Tracing Linguistic Heritage: Constructing a Somali-Italian Terminological Resource Through Explorers' Notebooks and Contemporary Corpus Analysis

Silvia Piccini, Giuliana Elizabeth Vilela Ruiz, Andrea Bellandi, Enrico Carniani

Istituto di Linguistica Computazionale "A. Zampolli"

Via Moruzzi, 1 – 56124 Pisa (Italy)

silvia.piccini@ilc.cnr.it, giulianaelizabethvilelaruiz@cnr.it, andrea.bellandi@ilc.cnr.it, enrico.carniani@ilc.cnr.it

Abstract

The aim of this contribution is to introduce the initial phases of constructing a Somali-Italian terminological resource that dates back to Italy's colonial expansion into Africa. Specifically, the terminological data were extracted from the notebooks authored by the Italian explorer Ugo Ferrandi (1852 - 1928) and published by the Società Geografica in 1903 under the title "Lugh. Emporio Commerciale sul Giuba". In order to develop Ferrandi's terminological resource, we employed Semantic Web technologies (RDF, OWL, and SPARQL) and embraced the Linked Open Data paradigm. This ensures the FAIRness of the data and enables the publication and sharing of our terminological resource within an open interconnected Web of Data, thus contributing to addressing the absence of Somali in the Linguistic Linked Data cloud. Whenever feasible, Ferrandi's lexicon entries have been linked and enriched with information derived from a Somali lexicon included in a contemporary Somali Corpus. This approach allows the synchronic corpus-related Somali lexicon to acquire historical depth, thereby illuminating the linguistic dynamics that have transpired over time and would otherwise have remained obscure.

Keywords: Somali language, computational terminology, Semantic Web

1. Introduction

Somali is the most widespread Cushitic language, spoken by about 21.8 million people primarily in Somalia, but also in Djibouti, Kenya, Ethiopia, and by a significant Somali-speaking diaspora in the Middle East, Europe, and North America. It belongs to the vast family of Afroasiatic languages, which also includes Ancient Egyptian, Semitic, Berber, and Chadic languages. Despite being widely spoken, Somali can be considered an under-resourced language due to the limited availability of annotated datasets and language resources for NLP tasks and generally for AI research. While considerable work remains to be done, during the past decades initiatives have been taken to develop resources for Somali, mostly focusing on automatic speech recognition (Biswas et al., 2019; Laryea & Jayasundara, 2020), lemmatization (Shafie Abdi & Muhidin Abdullahi, 2023), translation resources (Bonab, Allan, and Sitaraman, 2019; Duh et al., 2020) and information retrieval/sentiment analysis (Bahar & Ramaha, 2023).[1] The present work is intended as a further contribution in this direction. The aim is indeed to present a computational terminological resource in Somali and Italian. Specifically, terminological data were extracted from the notebooks written by the Italian explorer Ugo Ferrandi (1852-1928) during his stay in Lugh at the end of 19th century. This corpus, which dates to the first Italian colonialist expansion into Africa, is important for two key reasons. Firstly, it provides insight into an earlier stage of language and culture before the arrival of European powers. Secondly, these notebooks shed light on a historical culture that predominantly relied on oral tradition, with Somali adopting an official writing system only on 21 October 1972, when it was established as the official language of the Republic of Somalia. In addition, terminological data have been linked and enhanced, whenever feasible, with information taken from the contemporary computational lexicon included in the Somali corpus (Musse Jama, 2016). This integration adds a historical dimension to our resource, revealing, for example, which archaic or regional terms attested by Ferrandi have now become part of standard Somali and which terms have undergone changes in meaning over time.

The lexicon was built using technologies of the Semantic Web (RDF, OWL, and SPARQL), according to the Linked Open Data paradigm, to guarantee the data's findability, accessibility, interoperability, and reusability (Wilkinson et al., 2016). The main objective of our work is indeed to publish and share the terminological resource in an open interconnected Web of Data, that will allow Somali to be represented in the LOD (Linked Open Data) cloud. Once construction is complete, the resource will be made available on the CLARIN research infrastructure.

