

# Navigating Challenges of Multilingual Resource Development for Under-Resourced Languages: The Case of the African Wordnet Project

Marissa Griesel, Sonja Bosch

University of South Africa (UNISA)

Pretoria, South Africa

{griesm, boschse}@unisa.ac.za

## Abstract

Creating a new wordnet is by no means a trivial task and when the target language is under-resourced as is the case for the languages currently included in the multilingual African Wordnet (AfWN), developers need to rely heavily on human expertise. During the different phases of development of the AfWN, we incorporated various methods of fast-tracking to ease the tedious and time-consuming work. Some methods have proven effective while others seem to have little positive impact on the work rate. As in the case of many other under-resourced languages, the expand model was implemented throughout, thus depending on English source data such as the English Princeton Wordnet (PWN) which is then translated into the target language with the assumption that the new language shares an underlying structure with the PWN. The paper discusses some problems encountered along the way and points out various possibilities of (semi) automated quality assurance measures and further refinement of the AfWN to ensure accelerated growth. In this paper we aim to highlight some of the lessons learnt from hands-on experience in order to facilitate similar projects, in particular for languages from other African countries.

**Keywords:** multilingual wordnet; under-resourced languages; African languages

## 1. Introduction

The African Wordnet (AfWN) project has as its aim the development of wordnets for indigenous languages, including Setswana, isiXhosa, isiZulu, Sesotho sa Leboa, Tshivenda, Sesotho, isiNdebele, Xitsonga and Siswati. The most recent development phase is funded by the South African Centre for Digital Language Resources (SADiLaR)<sup>1</sup> and runs from 2018 to the end of February 2020, with an extension to 2022 currently under consideration. A next version of the AfWN language resources is currently being prepared for distribution and will be made available under the same conditions and on the same platform as the first versions for the initial five languages (UNISA, 2017). Also see Bosch and Griesel (2017) for a detailed description.

While the focus in the past was on the official South African languages, the project also strives to establish a network of projects across Africa with teams representing other African languages adding their own wordnets. In this presentation we hope to share some of the unique challenges and obstacles encountered during the development of a technically complex resource with very limited to no additional natural language processing (NLP) tools. We discuss examples of linguistic idiosyncrasies and suggest ways to represent these examples in a formal database such as a wordnet. Furthermore, we also look at some common pitfalls when using English (UK or US) as source language for manual resource development, opening the floor to further discussion between language experts, developers and users of African language resources so as to ensure the usefulness thereof within the rapidly expanding African human language technology (HLT) and digital humanities (DH) spheres.

## 2. Background

### 2.1 The AfWN Project

Ordan and Wintner (2007) as well as Vossen *et al.* (2016) describe two common methods used to develop wordnets, based on the number and size of additional resources available, the experience of the team and the underlying grammatical structure of the language being modelled. The *merge approach* is popular for new wordnets with ample additional resources such as bilingual dictionaries, descriptive grammars and text corpora. Wordnets are constructed independently from any existing wordnets as a stand-alone resource, after which a separate process is followed to align the newly created wordnet with the Princeton WordNet (PWN) (Princeton University, 2020; and Fellbaum, 1998). PolNet, a Polish wordnet (Vetulani *et al.* 2010) that is based on a high-quality monolingual Polish lexicon, is a good example of a project following this approach. In the case of less-resourced languages, the PWN can be used as template in which to develop a new wordnet. This is referred to as the *expand model* according to which the source wordnet, usually the English PWN, is translated into the target language, with the assumption that the new language shares an underlying structure with the PWN. The Croatian Wordnet is an example of a wordnet based on the expand model due to a lack of semantically organized lexicons (cf. Raffaelli *et al.*, 2008:350).

Typically, wordnets following this approach do not have access to many other digital resources and rely heavily on the linguistic knowledge of the development team. The AfWN project also followed the latter approach to build wordnets for the indigenous languages in a staggered but parallel manner. Initially, the project only included four languages – isiZulu, isiXhosa, Setswana and Sesotho sa Leboa – to allow the team to gain experience and to set up the infrastructure for further expansion. Once the project

<sup>1</sup> <https://www.sadilar.org/>

was established and more funding was secured, Tshivenda and later the remaining languages were added. The team also initially focussed on only providing usage examples to the basic synsets (including a lemma, part of speech and domain tags) and only during a third development stage

started adding definitions to the existing synsets<sup>2</sup>. Some of the languages had more resources available than others and the next section will give a brief overview of the different experiments performed to utilise as many available resources as possible.

