SardaNet: a Linguistic Resource for Sardinian Language

Manuela Angioni,
Franco Tuveri
CRS4
Center for Research and
Scientific Studies in Sardinia
Bld. 1, Piscina Manna,
Pula (CA), Italy.
{angioni,tuveri}@crs4.it

Maurizio Virdis,
Laura Lucia Lai,
Micol Elisa Maltesi
University of Cagliari
Cagliari, Italy
virdis@unica.it
llaura@gmail.com
micol-elisa@hotmail.it

Abstract

This paper describes the process of building SardaNet, a linguistic resource for Sardinian language including the different linguistic varieties in Sardinia. SardaNet aims at identifying the semantic relations between Sardinian terms, by manually mapping existing WordNet entries to Sardinian word senses. The work, still in progress, has been developed in collaboration with the University of Cagliari. After discussing some linguistic peculiarities, the paper presents the basic steps of the construction process, the method and the tools involved, the issues encountered during the development and the current version of SardaNet.

1 Introduction

Sardinian territory is characterized by a strong multilingualism, in which it is difficult to trace the precise boundaries between a variant and the other, each characterized by its phonetic, morphological and lexical features.

For a long time, the linguists have been trying to put the distinction between the different linguistic variants spoken in Sardinia, but there is not an unanimously shared theory.

SardaNet examines the Sardinian linguistic variants to which the "Legge regionale 26 del 1997" (Regional law 26 of 1997 for the preservation of linguistic minority) refers: Campidanese, Nuorese and Logudorese, and also the other not-Sardinian variants spoken in the island such as Sassarese, Gallurese, Tabarchino and Algherese.

The ultimate goal is the development of the semantic network related to WordNet[1] (Miller, 1995) and enriched with the peculiar terms and the concepts defined in the Sardinian languages.

SardaNet, in its first preliminary version, has been manually developed, starting with the set of 4689 Common Base Concepts[2] (CBCs) extended by BalkaNet (Tufis et al., 2004) in the Princeton WordNet 2.0 version, by inserting the corresponding terms in the Sardinian variants.

The work has been developed in collaboration with the University of Cagliari. Two trainees, coordinated by the Prof. Maurizio Virdis, leading expert on Sardinian studies, worked with us in this first phase of the project.

The remainder of this paper is organized as follows: Section 1 introduces the resource and the motivation behind, Section 2 presents an overview of the Sardinian language and its peculiarities and the dictionaries used to build SardaNet. Section 3 describes the method applied by the team involved, highlighting some emerging issues, while Section 4 depicts the building of the resource, the interface used, the LMF format and the mapping CILI. Finally, in Section 5 final remarks and future works directions are presented.

2 The Sardinian Language

Before starting the development of SardaNet, several discussions with Sardinian language

[1] Princeton University "About WordNet." WordNet. Princeton University. 2010. <http://wordnet.princeton.edu>
[2] http://globalwordnet.org/gwa-base-concepts/

experts of the University of Cagliari were conducted to better understand the key features of the language.

As it is reported in Virdis (2003a and 2003b), Sardinian is a neo-Latin language, which derives by the evolution of the Latin language, like Italian, French, Spanish, Portuguese and Romanian.

Compared to other neo-Latin languages, Sardinian evidences some peculiarities and even considerable diversities.

The written production in Sardinian language was in fact very plentiful, used in official and documentary written, or in patrimonial and juridical documents, and in particular the *condaghes*. The *condaghe*, from the medieval Sardinian term kondake (from the greek Κοντακιον), was a kind of administrative document used in the Sardinian *Giudicati* between the 11th and 13th centuries. The Condaghe of Santa Maria di Bonàrcado, more in details, allows having a valid source for the philological and glottological studies of the Sardinian language, and in particular, for the Arborensis area (Virdis, 1982).

As for the lexicon, Sardinian lexical heritage is an original amalgam of Latin, of ancient and modern Italian, of Catalan and Spanish, often with unique and distinctive creations and interpretations.

The Sardinian has long lived in a state of increasing marginalization by official uses and has been only restricted to familiar and colloquial use. It has been used mainly in low linguistic registers, or most of all for poetic literary compositions, and sometimes it has been also considered a language of marginal use compared to Spanish and Italian.

