

Options for a Lexicographic Treatment of Negation in Zulu

Sonja Bosch¹, Gertrud Faaß^{1,2}

¹University of South Africa / ²University of Hildesheim
PO Box 392, Unisa 0003, Pretoria, South Africa/Universitätsplatz 1, 31141 Hildesheim, Germany
boschse@unisa.ac.za, gertrud.faaß@uni-hildesheim.de

Abstract

Dictionaries of today should offer much more than just knowledge about single words, they should rather be regarded as language information tools. However, in most electronic dictionaries of today, complex morphological constructions are not considered, thus users of dictionaries are usually expected to analyse such complex words themselves and to query base forms. With such a task, language learners, especially beginners, are often out of their depth. The question arising now, in particular with regard to learners' dictionaries, is whether and how can we enable an electronic dictionary to analyse complex constructions providing information on their structure and on their meaning? Taking Zulu negation as an example of a complex morphological construction, we first examined the frequency of this phenomenon in the corpora available and found an impressive number of them. So in the latter part of this paper, we try to find options for a practical implementation in electronic dictionaries.

Keywords: negation, Zulu, corpus-based queries, lexicographic treatment, learners' dictionaries

1. Introduction

Negation is described by Crystal (1994:231) as “A process or construction in GRAMMATICAL and SEMANTIC analysis which typically expresses the contradiction of some or all of a sentence’s meaning.” As an important instrument of language use, one would therefore expect aspects of negation to be dealt with in dictionaries. However, as Dahl (1979) states, negation phenomena appear to be at the border of lexicon and grammar, thus, one could argue that grammatical issues are not a matter for lexicography. Electronic dictionaries, on the other hand, are nowadays seen rather as language information tools, that is, they are to contain and to present extra-lexicographic data about a language as well (cf. Prinsloo et al. 2012), so negation again comes into play. Kovarikova et al. (2012:827) point out that “The main advantages of such a dictionary - almost unlimited size, interconnectivity of entries, easy referencing both within the dictionary and to a corpus - can also be used to describe negation ... with all its aspects.” Although only very few works are available on the lexicographic treatment of negation, novel (paper) dictionary conventions for the handling of negative verbal morphemes in Northern Sotho are proposed by Prinsloo and Gouws (1996), while van Son et al. (2016) address the need for building a dictionary of affixal negations and regular antonyms.

In a language such as English, word formation rules are relatively straightforward in the sense that inflections and derivations are usually constructed by adding suffixes to a root. Therefore, written words are usually roots, or commence with stems, and these roots can be looked up with ease in an English dictionary. An agglutinating language such as Zulu, however, has a much richer morphological structure, comprising an extensive and productive system of affixation that “pushes” roots into the middle of a word. Just looking up a Zulu word therefore requires the ability of a language learner to take the word apart by stripping the prefixes and suffixes and identifying any morphophonological changes that took

place, in order to extract a stem that can be looked up in the dictionary.

The aim of this paper is therefore to investigate methods of enabling an electronic Zulu (learners') dictionary (with inflected forms) to analyse complex constructions, in this case negation phenomena, providing information on their structure and on their meaning (thereby supporting both perception and production). We firstly provide some background on negation in Zulu, followed by an investigation into the frequency of this phenomenon of negation as reflected in Zulu corpora. This is followed in section 4 by a description of negation as treated in existing Zulu dictionaries. In section 5 we suggest requirements for improved (electronic) Zulu bilingual (learners') dictionaries, and present options for a practical implementation in electronic dictionaries with detailed exemplification. These options are based on our findings in the foregoing two sections, and also include existing data and software. Finally, a conclusion and notes on future work are presented.

2. Background on Negation in Zulu

2.1 Orthography

Zulu follows a conjunctive orthography, which means that bound morphemes are attached to the words (unlike other South African languages e.g. the Sotho languages), and thus cannot occur independently as separate words.

Furthermore, the order of occurrence of morphemes is fixed, as in other agglutinating languages such as Turkish. Orthographic words are of a polymorphemic nature of affixes attached to the root or core of the word, while monomorphemic words are limited to the following parts of speech: ideophone, conjunction and interjection. It is also noteworthy that morphophonological changes may occur between lexical and surface levels.

Kosch (2006:42) emphasizes that mother-tongue speakers of a language are familiar with the structure and sound patterns of their language, and therefore intuitively select allomorphs that are conditioned by the relevant phonological rules. For language learners however, this

selection may seem unnatural and they need to learn consciously when certain sound changes need to be made. An example of a morphophonological change such as vowel elision in the formation of negatives is demonstrated in (1) where the vowel of the SC01_neg is deleted before the vowel initial VRoot -*akh-* in order to present as *akakhi* on the surface:

- (1) *a ka akh i*
 neg SC01_neg VRoot VEnd (neg)
 not 3rd person sg. build
 'he/she does not build / is not building'

2.2 Morphological Negation

Zulu is characterised, among others, by a rich morphological structure including a noun class system which classifies nouns into a number of noun classes, as indicated by noun prefix morphemes. Noun prefixes play a significant role in the morphological structure of the language in that they connect the noun to other parts of speech (e.g. verbs, adjectives, possessives and pronouns) in the sentence. This linking takes place by means of a system of so-called concordial agreement morphemes which are derived from the noun prefixes and usually bear a close resemblance to them.

