

African Wordnet: facilitating language learning in African languages

Sonja Bosch

Marissa Griesel

University of South Africa

boschse@unisa.ac.za



Global Wordnet Conference, 12 January 2018

Nanyang Technological University (NTU), Singapore

Outline

Aim of paper:

Exploring the interaction between **African Wordnet** and a basic free **online course for isiZulu**, as a first step towards moving to a more integrated computer-assisted language learning (CALL) system.

To be discussed:

- Background
- African Wordnet (AWN) project
- *LEarn To Speak an African Language (LetsAL)* courses
- Incorporating AWN in *LetsAL* as additional reference material
- Expanding AWN
- Conclusion and future work

Background [1]

- South Africa has **11 official languages** - 9 are African languages – belonging to Bantu language family.
- Multilingual scenario has given rise to SA National Development Plan 2030 (NDP) enjoining all South Africans to learn at least one indigenous language as part of **nation-building and social cohesion**.
- As from 2019 - all **education students** to include an African language in their Bachelors curriculum.
- Currently, very limited computer-assisted material available for language learners of African languages.

Background [2]

- Against this background we will investigate the usefulness of:
 - the **African Wordnet** (AWN) that has reached a stage of maturity where first steps towards an application can be attempted (even with less than 20 000 synsets per language).
- Application envisaged is **language learning**, and materials selected for this purpose are:
 - basic free online courses **LEarn To Speak an African Language** (*LetsAL*) - developed for first time language learners of five African languages, viz.
Setswana, isiXhosa, isiZulu, Sesotho sa Leboa and Sesotho.

African Wordnet Project [1]

- Languages involved are generally considered **resource scarce**.
- **Expand model** was therefore followed from the onset.
- Most of the development was of a **manual** nature – initially extended Common Base Concepts and Princeton Core Concepts used - more localised approach was needed, too many non-lexicalised concepts. Semi-automatic **linking** between bilingual dictionaries and PWN.
- Overall aim - to build **aligned wordnets** for languages spoken in *South Africa* as multilingual knowledge resources - could be extended to include wide variety of **related languages** also from other parts of Africa.
- **Five languages** currently included: isiZulu, isiXhosa, Setswana, Sesotho sa Leboa and Tshivenda.

African Wordnet Project [2]

Language	Synsets	Definitions	Usage examples	No. of words
Setswana (TSN)	15 803	3515	7203	86 071
IsiXhosa (XHO)	14 715	2198	7015	36 246
IsiZulu (ZUL)	10 782	2179	5112	40 597
Sesotho sa Leboa (NSO)	8412	1178	5253	44 753
Tshivenda (VEN)	4270	209	4270	12 925
<u>Total</u>	<u>53 982</u>	<u>9279</u>	<u>28 853</u>	<u>220 592</u>

African Wordnet Project [3]

- AWN still under active development – new round of funding, includes high-level technical support, secured via newly established **South African National Centre for Digital Language Resources (SADiLaR)** - focus on creation, management and distribution of digital language resources as well as applicable software related to all official languages of South Africa.
- Freely available as **open-source resource** under creative commons license:

Language Resource Management Agency

<https://rma.nwu.ac.za/>

Open Multilingual Wordnet (2.0)

<http://compling.hss.ntu.edu.sg/iliomw/omw>

Free Online Language Learning Courses [1]

- **Open Educational Resource (OER)** - created by University of South Africa and openly licensed to actively promote African languages in anticipation of 2010 FIFA Soccer World Cup.
- Used by **banks** and **private schools** as OER for language learning.
- First development of free online courses that focus on basic **language skills** for five African languages
<http://www.unisa.ac.za/sites/corporate/default/Unisa-Open/OER-@-Unisa/Learn-to-speak-an-African-Language>
- Further development of *LetsAL* on the cards – to include additional languages and more advanced levels.

Free Online Language Learning Courses [2]



- About
- Admissions ▾
- Colleges ▾
- Research
- Alumni
- Library
- News & Media
- Donate
- Contact us

[Home](#) / [Unisa Open](#) / [OER @ Unisa](#) / [Learn to speak an African Language](#)

- What are OERs?
- OER @ Unisa
- Finding OERs
- Open Access
- African & SA OER initiatives
- Contact us
- Learn to speak an African Language

Unisa Open



learn to speak an **African** language

WELCOME!

Select the language you wish to learn...