Currently, the most difficult phase for Sardinian and for dialects in general seems to be overcoming. People are less afraid to speak in dialect, sometimes rediscovering a pride in speaking the native national language. At Italian level, but even more within the European Community, the cultural, historical and anthropological value of European minority languages, such as Sardinian, is becoming more and more important. Political and cultural actions have been launched to save them, and even the regional politic in Sardinia tends to bring Sardinian language back to schools and proposes projects to perform that.

Sardinian is a particular language: there are a lot a variants in relation of the region considered such as Logudorese, Nuorese, Campidanese, and not native such as Sassarese, Catalano, Gallurese and Tabarchino, as depicted in Figure 1 that displays the geographical distribution of the varieties of Sardinian.

As explained in Virdis (1978), Sardinian language is spoken in Sardinia and only in Sardinia (excluding the large number of emigrants who carry, speak and practice Sardinian language outside of the island). But Sardinian language is not spoken in all Sardinia: in fact, it is necessary to exclude Gallura, where a Southern course dialect is spoken, Alghero where Catalan is spoken, and finally Carloforte and Calasetta where Ligurian is spoken.

The Sassarese has particular historical origins, born in the Middle Age at the time of Pisan-Genovese penetration as a free language due to a contact effect between two linguistic types: the Sardinian and the Italian continental one. Logudorese and Nuorese are mainly spoken in the northern sub-region of the island. Campidanese is the variety of the Sardinian language primarily spoken in South Central Sardinia.

Figure 1. Distribution of the linguistic varieties in Sardinia[3].

[3] The distribution of Sardinian dialects and sub-dialects (Virdis, 1988).

The language has never had a real unification and never linguistic variety above the others has been imposed. Sometimes who speaks a variant of north Sardinia has some difficulties in understanding a variant of south Sardinia and vice versa. So it is sometimes difficult to communicate.

Due to these peculiarities of Sardinian language, we decided, according to Prof. Maurizio Virdis, to insert in the Sardinian WordNet, for each WordNet entry, the indication of the language variation, considering them as synonyms, and following the expand approach.

2.1 The Sardinian Resources

In the construction of SardaNet we have considered different Sardinian lexical varieties, as shown in Figure 1, which present not only phonetic differences but also a multitude of exclusive lemmas: Logudorese, Nuorese, Campidanese, and the other languages spoken in Sardinia, as Sassarese, Catalano, Gallurese and Tabarchino. At present, Tabarchino and Algherese are not included in SardaNet.

Therefore, several dictionaries have been used, someone available only in paper format, mono linguistic or multi linguistic, mostly related to a single variant of Sardinian, others incorporating in a single dictionary the multiplicity of Sardinian variants.

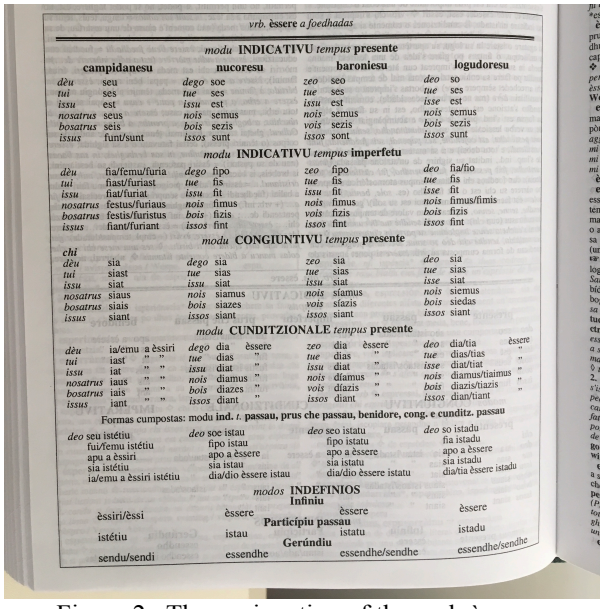
Among the first resources for the Sardinian language there are the Sardinian Campidanese - Italian dictionary written in 1832 by Vincenzo Raimondo Porru (Porru, 2002) and the "Vocabolariu Sardo-Italianu e Vocabolario Italiano-Sardo" a Sardinian-Italian, Italian-Sardinian dictionary, written by Giovanni Spano (Spano, 1998) in the period between 1851 and 1852.

