

## Somali Corpus: state of the art, and tools for linguistic analysis

Jama Musse Jama<sup>1</sup> | jama@jamamusse.com

Keywords: *Somali language; Corpus; Morphology; Concordance, Natural Language Processing (NLP).*

### Abstract

Developing IT resources for language mainly focuses on well-described languages with long-standing written traditions and with a large number of speakers. One of the main challenges for the languages with more recent written traditions is the lack of enough data for successful statistical approaches. This descriptive paper aims to present the state of the art of the construction of the Redsea Cultural Foundation's Somali Corpus (RCF-SC), and the development of a series of computer programs with which to analyze the corpus data for various purposes. The core of RCF-SC is unique in Somali speaking countries and wants to be, for Somali, a resource equivalent in quality to the British National Corpus. The first edition of the corpus is online at [www.somalicorpus.com](http://www.somalicorpus.com).

### 1. Introduction

Somali is one of the major East Cushitic languages of the Omo-Tana group and it has three major varieties: Af-Soomaali, also known as Standard Somali or Common Somali; Benaadir and Af-Maay or Maay<sup>1</sup>. Somali dialects are spoken by a population estimated at 16.5 million<sup>2</sup>, mainly living in the Horn of Africa (Somalia, Somaliland, Djibouti, Kenya, Ethiopia) and also widespread in Europe, The Middle East and North American countries, where the Somali Diaspora recently resettled. Somali is one of the best-documented Cushitic languages, with academic studies dating from before 1900<sup>3</sup>. Somali got its official orthography in 1972 when a Latin-based alphabet was finally selected by law after long contentious debates on the choice between various unofficial writing system, among them adaptations of the Arabic and Latin alphabets, or else newly invented alphabets<sup>4</sup>.

RCF-SC is an annotated and balanced Somali language corpus produced in two phases, using a combination of an automatic tagging system specifically developed according to the Somali grammatical rules, and subsequent manual corrections of the collected data. The RCF-SC corpus covers the major variety of the language, Standard Somali (or Af Soomaali), which is also the language of classical<sup>5</sup> poetry and the official instruction and working language of Somaliland, Somalia and the Somali Region of Ethiopia. The work is been implemented within a framework of collaboration between the University of Naples "L'Orientale" (Italy) and Redsea *Online* Culture Foundation (Somaliland), and the research is being carried out as part of doctoral thesis by the author.

This paper is organized as follows: in section two we describe the major starting up sources of data, to build a rich database used for several tools developed for Somali and the base for annotation process of the corpus. Section three briefly describes the tools implemented for different purposes in building and using the corpus. In section four we discuss the contents of the corpus and how we made it balanced in terms of language variation and time. In section five we present some statistical figures on corpus data and the state-of-the-art of the corpus as of today. We briefly present in section six some language peculiarities of Somali and observe them on the corpus, and finally in section seven, we present the future research activities on the corpus.

### 2. Source of the data

Corpora are compiled for different aims. One reason may be as support material for lexicographers, to compile dictionaries and spelling or grammar checkers. Another reason may be to build a basis for language acquisition or for studying the history of the language. It may also be as an empirical basis for NLP programs. RCF-SC wants to be a multi-purpose corpus for the Somali, even though a single corpus can never be used as a basis for all purposes. In the compilation of the RCF-SC corpus, the idea of "representativeness"<sup>6</sup> (statistical sense) in terms of variations of the Somali literature is replaced by the notion of "balance"<sup>7</sup>. The choice of the texts included into the corpus is therefore