The Somali-Italian terminological resource was created as part of a research project funded by the philanthropic organisation Fondazione RUT, in collaboration with the Istituto di Linguistica Computazionale "A. Zampolli" (ILC-CNR) and the Società Geografica Italiana (SGI). The primary objective of this collaboration is to enrich and facilitate access to valuable cultural materials stemming from 19th century expeditions in Somalia, such as geographical maps, photographs, artefacts, and travel literature housed in the SGI library.

---

[1] The purpose of the bibliography provided here is merely to highlight the increasing interest in the development of Somali resources; it does not claim to be comprehensive.

The remainder of this paper is structured as follows. After a brief introduction on the figure of Ugo Ferrandi and the notebooks he wrote during his expedition to Somalia, section 3 will be devoted to the methodologies and models used to construct the terminological resource. In section 4, some terminological entries and queries will be illustrated, and finally, in section 5, conclusions will be drawn.

2. Ugo Ferrandi's Notebooks

Decades after achieving unity in 1861, Italy embarked on a policy of colonial expansion primarily aimed at the African continent, first conquering Eritrea in 1882, followed by Somalia in 1889, and finally, after fierce resistance from indigenous peoples, Ethiopia in 1936, the last empire in Africa.

Explorers and travellers significantly contributed to fostering Italy's administrative and commercial penetration of the African continent, supported in their endeavours by geographical research centres emerging in Italy at that time, such as the Società Geografica Italiana founded in Florence in 1867 and the Società d'Esplorazione Commerciale in Africa established in Milan in 1869.

In this policy of commercial penetration, a key role was played by Ugo Ferrandi[2]. He was born in Novara on January 6, 1852, into a wealthy family of landowners. By the age of 22, he became a sea captain and embarked on merchant ships to the Red Sea and the Pacific, reaching as far as the southern Atlantic and North America. There is no further information about him until 1886 when he set foot in Africa for the first time as a member of the expedition led by Augusto Franzoj.

Despite the enterprise's catastrophic failure, Ferrandi chose to remain in Africa, initially as a commercial agent for the Bienenfeld Company in Aden and subsequently as an envoy for the Esplorazione Commerciale di Milano. Thus, he initiated a sequence of expeditions in Harrar, along the Juba River between Brava and Kisimajo, and Brava and Badera. In 1885, after taking part in Vittorio Bottego's expedition, he was bestowed with the rank of superintendent of the commercial station situated in Lugh, in the Benadir region (western Somalia). During his two-year stay, he turned the little settlement into a thriving commercial centre due to its advantageous location for commerce in East Africa. Furthermore, while staying there, he meticulously documented the material and immaterial culture of the Somali tribes settled in the area by compiling notebooks, which were later published in 1903 by SGI under the title "Lugh. Emporio commerciale sul Giuba"[3]. A comprehensive overview of the Somali realia at the dawn of the 20th century is thus offered, ranging from flora and fauna, to dwellings, wedding and funerary rites, customs, folklore, festivals, clothing, games, religion, superstitions, agriculture and livestock,

furnishings, social organisation, weapons, etc. Needless to say, Ferrandi's work constitutes a source of great historical, anthropological, ethnographic but especially linguistic value. The notebooks are indeed a veritable mine of terminological information: they contain a wealth of specialised terminology related to Somali culture, showcasing the language used by nomadic herders and farmers before European colonisation.
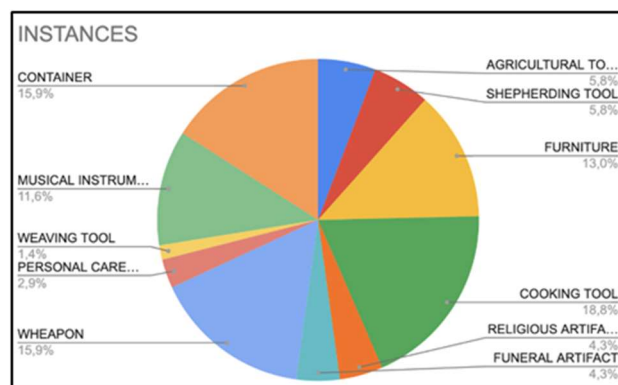

Figure 1. Classification of terms by semantic fields.