In the period between July 1934 and April 1947 Pietro Casu (Casu, 2002), dedicated many years to the collection of lexical materials, and wrote a manuscript, the "Vocabolario Sardo Logudorese - Italiano" (Sardinian Logudorese - Italian Dictionary), one of the most important works of Sardinian lexicography for the richness of the phraseology included.

More recently, the "Dizionario Etimologico Sardo" (DES) (The Sardinian Etymological Dictionary) (Wagner, 1964), written in three volumes, is certainly a fundamental work for the study of the Sardinian language. It contains the list of all the most relevant words of the Sardinian, which Wagner compares to investigate their source and their meaning. However, its consultation is sometimes complicated due to the incompleteness of the general indexes and the phonetic transcription of the lexical material.

The reprint of the DES dictionary edited by Giulio Paulis (Wagner, 2008) has enriched the indexes and has performed a thorough review of the texts, filling out some gaps in the original version. Paulis also wrote "Introduzione a Max Leopold Wagner, Fonetica storica del sardo" (Paulis, 1984), an introduction to the book, related to the historical phonetic of the Sardinian language, written by Max Leopold Wagner.

The "Dizionario Universale della Lingua di Sardegna" (Universal Dictionary of the Language of Sardinia) (Rubattu, 2001), in two volumes, allows instead a simpler and more immediate use. It contains the terms in the various linguistic varieties of Sardinia, Logudorese, Nuorese, Campidanese, Sassarese, Catalano, Gallurese and Tabarchino, whose distribution is shown in Figure 1, with correspondence in English, French, Spanish and German. It is also available on the Sardinian Digital Library[4].


vrb. èssere *a foedhadas*

modu **INDICATIVU** *tempus* **presente**

campidanesu	nucoresu	baroniesu	logudoresu
dèu seu	*dego* soe	*zeo* seo	*deo* so
tui ses	*tue* ses	*tue* ses	*tue* ses
issu est	*issu* est	*issu* est	*isse* est
nosatrus seus	*nois* semus	*nois* semus	*nois* semus
bosatrus seis	*bois* sezis	*vois* sezis	*bois* sezis
issus funt/sunt	*issos* sunt	*issos* sont	*issos* sunt

modu **INDICATIVU** *tempus* **imperfetu**

dèu fia/femu/furia	*dego* fipo	*zeo* fipo	*deo* fia/fio
tui fiast/furiast	*tue* fis	*tue* fis	*tue* fis
issu fiat/furiat	*issu* fit	*issu* fit	*isse* fit
nosatrus festus/furiaus	*nois* fimus	*nois* fimus	*nois* fimus/fimis
bosatrus festis/furistus	*bois* fizis	*vois* fizis	*bois* fizis
issus fiant/furiant	*issos* fint	*issos* fint	*issos* fint

modu **CONGIUNTIVU** *tempus* **presente**

chi			
dèu sia	*dego* sia	*zeo* sia	*deo* sia
tui siast	*tue* sias	*tue* sias	*tue* sias
issu siat	*issu* siat	*issu* siat	*isse* siat
nosatrus siaus	*nois* siamus	*nois* siamus	*nois* siemus
bosatrus siais	*bois* siazes	*vois* sfazis	*bois* siedas
issus siant	*issos* siant	*issos* siant	*issos* siant

modu **CUNDITZIONALE** *tempus* **presente**

dèu ia/emu a èssiri	*dego* dia èssere	*zeo* dia èssere	*deo* dia/tia èssere
tui iast " "	*tue* dias "	*tue* dias "	*tue* dias/tias "
issu iat " "	*issu* diat "	*issu* diat "	*isse* diat/tiat "
nosatrus iaus " "	*nois* diamus "	*nois* diamus "	*nois* diamus/tiaimus "
bosatrus iais " "	*bois* diazes "	*vois* dfazis "	*bois* diazis/tiazis "
issus iant " "	*issos* diant "	*issos* diant "	*issos* dian/tiant