- 
- 1 Somali is specifically collocated within the Lowland East Cushitic which comprises about twenty languages of the Cushitic family within the Afro-Asiatic language phylum. For details of the language classification for the Somali, see the article "The consonant phonemes of Proto-East Cushitic (PEC): a first approximation" by Sasse.
  - 2 See <http://www.ethnologue.com/language/som> [retrieved 21 November 2013].
  - 3 First academic writings on Somali include Praetorins (1870), Larajasse (1897), Reinisch (1900), Kirk (1903). Earlier data on Somali were published by European travelers who visited the Horn of Africa since the 18 century.
  - 4 For the history of writing of the Somali, see the articles "The introduction of a national orthography for Somali" and "The development of a national orthography in Somalia and the modernization of the Somali language" by Andrzejewski.
  - 5 We use the term "classical" for the poetic component of the Somali literature produced before 1930ies. With classical poetry we do mean the oral literature rooted in the society with a mainly nomadic style of life, and we call it as "classic" because almost all the rest of the subsequent poetry literature production has roots in it.
  - 6 There is widespread agreement among corpus linguists that, strictly speaking, representativeness in a statistical sense cannot be obtained for corpora, because of the difficulties associated with defining the underlying population. (see details Biber 1994). Moreover for a mainly oral society, like Somali speakers, a small percentage of the oral literature. is transcribed, and therefore it will take time to include enough data for the corpus.
  - 7 The idea of balance in corpus is intended as introduced in Kilgariff & Grefenstette 2003.

selected by the author on the base of quality of the writing in terms of language. Furthermore, to develop an annotation process for Somali, we had to build a container of as much Somali known words as possible, and we started with the dictionaries.

A major data source from which an initial word-list has been retrieved therefore was putting together lemmas from all existing Somali mono and bi-lingual dictionaries in circulation today<sup>8</sup>. Each dictionary was digitalized<sup>9</sup> and the text parsed automatically<sup>10</sup>. The information extracted for each lemma include the word itself, its grammatical context, and the definitions. This resulted in a collection of 47020 lemmas (word base) tagged with basic Part-Of-Speech (POS) references (i.e. Verbs, Nouns, etc.) The word list created in this manner has been revised<sup>11</sup>, normalized<sup>12</sup> and corrected for some minor spelling errors. On the top of this word list, we developed *somMorph*, a standalone finite-state Natural Language Processing (NLP) application, implementing noun and verb derivatives of Somali word base forms according to the rules defined in the accessible official sources published since the 1960s<sup>13</sup>. The core of the implementations is a rule-based morphological analyzer, specifically developed within the Somali grammatical context. The analyzer has a hard coded set of rules<sup>14</sup> that define different ways to expand and generate words derived from the base words regarding case (subject case, vocative, genitive), gender, number, nouns with suffixed articles, possessives, demonstratives or interrogatives, verb inflection for person, tense, mood, polarity, etc.

With *somMorph*, we generated from the 47020 base lemmas, all the derived forms, thus creating a database containing *c.a.* 1.2 million tagged words. We describe both the state of the art of the Somali Corpus, and specific computer tools developed for analyzing data and managing the corpus itself.

### 3. Tools

This database became the base for several tools, including the followings: 1) *somConcor*<sup>15</sup>, an application that creates concordance for any Somali text; 2) *somTagger*, a tiny automatic POS tagger for the Somali language; 3) *somIsearch*<sup>16</sup>, an advanced tools to query, for instance, a distance or proximity search, meaning that users can look for a pair of words within a given distance from each other; identifying root words, or *vice versa*, expanding the root and showing all its derivatives; 4) *Ubbo* – a Somali spelling checker<sup>17</sup>; 5) *e-Qaamuus. Redsea Somali Dictionary App*<sup>18</sup>, an electronic edition of the Somali dictionary that runs on smart-phone and other portable devices with Android operating system.

### 4. Contents

To build a balanced corpus, we selected representative authors of Somali language written literature, including books recently published by *Ponte Invisibile* ed. (today's major Somali language publisher; with around 20 selected texts); several other books published between 1970-80s, both school textbooks and general literature<sup>19</sup>; eight years of archive of *Geeska Afrika*<sup>20</sup> (Horn of Africa). All these data were analyzed with the above mentioned tools, and 1) each word identified within the database is tagged according to the rules; 2) each unidentified word found through this process is instead put on a new database; 3) the unrecognized words collected in this manner were then analyzed and if they really were missing words (and not just spelling errors or variants), they were added to the main dictionary and then applied the expander algorithm to generate their forms.