The two main forms that negation takes in the Bantu languages of Guthrie's so-called zone S (i.e. those of Southern Africa) including Zulu, are described by Gowlett (2003:636) as (i) the use of a preconcordial negative marker, with or without a concurrent suffixal marker; (ii) the use of a post-concordial negative marker, with or without a concurrent suffixal marker. This applies not only to verb constructions, but also to so-called copulative constructions that include adjectives and relatives¹.

According to Kosch (2006:106) the positive form of the verb is not clearly identifiably marked by affixes, while overt marking does occur in the negative form. Negativising strategies may vary in different moods such as the participial (sometimes referred to as the situative in grammatical descriptions) and the subjunctive mood; the imperative form of the verb, and tenses such as the past, perfect and future tenses, as illustrated in the following examples:

- (2a) Participial -
(uma) ehamba > engahambi 'if he goes/does not go'
 (2b) Subjunctive mood -
(ukuze) ahambe > angahambi 'so that he goes/does not go'
 (2c) Imperative form -
vala > ungavali 'close/do not close) (singular)
valani > ningavali 'close/do not close) (plural)
 (2d) Past tense -
bahamba > abahambanga 'they went/did no go'
 (2e) Perfect tense -
bahambe/bahambile > abahambanga 'they went/did no go'
 (2f) Future tense -
bazohamba > abazuhamba/abazukuhamba 'they will go/ will not go'

bayohamba > abayuhamba/abayukuhamba 'they will go/ will not go'

- (2g) Stative -
silele > asilele 'he/she/it is asleep/ is not asleep'

As evident in the examples above, with the exception of the stative form in (2g), one or more prefixes take a negative form in conjunction with a change in final suffix of verb, therefore Zulu verbal negation strategies can be summarised as showing either dyadic negation (3) or polyadic negation (4):

- (3a) *ngi- -hamb- -a*
 SC1p VRoot VEnd
 person Sg. go
 'I go'
 (3b) *a- ngi- -hamb- -i*
 neg SC1p VRoot VEnd (neg)
 not 1st person Sg. go
 'I do not go'
 (4a) *u- -ya- -hamb- -a*
 SC01 long pres tense VRoot VEnd
 3rd person sg go
 he/she goes
 'he/she goes/is going'
 (4b) *a- ka- -hamb- -i*
 neg SC01_neg VRoot VEnd (neg)
 not 3rd person sg. go
 'he/she does not go / is not going'

Over and above the regular negated constructions as shown in (3) to (4), we also find a number of additional rules for specific verbs or verb forms. Whereas passive verbs in the perfect and past tense suffix the negative suffix -*anga* (as in 5a), passive verbs in the present tense, for example, may not use the negative verbal ending -*i* when being negated, but retain the positive -*a*, as in (5b):

- (5a) *a- yi- -shay- w -anga*
 neg SC09 VRoot Pass VEnd (neg)
 not 3rd person sg. beat
 'It was not beaten'
 (5b) *a- ba- -thand- w -a*
 neg SC02 VRoot Pass VEnd
 not 3rd person pl. like Pass
 'They are not liked'

In Figure 1 we summarise verbal negativising strategies used in Zulu.

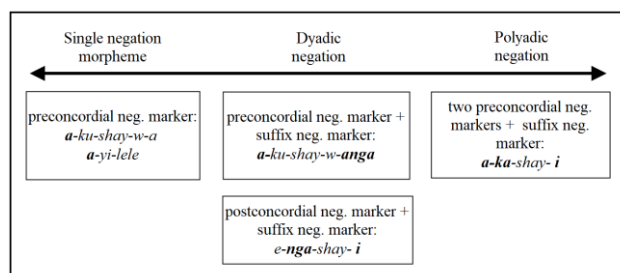


Figure 1: Continuum of Zulu verbal negation

¹ We will only be dealing with verbal negation in this paper.