Formas cumpostas: modu **ind.** *t.* **passau, prus che passau, benidore, cong. e cunditz. passau**

deo seu istétiu	*deo* soe istau	*deo* seo istatu	*deo* so istadu
fui/femu istétiu	fipo istau	fipo istatu	fia istadu
apu a èssiri	apo a èssere	apo a èssere	apo a èssere
sia istétiu	sia istau	sia istatu	sia istadu
ia/emu a èssiri istétiu	dia/dio èssere istau	dia/dio èssere istatu	dia/tia èssere istadu

modos **INDEFINIOS**

Infiniu			
èssiri/èssi	èssere	èssere	èssere
Partitzipiu passau			
istétiu	istau	istatu	istadu
Gerúndiu			
sendu/sendi	essendhe	essendhe/sendhe	essendhe/sendhe

Figure 2. The conjugation of the verb èssere.

Another reference dictionary is "Su Ditzionàriu de Sa Limba e de sa Cultura Sarda" (The Dictionary of the Sardinian Language and Culture) (Puddu, 2015), written entirely in Sardinian language with a partial matching of the

[4] Dizionario universale della lingua di Sardegna : I e II volume, Sardegna Digital Library: http://www.sardegnadigitallibrary.it/index.php?xsl=2435&s=17&v=9&c=4459&c1=Rubattu+Antoninu&n=24&ric=1&idtipo=0

words into five languages: Italian, English, French, Spanish and German. Puddu in his dictionary uses the linguistic variant defined as "Limba de Mesania" (Language of Mesania), a variant located beyond the external arborese border area, around the city of Sorgono, between the two macro-areas Logudorese and Campidanese, as shown in Figure 1.

The Figure 2 shows the conjugation of the verb *èssere* (in English *to be*) in some Sardinian languages as reported in Puddu (2015).

The information needed to build the language resource in the format required by the Global WordNet Association is not always included in the available dictionaries. As for the definitions, for example, Rubattu provides them only for verbs while Puddu puts them but written in the Mesania language.

An indispensable research manual for everyone interested in the Sardinian language and in Romance linguistics in general is the "Manuale di linguistica sarda" (Manual of Sardinian Linguistics) (Ferrer et al., 2017). It presents an overview of the problems of Sardinian linguistics with a detailed introduction of the current linguistic situation in Sardinia completed by a description both of the varieties of Sardinian itself and of the other languages spoken on the island.

3 Methodology

The work, still in progress, was performed manually taking advantage of the involvement and the linguistic expertise of some trainees belonging to the University of Cagliari, coordinated by the Prof. Maurizio Virdis.

The construction of the resource is based on the list of the 4689 Common Base Concepts expanded by BalkaNet from the initial set of 1024 Common Base Concepts developed in the European project EuroWordNet (Vossen, 1998)

SardaNet is created by using the *expand approach*, starting from the multilingual index and translating the English various synsets into the Sardinian language. This approach is more attractive since it maintains the multilingual index as the main structure and central repository of concepts and also allows to automatically using semantic relations already present in the English WordNet.

According to Bond et al. (2016), the majority of wordnets are based on the expand approach, exactly 28 out of 33 of the wordnets included in the OpenMultilingual Wordnet (OMW)[5].

3.1 Insertion and Validation

The collaboration with the trainees started by choosing the most suitable dictionaries and resources among those available as described in the Section 2.1. After then, an English term was selected and, for each of its meanings identified by a different synset ID and a gloss, the synonyms in the Sardinian language were assigned in all the variants considered.

The identification of Sardinian terms to be inserted into SardaNet in correspondence of the selected English terms was carried out through the consultation of the various Sardinian dictionaries. They display, besides the Sardinian term and its linguistic varieties, the definition of the same term in Italian or in Sardinian language and sometimes some examples of usage of the term, that help to understand its real meaning, and its translation into several languages, including Italian and English.

Each term inserted in SardaNet has been verified and confirmed by at least two people.

In case of discrepancy the team discussed in order to find an agreement and, when it was not possible, the terms involved were excluded.

3.2 Examples and Issues

During the building of the resource we have sometimes faced the problem of translation equivalence and the lack of correspondence of the Sardinian language with the about 5000 English senses.