For the time being, more than 400 terms have been manually extracted, covering the whole range of semantic fields listed above (Figure 1).

Some of these terms are also included in the glossary in the appendix of the notebooks, consisting of 300 terms that, according to Ferrandi, belonged to three languages spoken at the time in the village of Lugh: *Somali* (*s*), *Rahanuìn* (*r*), and "*Lughiano*" (*l*). Although this categorization does not align with the current linguistic classification of the Somali group, it may be reasonably assumed that by "*Rahanuìn*," Ferrandi was referring to the Maay dialect, described in Saeed (1982) as Central Somali. Instead, the explorer used the term "*Somali*" to refer to the dialect known as common Somali (Andrzejewski, 1971; Andrzejewski & Lewis, 1964), which later became the standard language of the Republic of Somalia due to its prominence. This dialect was used even before the arrival of colonial powers as a *lingua franca* to facilitate broader communication among the several Somali tribes. A more thorough investigation is needed to identify the dialect that Ferrandi called "*Lughiano*." It is important to emphasise that the linguistic data presented by Ferrandi require continuous and meticulous validation. The explorer documented words based on what he heard without a profound understanding of the local dialects. In addition, he often adopted incongruent spellings, given the absence of a writing system for a primarily oral language. Cultural aspects as well need to be "purged" of the stereotypes and prejudices that were prevalent in the highly simplified narrative of early explorers. The latter played a significant role in shaping a collective perception of Africa as a perilous

---

[2] For detailed information on Ugo Ferrandi, see Gavello (1975) et Marini (1991).

[3] Ugo Ferrandi's notebooks are available online at the following link: https://archive.org/details/lughemporiocomm00ferrgoog

and mysterious continent inhabited by wild communities in need of civilization. By stigmatising the Other, indeed, the Western world constructed its own positive identity (Mudimbe, 1988), enacting that dialectic mechanism known as "othering", according to the term coined by Spivak (1895).

3. Modelling the Computational Terminological Resource

The bilingual Italian-Somali lexicon can be technically defined as a termino-ontological resource, since the conceptual (ontological) and the linguistic (terminological) levels are separated although intimately linked, in accordance with paradigms and methodologies developed in recent decades (inter al. see Després & Szulman, 2008; Roche & Papadopoulou, 2019; Temmerman, 2022). The theoretical assumption on which this work is based is indeed that Terminology is a "twofold science", its specificity consisting precisely in the relation between language and specialised knowledge (Costa, 2013; Santos & Costa, 2015).

Without going into a complex issue that would be beyond the scope of this article, it is worth emphasising that unsurprisingly the distinction between the extralinguistic dimension of concepts and the linguistic level of senses has been strongly supported by the socio-cultural approach in terminology proposed by Diki-Kidiri (2008). Starting from a contrastive study of naming in African and European languages, he emphasises the methodological need to articulate terminological analysis along three axes: the signifier, the signified and the concept.

As previously underlined, in our resource the two levels – conceptual and linguistic – are described using two key Semantic Web technologies, i.e. the Web Ontology Language (OWL) and the Resource Description Framework (RDF).

3.1 The Linguistic Dimension

As far as the linguistic dimension is concerned, we adopted the OntoLex-Lemon model (McCrae et al., 2017), as it constitutes nowadays the de-facto standard for the publication of lexicons in RDF. The model is characterised by a modular structure, which allows for a detailed description of the linguistic characteristics of a term. Consistently with the theoretical assumptions expressed above, in OntoLex-Lemon the linguistic and the conceptual dimensions are kept separated. The concept, an extralinguistic entity designated by the signified, receives a formal description in an ontology outside the model[4]. The link between the lexical entry and the ontological concept is reified through the sense which is implemented by the class *ontolex:Lexical Sense*.

According to the model, each Italian and Somali lexical entry is defined as an instance of the class *ontolex:Lexical Entry*. The relations *ontolex:canonicalForm* and *ontolex:otherForm* link each lexical entry to its grammatical realisations that are described in detail (POS, gender, tense, etc.) and associated with a written representation. Each lexical entry is linked with one or more senses as in the case of polysemous words. The lexical sense, an instance of the class *ontolex:Lexical Sense*, is defined by a set of lexico-semantic relations expressing the paradigmatic relations among terms (hypernym, synonym, approximate synonym, and so forth). Each term both in Somali and in Italian is provided with a definition drawn from Ferrandi's notebooks. The definition is also given at the level of the conceptual entry. Somali lexical terms[5] and their Italian equivalents are linked by the property *ontolex:translatableAs*.