Moreover, there are also so-called defective verb forms such as *-sho* ‘say’ which take irregular negative suffixes, for example *-ongo* instead of the regular *-anga* in the past tense. Further defective verb forms are *-thi* ‘say; think’ and *-azi* ‘know’. They have an irregular verb ending *-i* which does not change when the verb is negated in the present or future tense, but is replaced by the negative suffix *-anga* in the past tense. See the following examples:

- (6a) *ba- -sh- -o*
SC02 VRoot VEnd
3rd person pl. go
‘They said’
- (6b) *a- ba- -sh- -ongo*
neg SC02 VRoot VEnd (neg)
not 3rd person pl. go
‘They did not say’
- (6c) *a- ba- -az- -i*
neg SC02 VRoot VEnd
not 3rd person pl. know
‘They do not know’
- (6b) *a- ba- -az- -anga*
neg SC02 VRoot VEnd (neg)
not 3rd person pl. know
‘They did not know’

2.3 Syntactic Negation

In English, verbs take different auxiliaries when forming the negative, there are so-called ‘is’ and ‘have’ forms. Thus, to correctly translate an identified negated verb form, we need to store the respective category of the (English) translation of each verb stem in the dictionary. In Zulu, such categories do not exist, we however find several lexicalized negation word forms used in the imperative as shown in (7), similar to negating strategies of e.g. English. We categorize these as “syntactic negation”.

- (7a) *mus- a- uku- -hamb- -a*
VRoot VEnd SC15 VRoot VEnd
do not (imp) cl15(inf)go
‘Do not go!’ (semantically stronger than simple negation)
- (7b) *yek- -a uku- -hamb- -a*
VRoot VEnd SC15 VRoot VEnd
stop (imp) cl15(inf) go
‘Do not go!!’ (semantically stronger than (7a))

3. Negation as Reflected in Zulu Corpora

It is well-known that in comparison to a language such as English for which corpora with billions of tokens are available, Zulu can be regarded as an under-resourced language (cf. Prinsloo, 2012:121, Quasthoff et al., 2016:89). To the best of our knowledge, there are only four Zulu corpora that are freely available:

- (a) The raw University of KwaZulu-Natal (UKZN) isiZulu National Corpus², containing about 19.5 million tokens (no publication found). Of this corpus, no sentences but a word frequency list is downloadable;
- (b) the raw Wortschatz Universität Leipzig Internet Corpus of Zulu (LC, Quasthoff et al. 2014) contains about 3.2 million tokens (2.77 million words);
- (c) the NCHLT isiZulu Annotated Text Corpus (2014), which is based on government web pages and contains about 46,000 tokens (39,869 words). This corpus is available in different formats, we chose the version annotated with parts of speech;
- (d) the UKWABELANA corpus (UK, Spiegler et al. 2010) containing about 21,400 words (no punctuation) which is very small by world standards, but is nevertheless also available in different formats. Again, we chose the version annotated with parts of speech.

For a better comparability and to simplify searches, we downloaded corpora (b), (c) and (d) and encoded them with the Corpus WorkBench (Evert and Hardie, 2011). In the case of the UKZN corpus, we made use of the word frequency list.

Table 1 shows the number of occurrences of the syntactic verb negation described above (*musa/musani/yeka/yekani* followed by a verb in the imperative). As in most corpora, we do not find many texts of the type “conversation” in which imperatives occur, thus these phenomena are not very frequent.

Type of negation	UKZN	LC	NCHLT	UK
<i>musa uk...a</i>	n.a. ³	69	3	0
<i>musani uk...a</i>	n.a.	7	0	0
<i>yeka uk...a</i>	n.a.	17	0	0
<i>yekani uk...a</i>	n.a.	2	0	0

Table 1: Frequency of occurrence of syntactic negation

A more frequent way of negating the imperative is the (semantically) weaker morphological negation form *unga...i* (singular) or *ninga...i* (plural) as described in example (2c). The frequency of occurrence of this strategy in the Wortschatz Universität Leipzig Internet Corpus of Zulu (~ 3,2 mio tokens) is 1,264 which is fairly high in comparison to Table 1.

Type of negation	UKZN	LC	NCHLT	UK
<i>unga...i</i>	13,824	1,140	11	28
<i>ninga...i</i>	1,955	124	0	1
Total	15,779	1,264	11	29

Table 2: Frequency of occurrence of *unga...i* and *ninga...i*

²<https://iznc.ukzn.ac.za/> [2017-12-25]

³ Not applicable because the available data of UKZN consists of wordlists and not sentences.

To find the cases of morphological negation, a number of scripts were developed which make use of regular expressions. These describe the different verb forms in their full paradigm of inflection. Taking the verb forms of *-thanda* ([to] like) as an example, we find the list of present tense indicative conjugation forms shown in (8a). The appropriate regular expression in (8b) encodes these forms, but does not include the root *-thand-*.