As expected, some technical terms do not have correspondence in Sardinian language.

The term *mouse,* as *a hand-operated electronic device (synset ID WN3.1 = 03799022, noun)*, does not have a corresponding sense in the Sardinian language. Other terms, i.e. *website, a computer connected to the internet that maintains a series of web pages on the World Wide Web,* could be translate with the Sardinian terms *giassu* (L) and *zassu* (N).

In general, terms belonging to specific domains, such as biology or chemistry, do not have an equivalent term in Sardinian. The English term *state,* as a chemical state of matter *(synset ID WN3.1 =14503199, noun)*, does not have a Sardinian equivalent sense.

The sense of the term *cell,* as *electric_cell*, *a device that delivers an electric current as the*

[5] http://compling.hss.ntu.edu.sg/iliomw/omw

result of a chemical reaction (*synset ID WN3.1 = 02994503, noun*), is not present in the available Sardinian dictionaries.

Sometimes the linguists have experienced difficulties to look for the right corresponding sense of some English concepts in WordNet. Despite of this, a concept such as *creating_from_raw_materials*, defined as *the act of creating something that is different from the materials that went into it* (*synset ID WN3.1 = 00910607*) could be simply translated in the Sardinian verb *bogai (C)*.

We observed that the key factor is that many English terms, especially neologisms and technical terms, often not have a correspondence in the Sardinian language. So frequently Italian terms are used instead. On the contrary, there are many common saying, as part of the juvenile language (Ferrer et al., 2017), that hardly find correspondence not only in English but also in Italian.

4 Building SardaNet

In order to build SardaNet, it was necessary to set up an interface able to display for each term in English its synonyms, the corresponding synset IDs, the POS, the definition, and allowing the terms to be included in each of the variants of the Sardinian language.

4.1 The application

Developed in PHP, the application allows the insertion of Sardinian terms starting from the English ones into different ways.

It is an evolution of a previous application developed for FreeWordNet, (Tuveri and Angioni, 2012a; Tuveri and Angioni, 2012b), a linguistic resource, still not released, based on WordNet. FreeWordNet was born as a possible extension of WordNet in Opinion Mining related context.

In FreeWordNet each synset is enriched with a set of properties related to adjectives and adverbs and has a positive, negative or objective value associated. The properties associated to each synset support a better identification of the sentiment expressed in relation to the domain and give more details about the relevant terms or the expressions having an opinion associated. SardaNet inherits the same properties from adjectives proposed in FreeWordNet but they have not been inserted in this first release.

The interface permits to insert any word in English starting from the about 5000 word senses in the set of CBCs.

The starting form, shown in Figure 3, offers 3 different options.

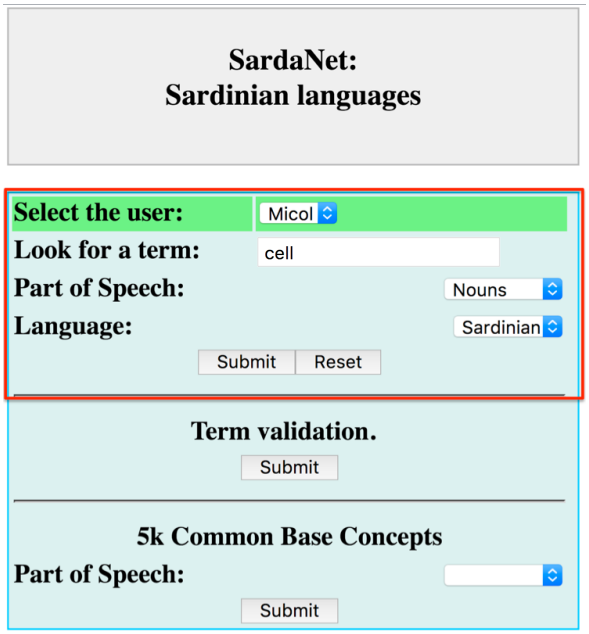


Figure 3. The access interface.

In the first one, followed by an use case, the user "Micol" can modify the entered information by indicating the term.

