which we called *predicative nouns.* These words belong to the category *noun* from a morphological point of view, but syntactically can only occur in predicative position and cannot be used in the canonical noun position (as head of a noun phrase that is subject or object, etc.).

Among the unresolved problems it is worth mentioning the semantic classification of the entries of the dictionary. In our opinion this aspect has not yet found a satisfactory solution, even in dictionaries belonging to an   old   and   consolidated lexicographic tradition, like those of our Western languages. It is no wonder, then, that the problem appears even more arduous to solve in dealing with a language that is the expression of a different culture, especially since it is not   yet   clear whether these semantic classifications should be internal to that language and that culture, at least for some aspects.

In conclusion, two aspects of the Somali lexical project should be re-emphasized. They were, in fact, mentioned at the beginning, but should be brought home once again as particularly relevant.

First of all, although the initial formulation of the project may seem rather theoretically *abstract*, especially considering the very practical needs of the Somali people, it is in fact particularly *concrete* and *user-oriented*. This may be clearly seen in the kind of theoretical and descriptive studies produced, studies which have lead to the compilation of a grammar and various dictionaries, including one for schools.

Secondly, the distinctive organizational characteristic of the project - i.e. pursuing research while training Somali linguists - has proven a success. This may be seen by the sheer number of Somalis involved in the project and by the names that appear on our numerous publications.

The essentially practical nature of our endeavour is therefore a reason for all of us, Italian and Somali researchers alike, to carry on with our project even now that Somalia is beset with strife. We may only hope that, soon, young Somalis will be returning to school where they will have, at their disposition, tools enabling them to better reflect on their language and culture. Hopefully they will realize that in spite of their undeniable regional and tribal differences, many other linguistic and cultural features unite them and allow them to consider themselves as a united people.

BIBLIOGRAPHICAL REFERENCES

Abdalla Omar Mansur
> (1989) *Le lingue cuscitiche e il somalo, Studi Somali 8*, (italian translation by C. Serra Borneto), Roma , MAE.

Abdalla Omar Mansur
> (in press) Il somalo dall'oralità alla scrittura, in *Lingua, Pensiero scientifico e Interculturalità*, Quaderni dell'Accademia Nazionale dei Lincei, Roma.

Abraham R.c.
> (1964) *Somali-English Dictionary*, London: University of London Press (reprinted 1968).

Agostini, F.
> (1981) Ragguaglio sui lavori del vocabolario somalo-italiano, in G. Cardona e F. Agostini, *Fonologia e Lessico, Studi Somali 1*, Roma, MAE.

Ajello, R.
> (1984) Il Focus nell'idioma degli Ashraaf di Shingaani, in A. Puglielli (1984b).

Andrzejewski, B.W.
> (1964) *The declension of Somali nouns*, London, School of Oriental & African Studies.

Andrzejewski, B.W.

   (1968) Inflectional characteristics of the so-called weak-verbs in Somali, in *African Language Studies* 9, pp. 1-51.

Andrzejewski, B.W.
   (1969) Some observations on hybrid verbs in Somali, in *African Language Studies* 9, pp. 1-51.

Andrzejewski, B.W.
   (1979) *The Case system in Somali*, London, School and Oriental African Studies.

Antinucci, F. e Axmed Faarax Cali "Idaajaa"
   (1986) Poesia orale somala: storia di una nazione, Studi Somali 7, Roma, MAE.

Antinucci, F. e Puglielli, A.
   (1981) Relative Clause construction in Somali: a comparison between the Northern and the Coastal Dialects, in *Proceedings of the Third International Hamito-Semitic Congress*, London.

Armstrong, L.E.
   (1934) The phonetic structure of Somali, *Mitteilungen das Seminars fur orientalischen Sprachen zu Berlin* 37-38, pp.116-161 [reprinted Westmead Farnborough, Gregg Press 1964].

Berghold, K.
   (1897) Somali-Studien, Zeitschrift fur afrikanische und ozeanische Sprachen, 3, pp.116-198.

Berghold, K.
   (1899) Somali-Studien, Wiener Zeitschrift fur die Kunde des Morgenlandes 13, pp.123-198.

Bruno, B.
   (1984) Note sui verbi di derivazione nominale in somalo, in A. Puglielli (1984b), pp. 115-137.

Cardona, G.R.
   (1981) Profilo fonologico del Somalo, in G.R.Cardona & F.Agostini (ed.) (1981), pp.5-26.

Cardona, G.R. & Agostini F. (ed.)
   (1981) Fonologia e Lessico, Studi Somali 1, Roma, MAE.

