

Towards an Integrated E-Dictionary Application – The Case of an English to Zulu Dictionary of Possessives

Sonja Bosch*, Gertrud Faaß¹

*University of South Africa, University of Hildesheim

E-mail: boschse@unisa.ac.za, faassg@uni-hildesheim.de

Abstract

This paper describes a first version of an integrated e-dictionary translating possessive constructions from English to Zulu. Zulu possessive constructions are difficult to learn for non-mother tongue speakers. When translating from English into Zulu, a speaker needs to be acquainted with the nominal classification of nouns indicating possession and possessor. Furthermore, (s)he needs to be informed about the morpho-syntactic rules associated with certain combinations of noun classes. Lastly, knowledge of morpho-phonetic changes is also required, because these influence the orthography of the output word forms. Our approach is a novel one in that we combine e-lexicography and natural language processing by developing a (web) interface supporting learners, as well as other users of the dictionary to produce Zulu possessive constructions. The final dictionary that we intend to develop will contain several thousand nouns which users can combine as they wish. It will also translate single words and frequently used multiword expressions, and allow users to test their own translations. On request, information about the morpho-syntactic and morpho-phonetic rules applied by the system are displayed together with the translation. Our approach follows the function theory: the dictionary supports users in text production, at the same time fulfilling a cognitive function.

Keywords: Zulu; less-resourced languages; learner’s dictionary, extralexicographic features; integrated e-dictionary; possessives

1 Introduction

This paper describes an integrated approach to e-dictionaries as learning tools informed by the Function Theory of Lexicography. Issues raised by information science as well as learners’ levels of knowledge in text production (cf. Bothma and Tarp 2012) are taken into consideration. The case of possessive constructions in Zulu will be used as example. We will explore ways of using an e-dictionary integrated in a learning environment to translate possessives from English to Zulu. We aim to enhance the language learning experience of users in a didactically valuable manner.

1 This project is part of the activities of Gertrud Faaß as an affiliated research fellow of the University of South Africa.

Our users in focus are not only language learners, but also others who do not strictly fall into one of the categories defined by the function theory, e.g. advanced speakers of the language possibly requiring knowledge about correct grammatical constructions, in this case possessives. We combine lexicographic and morpho-syntactic information and take particular extra-lexicographical situations and user needs into consideration.

To our knowledge this research on an integrated e-dictionary for Zulu is novel work. While e.g. Radtke and Heid (2012:798) propose an approach to address the needs of text production by making use of corpus data, and Prinsloo et al. (2012:292) offer graphical decision trees as a solution for explaining and translating Northern Sotho kinship terminology and copulative constellations to users, we go a step further by integrating morpho-syntactic and lexicographical information as well. We address text production by showing users how to form possessives and by offering information about how these are formed. In addition, we address the cognitive function by assisting users to understand the morphology of the other language (Gouws 2007:85). Though the e-dictionary described here in the long term may be developed into an intelligent computer assisted language learning (ICALL) system, this e-dictionary is not as yet, comparable to what is already developed for other Bantu languages, as e.g. described by Katushemerewe and Nerbonne (2013) for Runyakitara or Hurskainen (2009) for Swahili.

In the next section, a brief background on the morphological structure of Zulu will be given, followed by an exposition of the complexities encountered with the translation of possessives from English into Zulu. Thereafter the implementation of the integrated e-dictionary application will be discussed with detailed exemplification, followed by a conclusion and notes on future work.

2 Background of the morphological structure of Zulu

Zulu [zul] belongs to the Bantu languages which have a rich agglutinating morphological structure, based on two principles, namely the nominal classification system, and the concordial agreement system. According to the nominal classification system, nouns are categorized by prefixal morphemes, which for analysis purposes have been assigned class numbers. These noun class prefixes use concordial agreement to link the noun to other words in the sentence such as verbs, adjectives, pronouns, possessives etc. (cf. Poulos and Msimang 1998).