3.1.1 The Linking with the Somali Copus

When feasible, terms from Ferrandi's lexicon have been linked via the property *rdfs:sameAs* and subsequently enhanced with data extracted from the Somali lexicon included in the Somali Corpus created by Musse Jama (2016)[6]. The Somali Corpus has over seven million annotated words embedded in a grammatically verified text. It also provides search and analysis tools. This balanced and annotated Somali corpus underwent a two-stage compilation process. Initially, an automatic tagging system based on Somali grammatical rules was employed. Subsequently, manual corrections were made to refine the gathered data. The Somali corpus covers both prose and poetry literature and includes a lexicon that provides a concise overview of the linguistic findings obtained from corpus-based word analysis. This lexicon includes the word's frequency within specific sub-corpora; the etymology of the word; synonyms and antonyms; spelling variants; and definitions taken from reference dictionaries and translation resources.

Linking our termino-ontological resource with the corpus lexicon has not been possible without first converting the corpus lexicon to the OntoLex-Lemon model due to its proprietary format. The conversion procedure included an intermediary stage where the proprietary format has been converted into the CoNLL-U format, which is the standard often used in Universal Dependencies (UD) to annotate data at the sentence and word/token levels. A "miscellaneous" column (MISC) was used to store data that cannot be represented in the CoNLL-U format, such as translation or etymology. Developing this three-step process (proprietary format - CoNLL-U - OntoLex-Lemon) offers the benefit of creating a versatile

---

[4] The conceptual dimension can also be expressed through the class *Lexical Concept*, defined as "a mental abstraction, concept or unit of thought" and connected either to the lexical entry through the relation *ontolex:evokes* or to the lexical sense through the relation *ontolex:lexicalisedSense*.

[5] For entries in Ferrandi's lexicon, the original spellings chosen by the explorer have been retained to facilitate

interested scholars in reconstructing the phonetic characteristics of Somali from that period, where feasible. However, it is essential to acknowledge the challenge of this endeavour owing to the lack of a standardised alphabet at that time and the inconsistencies present in Ferrandi's work.

[6] The corpus data have been provided by Musse Jama, who is actively involved in the research project sponsored by the RUT Foundation.

conversion tool from CoNLL-U to OntoLex-Lemon. This tool can be applied to any resource in CoNLL-U format, making it a valuable asset. Figure 2 illustrates the workflow of the project.
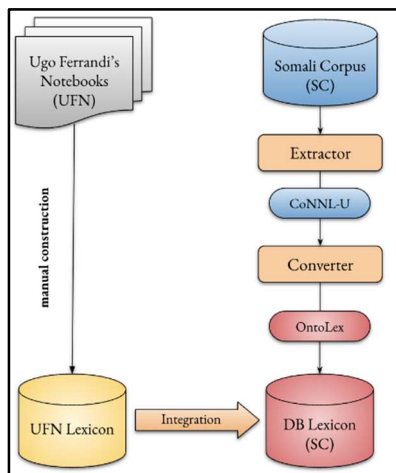

Figure 2. The workflow of the project.

The two resources are kept physically separated but intimately connected, complementing each other. This allows the synchronic corpus-related Somali lexicon to be given historical depth, thus shedding light on the linguistic dynamics that have transpired over time and that would otherwise have remained obscure.

3.2 The Conceptual Dimension

The sense of each lexical entry is connected via the *ontolex:reference* relation to a concept outlined in an OWL ontology that formally describes the prevailing conceptualization of the world in Somalia during the early 20th century. The structure of this ontology draws heavily from the SIMPLE lexical model introduced by Lenci et al. (2000) and proven effective in organising specialised lexicons (Piccini et al., 2013). Rooted in the core tenets of generative lexical theory established by Pustejovsky (1995), this model adeptly captures the multi-dimensionality of concepts through the Qualia structure. The latter, with its four roles (formal, constitutive, telic, and agentive), makes it possible to express orthogonal dimensions of a concept's meaning, thus going beyond the hierarchical subsumption relationships.