In the second, the user "Micol" will verify and confirm the items previously entered by another user, by simply submitting the button related to the "Term Validation". The application allows to show only the terms to be validated, that is, those entered by a person and that needs to be confirmed by at least another person. During this process, the user can also erase incorrectly entered terms or variants.

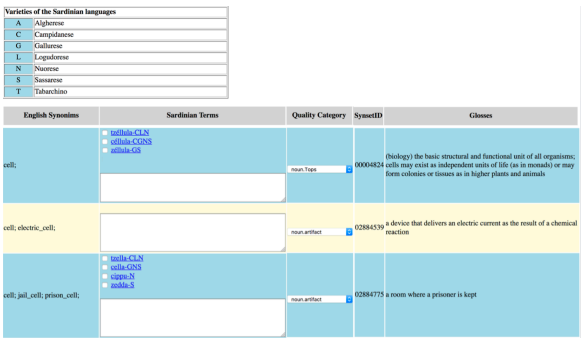


Figure 4. Mapping cell-related terms.

In the third option, the set of the 4689 CBCs is displayed. The user can select each term in English and insert the corresponding terms in the

Sardinian language. The interface will show all the synsets related to the selected word in English in term of polysemy, the definition, the synset ID and the gloss, as the Figure 4 shows. New terms entered or confirmed are saved in the database.

In the Figure 3 the access form to SardaNet is shown. The user "Micol" looks for the noun "cell" in SardaNet.

Figure 4 shows the already inserted Sardinian terms, having possible cellular related meanings, and the information provided by WordNet 3.1. The inserted values can also be changed.

By selecting one of the links referring to the noun "cell", the user can insert new corresponding nouns in Sardinian language, delete or edit the existing ones, for example by modifying the associated language varieties, as shown below in Figure 5.

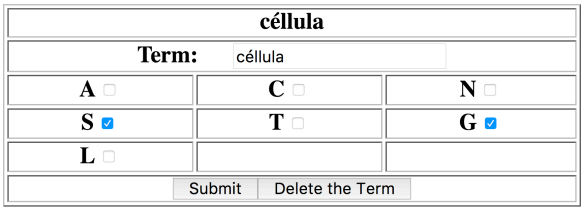


Figure 5. How to modify the language varieties of a term.

4.2 Format and CILI Mapping

The used base concepts include the WordNet synsets in version WN2.0. It was therefore necessary to mapping the synsets from this version to WN3.1 version. This work has been carried out in two steps, from version WN2.0 to WN3.0, from WN3.0 to WN3.1.

The mapping from version WN2.0 to WN3.0 of WordNet was done thanks to the work performed by Tufiş et al. (2011), through the resource available at http://nlptools.racai.ro/.

The mapping from the WN3.0 to the WN3.1 WordNet version has been made possible thanks to the work of John McCrae, through the *git* project: https://github.com/globalwordnet/ili.

Thanks to his work we put together entirely the mapping WN2.0 - WN3.0 - WN3.1, ILI indexes included.

The starting set of about 5000 CBCs is expressed in WN2.0 and a mapping to the WN3.1 version and to the associated ILI indexes has been necessary.

A subset of them has been used in the first edition of SardaNet and we decided to define the resource, right from the beginning, in the LMF (Lexical Markup Framework) format, as required by the Global WordNet Association.

The mappings are also available on request in the SQL format too.

We try to remain as faithful as possible to the CILI, Collaborative Interlingual Index (Bond et al., 2016), but the building of the LMF evidenced almost an additional requirement related to the SardaNet linguistic variants that has been indicated adding the "Tag" element to the "Lemma" item, built in the following way:

```
<Lemma writtenForm="bóitu" partOfSpeech="n">
    <Tag category="variants">AC</Tag>
</Lemma>
```

Figure 6 shows a portion of SardaNet in LMF format with the addition of the linguistic variant.

```
<LexicalEntry id="wn31-céllula-n">
  <Lemma writtenForm="céllula" partOfSpeech="n">
    <Tag category="variants">ACGL</Tag>
  </Lemma>
  <Sense id="wn31-00006484-n" synset="wn31-00006484-n">
    <SenseRelation target="wn31-00327929-a" relType="derivation"/>
    <SenseRelation target="wn31-02696036-a" relType="derivation"/>
  </Sense>
</LexicalEntry>
```

Figure 6. The LMF format with the addition of the Sardinian variant.