Some degree of morpho-phonological complexity occurs which can mainly be ascribed to the phonological phenomena at morpheme boundaries resulting from the conjunctive orthography of the Zulu language. These phenomena are mostly predictable and rule-based. Because of its complexity, the construction of the possessive constitutes a particular challenge for language learners. First, information on the noun class of the possession and the possessor is required in order to determine the connecting element: it can either be a “regular” class specific possessive concord containing the possessive marker *a*, or if the possessor is a noun belonging to class 1a, then a special possessive marker

ka is used (cf. Poulos and Msimang 1998:146). After selecting the appropriate connecting element, the speaker has to know the morpho-phonetic rules for cases where the connecting element and the possession are to be fused. See some examples in the following table:

| English possessive construction | Zulu possession (class) | connecting element | Zulu possessor (class) | Translation | Rule |
|--|-------------------------|--------------------|------------------------|------------------------|--|
| Sipho's cats lit. 'cat's of Sipho' | amakati (6) | ka- | uSipho (1a) | amakati kaSipho | fuse connecting element with possessor by deleting initial vowel <i>u-</i> |
| old women's box lit. 'box of old women' | ibhokisi (5) | la- | izalukazi (8) | ibhokisi lezalukazi | fuse connecting element with possessor by means of vowel coalition <i>a+i > e</i> |
| mother's bread lit. 'bread of mother' | isinkwa (7) | sika- | <i>umama</i> (1a) | isinkwa sikamama | fuse connecting element with possessor by deleting initial vowel <i>u-</i> |

Table 1: Examples of Zulu possessive constructions.

When a possessor is to be replaced by a pronoun, its noun class again plays an important role. For instance when translating the phrase 'their box' (lit. 'box of them') referring to 'old women' as possessor, the connecting element or possessive concord of the possession ('box') class 5 *la-* is prefixed to the (abbreviated) pronoun of class 8 *zo*, resulting in *ibhokisi lazo*.

Examples of even higher complexities are highlighted in the implementation section below.

3 Challenges for learners

The microstructure of bilingual dictionaries usually contains source language lemmas and inter alia their respective translation equivalent in the target language. The user needs in terms of the function theory could be on the level of text reception², text production or cognitive needs in order to understand the construction. Occasionally a language learner might want to query short phrases, such as possessive structures e.g. "Sipho's food", or "(the) doctors' medicine" without having to look up the possessor, the possession as well as the grammatical rules on how to formulate the construction.

Although Computer Aided Language Learning (CALL-) applications offer such possibilities, these applications usually only have restricted dictionaries that contain solely what is described in the learning material accompanying them. This is usually sufficient, as the learner can learn about the rules

² The work described here, however, does not cater for receptive user needs as this issue is foreseen to be taken into account at a later stage.

for producing the constellation in the target language and only has to use a bilingual dictionary to look up translations for the words (s)he needs to know.

But what if even simple constellations show a huge variety in forms? In languages like German, English or French, there are only three parameters to take into account when formulating a possessive: person, number, and gender, the patterns are not too difficult to learn and there are few exceptions. For Zulu, a fourth parameter replaces the 3 choices of gender with up to 16 choices of noun class, some of which are illustrated in Table 1. Language learners hence struggle for quite a while to learn full paradigms for each of the day-to-day constellations. In addition, the above mentioned morpho-phonetic phenomena need to be memorized and kept track of as well. One of the constellations that is very challenging to absorb is the possessive construction, as knowledge of the following facts is required: (i) class of the possessor; (ii) class of the possession; (iii) the concord that is to be used; (iv) generic morpho-phonological rules.

To give an idea of the complexity of possessive constructions, in some Zulu learner guides, the description of the possessive construction is spread over an average of four pages which “can be very user-unfriendly or time consuming for him/her to find the right information by having to read entire sections of grammatical descriptions.” (Prinsloo and Bosch 2012:300).

4 Current State of Implementation

Our system architecture comprises three parts: (1) an sql-database containing the dictionary (lexicographic content part); (2) morpho-phonological rules (linguistic part) and, lastly, (3) programming scripts that build the Graphical User Interface (GUI) and build the web page, i.e. the interface between database and user (the presentation part), cf. figure 1.

We opted for using the XAMPP³-environment for developing the sql-database, because it assists when designing and filling databases with data and also simplifies the creation of websites. *php* is a dynamic scripting language specifically designed to develop websites and is part of the XAMPP environment. It offers phpMyAdmin, an sql-interface, i.e. sql-commands can be executed on a respective database with *php* and allows the results to be written into variables that can then be displayed on a website. After the development has been finalized, the php scripts and the database can easily be ported onto an apache webserver. Our test interface can be found on www.uni-hildesheim.de/iwist-cl/projects/eZulu/login.php, access is currently restricted to known users⁴.