- 
- 8 Existing Somali dictionaries taken into consideration are: (1) Yaasiin Cusmaan Keenadiid. *Qaamuuska Afsoomaaliga*, Wakaaladda Madbacadda Qaranka, Muqdisho, 1976; (2) Annarita Puglielli, Francesco Agostini e Ciise Moxamed Siyaad (a cura di), *Dizionario Somalo-Italiano*, Cooperazione Italiana allo Sviluppo, Roma, 1985; (3) Annarita Puglielli (ed.), *Dizionario italiano-somalo*, Carocci, 2010; (4) Annarita Puglielli iyo Cabdalla Cumar Mansuur (Eds), *Qaamuuska Af-soomaaliga*. Centro Studi Somali, Università Roma Tre, 2012; (5) Aadan Xasan Aadan (a cura di), *SOO-MAAL, Qaamuus Afsoomaali*, Machadka Afafka ee Jabuuti (ILD) / Naadiga Qalinleyda iyo Hal'abuurka Soomaaliyeed (SSPEN), 2013 (this last one is being uploaded and partial data was so far digitalized);
  - 9 Some of the dictionaries were got in electronic edition with permission to use for this purpose. We are grateful to the institutions like University of Rome Three for this permission. Other dictionaries were first scanned, and with OCR tool produced a file format. There has been a subsequent manual correction of the scanned documents.
  - 10 For each dictionary, a specific script written in *perl* was developed for parsing its files, extracting data from them and normalizing data before saving them to the database.
  - 11 The revision process has been carried out by the author with the support of two volunteers based in Hargeysa, Somaliland. A *workflow* process managed with web-based application is developed specifically for this purpose. The author checked and approved all the revisions done by the two volunteers.
  - 12 Normalization included linking all different definitions given by different dictionaries for a specific lemma. Other normalization actions included putting together different variants in the writing of one lemma (*dawaco* vs *dacawo*).
  - 13 See the article “Rule based morphological parser for Somali”, Jama Musse Jama.
  - 14 The rules are defined, among the other official sources, in Andrzejewski [1964, 1979], Keenadiid [1976], Banti [1988], Saeed [1993] and Mansur and Puglielli [1999].
  - 15 See the article “Analyzing the unexplored field of the Somali discourse in the text form”, Jama Musse Jama.
  - 16 See [www.redsea-online.com/soomaali](http://www.redsea-online.com/soomaali)
  - 17 See [www.redsea-online.com/ubbo](http://www.redsea-online.com/ubbo)
  - 18 See [www.redsea-online.com/e-qaamuus](http://www.redsea-online.com/e-qaamuus)
  - 19 For bibliographical details of the selected text, see appendix 2
  - 20 *Geeska Afrika* is the most widespread and circulated daily newspaper published in Hargeysa is.

### 5. State of the art of the Somali Corpora

The above described process has finally created a balanced corpus of around 1.5 M tagged tokens, currently divided into ten sub-corpora, which are: 1) poetry<sup>21</sup> 2) traditional songs<sup>22</sup> 3) proverbs<sup>23</sup> and riddles<sup>24</sup> 4) traditional tales<sup>25</sup> 5) fiction, 6) non-fiction<sup>26</sup>, 7) newspapers<sup>27</sup>, 8) drama<sup>28</sup> and 9) sciences<sup>29</sup>. The corpus is distributed over time (from early poetry to contemporary literature production) and over the following text types: literary texts (40%), scientific (maths) literature (3%), other non-fiction (27%), oral poetry and prose transcribed (30%).

There is a web based front end with a simple user interface to provide a word information system on the basis of the above mentioned database and the corpus<sup>30</sup>. The input for word profiling is the lemmatized Somali word, and the result page of each search is divided in several sections: 1) definition(s) of the searched word, its possible spelling variations, and the dictionary(ies) that define it; 2) synonyms and antonyms 3) depending on the selected sub-corpora, the result also includes some examples of the use of the lemma and its derivations.