-  Northern Sotho
-  Tswana
-  Zulu
-  Southern Sotho
-  Xhosa

Free Online Language Learning Courses [3]

- **Basic lessons covering 10 general themes:**
 - greetings and courtesies; asking for help; numbers, days, months, seasons; question words, quantities, weather; banks, taxis and restaurants; transportation and finding your way; touring and socializing; at the filling station; the human body and ailments; shopping and sport;
 - list of appropriate vocabulary and phrases accompanied by translations;
 - short dialogue;
 - video to contextualize the content;
 - noteworthy cultural customs explained.

Free Online Language Learning Courses [4]

- Compared to similar products for well-known world languages, the following aspects are still lacking:
 - inclusion of **real-time interaction** between learners;
 - games and **fun content** for assessment purposes;
 - links to other **external resources**.
- The AWN (as one of very few freely available African language resources) could play a central role in future improvements to these courses, in particular with regard to external resources.

Incorporating AWN in *LetsAL* [1]

- Related study: Winiwarter (2011) describes a **collaborative language learning and instruction environment (COLLIE)** for Japanese foreign language learners. This system combines advanced NLP tools such as machine translation and complex analysers with the **English PWN** and the **Japanese Wordnet** to provide learners with translations of Japanese web pages, as well as detailed information on the word level.
- Incorporating all **available NLP resources** into a CALL system offers learners an integrated application to enhance their understanding of the subject matter.

Incorporating AWN in *LetsAL* [2]

- Focus in this presentation - **isiZulu**.
- NLP support still limited for isiZulu but it is important that **resources that are freely available** at this stage, are integrated.
- With elements of a wordnet, learner decision support can be improved via **semantic relations**, e.g. synonymy and antonymy, **usage example sentences** and **definitions**.
- IsiZulu is known as **tone language** - pitch variation plays a role in conveying lexical as well as grammatical distinctions.
- Tone not marked in **standard orthography** of isiZulu, therefore language learners are often confronted with a difficult choice between two meanings of (seemingly) the same word.

Incorporating AWN in *LetsAL* [3]

- How tone distinguishes meaning:

íthàngá (thigh) HLH

íthàngà (pumpkin) HLL

ínyàngá (moon; month) HLH

ínyàngà (herbalist, herb doctor) HLL

-*dúmà* (be tasteless) HL

-*dù má* (roar, be famous) LH

Incorporating AWN in *LetsAL* [4]

- ***LetsAL* environment** can be enriched with **AWN data** to assist learners to understand not only the broad meaning of words in context, but can also point out subtle differences and potential language specific pitfalls (orthography).
- A selection of examples will show possibilities for **added support** from existing AWN data for disambiguating context, and useful additional information:
 - Example words already occur in the *LetsAL* course material, but only with the **English translation** and no further information on meaning nuances in isiZulu;
 - We **combined information** from the isiZulu AWN and the relevant information from the English PWN.

Incorporating AWN in *LetsAL* - Example 1

POS: n; ID PWN 2.0: ENG20-05243922-
n; ID PWN 3.1: 05569882

Synonyms: *ithanga* :1

Definition: *isitho somuntu*

esiphakathi kwedolo nenqulu

Usage: *umdlalikazi watheleka*

emcimbini wama-Grammy esho

ngelokwe eliveza lonke ithanga

Domain: anatomy

Synonyms: thigh :1

Definition: the part of the leg
between the hip and the knee

Domain: anatomy

POS: n; ID PWN 2.0: ENG20-07263505-
n; ID PWN 3.1: 07751486

Synonyms: *ithanga* :2

Definition: *isitshalo esimila*

phansi esinombala ophuzi

Usage: *kanti-ke abangani*

abathathu bapheka ithanga

elikhulu kakhulu

Domain: gastronomy

Synonyms: pumpkin :2

Definition: usually large pulpy
deep-yellow round fruit of the
squash family maturing in late
summer or early autumn

Domain: gastronomy

Incorporating AWN in *LetsAL* - Example 1

POS: n; ID PWN 2.0: ENG20-14348156-n; ID PWN 3.1: 15234209

Synonyms: *inyanga* :1

Definition: *isikhathi sezinsuku ezingamashumi amathathu*

Usage: *inyanga yesibili manje*

Usage: *ngiphumule inyanga eyodwa ngemuva kwama-Olympics*

Domain: time_period


Synonyms: calendar month :1, month :1

Definition: one of the twelve divisions of the calendar year

Usage: he paid the bill last month

Domain: time_period

Incorporating AWN in *LetsAL* – Example 2(a,b,c)