3 <http://www.apachefriends.org/de/xampp.html>

4 Please contact G. Faaß if you should like to try out the test interface

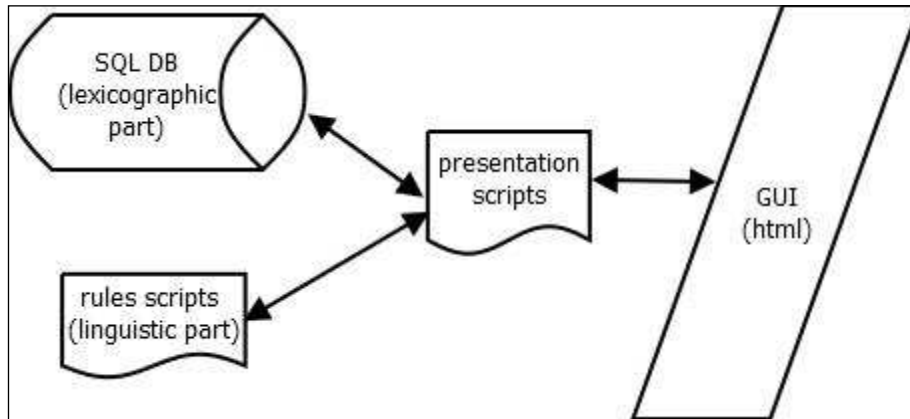


Figure 1: System Architecture.

For a start, our database is kept simple and currently contains representative data of all noun classes necessary to perform the given task. We are still working on the design of the website, as it should fulfil the users' needs in terms of the relevant issues of the function theory described above. Currently, we focus on an accurate implementation of the rules and a user-friendly way to explain their application to those users who request such an explanation by ticking the checkbox.

Though the website design is preliminary, we show screenshots in Figures 2, 3 and 4 in order to demonstrate our implementation principle: keep it simple and intuitive. In the header, the user is asked what kind of information (s)he requires over and above the translation which is done in all cases. Checkbox 1 on the left asks whether the user would like to have a pronominal version of the translation (i.e. the possessor noun being replaced by the pronoun belonging to the same noun class) – in addition to the full translation as Zulu uses different pronouns dependent on the noun class they belong to, hence the pronouns “it” or “they” has several translation equivalents, dependent of the noun class of the noun that it stands for. If we ask the user what possessor the pronoun stands for, it will be possible to identify the noun class of this noun so that the tool can choose the right pronoun.



Figure 2: Start page: user requires translation of “food of person” plus rules plus pronominal translation of the possessor.

Ticking checkbox 2 on the right hand side will lead to an additional explanation how the resulting Zulu translation was formed concerning the elements and the rules applied. These additional explanations are currently still kept brief, but we plan to link these to other pages where the applied rules will be described in more detail and with additional examples.

Figure 3 demonstrates the result of the query shown in Figure 2: A dynamic webpage which is divided into 4 blocks is visibly distinguishable: The 1st block, i.e. the header consists of static information on the page as such (“Zulu e-Dict test version”), while the 2nd block – which appears in all cases of use – repeats the English expression entered (“food of person”⁵). It then shows the word-by-word translation of possession and possessor (*ukudla* and *umuntu*), and also the possessive concord that is to be used (*kwa*). Since the checkbox requiring the pronominal form was ticked, it appears as well (*ukudla kwakhe*).

Ticking the checkbox “Do you require an explanation of the rules applied” results in the popping up of a 3rd block appears (identified by a darker blue background). Here, the English expression is repeated, however now, possession and possessor are clearly categorized as such. This 3rd block then also informs the user that the elements of the English construction “food of person” are *ukudla kwa*, and *umuntu* in Zulu. In addition to the translation of each of the elements, the rules forming the correct translation equivalent, *ukudla komuntu*, are briefly explained. A second paragraph describing the rules of forming the pronominal form is added, because the respective checkbox was ticked.

Zulu e-Dict test version
Translation of English possessives into Zulu

food of person is translated as *ukudla komuntu* pronominal form: *ukudla kwakhe*

| | possession | possessive concord | possessor |
|---------|-------------------|--------------------|-------------------|
| English | food | | person |
| Zulu | ukudla (class 15) | kwa | umuntu (class 01) |

rules applied:
possession is of class 11 or 15 and possessor is not of classes 01a/02a
rule: vowel coalescence (.a + u.. becomes .o..)
intermediate result: ukudla kwomuntu
rule 2: semi vowel elision (.w+vowel becomes .vowel)

rule for forming the pronominal version:
(*kwa* merges with *khe* (possessive pronoun of class 01));

Next query? Go to [query](#) or [logout](#).

© 2013 2014 University of Hildesheim (UHI) and University of South Africa (UNISA)

Figure 3: Webpage resulting from query shown in Figure 2.