### 6. Some interesting findings

In analyzing the unexplored field of Somali discourse in its text form, the RCF-SC proved to be an important tool, e.g., while parsing dictionaries, because we were able to easily identify the missing lemmas in the main Somali mono-lingual dictionaries. Another useful application of the corpus is to identify pairs of Somali words with same spellings but different meanings and, sometimes, also different pronunciations. In several cases they require different root lemmas. When all expected data will have been included, we will be able also to identify the words used a certain poet or writer, but never used by other renown poets or writers, and to learn the history of Somali words. The corpus along with its database, is already useful for Somali lexicographers in order to quickly choose appropriate examples for their lemmas.

### 7. Conclusion and Future developments

The main idea of this paper is to present the state-of-the-art in the development of first Somali corpus. We described how the corpus is being implemented and provided some statistical figures on the current contents of the corpus. We also showed some linguistic features of Somali that can be observed in the corpus

#### 7.1 MT Somali-English-Somali

In the process of collecting data for the Somali Corpus, we also digitalized a good deal of parallel text (Somali – English). It mainly consists of translations into Somali of the short stories from western classics, essays and Somali poetry, novels and drama translated into English. Most of them are books published by the *Ponte Invisibile Ed.* within the *Iswaydaarsi* (Exchange) column. It also includes official documents like the Declaration of Human Rights, the constitution of the Republic of Somaliland; mathematics terminology as adopted by the Somali Mathematics Curricula for secondary schools; a collection of bilingual (Somali-English) press releases by several Somali organizations. The idea is to generate from this balanced content in terms of language variations, enough parallel segments to produce a good quality MT engine for English-Somali and Somali – English translations.

#### 7.2 Corpus and tool enrichment

We expect to complete the construction of the database and developing tools for updating the corpus and querying against it. For example, we want to extend and make more intuitive the results of the word profile consultation. We want to finalize the algorithms for Automatic POS tagging for Somali, and the procedure for generating the concordance and other linguistic analyses of the Somali text. In this regard, there is an embryonic algorithm for a text analyzer which uses inductive logic programming for identifying the root of a Somali word through a learning process.

#### 7.3 Further development: advanced linguistic analysis

Within the word profile results, we want to indicate the syntactic context of the searched lemma: the relationship that a word has with other words, and what are the words that are associated with the displayed variation of the searched word (e.g., collocations or collocates). In the advanced search we also want to locate compound words in which the searched

---

21 For biography details of the selected poets, see appendix 1

22 Collection of traditional songs transcribed by the author

23 Includes proverbs from George Kaptches; other collections of proverbs

24 Includes text from different sources (see appendix 2 for details)

25 Includes text from different sources (see appendix 2 for details)

26 Includes all Somali books in Curisyo series.

27 Eight years of Geeska Afrika (Horn of Afrika) daily newspaper.

28 Three short plays written by Cabdiraxmaan Yuusuf Cartan published in *Baqayo roqad* (Ponte Invisibile 2010); *Shabeel Naagood* by Xasan Sheekh Muumin; *Wadhafiyo Shimbiro war isuma hayaan* by Cismaan Aadan Xuseen and edited by Muuse Xaaji Saciid Muumin; *Masiibadu Adduunyada iyadaa u Macallin ah* (1983) by Xasan Ganey, edited by Jama Iswiidhen; Some excerpts from *Kalahaab iyo kalahaad* play by Cali Sugulleh (play performed in 1966; text taken from "Modern and traditional aspects of Somali drama" by B.W.Andrzejewski.

29 Includes *Aasaaska Xisaabta* – Basics of Mathematics and Mathematical terminology by Jama Musse Jama (Ponte Invisibile, 2014).

30 See [www.redsea-online.com/soomaali](http://www.redsea-online.com/soomaali)

word is an element of the compound. In particular, compounds that begin or end with the searched word. Finally, given a specific sub-corpus, we want to understand what are the favoured syntactic relations of the searched word, i.e.:

- a) for nouns (N): Subject Of, Direct/Oblique Object Of, Predicate Noun Having As Subject in a main clause (i.e., occurs in the context NP *waa* \_\_), Predicate Noun Having as Antecedent in a Relative Clause (i.e., occurs in the context N \_\_ *ah*), etc.;
- b) for verbs (V): Has As Subject, Has As Direct/Oblique Object, Co-Occurs With Directional Particles (*Wada, Kala, Sii, Soo*), etc.