Synonyms: inyanga :1	Synonyms: inyanga :1	Synonyms: inyanga :1
Synonyms: calendar month :1, month :1	Synonyms: moon:1	Synonyms: herbalist :1, herb doctor :1
		
	Related meaning	Unrelated meaning
Domain: time_period	Domain: astronomy	Domain: medicine

Incorporating AWN in *LetsAL* - Example 2(a)

POS: n; ID PWN 2.0: ENG20-14348156-n; ID PWN 3.1: 15234209

Synonyms: *inyanga* :1

Definition: *isikhathi sezinsuku ezingamashumi amathathu*

Usage: *inyanga yesibili manje*

Usage: *ngiphumule inyanga eyodwa ngemuva kwama-Olympics*

Domain: time_period

Synonyms: calendar month :1, month :1

Definition: one of the twelve divisions of the calendar year

Usage: he paid the bill last month

Domain: time_period

Incorporating AWN in *LetsAL* - Example 2(b)

POS: n; ID PWN 2.0: ENG20-08772174-n; ID PWN 3.1: 09381255

Synonyms: *inyanga* :1

Definition: *isathalayithi yoMhlaba ekhanyisa esibhakabhakeni*

Usage: *inyanga iphuma ebusuku*

Usage: *wabona inyanga eyisiliva iqala ukuphakama phezulu esibhakabhakeni*

Domain: astronomy

Synonyms: moon :1

Definition: the natural satellite of the Earth

Usage: the average distance to the moon is 384,400 kilometers

Usage: men first stepped on the moon in 1969

Domain: astronomy

Incorporating AWN in *LetsAL* - Example 2(c)

POS: n; ID PWN 2.0: ENG20-09516232-n; ID PWN 3.1: 10191128

Synonyms: *inyanga* :1

Definition: *umuntu onolwazi lokwelapha izifo ngemithi*

Usage: *uGcabashe umbikele ukuthi akaphilile udinga ukubona inyanga*

Usage: *inyanga ithi ifuna ukuthenga umuthi omhlophe*

Domain: medicine

Synonyms: herbalist :1, herb doctor :1

Definition: a therapist who heals by the use of herbs

Domain: medicine

Incorporating AWN in *LetsAL* - Example 3

POS: v; ID PWN 2.0: ENG20-02472355-v; ID PWN 3.1: 02553283

Synonyms: *lekelela* :1, *siza* :1

Definition: *ukwelekelela noma ukusiza ekwenzeni okuthile*

Usage: *le nhlangotho ilekelela ekufundiseni izingane eziqhamuka emakhaya*

Usage: *iholo lakhe lisiza ukuthi akwazi ukuphilisa umndeni*

Synonyms: help :1, assist :1, aid :1

Definition: give help or assistance; be of service

Usage: Everyone helped out during the earthquake

Usage: Can you help me carry this table?

Usage: She never helps around the house

Incorporating Awn in *LetsAL* - Example 4(a-b)

POS: a; ID PWN 2.0: ENG20-00589776-a;
ID PWN 3.1: 00622052

Synonyms: *luhlaza* :4

Definition: *ukudla okungavuthiwe*

Usage: *Babengawusebenzisi umlilo, babemane bayifukuthe bayidle luhlaza inyama*

Domain: gastronomy

-->> [**near_antonym**] *vuthiwe*:1

Synonyms: *raw* :3

Definition: not treated with heat to prepare it for eating

Domain: gastronomy

-->> [**near_antonym**] *cooked*:1

POS: a; ID PWN 2.0: ENG20-00586933-a;
ID PWN 3.1: 00618376

Synonyms: *vuthiwe* :1

Definition: *kuphekiwe kulungele ukudliwa*

Usage: *ukudla okuvuthiwe kundayiswa emgwaqweni*

Domain: gastronomy

-->> [**near_antonym**] *raw*:3

Synonyms: *cooked* :1

Definition: having been prepared for eating by the application of heat

Domain: gastronomy

-->> [**near_antonym**] *raw*:3

Incorporating AWN in *LetsAL* - Example 4(c-d)