We make use of the checkboxes mainly to show only the relevant data on the resulting page, i.e. the required units of information as they were defined by the user (information on demand, cf. Bothma and Tarp 2012:90). This is in line with Tarp’s (2009:26) claim that “The study of access routes, i.e. the rapid and easy access to the relevant needs-adapted data, is of utmost importance to lexicography.” Figure 4 demonstrates what happens if both checkboxes remained unticked.

5 We also consider offering the shorter form, i.e. “person’s food” with the final design of the website.

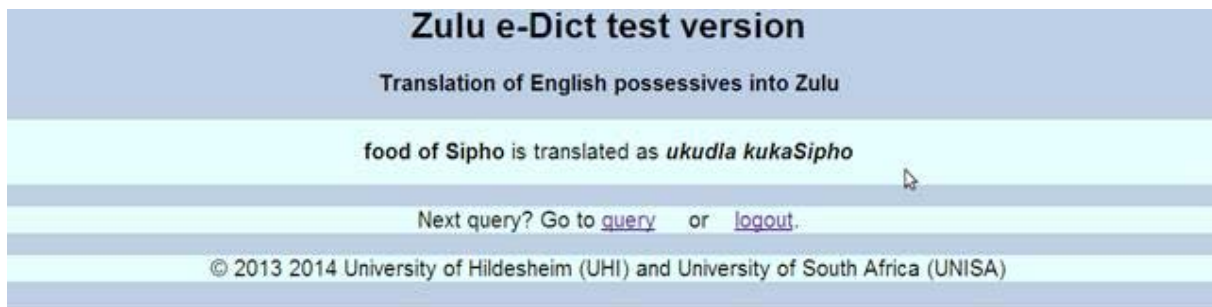


Figure 4: Webpage resulting from the query: “food of Siphoh”, no checkboxes ticked.

In this version of the website, no linking information to study guide pages/chapters is given, however as it is planned to become part of a free online course (cf. University of South Africa, 2010), this will be added then. As mentioned earlier on, the lexicon presently only contains a limited number of nouns (approximately 500 entries) that may be entered. However, we are currently working towards a second version of the database that will contain the vocabulary described in the study material used in the online course. The third version is then planned to become part of an e-dictionary of the African languages currently developed in the framework of the SeLA⁶ project which will translate single words and other morpho-syntactic constructions in both directions

5 Challenges and Plans for further Implementation Steps

What we still consider a problem is the case where a possessive concord ends in *-a* while the possessor begins with *-o*. In this case – restricted to class 2a noun possessors – there are two ways of forming the Zulu possessive, as demonstrated by Figure 5: The first rule, when applied, merges the concord with the possessor and thereby deletes the vowel (“vowel elision”) leading to *yodohotela* while the second, alternative rule adds *-w-* (“semi-vowel insertion”) resulting in the constellation *yawodohotela*. As both constellations occur frequently in a Zulu corpus⁷, we cannot decide prescriptively which of the constellations should be used. This is a problem in the light of the productive function that this e-dictionary should fulfil: the user might be confused not knowing which of the two constellations is “better”. Currently, we need to offer both alternatives, without distinctive features, as more corpus based research will be necessary to determine whether there are contextual factors that give preference to one of the constellations above the other.

An open challenge is the didactic element that is foreseen: we would like give users the opportunity to test their knowledge of Zulu, i.e. we will allow them to enter a translation equivalent that they

6 Scientific e-Lexicography for Africa, cf. Heid (2012) and <http://www.uni-hildesheim.de/iwist-cl/projects/sela/>

7 We are very thankful to Prof. DJ Prinsloo for giving us access to the Zulu corpus at the University of Pretoria, cf. Prinsloo and de Schryver (2005).

deem correct in one of the next versions of the dictionary. The system is then to compare the user's input with the correct output and, in a first stage, it will inform the user whether his/her input was correct or not. However, we will ask each user whether (s)he will give his/her consent to store the given data. If this consent is given, the input plus the attempted and the correct output will both be stored in a log file. We hope that these data will help to investigate typical learners' mistakes and incorrect generalizations. Depending on the output, we hope that in future the system can be enabled to dynamically offer appropriate feedback. These feedback messages, like the log file, are not planned to be personalized, but will only depend on the given inputs.