(8a) *angithandi, awuthandi, asithandi, anithandi, akathandi, abathandi, awuthandi, ayithandi, alithandi, wathandi, asithandi, azithandi, ayithandi, azithandi, aluthandi, abuthandi, akuthandi*

(8b) `(a[bkw]a[angi]a[lnsy]i[a[blkw]u].)+i`

Thus we plan to find and count all negated verb forms following the regular conjugation pattern in the corpora. We do not differentiate between upper and lower case letters, but we exclude forms matched by the regular expression of which we know that they are not negated verbs. There are, for example, also deverbative nouns beginning with *aba-* and ending in *-i* (e.g. *abafazi*), of which we generated a stop list. We also exclude relative and adjective constructions like *ababanzi* or *abaningi*. However, ambiguous forms like *ababhali* (verb as well as deverbative noun) remain in the query as noise (see Annexure A for a list). Lastly, we use `pos="v"` as a selection condition (only for the corpora where parts of speech are annotated). The results are found in Table 2 and they show that negated verb forms are a frequent matter (at least in written text) worth describing in more detail in dictionaries.

Table 3 also shows that there would be sufficient data in the corpora for finding examples to be linked to the entries of dictionaries.

Type of negation	UKZN	ZULU	NCHLT	UK
imperative	9,534	1,545	11	28
present tense	121,554	20,967	105	123
participial/subjunctive	86,471	11,504	83	129
recent past	15,483	1,869	10	20
recent past continuous	3,939	554	1	7
remote past	13,758	1,711	10	15
recent past remote cont.	7,832	486	0	26
recent past perfect	163	25	0	0
remote past perfect	506	62	0	2
future tense	13	0	0	0
future tense continuous	(no. of aux verbs)	24	0	0
future tense perfect	5,022 ⁴	0	0	0

⁴ These forms are generated with a preceding auxiliary word (*asobe, basobe* etc.). As there's only a word list in UKZN available, we search for those.

Total	264,275	38,747	220	350
no. of words	19,553,511	2,771,207	39,867	21,416
% verb negations	1.35	1.40	0.55	1.63

Table 3: Frequencies of occurrences of morphological verb negation

4. Negation as Reflected in Zulu Dictionaries

In this section we address the treatment of negation in a variety of Zulu dictionaries ranging from paper to online dictionaries and compare it to the findings of negation as reflected in the available corpora discussed in the foregoing section. Dictionaries are fundamental resources for language learning, however, lexical resources for Zulu are still very limited, and machine-readable lexicons are not freely available.

In Table 4, we show how some well-known Zulu paper dictionaries, namely the bilingual general dictionary of Doke et al. (2005), the bilingual learners' dictionaries of Dent and Nyembezi (1969) and of De Schryver (2010), and the monolingual general dictionaries of Nyembezi (1992) and Mbatha (2006) deal with the negation phenomena of Zulu verbs. It is conspicuous that negation is treated inconsistently in the various dictionaries.

Dictionary	morph. negation	syntactic negation	outer matter
Doke et al.	Yes	Yes	Notes / Tables
Dent & Nyembezi	Yes, two examples provided with 'not'	Yes	No info on negation
De Schryver (ed)	Occasional examples, Textboxes	Yes, Textbox	Mini-Grammar
Nyembezi	No	Yes	No info on negation
Mbatha	No	only <i>phinde</i>	No info on negation

Table 4: Negation in printed dictionaries

Doke et al. (2005) list syntactic negation by means of the two (auxiliary) verb stems *musa* 'don't' and *yeka* 'leave off; stop; let go'. In the case of *musa*, the plural *musani* is also listed, as well as the information for the user that this verb is used to form negative imperatives 'don't; you mustn't'. The outer matter also contains notes and tables dealing with negation.

In a scholar's dictionary such as that of Dent and Nyembezi (1969) one would expect some outer matter information on negative constructions to guide scholars. The following is the only information available: the two (auxiliary) verb stems *musa* 'don't' and *yeka* 'leave off;

stop; let go' as well as the conjunctive *phinde* 'never' are included in the Zulu-English side of the dictionary, while a lookup under 'not' on the English-Zulu side, actually provides two negated verb constructions *angiboni* 'I do not see' and *asibonanga* 'we did not see'.

In De Schryver's (2010) bilingual school dictionary, morphological as well as syntactic negation are included in the dictionary with occasional examples and textboxes referring the user to the mini-grammar in the outer matter that contains tables of negative forms. This is illustrated in Figure 2 below.

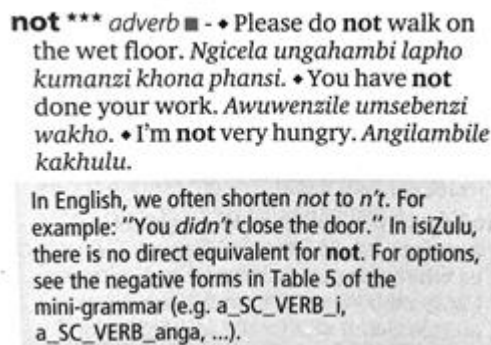


Figure 2: Example of textbox (De Schryver, 2010:431)