POS: a; ID PWN 2.0: ENG20-01442460-a;
ID PWN 3.1: 01497045

Synonyms: *luhlaza* :5

Definition: *okungavuthiwe kwezithelo*

Usage: *abanye badla izithelo eziluhlaza*

-->> [**near_antonym**] *vuthiwe*:2

Synonyms: *green* :3, *unripe* :1,
unripened :1, *immature* :4

Definition: *not fully developed or
mature; not ripe*

Usage: *unripe fruit*

Usage: *fried green tomatoes*

Usage: *green wood*

-->> [**near_antonym**] *ripe*:1, *mature*:4

POS: a; ID PWN 2.0: ENG20-01441835-a; ID
PWN 3.1: 01496321

Synonyms: *vuthiwe* :2

Definition: *okulungele ukudliwa noma
ukuphuzwa*

Usage: *izitshalo zikabhekilanga zivuthiwe
lapho umbala ngemuva kwesihloko
ushintsha*

-->> [**near_antonym**] *luhlaza*:5

Synonyms: *ripe* :1, *mature* :4

Definition: *fully developed or matured and
ready to be eaten or used*

Usage: *ripe peaches*

Usage: *full-bodies mature wines*

-->> [**near_antonym**] *green*:3, *unripe*:1,
unripened:1, *immature*:4

Incorporating AWN in *LetsAL*: planned implementation

- Value-added website, showing a pop-up window included in ***LetsAL* environment** with information for isiZulu pertaining to:
 - Link to PWN and isiZulu AWN entry, gives learner access to full synset;
 - Domain, usage example(s) and definition(s) are also shown;
 - Other disambiguating or relevant relations could be included, e.g. hypernym (superordinate).

Zulu: Theme 9 - The human body and ailments

The human body

I try

Elbow /elbows	Indololwane / izindololwane
Hip /hips	Inqulu / izinqulu
Back	Umhlane / imihlane
Knee / knees	Idolo / amadolo
Buttock / buttocks	Isinqe / izinqe
Thigh /thighs	<u>Ithanga</u> / amathanga
Calf /calves	Isitho / izitho
Leg /legs	Umlenze / imilenze
Toe /toes	Uzwani / izinzwani

thigh
ithanga

Domain: anatomy

Definition: the part of the leg between the hip and the knee

Superordinates: limb
isitho

Usage: *umdlalikazi wathleka emcimbini wama-Grammy esho ngelokwe eliveza lonke ithanga*

Back

Continue

Zulu: Theme 9 - The human body and ailments

The human body

I try

Elbow /elbows	Indololwane / izindololwane
Hip /hips	Inqulu / izinqulu
Back	Umhlane / imihlane
Knee / knees	Idolo / amadolo
Buttock / buttocks	Isinqe / izinqe
Thigh /thighs	<u>Ithanga</u> / amathanga
Calf /calves	Isitho / izitho
Leg /legs	Umlenze / imilenze
Toe /toes	Uzwani / izinzwani

Back

Continue

ithanga “pumpkin”
Domain:gastronomy

thigh
ithanga

Domain: anatomy

Definition: the part of the leg between the hip and the knee

Superordinates: limb
isitho

Usage: *umdlalikazi watholeka emcimbini wama-Grammy esho ngelokwe eliveza lonke ithanga*

Expanding the AWN with *LetsAL* data [1]

- Our previously developed **linking protocol** was used to semi-automatically speed up creation of new synsets for isiZulu AWN with data from *LetsAL*.
- List of **unique words** was first extracted (function words removed): 486 words in total; of which 94 words already included in isiZulu AWN.
- English stems with their associated synsets were extracted from PWN and limited to those with attributes “**sense:1**”.
- Linguist assessed **validity** of each link before inclusion in AWN.
- In a next pass more depth will be aimed for.

Expanding the AWN with *LetsAL* data [2]

POS category	Lemmas in <i>LetsAL</i>	Possible links identified	Correct matches added to AWN
Nouns	259	130	119
Verbs	126	51	46
Adjectives	9	7	7
TOTAL	392	188	172

Conclusion & future work

- *LetsAL* infrastructure lends itself well to inclusion of data from AWN as additional on-line **reference source**.
- **AWN contribution** of sense differentiation, definitions and translation equivalents, can assist language learners to gain valuable insight into language specific complexities as well as an expanded vocabulary.
- The work proposed here for isiZulu can be **replicated** for all African languages for which a wordnet and a *LetsAL* course currently exist.
- Growing the AWN is a priority and the comparison with *LetsAL* presented here will serve as valuable input for the **next phase** of development.

**Thank you
Enkosi
Siyabonga
Re a leboga
Ro livhuwa**



Acknowledgements

The authors would like to thank the South African National Research Foundation (NRF), National HLT Network, Department of Arts and Culture, the University of South Africa's Women-in-Research Fund as well as the South African Centre for Digital Language Resources for providing funding in the various phases of the AWN project.