## 2.2 Limited Available Resources for Some Languages

As reported in Griesel & Bosch (2014) initial manual development of the wordnets was a time consuming and tedious process. Not only were the team still learning the finer details of this type of language resource development, but linguists had to choose which synsets to translate without much help from electronic resources and only added roughly 1000 synsets per language per year. It was clear that more creative ways to speed up the development would have to be implemented if the project were to grow to a useful size within a practical time frame. It is also important to note again that almost all the linguists and language experts making up the AfWN team were working on this project on a part-time basis. Any degree of fast-tracking would therefore also be beneficial in easing their workload.

One experiment included using very basic bilingual wordlists found on the internet to identify synsets in the PWN that could be included in the AfWN semi-automatically. It involved matching an English term with the most likely PWN synset and then extracting the applicable English information such as a definition, usage example and classification tags from that synset into a spreadsheet with the African language translation of the lemma. Linguists then could easily translate these sheets before they were again included in the wordnet structure in the same position as the identified PWN synset. Griesel & Bosch (2014) give a complete overview of the resources that were used for Setswana, Sesotho sa Leboa, Tshivenda, isiXhosa and isiZulu to add just over 8000 new synsets to the AfWN.

Unfortunately, this method could not be followed for all languages as a basic resource such as freely available, digital, bilingual wordlists do not even exist for all the South African languages. Linguists have to rely solely on their own knowledge, underpinned by commercial (hardcopy) dictionaries and private databases. For these language teams, working in groups with constant communication between the linguists was essential as they performed the mostly manual development task.

## 3. The SILCAWL List

### 3.1 SILCAWL List as Alternative to Other Seed Lists

Another key challenge for the AfWN project was deciding on which concepts to include at which stage of

development. It may seem logical to move alphabetically through the English source data and simply translate every synset but taking into consideration the capacity available in the project, this decision becomes less trivial. As mentioned previously, the project depends heavily on part-time team members and also on securing funding for limited periods of time. To translate all 250 000 synsets in the PWN would therefore take years and the AfWN would not be very useful for further NLP applications until the complete A – Z translation has been performed. As we later discuss in section 5, many lexical gaps exist between the PWN and the African languages and including only synsets also found in (American) English would result in a very flat meaning representation in the AfWN, with many concepts unique to the African context being omitted.

At the onset of the AfWN project, the team followed the example of many other wordnet projects such as the Catalan wordnet (Benítez *et al.* 1998) and the IndoWordnet project (Prabhu *et al.* 2012) and started with the translation of the so called “common base concepts” (CBC; created in the BalkaNet project<sup>3</sup>). This list is regarded as the building block for common semantic relations and is derived from comparing frequency lists for all of the Balkan languages included in that project to find the common set of 5 000 concepts to use as seed list (Weisscher, 2013). However, as discussed in Griesel *et al.* (2019) it soon became apparent that this Eurocentric list would not be ideal for further use in the AfWN project as it contained many concepts that were not lexicalised in the African languages.

Upon further research, the development team decided to employ the SIL Comparative African Wordlist (SILCAWL), which was compiled in 2006 by Keith Snider (SIL International and Canada Institute of Linguistics) and James Roberts (SIL Chad and Université de N'Djaména). This bilingual English-French wordlist includes 1 700 words compiled after extensive linguistic research in Africa. An interesting comparison between the usefulness of the CBC and the SILCAWL lists for expansion of the AfWN is drawn in Griesel *et al.* (2019) indicating that the SILCAWL list to be much better suited to the needs of the AfWN. The most significant enhancement is observed against the background of localisation where the content (of the entries) is lexicalised within an African environment, thereby guarding against datasets that may perpetuate culturally and cognitively biased language resources. This list was therefore used, not only to expand the five languages that formed part of the first two development stages, but also as starting point for the remaining four languages added in the most recent third stage. Xitsonga, Sesotho, Siswati and isiNdebele would therefore include as their first synsets entries from this more localised list.