Although Nyembezi (1992) lists *musa* and *yeka* as (auxiliary) verb stems with the meanings 'do not; stop doing' there are no examples provided, and no description of any negation in the outer matter. The same applies to Mbatha's (2006) monolingual dictionary. In fact, syntactic negation in this dictionary is limited to the auxiliary conjunctive *phinde* 'never'.

isiZulu.net (2018) functions as a Zulu-English online dictionary that also offers morphological decomposition without the need of stem identification before a word is looked up. Prinsloo (2012:135) describes isiZulu.net as "probably the most sophisticated online dictionary for the Bantu languages." A fairly high amount of back matter is offered in the form of grammar and verb conjugation tables (which we used for developing the regular expressions in section 3). However, the only negative morphemes that occur in the tables are the first person singular subject concord *-ka-*, and *-zu-* the negative form of the future tense morpheme. isiZulu.net (2018) already offers a translation for negated verbs, e.g. *angihambi* is translated as 'I do not go'.

The individual analyses of lookups present automatic morphological decomposition, which in the case of negative verb forms decomposes the prefixes, i.e. the negative morpheme and subject concord, but the negative suffixes are only decomposed selectively, e.g. those of the past tenses. Nevertheless, learners of the language can use this information as a pattern for producing other negated verbs. Figure 3 shows three respective analyses by isiZulu.net (2018).

So far, we do not see a sufficient treatment of negation in Zulu in the major (paper) dictionaries, except maybe the Oxford learners' dictionary of De Schryver (2010). This dictionary, however, is rather small and addresses mainly

learners. We are also not informed whether there are still newer editions of the existing printed dictionaries of Zulu planned. However, for such, we would suggest adding a number of textboxes which describe at least the negation forms of highly frequent verbs and rules for forming irregular (defective) forms. Syntactic negation should at least be mentioned with the respective auxiliaries adding examples of their use. Respective back matter information in the form of conjugation tables and/or mini-grammars should be added to all bilingual dictionaries.

abazanga [aba'za.ŋa] ← <i>a + ba + azi + anga</i> <i>Recent or Remote Past Negative [←azi (v.)]</i> <i>SC: ba- (cl. 2, cl. 2a)</i> they did not know
angihambi [aŋi'ha.mbi] ← <i>a + ngi + hamb</i> <i>Present Tense Negative [←hamba (v/i.)]</i> <i>SC: ngi- (1p sg.)</i> I do not go
akalambile [aɣalam'bi.le] ← <i>a + ka + lambile</i> <i>Present Tense Negative [←lambile (v/stat.)]</i> <i>SC: ka- (cl. 1, cl. 1a)</i> he is not hungry she is not hungry it is not hungry

Figure 3: isiZulu.net (2018) analyses of *abazanga*, *angihambi* and *akalambile*

5. Requirements for Improved (Electronic) Zulu Bilingual (Learners') Dictionaries

It is not known how negated verbs in isiZulu.net (2018) are analysed and we do not wish to speculate. In general, however, we do not think that changing the data model of the dictionary's database is a solution because morphological negation is a dynamic process of word formation. We rather see a query processor first checking whether the word queried by the user is contained in the database. If that is not the case, an analysis of the word must take place. In our view, there are two possible options for such an analysis tool when extending electronic dictionaries so that negated verb forms can be queried:

- Implementing a rule-based component on the basis of regular expressions as it was done for a few examples in the Zulu Learners' dictionary (Faaß and Bosch 2016); this method could be enhanced by utilizing a dictionary of affixal negations as suggested by van Son et al. (2016).
- Adding ZulMorph⁵, the Finite State Morphological Analyser for Zulu as described in Bosch and Pretorius (2016:11) as a component of the dictionary.

The implementation of ad-hoc rules as described for option (a) would offer an opportunity to select and show the most probable analysis of a word form and to add

⁵A demo version of ZulMorph is accessible at <http://gama.unisa.ac.za/demo/demo/zulmorph>

didactic information for language learners, i.e. the users of the dictionary for instance by adding a link to adequate online lectures concerned with negation or by giving additional explanations on special cases. Such a component could be limited to the vocabulary and morphology addressed in the teaching materials as suggested by Antonsen (2013) for cases of morphologically complex indigenous languages that do not have morphological analysers. Instead of putting all the knowledge and the processing in one component, one could alternatively use the dictionary of affixal negation as proposed by van Son et al. (2016) as a model and compile a new dictionary of isiZulu affixal negations with data of the Zulu wordnet which is based on the English Princeton WordNet (cf. Bosch and Griesel, 2017). Another option would be to feed such a dictionary with data from the part-of-speech ontology implemented by Taljard et al. (2015). The result would become a knowledge base of which a processing component could make use of. Such an additional dictionary could also contain additional information on regular antonyms, again taken from wordnet data, e.g. *bonakala* ‘appear’ vs. *nyamalala* ‘disappear’. However, implementing morphological rules to reproduce the natural processes of negation is an effort already performed with the existing finite state transducer (FST) machine and by adding such rules and extra data to a dictionary we would in a way re-invent the wheel. We hence rather look at ways and means to add the FST machine as a module to the dictionary. Here we are however facing the first challenge, namely that in the case of ambiguous words, the analyser returns multiple analyses: just for a rather simple verb like *abahambi* ‘they do not walk’, the FST offers five different analyses, *ungathi* ‘you/it do(es) not say’ even results in as many as 24 analyses. A solution for this problem could be an often-used and reliable method to reduce the number of analyses: the application of Optimality Theory (OT) (Archangeli and Langendoen, 1997) on this FST, i.e. by ranking its paths in order to find the most probable one. Such task would also be useful for instance for developing a parser or when making use of the FST for tagging, etc.