## 8. References

Andrzejewski B. W. (1964), *The Declensions of Somali Nouns*, School of Oriental and African Studies, University of London, London.

Andrzejewski B. W. (1974), *The introduction of a national orthography for Somali*, African Language Studies, XV, 1974, pp. 199-203.

Andrzejewski B. W. (1978), *Modern and traditional aspects of Somali drama* in Folklore in the modern world, edited by Richard M. Dorson (London and the Hague, Mouton, 1978, on p.89)

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

Andrzejewski B. W. (1979), *The development of a national ortography in Somalia and the modernization of the Somali language*, Horn of Africa, 1, 3, 1978, pp. 39-45.

Banti G. (1988), *Reflections on Derivation from Prefix-conjugated Verbs in Somali*, In Puglielli A. (ed), Proceedings of the Third International Congress of Somali Studies, pp. 43-95, Roma.

Biber D. (1994), *Representativeness in corpus design*, in: Zampolli, Antonio/Calzolari, Nicoletta/Palmer, Martha (eds.): Current Issues in Computational Linguistics: In Honour of Don Walker (=Linguistica Computazionale IX-X). Pisa: Giardini/Dordrecht: Kluwer, pp. 377-407.

Kapchits, G. (2011), *Soomaali Been Ma maahmaahdo*, Ponte Invisibile Ed.; Pisa.

Kilgarriff, A. and Gregory G. (2003). *Introduction to the special issue on the web as corpus*. Computational Linguistics 29:3, 333-348.

Kirk, J. W. C. (1903) *Notes on the Somali language, with examples of phrases and conversational sentences*, Henry Frowde, London.

Jama Musse Jama (2013), *Rule based morphological parser for Somali*, Dhaxalreeb, Vol. III-2013, November 2013, Pisa.

Jama Musse Jama (2013), *Analyzing the unexplored field of the Somali discourse in the text form*, Dhaxalreeb, Vol. III-2013, November 2013, Pisa.

Larjasse E. de, Sampont C. (1897), *Practical grammar of the Somali language with a manual of sentences*, Kegan Paul, Trench, Trubner & Co., London.

Praetorins F., *Ueber die Somalisprache* in Zeitschrift der Deutschen Morgenländischen Gesellschaft, Band 24. Wiesbaden: Harrassowitz. 145-171. 1870.

Reinisch L. (1900), *Die Somali-Sprache. I: Texte*, Alfred Holder, Wien.

Reinisch L. (1902), *Die Somali-Sprache. II: Wörterbuch*, Alfred Holder, Wien.

Reinisch L. (1903), *Die Somali-Sprache. III: Grammatik*, Alfred Holder, Wien.

Saeed J. I., *Somali Reference Grammar*. Dunwoody Press, Kensington; 1993.

Sasse, Hans-Jürgen. 1979. The consonant phonemes of Proto-East Cushitic (PEC): a first approximation, Afroasiatic

Linguistics 7.1.1-67.

Yaasiin Cismaan Keenadiid, *Qaamuuska Af-Soomaaliga*, Firenze; 1976.

*Appendix 1: Biographies of authors included (or books/poetry being digitalized)*