### 3.2 Translation Procedure

In an effort to fast-track development, it was decided to first add (South African) English definitions and usage examples to the SILCAWL list and then to translate the data into the African languages. The first step would be done by an English lexicographer and expert translators

<sup>2</sup> See Bosch and Griesel (2017) as well as Griesel *et al.* (2019) for a detailed description of the development process followed thus far.

<sup>3</sup> See <http://www.dblab.upatras.gr/balkanet/>

rather than our core project team, where possible, allowing the different tasks to run simultaneously and thereby saving time.

The SILCAWL list only contains an English and a French lemma with very little information by which to disambiguate the implied meaning. In order to maintain the mapping to the PWN as far as possible, the first step was to determine which of the lemmas are included in the PWN and to extract all possible synsets for each lemma. Each candidate synset was then scrutinised manually by the development team and the best possible meaning representation selected from the possible senses. The PWN ID, definition and usage example (where available) were also added to the SILCAWL list. 41 SILCAWL lemmas were however not found in the PWN at all and the definitions or usage examples for many of the other lemmas needed revision in order to create a standardised, localised English dataset that could be translated to the African languages.

The project team, who are experienced wordnet developers after more than 13 years in the AfWN, next used the resulting translations to create synsets, complete with semantic relations in WordnetLoom (cf. Naskret et al., 2018), an open wordnet editor, with elaborated visualization for wordnet structures. The project team would also still work in teams of at least two language experts for each language so as to perform manual verification and quality assurance on the AfWN content throughout the development process.

An AfWN style guide was drawn up and sent to both the English lexicographer as well as the African language translators. This document included details on the translation and formatting of usage examples and definitions, including guidelines on the following aspects:

- No sentence initial capitalisation or punctuation at the end of a sentence is to be included;
- A specific tag should be used to reference definitions taken from the Open Educational Resource Term Bank (OERTB<sup>4</sup>);
- Examples of well formulated definitions and usage examples;
- A reminder not to include any usage examples from proprietary sources such as dictionaries;
- The lemma or head word of a synset also needs to be included in the usage example, but not in the definition;
- etc.

## 4. Evaluation of Translations

During discussions with linguists regarding the new synsets created from the expanded and translated SILCAWL list, many language-specific as well as general concerns were raised. The most notable two categories of concerns were those of a technical nature where the style guide was not adhered to or where mandatory fields were filled incorrectly, as well as issues that had to do with differences between the English source language and the nine African

target languages. Some examples of each category as well as important decisions made are discussed below.

### 4.1 Technical Errors

Smrz (2004) as well as Miháltz *et al.* (2008) describe several ways to perform automatic and semi-automatic quality assurance on wordnets. These heuristics involve structural checks such as making sure only valid values are entered into specified fields (for instance for the POS, SUMO and MILO domains and semantic relations) which need to be referred back to a language expert for revision, as well as formatting checks (for instance eliminating sentence initial capitalisation or sentence ending punctuation) which could be solved automatically. The development team also began initial experiments to incorporate many of these checks/corrections in simple SQL queries or scripts which will result in a more cohesive and standardised resource. Figure 1 shows some of the basic errors found in the isiZulu wordnet, including capitalisation and punctuation mistakes, duplicate usage examples and English usage examples in the African language field.

Figure 1. Automatic extraction of errors in the isiZulu wordnet.

It is envisioned that a language independent quality control pipeline could be established to incorporate these automatic and semi-automatic corrections. A simple user interface built on top of such a pipeline could present problematic synsets/fields to a language expert one at a time with options to accept an automatically generated correction or reject and manually correct possible errors. The second category of errors, namely language specific decisions, would be more complicated to identify automatically and would almost always require human intervention to solve.

### 4.2 Language Specific Decisions

#### 4.2.1 Euphemisms

The African languages often make use of euphemisms to refer to taboo terms, especially terms related to the human body. One such example in Xitsonga is for the concept of “breaking wind/farting” where the biological translation would be *tamba* but the preferred euphemism is *humesa moyo* – literally translated as “to kill an insect”. In isiZulu, the biological term for “clitoris” is *umsunu*, however, *ubhontshisi*, the euphemism literally meaning “bean”, is preferred. Discussions with the translators and the wordnet experts made it clear that, although the scientific term exists, it is very rarely used and considered vulgar and inappropriate language in most contexts. The team therefore decided to include both terms – the taboo and the euphemism – with a tag marking them as such in the wordnet.