Another challenge is that of underspecification: When querying the verb form *ayibaleki* in ZulMorph, there are 12 analyses delivered⁶. For the verb root *-bal-* identified in (9) and (10), ZulMorph finds two valid analyses: The verb root *-bal-* means “count”; here it is extended with the neuter extension *-ek-* changing its meaning to the intransitive “be countable”. No object concords occur in these analyses.

- (9) a[NegPre]
i[SC][4]
bal[VRoot]ek[NeutExt]
i[VTNeg]
“they are not countable”

- (10) a[NegPre]
i[SC][9]
bal[VRoot]ek[NeutExt]
i[VTNeg]
“he/she/it is not countable”

In (11) and (12), the intransitive root *-balek-* ‘run away’ is identified. Again, no object concord is identified; the analyses are therefore both valid.

- (11) a[NegPre]
i[SC][4]
balek[VRoot]
i[VTNeg]
“they do not run away”
(12) a[NegPre]
i[SC][9]
balek[VRoot]
i[VTNeg]
“he/she/it does not run away”

Analyses (13) to (20) can be ignored because the identified base verb root *-al-* ‘deny; refuse; reject’ contains an object concord together with the neuter extension *-ek-* which in each case, changes the verb’s valency⁷.

- (13) a[NegPre]
i[SC][4]ba[OC][2]
al[VRoot]ek[NeutExt]
i[VTNeg]*
(14) a[NegPre]
i[SC][9]ba[OC][2]
al[VRoot]ek[NeutExt]
i[VTNeg]*
(15) a[NegPre]
i[SC][4]bu[OC][14]
al[VRoot]ek[NeutExt]
i[VTNeg]*
(16) a[NegPre]
i[SC][9]bu[OC][14]
al[VRoot]ek[NeutExt]
i[VTNeg]*
(17) a[NegPre]
i[SC][9]bu[OC][14]
alek[VRoot]
i[VTNeg]*
(18) a[NegPre]
i[SC][4]ba[OC][2]
alek[VRoot]
i[VTNeg]*
(19) a[NegPre]
i[SC][4]bu[OC][14]
alek[VRoot]
i[VTNeg]*

⁶ The results of ZulMorph were sorted here by the verbal roots identified; numbers and carriage returns were inserted for a better overview.

⁷ So far, ZulMorph is not informed about the valencies of verbs.

(20) a[NegPre]
i[SC][9]ba[OC][2]
alek[VRoot]
i[VTNeg]*

When examining the valid analyses for *ayibaleki*, we find that ZulMorph identifies the following two verb roots as shown in (21) and (22):

- (21) *-bal-* ‘count; calculate’
(22) *-balek-* ‘run away; escape; flee’

Using the Oxford Bilingual School Dictionary (De Schryver, 2010) as guideline with regard to corpus frequencies of verb stems, the most likely verb root in the above list is *-balek-* (two stars - the second group of most frequently used headwords) followed by *-bal-* (one star - the third group of most frequently used headwords). In the corpora consulted, as described in section 3, we investigated the present tense forms (short, long and negative form plus the forms of participial, and subjunctive mood) of *-balek-* and found the occurrences shown in Table 5. We do not know which corpus was used to generate the frequency lists for the Oxford School Dictionary, however our data differs slightly from that of the Oxford School Dictionary.

Verb root	UKZN	LC	NCHLT	UK
<i>-balek-</i>	1,919	166	0	8
<i>-bal-</i>	2,500	233	1	0

Table 5: Frequencies of occurrences of the present tense forms of *-balek-* and *-bal-*

Methodologically, a script working with respective regular expressions (described in the constraints above) which are informed about verb frequencies could determine that *ayibaleki* is a negated verb form with the roots *-balek-* or *-bal-*, of which the more frequent one is the preferred one and should be shown first. We are fully aware of the fact that *ayibaleki* might be a straightforward case, however, we see such a “picking” of the relevant repetitive parts of the analyses as a feasible option when connecting the finite state transducer to an electronic dictionary.

For syntactic negation, that is - from a technical perspective - for analysing and translating word sequences showing negation elements like *musa* or *yeka*, a word-based dictionary will most probably not be capable of offering the correct translation. For translating sequences, we are in need of a parser and/or a machine translation tool.