Zulu e-Dict test version
Translation of English possessives into Zulu

medicines of doctors is translated as *imithi yodokotela / imithi yawodokotela* pronominal form: *imithi yabo*

| | possession | possessive concord | possessor |
|---------|--------------------------|--------------------|------------------------------|
| English | medicines | | doctors |
| Zulu | <i>imithi</i> (class 04) | <i>ya</i> | <i>odokotela</i> (class 02a) |

rules applied:
possession is not of class 11 or 15 and possessor is of class 02a
rule: vowel elision (*a* + *o* . becomes *o* .)
alternative rule: semi-vowel insertion (*a* + *o* . becomes *awo* .)

rule for forming the pronominal version:
(*ya* merges with *bo* (possessive pronoun of class 02a)):

Next query? Go to [query](#) or [logout](#)

© 2013 2014 University of Hildesheim (UHI) and University of South Africa (UNISA)

Figure 5: Webpage resulting from a query with a noun of class 2a as possessor (usually beginning with *o-*) and a noun of a class where the possessive concord ends in *-a* as possession.

6 Conclusion and Future Work

We believe that this prototype e-dictionary with integrated morpho-syntactic information will go a long way to address user needs in specific user situations, in particular with respect to complicated issues such as the translation of possessives from English to Zulu. A next stage of development will allow for a bi-directional translation, where Zulu possessives given as input will be translated into English. This method will serve as a starting point for adding further constructions, while the vocabulary contained will be extended continuously. A parallel activity will be a thorough evaluation of the system; so far, the performed random sample tests have not discovered any errors.

At a later point, we also foresee an expansion of the prototype to closely related Bantu languages such as Xhosa, Swati, Southern Ndebele and Zimbabwe Ndebele. The prototype will also be testing for usability and user preferences, and therefore a test access has been made available for interested users.

7 References

- Bothma, T. and Tarp, S. (2012). Lexicography and the Relevance Criterion. *Lexikos* 22, pp. 86-108.
- Gouws, R. (2007). A Transtextual Approach to Lexicographic Functions. *Lexikos* 17, pp. 77-87.
- Heid, U. (2012). SeLA – a new project on electronic lexicography. *Lexicographica* 28(1), pp. 437-440.
- Hurskainen, A. (2009). Intelligent computer-assisted language learning: Implementation to Swahili. Technical Reports in Language Technology Report No 3, University of Helsinki, Finland. <http://www.njas.helsinki.fi/salama>. [09/04/2014]
- Katushemererwe, F. and Nerbonne, J. (2013). Computer-Assisted Language Learning (CALL) in support of (re-) learning native languages: The case of Runyakitara. *Computer-Assisted Language Learning*, 1-18 DOI: 10.1080/09588221.2013.792842 [09/04/2014]
- Poulos, G. and Msimang, C.T. (1998). *A linguistic analysis of Zulu*. Pretoria: Via Afrika.
- Prinsloo, D.J. and Bosch, S. (2012). Kinship terminology in English-Zulu/Northern Sotho dictionaries - a challenge for the Bantu lexicographer. In: *Proceedings of the 15th EURALEX International Congress*. 7-11 August (2012), Oslo, pp. 296-303. Oslo: Representralen, UiO. ISBN 978-82-303-2095-2
- Prinsloo, D.J. and de Schryver, G-M. (2005). Managing eleven parallel corpora and the extraction of data in all official South African languages. In W. Daelemans, T. du Plessis, C. Snyman & L. Teck (eds.) *Multilingualism and Electronic Language Management*. Proceedings of the 4th International MIDP Colloquium, 22-23 September 2003, Bloemfontein, South Africa (Studies in Language Policy in South Africa 4), pp. 100–122. Pretoria: Van Schaik Publishers. Available: <http://tshwanedje.com/publications/11Par-Cor.pdf> [28/10/2013]
- Prinsloo, D.J., Heid, U., Bothma, T., and Faaß, G. (2012). Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22 pp. 290-320
- Radtke, J. and Heid, U. (2012). Word formation in electronic language resources: state of the art analysis and requirements for the future. In: *Proceedings of the 15th EURALEX International Congress*. 7-11 August 2012, Oslo, pp. 794-802. Oslo: Representralen, UiO. ISBN 978-82-303-2095-2
- Tarp, S. 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. In Bergenholtz, H., Nielsen, S. & Tarp, S. (eds), *Lexicography at a Crossroads Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Linguistic Insights 90, pp.17-32. ISBN 978-3-03911-799-4 br
- University of South Africa Free Online Course. 2010. Available: http://www.unisa.ac.za/free_online_course/ [26/08/2013].