Bootstrapping Siswati lexical resources from isiZulu

Roné Wierenga
NLP Research Group
CSIR
Pretoria, South Africa
rwierenga@csir.co.za

Laurette Marais
NLP Research Group
CSIR
Pretoria, South Africa
lmarais@csir.co.za

Ilana Wilken
NLP Research Group
CSIR
Pretoria, South Africa
iwilken@csir.co.za

Abstract

IsiZulu and Siswati are closely related languages that share significant morphosyntactic characteristics. Systematic differences between these languages have been identified at the phonological and morphosyntactic levels. Due to the resource-scarce status of these languages, this similarity has led to bootstrapping of computational language resources at the morphological and syntactic levels. In this work, we investigate the feasibility of adapting lexical items in a computational lexicon from isiZulu to Siswati. We use Grammatical Framework resource grammars for both languages to analyse and transform lexical items, which are then evaluated against a parallel term list. An iterative process yields a success rate of 70.5%, indicating that this approach is largely viable as a means of significantly reducing the manual effort needed to develop lexicons for computational resources for Siswati.

1 Introduction

Siswati is a Southern Bantu language of the Nguni language family. Many Nguni languages, including isiZulu and Siswati, exhibit a high degree of mutual intelligibility, which facilitates the adaptation of linguistic resources developed for one language to another with relatively minor modifications (Bosch et al., 2008; Marais et al., 2024).

Both Siswati and isiZulu are recognied as official languages in South Africa; however, census data indicates that isiZulu is the most widely spoken language in the country, with over 15 million speakers, whereas Siswati is a comparatively smaller language group, with approximately 3 million speakers in South Africa and around 1 million speakers in the Kingdom of Eswatini (Moors et al., 2018). Consequently, significantly more linguistic resources have been developed for isiZulu than for Siswati, despite both languages being classified as under-resourced. Marais et al. (2024) demonstrated that syntactic resources can be successfully bootstrapped from isiZulu to Siswati due to their shared linguistic features, such as conjunctive orthography and morphophonological affixing (refer to Taljaard and Bosch (1988) for an outline of isiZulu grammar and Taljaard et al. (1991) for Siswati). Building on this research, the present study employed semi-automatic methods to derive Siswati computational lexicons from isiZulu ones. We show how parallel Grammatical Framework Resource Grammars for isiZulu and Siswati enable the bootstrapping process by enabling morphosyntactic analysis of isiZulu terms, as well as generation of parallel Siswati terms. An evaluation is performed by comparing the automatically generated terms with those found in parallel dictionaries for isiZulu and Siswati.

2 Morphophonological and Lexical Differences Between isiZulu and Siswati

As previously mentioned, Siswati and isiZulu are closely related Bantu languages. The grammars of both languages are governed by two fundamental principles: (i) nominal classification, which involves a system of noun classes, and (ii) the conjugation of various word categories to ensure concordial agreement. Typically, nouns consist of a noun class prefix and a root or stem. Noun prefixes determine the noun class, which functions as a grammatical gender system, grouping nouns into specific classes. Each noun class is associated with systematic parameters that regulate grammatical agreement. Additionally, noun classes are numbered, and the noun class systems of isiZulu and Siswati are largely similar.

Both languages also employ a structural element often referred to as an agreement morpheme, which formally marks the relationship between the noun and all other words that share a semantic-syntactic relationship with it, ensuring grammatical (or con-

cordial) agreement. Consider the following isiZulu example:

(1) *Ukudla kumnandi.*
'The food is delicious.'

In example (1), the noun class prefix *uku-* establishes agreement with the verbal prefix *ku-*, which expresses third-person singular in the present tense. This type of morpheme is referred to as subject concord. These concord morphemes can mark subject-verb agreement and express semantic functions such as possession, vocation, negation, and plurality. Marais et al. (2024) summarise key systematic morphophonological differences between isiZulu and Siswati as outlined in existing literature. For clarity, a brief overview is provided below:

Although the lexicons of isiZulu and Siswati are very similar, certain lexical items, including loanwords, exist in one language without existing in the other. These differences in the two languages are not necessarily systematic and it is not necessarily possible to identify which loan words or lexical items are likely to occur in both languages or where the lexicons of these languages diverge.