6. Conclusion and Future Work

In this paper, we examined the linguistic phenomena of morphological and syntactic verbal negation in Zulu. These are not very prominently discussed in printed dictionaries though they are difficult to (de-)construct for learners. In the only existing electronic dictionary providing a good coverage, isiZulu.net (2018), such

negation is handled appropriately, however, as no publications exist, we can only speculate on how this implementation was done.

Verbal negation occurs frequently in the existing corpora of the language, we may thus assume that learners are confronted with verbal negatives quite frequently, especially in reception (for example in newspaper texts that were collected in the UKZN corpus). We hence provide suggestions on enhancing presentations in printed dictionaries, for example by making more extensive use of textboxes illustrating the linguistic phenomena in question (cf. Gouws and Prinsloo, 2014).

As Prinsloo et al. (2012) rightly state: “there are numerous complex situations where users need more detailed support than currently available in e-dictionaries, to make valid and correct choices”. The proposal of Kovarikova et al. (2012) to interconnect affirmative and negative forms individually via referencing tools in e-dictionaries is a valid proposal too. We thus offer suggestions (including the incorporation of existing data and software) on how to enhance electronic dictionaries as language information tools so that they can handle at least the morphological negation phenomena appearing in Zulu and its related languages.

Although we only pay attention to negation in Zulu in this paper, this approach may lay the foundation for the lexicographic treatment of further complex constructions in Zulu, as well as negation in electronic dictionaries for the other four Nguni languages that are closely related to Zulu.

7. Acknowledgements

The various phases of research activities related to this paper were supported by the Scientific eLexicography for Africa project; the South African Centre for Digital Language Resources (SADiLaR); and the South African National Research Foundation. We also acknowledge the useful feedback from the anonymous reviewers.

8. Bibliographical References

- Antonsen, L. 2013. Why ICALL for indigenous languages?
https://en.uit.no/tavla/artikkel/sub?sub_id=354114&p_document_id=320146#antonsen
(accessed on 15/01/2018)
- Archangeli, D. and Langendoen, D.T. (eds.). 1997. *Optimality Theory: An Overview*. Oxford: Blackwell Publishing.
- Bosch, S. and Griesel, M. 2017. Strategies for building wordnets for under-resourced languages: the case of African languages. *Literator* 38(1), a1351.
<https://doi.org/10.4102/lit.v38i1.1351>
- Bosch, S.E. and Pretorius, L. 2017. A computational approach to Zulu verb morphology within the context of lexical semantics. *Lexikos* 27:152-182.
- Crystal, D. 1994. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishers.
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17:79-106.

- De Schryver, G.-M. (ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English*. Cape Town: Oxford University Press Southern Africa.
- Dent, G.R. and Nyembezi, C.L.S. 1969. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter and Shooter.
- Doke, C.M., Malcolm, D.M., Sikakana, J.M.A. and Vilakazi, B.W. 2005. *English-Zulu, Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Evert, S. and Hardie, A. (2011). Twenty-first century Corpus WorkBench: Updating a query architecture for the new millennium. *Proceedings of the Corpus Linguistics 2011 Conference*. University of Birmingham, UK.
- Faaß, G. and Bosch, S. 2016. An Integrated e-Dictionary Application - the case of an open educational trainer for Zulu. *International Journal of Lexicography* 29(3):296-310.
- Gouws, R. and Prinsloo, D. J. 2014. Thinking out of the box. In: *Proceedings of the 16th EURALEX International Congress*, pages 501-511. Bolzano/Bozen, Italy
- Gowlett, D. 2003. Zone S. In Derek Nurse and Gérard Philippson (eds) *The Bantu Languages*. Pp. 609-638. London and New York: Routledge.
- isiZulu.net Zulu-English Dictionary. 2018. <https://isizulu.net/> (accessed on 15/01/2018)
- Kosch, I.M. 2006. *Topics in Morphology in the African Language Context*. Pretoria: Unisa Press.
- Kovarikova, D., Chlumská, L. and Cvrček, V. 2012. What belongs in a dictionary? The Example of Negation in Czech. In: Ruth Vatvedt Fjeld & Julie Matilde Torjusén (Eds.): *Proceedings of the 15th EURALEX International Congress*. 7-11 August 2012, Oslo, pp 822-827. Oslo: Reprosentralen, UiO. ISBN 978-82-303-2095-2. Accessed on 14 March 2016.
- Mbatha, M.O. (ed.). 2006. *Isichazamazwi SesiZulu*. Pietermaritzburg: New Dawn Publishers.
- Nyembezi, S. 1992. *Isichazamazwi sanamuhla nangomuso*. Pietermaritzburg: Reach Out Publishers.
- Prinsloo, D. J. 2012. Electronic Lexicography for lesser-resourced languages. In Sylviane Granger & Magali Paquot (Eds): *Electronic Lexicography*. Pp. 119-144. Oxford: Oxford University Press.
- Prinsloo, D.J. and Gouws, R.H. 1996. Formulating a new dictionary convention for the lemmatization of verbs in Northern Sotho, *South African Journal of African Languages* 16(3):100-107.
- Prinsloo, D. J, Heid, U., Bothma, T. and Faaß, G. 2012. Devices for information presentation in Electronic Dictionaries. *Lexikos* 22:290-320.
- Quasthoff, U., Goldhahn, D. and Bosch, S. 2016. Morphology Learning for Zulu. Claudia Soria et al. (eds.) *Workshop CCURL 2016 - Collaboration and Computing for Under-Resourced Languages - 'Towards an Alliance for Digital Language Diversity'*. 23 May 2016. 10th International Conference on Language Resources and Evaluation, Portoroz, Slovenia. Pp.89-95.
- Spiegler, S., van der Spuy, A. and Flach, P.A. 2010. Ukwabelana - an open-source morphological Zulu corpus. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, August 2010.
- Taljad, E., Faaß, G., and Bosch, S.E. 2015. Implementation of a Part-of-Speech Ontology: Morphemic Units of Bantu languages. *Nordic Journal of African Studies* 24(2):146-168.
- Van Son, C., van Miltenburg, E. and Morante, R. 2016. Building a Dictionary of Affixal Negations. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 49-56 (ExProM 2016). Osaka, Japan.