Cerulli, E.
   (1957-59) *Somalia: scritti vari editi e inediti* , Roma, Istituto Poligrafico dello Stato.

Ciise Mohamed Siyaad
   (1985) Favole somale, Studi Somali 6, Roma, MAE.

Da Palermo, G.M.
   (1915) Dizionario della lingua somala, Asmara.

De Mauro, T.

(1980) *Guida alle parole,* Roma, Editori Riuniti.

Farnetani, E.
    (1981) Dai tratti ai parametri: introduzione all'analisi strumentale della lingua somala, in G.R.Cardona & F.Agostini (ed.) (1981).

Gebert, L.
    (1986) Focus and word order in Somali, in *Afrikanistiche Arbeitpapiere* 5, pp.42-69.

Gebert, L. e Abdalla Omar Mansur
    (1984) Struttura del Focus in Dabarre, in A. Puglielli (1984b).

Giannattasio, F.
    (1988) Strumenti musicali, in A. Puglielli (ed.) *Aspetti dell'espressione artistica in Somalia*, Roma, Università di Roma "La Sapienza".

Hyman, L.M.
    (1981) L'accento tonale in somalo, in G.R.Cardona & F.Agostini (ed.) (1981).

Moreno, M.M.
    (1955) Il somalo della Somalia. Grammatica e testi del benadir, darod e dighil, Roma, Istituto Poligrafico dello Stato.

Panza, B.
    (1974) *Af Soomaali*, Firenze, Le Monnier.

Philibert, C.
    (1976) Petit Lexique Somali-Francais, Paris.

Puglielli, A.
    (1981) Sintassi della lingua somala, Studi Somali 2, Roma, MAE.

Puglielli, A.
    (1984a) La derivazione nominale in somalo in A.Puglielli (ed.) (1984b), pp. 1-52.

Puglielli, A. (ed.)
    (1984b) Aspetti morfologici, lessicali e della focalizzazione, Studi Somali 5, Roma, MAE.

Puglielli, A. e Abdalla Omar Mansur
    (1993) Grammatica della lingua somala (manuscript).

Puglielli, A. & Ciise Mohamed Siyaad
    (1984) La flessione del nome in A.Puglielli (ed.) (1984b), pp. 53-112.

Reinisch, L.
    (1900-1902) Die Somali-Sprache; II Worterbuch Somali -- Deutsch, Deutsch Somali, Wien, A. Holder.

Schleicher, A.W.
    (1892) Die Somali-Sprache. Erster Text: Texte, Lautlehre, Formenlehere und Syntax, Berlin, Th. Frohlich.

Squillacciotti, M. (ed.)

        (1992) *Produzione e riproduzione nel gruppo domestico. Saggi di ricerca in un villaggio somalo*, Siena, Facoltà di Lettere e Filosofia (new edition in Studi Somali 10, in press).

Stepancenko, D.I. e Moxamed X. Osman

        (1969), Kratkij Somali-Russkij i Russko-Somali, Slovar, Moskva.

Svolacchia M., Mereu L. e Puglielli, A.

        (in corso di stampa) Aspects of Discourse Configurationality in Somali, in K. Kiss (ed.) *Discourse Configurational Languages*, London, O.U.P (in press).

Von Tiling, M.

        (1918) Die Vocale des bestimmten Artikels im Somali, *Zeitschrift fur Kolonialsprachen* 9, pp. 132-66.

Von Tiling, M.

        (1919) Adjektiv-Endungen im Somali, *Zeitschrift fur Kolonialsprachen* 10, pp. 208-40.

Von Tiling, M.

        (1925) Somali-Texte und Untersuchungen zur Somali-Lautlehre *Zeitschrift fur Eingeborenensprachen* 18, pp. 1-156.

Yaasiin C. Keenadiid

        (1976) Qaamuuska Af-Soomaaliga, Firenze.

NOTE: the bibliographical references do not include every publication produced within the Somali project by the numerous scholars who have participated in it; only the articles most relevant to this presentation have been cited.

[1] A particular set of character by Osman Keenadid based on the arabic alphabet.

[2] For a more complete description of the pre-1972 linguistic situation in Somalia, see Abdalla Omar Mansur (in press).

[3] See bibliography.

[4] From a linguistic point of view these are very interesting cases because if they should become stable in the language they could even provoke changes in its structural system.

[5] "The result of the operation was largely satisfactory: the fact that we had several attestations, allowed us to verify the validity of many words in different periods of time, and to integrate the various acceptations".