3 Methodology

Figure 1 shows an overview of our methodology, which consists of three distinct phases. These phases are indicated by arrows of different colours: light green for Phase 1 (Term list compilation), light blue for Phase 2 (Lexicon adaptation) and dark blue for Phase 3 (Evaluation).

3.1 Phase 1: Term list compilation

The term list was compiled by manually capturing lexical items from *The Official Foundation Phase CAPS English-isiZulu Dictionary* and *The Official Foundation Phase CAPS English-Siswati Dictionary*, which are picture dictionaries developed by the National Lexicography Units of South Africa for Grades R to 3. A parallel list of terms represented by single words in both isiZulu and Siswati was extracted. An overview of the work done in the phase is given in Section 4.

3.2 Phase 2: Lexicon boostrapping

Phase 2 involved the the utilisation of the isiZulu Resource Grammar (RG) to facilitate the identification of roots and stems within the isiZulu term list by parsing the isiZulu words. Having identified systematic phonological and orthographic differences between isiZulu and Siswati from the literature, a Python script was developed that adapts isiZulu roots and stems according to these differences. Finally, the Siswati RG was used to generate (via linearisation) Siswati terms using the adapted roots and stems. This process is discussed in Section 5.

3.3 Phase 3: Evaluation

The final phase was to evaluate the lexical items generated via the Siswati resource grammar by comparing them to the manually captured term list developed during Phase 1. The main aim of this phase was to quantify the success of the bootstrapping approach via adaptation, as well as to categorise and quantify the errors produced. In this way, insights could be gained about the feasibility of the bootstrapping approach for development of Siswati lexica from isiZulu sources in future, with reference to the amount and the nature of post-editing required to obtain a high quality resource. The analysis and evaluation of the process is discussed in Section 6.

4 Phase 1: Term list compilation

One of the parallel resources used for this project is the Official Foundation Phase Curriculum and Assessment Policy Statement's (CAPS) bilingual picture dictionary series published by the South African National Lexicography Units (NLUs) for each of South Africa's official languages. For this project, the English-isiZulu (Mbatha et al., 2018) and English-Siswati (Lubisi et al., 2018) dictionaries were used. The picture dictionaries use English as the pivot language to facilitate lexical acquisition in Grades R, 1, 2, and 3, collectively known as the foundation phase in the South African education system. Children in this phase are typically between the ages of six and nine. The picture dictionaries constitute a parallel term list in 11 of South Africa's official languages (excluding South African Sign Language). With English as the pivot language, each term corresponds to a single English word, but is often expressed in the other languages using multi-word terms.

Since these dictionaries were developed by the NLUs under the Pan South African Language Board (PanSALB), we can assume that the lexicons were compiled by (i) native speakers of each language and (ii) teams of linguistic and language-practice experts specialising in various sub-fields. Given that many South African Bantu languages,

Morphophonological difference	**isiZulu**	**Siswati**
Alphabet and click omission	/c/, /q/ and /x/	/c/
Consonant substitution or addition	/z/ /th/ and /t/ /th/ and /t/ /d/ /d/ /mp/ /nk/	/t/ /tf/ before /o/, /u/ and /w/ /ts/ before /a/, /e/ and /i/ /df/ before /o/, /u/ and /w/ /dz/ before /a/, /e/ and /i/ /mph/ /nkh/
Pre-prefix vowel deletion, addition and substitution	Noun class prefix consists of a consonant-vowel sequence (also referred to as the basic prefix), preceded by a so-called augment (also referred to as class pre-prefix), a preceding copy vowel that fulfils different grammatical functions, e.g., definiteness and specificity, and is subject to morphophonological processes such as vowel deletion and coalescence.	This augment is only present in classes 1, 3, 4, and 6 and in class 9 where it precedes a nasal consonant.
Relative construction and concords	a-	la-
Orthography	Demonstratives are written disjunctively from the noun that follows.	First position demonstrative ('this/these') is written conjunctively with the following noun.
Imperative	Monosyllabic verb stems, yi- or i- are prefixed or -na suffixed to the stem for the imperative directed at one person.	-ni is suffixed to monosyllabic verb stems for imperatives directed at one person.

Table 1: Systematic morphophonological differences between isiZulu and Siswati.