<sup>4</sup> See <http://oertb.tlterm.com/about/>

#### 4.2.2 Lexical Gaps or Lexicalisations Between English and the African Languages

A typical example of lexical gaps existing in the PWN, is the intricate system of kinship terms in the African languages that needs to be made provision for in the AfWN. The following table provides a few examples that demonstrate how the English kinship relations “uncle” and “aunt”, as well as the “in-laws” need to be expanded for the target languages isiZulu and Sesotho sa Leboa in the AfWN (also cf. Griesel et al., 2019):

SILCAWL	ISIZULU	SESOTHO SA LEBOA
<b>BLOOD RELATIONS</b>		
0348 father's brother (uncle)	<i>ubabamkhulu</i> (big father) 'father's elder brother' <i>ubabomncane</i> (small father) - 'father's younger brother'	<i>ramogolo</i> 'father's elder brother' <i>rangwane</i> 'father's younger brother'
0351 father's sister (aunt)	<i>ubabekazi</i> (female father) 'father's sister'	<i>rakgadi</i> 'father's sister'
0349 mother's brother (uncle)	<i>umalume</i> (male mother) 'mother's brother'	<i>malome</i> 'mother's brother'
0350 mother's sister (aunt)	<i>umamekazi</i> (female mother) or <i>umame</i> 'mother's sister'	<i>mmamogolo</i> 'mother's elder sister' <i>mmane</i> 'mother's younger sister'
<b>MARRIAGE RELATIONS</b>		
0365 father-in-law	<i>ubabezala</i> 'father-in-law' used by Zulu-speaking woman <i>umukhwe</i> 'father-in-law' used by Zulu-speaking man	<i>ratswale</i> 'father-in-law'
0366 mother-in-law	<i>umkhwekazi</i> 'mother-in-law' used by Zulu-speaking man <i>umamezala</i> 'mother-in-law' used by Zulu-speaking woman	<i>mmatswale / mogwegadi</i> 'mother-in-law' (man speaking – dialectal) <i>mmatswale</i> 'mother-in-law' (woman speaking)
0367 brother-in-law	<i>umfowethu</i> 'husband's brother' <i>umkhwenyawethu</i> 'sister's husband' (man speaking) <i>umlamu</i> 'wife's brother' <i>umkhwenyana</i> 'sister's husband'	<i>molamo, sebara</i> 'sister's husband' (man and woman speaking) <i>molamo, sebara</i> 'wife's brother' (man speaking)

	(woman speaking)	
0368 sister-in-law	<i>udadewethu</i> 'husband's sister' <i>umakoti, umlobokazi, umkami</i> 'brother's wife' (man speaking) <i>umlamu</i> 'wife's sister' <i>umakoti womfowethu, umakoti womnewethu</i> 'brother's wife' (woman speaking)	<i>mogadibo</i> 'husband's sister' / 'brother's wife'

Table 1. Lexical gaps between the source language English and the target languages isiZulu and Sesotho sa Leboa.

With regard to lexicalisation in the African languages, an example in isiZulu is the verb *finya* “blowing the nose”. This example of lexicalisation prevents the noun *ikhala* “nose” from featuring in the usage example:

1. nose ENG20-05278188-n *ikhala* “blow your nose after you sneeze”  
*finya emuva kokuthimula*

Translating from English to isiZulu without knowledge of the wordnet structure and the stipulated guideline that the usage example needs to include the lemma, results in a semantically acceptable sentence but would confuse a user of the wordnet. In other words, a more suitable usage example should be suggested by the linguist, e.g.

2. “the boxer injured his nose”  
*umlobi wesibhakela walimala ekhaleni lakhe*

Numerous examples of concepts that are not lexicalised in the African languages were also encountered. Linguists who were unfamiliar with wordnet development and deemed it necessary to adhere stringently to the CBC list then included descriptions of these terms comprising up to 7 words as the lemma, rather than choosing a more suitable PWN sense or omitting the synset completely in the African language. This took valuable time and led to frustration on the side of the linguists as they were constantly busy coining new descriptions rather than adding more frequently used concepts to the wordnet. The English concept of a “complication” (ENG20-13271751-n; any disease or disorder that occurs during the course of (or because of) another disease) was for instance translated as *izinkinga zokugula ezidalwa ukuba khona kokunye ukugula*, literally meaning “disease problems caused by the presence of another disease”.