The SIMPLE ontology consists of 139 concepts structured in a hierarchy with 6 levels of depth. These concepts are interconnected through an extensive web of relationships, also influenced by the Qualia structure, and categorised into formal relationships (*is-A*), constitutive relationships (*isPartOf*, *hasAsPart*, *location*, *madeOf*, *produces*, etc.), telic relationships (*purpose*, *objectOfTheActivity*, *usedFor*, etc.), and agentive relationships (*resultOf*, *causedBy*, *derivedFrom*, etc.). The SIMPLE ontology, designed for general rather than domain-specific lexicon, serves as a foundational ontology, its concepts representing the highest-level nodes in the hierarchy that are further specialised to effectively represent

Ferrandi's specific domains. A fundamentally top-down approach is employed to refine and extend the model in light of the specific issues raised by the data.4

4. Example and Queries

For illustrative purposes, the entry *fandal* "spoon" is reported below (Figures 3 and 4). In Figure 3, the RDF formalisation of the linguistic component is presented, while in Figure 4, the OWL formalisation of the concept, carried out using the Protégé ontology editor, is illustrated.

```
fandal_entry a ontolex:Word ;
      ontolex:canonicalForm [
            ontolex:written Rep "fandal"@som ]  ;
      ontolex:sense :fandal_s1 ;
      rdfs:seeAlso fandhaal_entry .

:fandal_s1 a ontolex:LexicalSense ;
      skos:definition "Cucchiaio di legno utilizzato
            per mescolare i grandi di caffè (bun) messi
            a friggere nel burro. Lo strumento era
            utilizzato anche dai Lughiani e dai
            Kahanuin  per prendere un po' del burro
            di frittura dei bun, con il quale si
            ungevano le mani recitando la Fatah.
            (Ferrandi 1903: 251)"@it ;
      ontolex:reference onto:FANDHAAL .

fandhaal_entry a ontolex:Word ;
      lexinfo:partOfSpeech lex info:noun ;
      lexinfo:gender lexinfo:masculine ;
      ontolex:canonicalForm [
            ontolex:written Rep "fandhaal"@som ]  ;
      ontolex:sense :fandhaal_s1 .

:fandhaal_s1 a ontolex:LexicalSense ;
      skos:definition "Qaaddo qori ka samaysan."@som ,
            "Wooden spoon."@en
            "Utensile da cucina tradizionale.
            Cucchiaio di legno."@it ;
      ontolex:reference onto:FANDHAAL.
```

Figure 3.The RFD entry of fandal "spoon".

As depicted in the ontological description, the spoon was the sole utensil utilised by the Somalis during that period, prevalent across the entire region.
As evidenced by the accompanying images[7], these objects varied in size and shape, showcasing diverse levels of craftsmanship, ranging from simple wooden ladles devoid of embellishments to intricately carved small spoons made from sturdy hardwood with finely detailed handles. Their usage was significantly distinct from what we can imagine today, as they were exclusively used to mix fried coffee beans in butter or on special occasions to serve honoured guests. They were not used for eating. As it emerges from the Somali corpus, the term *fandhaal* currently denotes the traditional wooden spoon, while the term *mulqaacad* refers to the commonly used steel cutlery.

---

[7] The first image is taken from the SGI archives, the second from Grottanelli (1968).

Figure 4.The ontological formalisation of <SPOON>

Although not all terms and concepts have been formalised to date, it is already possible to appreciate the benefits associated with such formal structuring. Indeed, it is possible to perform queries that take into account either the linguistic dimension, or the conceptual dimension, or both in combination. For instance, through the SPARQL query language, the user can identify how many Swahili or Arabic terms are present in the resource and in which semantic fields they are more concentrated. The following query (Figure 5) is aimed, for example, at identifying all terms that designate artefacts and were borrowed from Swahili.

```
SELECT ?wr
WHERE {
    ?le ontolex:sense [ ONTO:ARTIFACT ] ;
        lexinfo:etymology ?etymology ;
        ontolex:lexicalForm [ ontolex:writtenRep ?wr ] .
    ?etyLex lime:entry ?etymology ;
        dc:language <http://www.lexvo.org/page/iso639-3/swa>
```

Figure 5. A SPARQL query combining the conceptual and linguistic levels.