*Written prose literature:*

Aamina Xaaji Yaaxeen (? , ? - , )  
Axmed Cali Abokor (? , ? - , )  
Axmed Cartan Xaange (Laascaanood, ? - , )  
Axmed Faarax Cali "Idaajaa" (Dhuusamareeb, 1948 - , )  
Axmed Ibraahin Cawaale (Hargeysa, ? - , )  
Axmed Sheekh Jaamac (Burco, - , )  
Cabdillaahi Cawed Cige (Burco, 1968 - , )  
Cabdillaahi Sheekh Xuseen "Hantiwadaag" (? , ? - , )  
Ciise Maxamed Siyaad (? , ? - , )  
George Kapchits (Moosko, 1939 - , )  
Ibraahin Yuusuf Axmed "Hawd" (Burco, 1967 - , )  
Jaamac Cumar Ciise (Boosaaso, 1932 - , )  
Jama Musse Jama (Hargeysa, 1967 - , )  
Maxamed Daahir Afrax (Garowe, ? - , )  
Maxamed Faarax Jaamac "Cawl" (Laasqorey, 1937 - Beledweyn, 1991)  
Maxamed Xasan Cismaan (? , ? - , )  
Muuse Xaaji Ismaaciil Galaal (Cayn, 1920 - Jeddah, 1980)  
Rashiid Maxamed Shabeelle (? , ? - , )  
Rashiid Sheekh Cabdillaahi Axmed (Qoolcaday, 1940 - , )  
Shire Jaamac Axmed (? , ? - ? , ?)  
Siciid Jaamac Xuseen (Ceerigaabo, 1940 - , )  
Xasan Sheekh Muumin (Boorame, 1930 - Oslo, 2008)  
Xuseen Sheekh Axmed "Kaddare" (Muqdishu, ? - , )  
Yaasiin Cismaan Keenadiid (Hobyo, 1919 - ? , ?)  
Yuusuf Maygaag Samatar (Hargeysa, ? - ? , ?)

*Written and oral poetry literature:*

Aadan Axmed "Afqallooc" (Ceerigaabo, 1871 - Ceerigaabo, 1941)  
Axmed Diiriye Diiriye "Qaasin" (Mandheera, 1933 - Hargeysa, 2006)  
Axmed Saleebaan Bidde (Hargeysa, 1943 - , )  
Cabdillaahi Macallin Axmed Dhodaan (Doolo, 1941 - , )  
Cabdillaahi Maxamed Xirsi "Salaan Carrabay" (Beer, 1950 - Burco, 1943)  
Cabdillaahi Suldaan Timacadde (Gebiley, 1920 - ? , 1973)  
Cali Aadan "Cali-Dhuux" (Laascaanood, 19c - ? , 20c)  
Cali Jaamac Haabiil (Berbera, 1856 - ? , ?)  
Cali Sugulleh "Duncarbeed" (Oodwayne, 1936 - , )  
Cali Xuseen Xirsi (Burco, 1913 - ? , 1976)  
Cilmi Bowndheri (Hargeysa, 1908 - Berbera, 1941)  
Faarax Nuur (Hargeysa, 1855 - ? , 1930)  
Faarax Shuuriye (Mogadishu?, ? - , )  
Ismaaciil Mire (Laasadaar, 1862 - ? , 1951)  
Maxamed Cabdille Xasan (Buuhoodle, 1856 - Iimey, 1921)  
Maxamed Ibraahin Warsame "Hadraawi" (Burco, 1943 - , )  
Maxamed Liibaan Jadeer (Hargeysa, fl. 19th century - ? , ?)  
Maxamed Nuur Fadal (Berbera, 1885 - ? , ?)  
Maxamed Xaashi Dhamac "Gaariye" (Hargeysa, 1949 - Oslo, 2012)  
Muuse Xaaji Ismaaciil Galaal (Cayn, 1920 - Jeddah, 1980)  
Qamaan Bulxan (Dhagaxbuur, 19c - ? , 20c)  
Raage Ugaas Warfaa (Dhagaxbuur, 1810 - ? , ?)  
Saahid Qamaan (Dhagaxbuur, 19c - ? , 20c)  
Sheekh Cali Cabdiraxmaan (Sheekh Cali Majeerteen) (Mogadishu?, ? - , )  
Sheekh Liibaan (Mugud/Galguduud, - , )  
Xasan Sheekh Muumin Gorod (Boorame, 1930 - Oslo, 2008)

## Appendix 1: List of books included into the corpus

### Poetry:

- ▲ A list of 130 renowned poets including Maxamed Ibraahim Warsame “Hadraawi”, Sayid Maxamed Cabdille Xasan, Dhoodaan, Gaarriye and others are included with complete works. Other incomplete works from other poets are also included. See the list on [www.somalicorpus.com](http://www.somalicorpus.com).