#### 5. Suggestions for Improvement

Given the types of stumbling blocks and language specific idiosyncrasies observed throughout the development process, including quality assurance, the project team

suggests the following improvements in the protocol. Some of these aspects were immediately implemented while some will require future work.

One of the first measures to improve the translated data to better fit the wordnet application, is to make sure that the English lexicographer as well as the African language translators are well informed about the ultimate use of their work. The style guide was expanded to include updated instructions and examples of suitable definitions and usage examples. A section was also added on quality assurance and the types of errors to be especially mindful of. Since the linguists all work in the AfWN project on a part-time basis, the team is constantly growing to include more linguists or replace those who no longer have time available. Continuous training of new linguists at the hand of the extended style guide is therefore more effective as well.

Adding morphological analysis or lemmatisation in the pipeline for purposes of quality assurance, for instance in order to verify that the lemma or head word of a synset is included in the usage example, requires further experimentation but will greatly reduce the amount of confusing usage examples. In direct searches, the lemma or head word can easily be obscured by inflection and morphophonological alternations, particularly in conjunctively written languages. For instance, in the following examples:

3. isiZulu  
thigh ENG20-05243922-n *ithanga*  
“she has a huge bruise on her thigh”  
*unomhuzuko omkhulu ethangeni lakhe*

The noun *ithanga* “thigh” is used in the locative form in the usage example, viz. *ethangeni*.

4. isiXhosa  
perspire, sweat ENG20-00065374-v *bila*  
“exercise makes one sweat”  
*ezemithambo ziyambilisa umntu*

The verb *ziyambilisa* “it causes one to sweat” is used with the causative suffix or verb extension *-is-*.

As a future goal, the team is also planning to include (semi) automatic quality assurance measures directly into the development interface. Morphological analysis as mentioned above, spelling correction, checking for empty fields and allowed categories can all be done in-line. Suggestions/prompts before saving can be added to the interface as a final step before a linguist signs off on a specific synset. We further envision enhancing the interface with improved internal communication so that a linguist can send comments on a synset directly to a team member for verification. Having a full record of the (linguistic) decisions made will also help improve the protocol for development and will offer valuable insights to new wordnet projects so that there is no need to “reinvent the wheel”.

## 6. Conclusion

All data developed in the AfWN project will be made available under a Creative Commons license<sup>5</sup> via the SADiLaR language resource repository with the hope that it can increase NLP development particularly for the African languages. It is important for users of the data to be aware of certain linguistic and technical decisions made during development so that they can also make provision for certain aspects in their systems.

Since so many wordnets for under-resourced, linguistically complex languages follow the expand method for wordnet development and rely heavily on the English source data as in the PWN, it is further important to document the lexical gaps and applicable differences between languages. We hope that by doing so in the project documentation and in publications, that we can facilitate the accelerated growth of the AfWN to include languages from other African countries.

## 7. Acknowledgements

The African Wordnet project (AfWN) was made possible with support from the South African Centre for Digital Language Resources (SADiLaR), a research infrastructure established by the Department of Science and Technology of the South African government as part of the South African Research Infrastructure Roadmap (SARIR).

The authors would also like to acknowledge and thank the linguists and technical project members involved in the AfWN project. A list of significant contributors is available on the project webpage (<https://africanwordnet.wordpress.com/team/>).