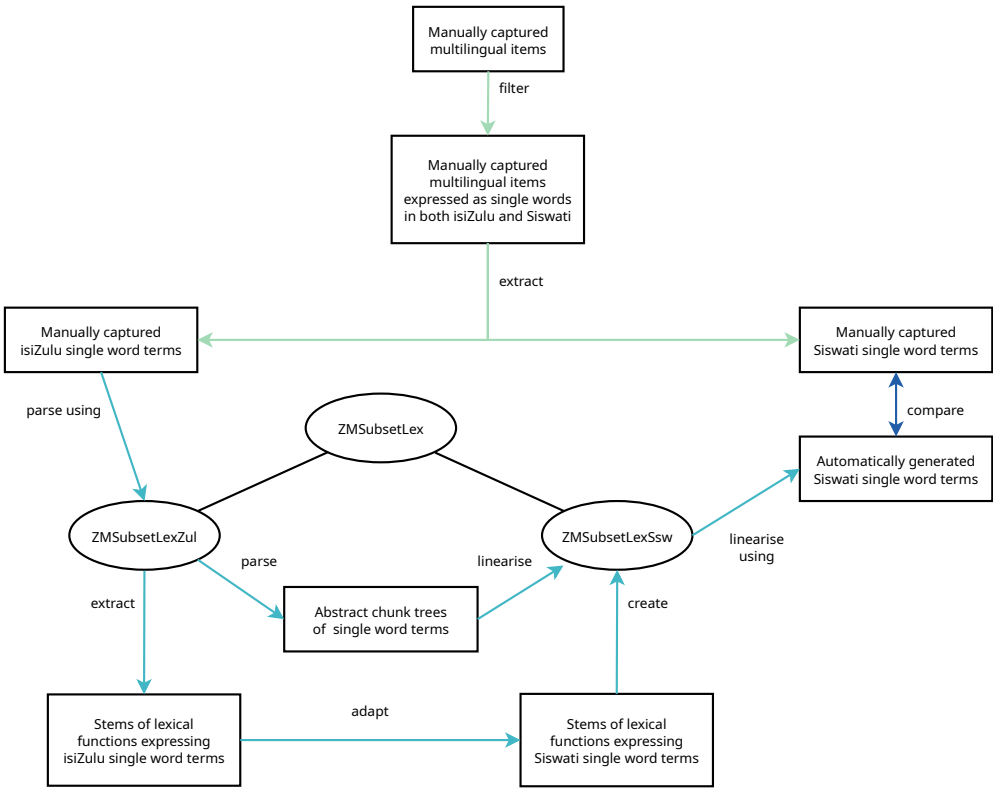

Figure 1: Methodology overview

including Siswati, remain under-resourced, certain terminologies do not yet exist (refer to Khumalo and Nkomo (2022) for a discussion on the challenges of developing terminology in isiZulu). Thus, generally speaking, there are often uncertainties regarding the validity of specific translations due to dialectal variation or the absence of standardised forms for specific terms (Khumalo and Nkomo, 2022). However, these particular picture dictionaries can be considered verified and relatively reliable, since any standardisation issues or dialectal disparities would have been resolved by the respective NLUs. The process to capture terms from the dictionaries for each language involved intensive manual work and was done on a Microsoft Excel spreadsheet. First, all English terms were captured by hand and categorised according to the preset themes and categories as set out in the dictionaries. For each English term, a tag was added to indicate if the term is a noun, verb, adjective, adverb or a sentence, as indicated in the dictionaries. The isiZulu and Siswati translations of each term were added next to the English. For several English terms, more than one translation was provided in either isiZulu or Siswati. In those instances, cells were added to the spreadsheet to ensure the translations still aligned across the three languages. For the English-isiZulu list, a total of 1990 terms were recorded, and 1917 terms for the English-Siswati list. These terms were mostly nouns, followed by verbs, adjectives, sentences, and adverbs.

From this term list, all terms that comprised single words in both isiZulu and Siswati was extracted. The reasons for this is that in order to develop parallel Grammatical Framework (GF) lexicons, a reliable list of meaning-equivalent isiZulu and Siswati words is required for the bootstrapping process.

5 Phase 2: Lexicon bootstrapping

The isiZulu resource grammar has a large lexicon with several thousands of lexical items, and the work described in this paper is aimed at the eventual development of a similarly sized Siswati lexicon.