However, in the first release of the resource in the LMF format we leave out the glosses in the Sardinian language. In fact, there is a scarcity of Sardinian online dictionaries and the most complete available ones not always have a gloss or a sentence defining the usage of the specific term.

An automatic process is not always possible, so we will probably proceed manually with the help of students of Linguistics.

Currently we are also not able to calculate the frequency of use of the Sardinian terms for each synset. So, in the first release of SardaNet it is not provided.

4.3 Current Status of SardaNet

The quantitative data pertaining to the Sardinian WordNet are summarized in the tables below.

Table 1 shows the couple of synsets and Sardinian terms inserted into SardaNet, the total number of distinct synsets and the number of distinct Sardinian terms.

As you can notice, SardaNet includes a lot of terms for each synset. It is due both to several synonyms related to each sense, typical of the

Sardinian language, and to the presence of the several variants of the language.

Synsets -Terms	Distinct Synsets	Terms
21025	1601	9899

Table 1. Synsets and terms in SardaNet.

The following Table 2 reports the distribution of terms and synsets in SardaNet for each part of speech (POS), referred to the number of couple of synsets and Sardinian terms, the number of validated synsets and the number of Sardinian terms.

POS	Synsets Terms	Distinct Synsets	Terms
Nouns	18885	1452	8920
Adjectives	1657	130	685
Verbs	483	19	294

Table 2. Distribution of synsets and terms for POS.

The results above show the prevalence of nouns among the other parts of speech. Translating English nouns into Sardinian nouns seems to be more intuitive and immediate and involves fewer problems than verbs, adjectives and adverbs. The resource does not yet include any adverb.

Variant	Synsets Terms	Distinct Synsets	Terms
C	5622	1584	2787
G	4097	1550	1947
L	8219	1590	3828
N	5738	1581	2738
S	3076	1560	1505

Table 3. Distribution of the Sardinian variants in SardaNet.

As depicted in Table 3, among all the Sardinian variants, Logudorese, Nuorese, Campidanese, Gallurese and Sassarese include in SardaNet the largest number of terms, while Algherese and Tabarchino are not yet considered in the resource.

Despite the application does not calculate the correct percentage of the coverage across the dialects, we found that the total coverage of validated terms in SardaNet is about 16,5%, 775 senses on 4689 of the CBCs. Nevertheless SardaNet contains a total of 1601 senses, 826 not included in the CBCs. These senses come out because the application shows, for each sense included in the CBCs, all the senses related by the polysemy property.

5 Conclusions and Future Works

In its first release, SardaNet includes only a partial set of the 4689 Common Base Concepts expanded by BalkaNet from the initial set of 1024 Common Base Concepts developed in the European project EuroWordNet. So we are first planning to complete all the senses available in the set of CBCs. Although there are many terms, common saying and phrases typical of Sardinian language, they are not currently present in SardaNet. We leave out the glosses in the Sardinian language and the frequency of use of the Sardinian terms for each synset, even if they are required in the LMF format.

Further works will include both the glosses and the frequency of the terms, that will be calculated both manually, using the available dictionaries, and automatically by means of a corpus of Sardinian documents. We are also taking into account to enrich SardaNet with new terms, currently not included in the English WordNet, but characteristic of the Sardinian language.

Acknowledgments

We wish to thank Georgia Sanna for the first joint analysis about the possibility of creating a wordnet for the Sardinian language.

References

Pietro Casu. 2002. *Vocabolario sardo logudorese-italiano*. Ed. Giulio Paulis, Nuoro, Ilisso, ISBN 88-87825-36-X.

Francis Bond, Piek Vossen, John McCrae, Christiane Fellbaum. 2016. *CILI: the Collaborative Interlingual Index*, in: Proceedings of the 8th Global WordNet Conference 2016 (GWC2016) in Bucharest, Romania, January 27-30.

Eduardo Blasco Ferrer, Peter Koch, Daniela Marzo. 2017. *Manuale Di Linguistica Sarda*. Ed. de Gruyter Mouton. ISBN 3110274507, 9783110274509

George A. Miller. 1995. *WordNet: A Lexical Database for English*. Communications of the ACM Vol. 38, No. 11: 39-41.