Among the set of 85 terms related to artefacts, a few indeed have origins in the Swahili language, such as *tana* "comb" from the Swahili word *tana* "to comb" (< *kitana* "small comb"); *jembe* "small hoe"; and *parapanda*, a musical instrument most probably an oboe of non-African origin (Grottanelli, 1976[8]).

---

[8] The remarkable resemblance between the object and an Indonesian breathing instrument, which was imported from China, suggests the possibility that both artefacts share a common origin in an extremely eastern variety of oboe.

Such linguistic nuances serve as tangible evidence of the spread of language, culture, and artefacts from Swahili-speaking peoples into the southern region of Somalia. The significance of these borrowed linguistic terms lies in their ability to unveil the integration of novel objects into the society, to reveal cultural evolution or changes in everyday life, such as the incorporation and assimilation of new material over time. An alternative general query might provide insight into the extensive utilisation of wood as a material. The Somalis, as nomadic pastoralists, developed a material culture that catered to their lifestyle of constant mobility. This led to a preference for objects that were lightweight, portable, and crafted from sturdy materials, such as wood and woven fibres, rather than ceramics.

5. Conclusion

This article presents the development of a Somali-Italian termino-ontological resource focusing on the terms extracted from the notebooks of the Italian explorer Ugo Ferrandi. Despite the importance of handling data with care as previously highlighted, this termino-ontological resource will allow researchers to delve into the terminological and conceptual landscape of Somalia during the early 20th century, providing a deeper insight into a world that, characterised by a robust oral tradition, was at risk of fading into obscurity.

The development of this resource is part of a larger project scheduled for completion in 2025. Once compiled, the data will be accessible on the CLARIN platform through advanced queries as well as specialised visualisation tools.

6. Acknowledgments

7. Bibliographical References

Andrzejewski, B. W. (1971). The Role of Broadcasting in the Adaptation of the Somali Language to Modern Needs. In W. H. Whiteley (Ed.), *Language Use and Social Change: Problems of Multilingualism with Special Reference to Eastern Africa.* Studies Presented and Discussed at the Ninth International African Seminar at University College, Dar es Salaam. Oxford University Press, London, pp. 262-73.

Andrzejewski, B. W. and Lewis, I. M. (1964). Somali Poetry: An Introduction. Clarendon Press, Oxford, 1st edition.

Badel, A. M., Zhong ,T., Tai, W., and Zhou, F. (2023). Somali Information Retrieval Corpus: Bridging the Gap between Query Translation and Dedicated Language Resources. In H. Bouamor, J. Pino, and

K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages* 7463-7469, Singapore. Association for Computational Linguistics (ACL).

Bahar, K., and T.A Ramaha, N. (2023). Exploring Somali Sentiment Analysis: A Resource-Light Approach for Small-Scale Text Classification. *International Conference on Applied Engineering and Natural Sciences* 1 (1): 620-28. https://doi.org/10.59287/icaens.1069.

Biswas, A., Menon, R., van der Westhuizen, E., and Niesler, Th. (2019). Interspeech Improved low-resource Somali speech recognition by semi-supervised acoustic and language model training. In *Proceedings of Interspeech,* pages *3008-3012, Graz, Austria*.

Bonab, H., Allan, J., & Sitaraman, R. (2019, September). Simulating CLIR translation resource scarcity using high-resource languages. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, *(ICTIR)*, pages 129-136, Santa Clara, CA, USA, October 2-5.

Chiarcos, Ch., Nordhoff, S. and Hellmann, S. (2012). Linked Data in Linguistics. Springer, Heidelberg.

Cimiano, P., McCrae, J. P. and Buitelaar, P. (2016). Lexicon Model for Ontologies: Community Report. W3C Community Group Final Report.

Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, 29: 29-42.