The artifact in view, therefore, is a GF lexicon module that can be used in conjunction with the Siswati resource grammar. Such a lexicon is based on specifying the noun stems and verb roots, with the grammar itself handling all the applicable morphology. Therefore, a list of Siswati stems and roots must be developed to serve as the basis for a GF lexicon.

In this section, we describe the process of leveraging the parsing and linearisation capabilities of

Lexical category	Occurrences	%
Nouns	756	73%
Verbs	239	23%
Adjectives	30	3%
Adverbs	13	1%
Total	1038	

Table 2: Lexical categories of words in final term list

the GF runtime (Angelov and Ljunglöf, 2014), in conjunction with the isiZulu and Siswati resource grammars, to extract stems and roots from the isiZulu single word terms, to adapt these to Siswati, and then to automatically generate Siswati single word terms using a script based on the known systematic differences set out in Section 2.

5.1 Parsing isiZulu words

A list of about 1500 manually captured isiZulu-Siswati single word term pairs served as the starting point, including nouns, verbs and so-called adjectives. In reality, true adjectives in isiZulu and Siswati are a small, closed class of which the roots can easily be enumerated. English adjectives are almost always translated to isiZulu (and Siswati) using predicate-based qualificatives. In fact, an analysis of the isiZulu Wordnet (Griesel et al., 2019) revealed that 98% of terms captured by translating English adjectives to isiZulu were morphologically complex, predicate-based constructions involving a verb or noun (Marais and Pretorius, 2023).

The isiZulu resource grammar (ZRG) was used to perform a morphosyntactic analysis of each isiZulu word. This was done by employing the ZRG as a morphosyntactic parser. Around 400 items could not be parsed, mostly due to being proper nouns or loan words that are not included in the large isiZulu lexicon. However, a total of 1138 lexical items were successfully parsed, representing a success rate of about 75% and providing a suitably representative basis for the adaptation experiment. A breakdown of the composition of the final term list according to lexical category is given in Table 2.

As an example of the output of the parsing process, and to illustrate how English adjectives that were translated as qualificatives were handled, consider the syntax tree obtained by parsing the isiZulu token *okuthambile*, as shown in Figure 2. Apart from the leaf nodes, each node label consists of a function name and a syntactic category, separated by a colon.

The tree represents a relative sentence chunk involving a so-called stative verb, which may be translated to English in a somewhat syntactically equivalent way as '(that) which has become soft', but is more naturally translated simply as '(that) which is soft'. The presence of the pronoun (`Pron`) in the tree is due to the fact that the initial morpheme *oku-* is the relative agreement morpheme of class 17, which is often used to indicate indefinite agreement in the absence of a specific noun. The English adjective was therefore captured by the dictionary developer using a relative construction with a stative verb and a particular choice of agreement information (it should be noted that in other cases, different choices for agreement were made). It is this stative verb that our process is aimed at adapting from isiZulu to Siswati.

The leaf nodes of the tree represent the morpheme-based substrings that make up the word. In fact, one of them is the root of the verb that we are seeking to extract in order to adapt to Siswati, namely *thamb*, which means 'to become soft'. This information is captured in the ZRG in its lexicon module in the following way:

```
become_soft_V : V ;           -- abstract
become_soft_V = mkV "thamb" ; -- concrete
```

In this code, `become_soft_V` is the name of an abstract lexical function[1] that produces an expression of type `V`, representing a verb. This is defined in the abstract module. The concrete module contains the logic for how this verb should be represented in isiZulu. Hence, `mkV` is an operation for generating a record that represents an isiZulu verb, and it uses the string "`thamb`", the root, to do this.

The Siswati resource grammar (SRG) also has an operation `mkV`, which similarly accepts a root string as argument and produces the correct record representing a Siswati verb. We know that a systematic change that replaces /th/ with /ts/ in this context would yield the correct root *tsamb*. Hence, our goal for this example would be to produce the following line of code to be included in a Siswati lexicon module:

```
become_soft_V = mkV "tsamb" ; -- concrete
```

This process is discussed in the next section.

[1]While the function name has been chosen to simplify discussion of the GF code for English readers, it has no computational significance.