Giulio Paulis. 1984. *Introduzione a M.L. Wagner, Fonetica storica del sardo (trad. italiana di Historische Lautlehre des Sardischen, Halle, Niemeyer, 1941)*. Casteddu/Cagliari, Gianni Trois Editore, pp. VII-CX.

Vincenzo Raimondo Porru. 2002. *Nou dizionariu universali Sardu - Italianu.* A cura di Lörinczi Marinella. Ilisso Edizioni, Nuoro.

Mario Puddu. 2015. *Ditzionàriu de sa limba e de sa cultura sarda.* 2896 p., ed. Condaghes, 2 ed., Collana: Ainas.

Antoninu Rubattu. 2001. *Dizionario universale della lingua di Sardegna.* EDES. Collana: Lingua e letteratura. 2 voll., 2254 p., EAN: 9788886002394.

Giovanni Spano. 1998. *Vocabolariu Sardo-Italianu e Vocabolario Italiano-Sardo.* 4 vol. Ed. Giulio Paulis, Nuoro, Ilisso.

Dan Tufiş, Dan Cristea, Sofia Stamou. 2004. *BalkaNet: Aims, methods, results and perspectives. A general overview.* Science and Technology 7(1/2): 9–43.

Dan Tufiş, Radu Ion, Verginica Barbu Mititelu, Elena Irimia, Dan Ştefănescu, Cătălin Mihăilă. 2011. *Extending and completing the Ro-WordNet lexical ontology by eliminating the existing semantic conflicts and by validating the differential semantics model based on Ro-WordNet.* Academic report (in Romanian). Bucharest, Romania.

Franco Tuveri, Manuela Angioni. 2012a. *Definition of a Linguistic Resource for Opinion Mining.* Proceedings of the 9th International Workshop on Natural Language Processing and Cognitive Science – NLPCS - SciTePress - june 2012. ISBN: 978-989-8565-16-7

Franco Tuveri, Manuela Angioni. 2012b. *A Linguistic Approach to Feature Extraction Based on a Lexical Database of the Properties of Adjectives and Adverbs.* Proceedings of the 6th International Global Wordnet Conference Tribun EU pages 365-370 Christiane Fellbaum, Piek Vossen Global WordNet Association. ISBN: 978-80-263-0244-5

Maurizio Virdis. 1978. *Fonetica del dialetto sardo campidanese*. Cagliari, Edizioni della Torre.

Maurizio Virdis. 1988. *Sardisch: Areallinguistik (aree linguistiche).* In: Lexikon der Romanistischen Linguistik, vol. IV, Italienisch, Korsisch, Sardisch. A cura di Günter Holtus, Michael Metzeltin, Christian Schmitt, Tübingen, Niemeyer, 1988, pp. 897-913.

Maurizio Virdis. 1982. *Note sui dialetti dell'area arborense e la lingua del Condaghe di Santa Maria di Bonarcado.* In il Condaghe di Santa Maria di Bonarcado, riedizione del testo di Enrico Besta a cura di Maurizio Virdis, Oristano, S'Alvure.

Maurizio Virdis. 2003a. *La lingua sarda fra le lingue neolatine: storia uso e problemi, in: La lingua e la cultura della Sardegna.* Rapporto del Convegno internazionale. La lingua e la cultura della Sardegna. Tokyo, 9-10 maggio 2003. (pp. 15-24). TOKYO: Waseda University (JAPAN).

Maurizio Virdis. 2003b. *Tipologia e collocazione del sardo tra le lingue romanze.* In «IANUA. REVISTA PHILOLOGICA ROMANICA (on line).», 4.

Piek Vossen (ed.). 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks.* Dordrecht: Kluwer Academic Publishers.

Max Leopold Wagner. 1960-64. *D.E.S. - Dizionario etimologico sardo, DES.* 3 Vol. Heidelberg, Winter.

Max Leopold Wagner. 2008. *D.E.S. - Dizionario etimologico sardo*. 2 Vol. Ed. Giulio Paulis, Nuoro, Ilisso. ISBN 978-88-6202-030-5.