### Proverbs and riddles:

- ▲ Kapchits, *Soomaali Been Ma Maahmaahdo*
- ▲ Other proverbs and riddles collected by the author from oral sources

### Traditional tales:

- ▲ Kapchits, *Somali tales*.
- ▲ Xange, *Sheekooyin*.
- ▲ Muuse Galaal, *Sheekooyin*.
- ▲ Sheeko iyo Shaahid short stories collection.
- ▲ Amina Xaaji Yaaxeen, *Suugaanta Dhallaanka, Akaadmiyadda Dhaqanka, Xamar*.

### Fiction:

- ▲ Cabdillaahi Sheekh Xuseen “Hantiwadaag”, Ayaan daran, 1981 (used second edition reprinted in 2003)
- ▲ Ibraahim Yuusuf Ahmed “Hawd”, *Aanadii Negeeye*, Ponte Invisibile Ed. (redsea-online), Pisa, 2007. A long novel.
- ▲ Ibraahim Yuusuf Ahmed “Hawd”, *Madxafkii waallida*, Ponte Invisibile Ed. (redsea-online), Pisa, 2014. A collection of short stories.
- ▲ Kapchits G., *Sheekooyin Male-awaal ah*,
- ▲ Jama Musse Jama, *Mahdi*, unpublished long novel.
- ▲ Maxamed Daahir “Afrax”, *Maanafaay*, African Triangle, 1993. A long novel (used also 1979 edition)
- ▲ Maxamed Daahir “Afrax”, *Talaseeg*, Unpublished short story, 2003.
- ▲ Siciid Jaamac Xuseen, *Shufbeel*, Ponte Invisibile, Pisa, 2010. A collection of short stories.
- ▲ Siciid Jaamac Xuseen, *Safar aan jiho lahayn*, Ponte Invisibile, Pisa, 2013. A collection of short stories.
- ▲ Xuseen Sheekh Axmed “Kaddare”, *Waasuge iyo Warsame: socdaalkii 30ka maalmood*, Akeedamiyada Cilmiga iyo Fanka, Muqdishu, 1983. A short novel.

### Non-fiction

- ▲ Jaamac Muuse Jaamac, *Gobannimo bilaash maaha*, Ponte Invisibile Ed. (redsea-online), Pisa, 2007.
- ▲ Maxamed Daahir Afrax, *Dal Dad Waayey iyo Duni Damiir Beeshay: Soomaaliya Dib ma u Dhalan Doontaa?*, 2004
- ▲ Rashiid Sheekh Cabdillaahi Axmed, *Adduun iyo taladii*, Ponte Invisibile Ed. 2010, Pisa.
- ▲ Siciid Jaamac Xuseen, *Shufbeel*, Ponte Invisibile Ed. 2010, Pisa.
- ▲ Siciid Jaamac Xuseen, *Safar aan jiho lahayn*, Ponte Invisibile Ed. 2010, Pisa.

### Newspapers

- ▲ Geeska Afrika – issues 1-608, years 2006-2013.

### Drama

- ▲ Cabdiraxmaan Yuusuf Cartan, *Baqaya roqad*, Ponte Invisibile Ed. (redsea-online), Pisa, 2010.
- ▲ Xasan Sheekh Muumin, *Shabeel Naagood*,
- ▲ Xasan Cabdillaahi Xasan “Ganey”, *Wadhaf iyo Shimbiro war isuma hayaan*, edited by Jama Iswiidhen
- ▲ Cali Sugulleh, *Kalahaab iyo kalahaad*

### Sciences, agriculture and technology

- ▲ Axmed Ibraahin Cawaale, *Environmental crises*
- ▲ Jama Musse Jama, *Aasaaska Xisaabta*

---

<sup>i</sup> Jama Musse Jama (aka Jaamac Muuse Jaamac) is a mathematician and computational linguistics specialist based in Pisa, who is also the Director of the Redsea Online Cultural Foundation. He has PhD in African Studies with particular interest in the Somali language and literature, and at the Oriental University of Naples, Italy, he developed and currently directs the Somali Corpus (see [www.somalicorpus.com](http://www.somalicorpus.com)).