Desprès, S., and Szulman, S. (2008). Réseau terminologique versus Ontologie. In *Actes de la deuxième conférence Toth*, *(Toth 2008)* pages 17-34, Annecy, France, june.

Diki-Kidiri, M. (Ed.) (2008). Le vocabulaire scientifique dans les langues africaines, Pour une approche culturelle de la terminologie. Karthala, Paris.

Duh, K., McNamee, P., Post, M., & Thompson, B. (2020). Benchmarking Neural and Statistical Machine Translation on Low-Resource African Languages. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, pages 2667-2675, Marseille, France.

Gavello, A.M. (1975). Ugo Ferrandi, esploratore novarese. Tipografia La Cupola, Novara, 1st edition.

Grottanelli, V. L. (1968). Somali Wood Engraving. *African Arts,* vol. 1, n° 3. UCLA James S. Coleman African Studies Center, pp. 8-13+72-73+96.

Grottanelli, V. L. (1976). Gerarchie etniche e conflitto culturale: saggi di etnologia nordest-africana. Milano, Franco Angeli Editore.

Laryea, J., and Jayasundara, N. (2020). Automatic Speech Recognition System for Somali in the interest of reducing Maternal Morbidity and Mortality. Thesis Högskolan Dalarna, Mikrodataanalys. http://urn.kb.se/resolve?urn=urn:nbn:se:du-34436

Lenci, A. et alii (2000). SIMPLE: A General Framework for the Development of Multilingual Lexicons. *International Journal of Lexicography*, 13(4):249-263.

Marini, E. (1991). Un novarese in Somalia: Ugo Ferrandi (1852-1928): una documentazione inedita. *Bollettino storico per la provincia di Novara 82*.

McCrae, J. P., Bosque-Gil J., Gracia J., Buitelaar P., and Cimiano P. (2017). The Ontolex-Lemon model: development and applications. In *Proceedings of the Electronic Lexicography of the 21st Century conference (eLex 2017)* pages 19-21, Leiden, the Netherlands, September.

Mudimbe, V. Y. (1988). The Invention of Africa: Gnosis, Philosophy, and the Order of Knowledge. Indiana University Press, Indianapolis, 1st edition.

Musse Jama, J. (2016). A Syntactically Annotated Corpus of Somali Literature, PhD thesis [unpublished], University of Naples "L'Orientale". See www.somalicorpus.com.

Piccini, S., Ruimy, N, and Giovannetti, E. (2013). Le lexique électronique de la terminologie de Ferdinand de Saussure : une première. In D. Trotter, A. Bozzi, C. Fairon (Eds.), *Actes du XXVIIe Congrès international de linguistique et de philologie romanes*, Nancy, 15-20 juillet 2013. Section 16 : Projets en cours ; ressources et outils nouveaux. Nancy, ATILF, 255-267. http://www.atilf.fr/cilpr2013/actes/section-16.html

Pustejovsky, J. (1995). The Generative Lexicon. The MIT Press, Cambridge MA.

Roche, C., and Papadopoulou, M. (2019). Mind the Gap: Ontology Authoring for Humanists. *Joint Ontology Workshops*.

Saeed, J. (1982). Central Somali: A grammatical outline. *Afroasiatic linguistics* 8:2. Undena, Malibu.

Santos, C., and Costa, R. (2015). Domain specificity. In *Handbook of terminology* Vol. 1. John Benjamins Publishing Company, Amsterdam/ Philadelphia,153-179.

Shafie Abdi, M., and Muhidin Abdullahi, M. (2023). Lexicon and ruled-based word lemmatization approach for the Somali language. In *Proceedings of the 4th Workshop on African Natural Language Processing* (*AfricaNLP@ICLR 2023*), Kigali, Rwanda, May 1.

Spivak, G. G. (1985). The Rani of Sirmur: an essay in reading the archives. *History and Theory*, 24, 3 (1985), 247-272.

Temmerman, R. (2022). Units of Understanding in Sociocognitive Terminology Studies. In P. Faber, and M. C. L'Homme (Eds.), *Theoretical Perspectives on Terminology: explaining terms, concepts and specialized knowledge*. John Benjamins, Amsterdam, the Netherlands, 331-52.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), 1-9.