9. Language Resource References

- NCHLT isiZulu Annotated Text Corpora. 2014. <https://rma.nwu.ac.za/index.php/resource-catalogue/isizulu-nchlt-annotated-text-corpora.html>. ISLRN 729-409-508-086-4.
- UKWABELANA http://www.cs.bris.ac.uk/Publications/pub_master.jsp?id=2001224
- University of KwaZulu-Natal. 2018. IsiZulu National Corpus. <https://iznc.ukzn.ac.za/>
- Wortschatz Universität Leipzig. 2018. Corpus: Zulu (zul_mixed_2016). <http://wortschatz.uni-leipzig.de/de>
- ZulMorph. n.d. Finite state morphology demo. <http://gama.unisa.ac.za/demo/demo/zulmorph>

Annexure A

Frequencies of occurrences of negated verbs which might also be deverbative nouns (included in the corpus query results displayed in Table 3).

Ambig.Deverb.N.	UKZN	ZULU	NCHLT	UK
<i>ababhali</i>	346	54	0	
<i>ababukeli</i>	219	62	0	
<i>ababulali</i>	256	19	0	
<i>abaculi</i>	2,018	264	0	
<i>abacwaningi</i>	170	70	2	
<i>abadayisi</i>	136	27	0	
<i>abadidiyeli</i>	105	15	0	
<i>abadlali</i>	5,979	678	3	1
<i>abafundi</i>	7,143	1,743	59	1
<i>abafundisi</i>	189	55	3	1
<i>abagadli</i>	277	31	0	
<i>abagibeli</i>	499	197	0	
<i>abagijimi</i>	226	14	0	
<i>abagququzeli</i>	119	25	0	
<i>abahlali</i>	497	183	1	
<i>abahlaseli</i>	74	37	0	
<i>abahleli</i>	369	73	0	
<i>abahloli</i>	169	53	4	
<i>abahluleli</i>	40	9	0	
<i>abahluzi</i>	55	3	0	
<i>abaholi</i>	2,121	487	0	
<i>abakaki</i>	5	1	0	
<i>abakhongi</i>	296	10	0	
<i>abalaleli</i>	356	103	0	
<i>abalandeli</i>	2,692	443	0	
<i>abalimi</i>	491	583	3	

<i>abalingisi</i>	370	57	0	
<i>abalobi</i>	62	13	0	
<i>abalози</i>	44	4	0	
<i>abameli</i>	381	95	2	
<i>abangani</i>	1,125	159	0	
<i>abanini</i>	49	28	0	
<i>abaphathi</i>	1,194	313	2	
<i>abaqashi</i>	259	124	0	
<i>abaqegeshi</i>	557	65	0	
<i>abasakazi</i>	525	157	0	
<i>abasebenzi</i>	2,791	1,546	25	
<i>abaseshi</i>	148	39	0	
<i>abashayeli</i>	657	235	0	1
<i>abasiki</i>	207	26	0	
<i>abasizi</i>	84	19	0	
<i>abathakathi</i>	237	9	0	
<i>abathandi</i>	627	92	0	
<i>abathengi</i>	412	159	0	
<i>abaxhasi</i>	287	67	0	
<i>abazali</i>	3,798	609	8	5
<i>abefundisi</i>	607	40	0	1
<i>abelusi</i>	95	6	0	
Total	39,363	9,101	112	10