## 8. Bibliographical References

- Benítez, L., Cervell, S., Escudero, G., López, M., Rigau, G. and Taulé, M. (1998). Methods and tools for building the Catalan Wordnet. In ELRA (ed.). *Proceedings of the First International Conference on Language Resources and Evaluation (LREC'98)*, Granada, Spain, May 28–30. Available at <http://www.cs.upc.edu/~escudero/wsd/98-lrec.pdf>.
- Bosch, S. and Griesel, M. (2017). Strategies for building wordnets for under-resourced languages: the case of African languages. *Literator* 38(1), a1351. Available at <https://doi.org/10.4102/lit.v38i1.1351>
- Fellbaum, C. (ed). (1998). *Wordnet: An electronic lexical database*. The MIT Press, Cambridge, Mass.
- Griesel, M. and Bosch, S. (2014). Taking stock of the African Wordnet project: 5 years of development. In Fellbaum, C. et al. (eds.) *Proceedings of the Seventh Global WordNet Conference 2014 (GWC2014)*, pp. 148-153. Tartu, Estonia. Available at [http://gwc2014.ut.ee/proceedings\\_of\\_GWC\\_2014.pdf](http://gwc2014.ut.ee/proceedings_of_GWC_2014.pdf)
- Miháltz, M., Hatvani, C., Kuti, J., Szarvas, G., Csirik, J., Prószéky, G. and Váradi, T. (2008). Methods and results of the hungarian wordnet project. In: Tanács, A., et al. (eds). *Proceedings of the 4th Global WordNet Conference (GWC2008)*, pp. 311-320. Szeged,

<sup>5</sup> See <https://creativecommons.org/licenses/>

- Hungary. Available at <http://www.inf.u-szeged.hu/projectdirs/gwc2008/>
- Naskręt, T., Dziob, A., Maciej, P., Maciej, S., Chakaveh and Branco, A. (2018). WordnetLoom - a Multilingual Wordnet Editing System Focused on Graph-based Presentation. In: Fellbaum, C, *et al.* (eds). *Proceedings of Ninth Global WordNet Conference 2018 (GWC2018)*. Nanyang Technological University (NTU), Singapore. Available at <http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/gwc-2018-proceedings.pdf>
- Ordan, N. and Wintner, S. (2007). Hebrew WordNet: A test case of aligning lexical databases across languages. *International Journal of Translation, special issue on Lexical Resources for Machine Translation* 19(1), 39–58.
- Prabhu, V., Desai, S., Redkar, H., Prabhugaonkar, N., Nagvenkar, A. and Karmali, R. (2012). An efficient database design for IndoWordNet development using hybrid approach. *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP) at COLING 2012, Mumbai, December*, pp. 229–236.
- Raffaelli, I., Tadić, M., Bekavac, B. and Agić, Z. (2008). Building Croatian WordNet. In: Tanács, Attila, *et al.* (eds). *Proceedings of the 4th Global WordNet Conference (GWC2008)*, pp. 311-320. Szeged, Hungary. Available at <http://www.inf.u-szeged.hu/projectdirs/gwc2008/>
- Smrz, P. (2004). Quality Control and Checking for Wordnets Development: A Case Study of BalkaNet. In *Romanian Journal of Information Science and Technology Special Issue*, volume 7, No. 1-2.
- Snider, K. and Roberts, J. (2006). *SIL Comparative African Wordlist (SILCAWL)*. Available at [https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Snider\\_silewp2006-005.pdf](https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Snider_silewp2006-005.pdf)
- Vetulani, Z., Kubis, M. and Obrębski, T. (2010). PolNet – Polish WordNet: Data and tools. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, *et al.* (eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 17–23, pp. 3739–3797.
- Vossen, P., Bond, F. and McCrae, J. (2016). Toward a truly multilingual GlobalWordnet Grid. In Mititelu, V. *et al.* (eds.). *Proceedings of the Eighth Global WordNet Conference 2016 (GWC2016)*. Bucharest, Romania. Available at <http://www.racai.ro/p/gwc2016/>
- Weisscher, A. (2013). Global Wordnet Association base concepts. Available at <http://globalwordnet.org/gwa-base-concepts/>

## 9. Language Resource References

- Princeton University. (2020). WordNet – A lexical database for English. Available at <https://wordnet.princeton.edu/>
- University of South Africa (UNISA). 2017. Wordnets for isiXhosa (ISLRN 484-200-848-869-7), isiZulu (ISLRN 884-495-382-918-6), Setswana (ISLRN 609-257-955-359-4), Sesotho sa Leboa (ISLRN 639-836-685-202-8) and Tshivenda (ISLRN 193-390-599-634-9). African Wordnet Project, distributed via SADiLaR, 1.0. Available at <https://repo.sadilar.org/handle/20.500.12185/7/browse?type=project&value=African+Wordnet+Project>