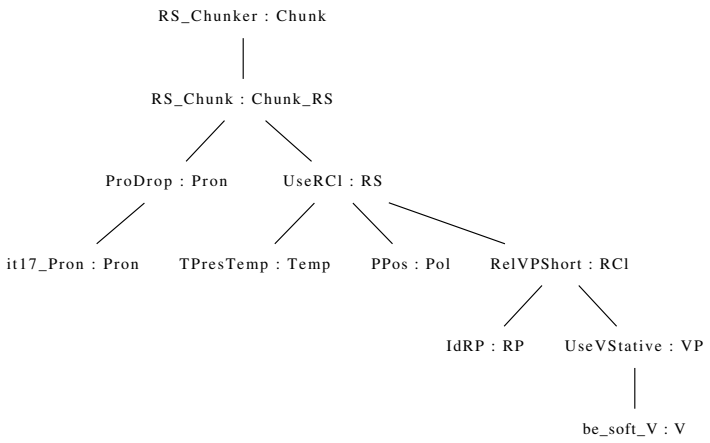

Figure 2: Abstract tree obtained by parsing the isiZulu token *okuthambile*

5.2 Automatic adaptation of stems and roots

Once the morphosyntactic trees were obtained and the roots and stems within them identified, a systematic adaptation could be implemented to facilitate the compilation of a Siswati lexicon module. Since the adaptation approach was applied directly to roots and stems, there was no need to account for additional contextual factors. Consequently, the adaptation solution functioned effectively as a find-and-replace mechanism, as root- or stem-adjacent morphemes had already been stripped from the input. In fact, our process captures roots and stems in their lemmatised forms. The replacement rules shown in Table 3 were applied in the specified order.

This solution has certain limitations. In its current form, it does not distinguish between loanwords and other lexical items. Since loanwords do not always conform to the same phonological rules as native lexical items, it is possible that the substitution rules may not apply consistently to them. However, no existing module or dictionary systematically differentiates loanwords from other lexical items in either isiZulu or Siswati. As a result, this distinction would have to be made manually.

Additionally, the current approach does not account for exceptions to the defined substitution rules. While efforts were made to identify systematic exceptions or deviations, incorporating such exceptions led to unintended errors in the generation of other lexical items. Consequently, no systematic exceptions could be handled without compromising the overall accuracy of the transformation process.

Both of these limitations present opportunities for further research. Future work could explore

isiZulu form	**Siswati form**
q	c
x	c
z	t
tho	tfo
thu	tfu
thw	tfw
to	tfo
tu	tfu
tw	tfw
tha	tsa
the	tse
(thi)[a-zA-Z]	tsi
do	dvo
du	dvu
dw	dvw
da	dza
de	dze
di	dzi
(d)	dz
(th)[ou]	tf

Table 3: Replacement rules for converting isiZulu to Siswati

methods for distinguishing loanwords, potentially through corpus-based analyses or machine learning approaches that focus on identifying phonological patterns. Similarly, refining the substitution rules to accommodate exceptions without introducing errors could enhance the accuracy and applicability of this approach.

5.3 Linearisation of Siswati surface forms

Once the stems were adapted automatically, a Siswati lexicon module was generated as described in Section 5.1. This allowed us to use the SRG, in conjunction with the new lexicon module, to re-linearise the tree shown in Figure 2. The result for our example is the string "lokutsambile", adapted, via the parallel grammars, from "okuthambile".

At this point, it is possible to perform a direct comparison between the generated string and the string found in the original dictionary. If it is identical, we know that the stem was adapted successfully. However, in the case of our example, the Siswati dictionary contains "tsambile", instead of "lokutsambile". The difference is that the generated string matches the morphosyntactic form of the isiZulu dictionary term, while the same term was included in the Siswati dictionary in a slightly different form. The isiZulu term represents a fully realised relative construction using the unspecified agreement value of class 17, while the Siswati term represents a somewhat lemmatised form that lacks any verb prefixes. However, it is clear that the adaptation of the stem succeeded, and that the correct entry in the GF Siswati lexicon was generated, and hence the evaluation process should reflect this.

To account for these differences in morphosyntatic form between the two dictionaries, in all cases where generated Siswati strings failed a direct comparison with the isiZulu term, the Siswati dictionary term was itself parsed using the newly created lexicon. In our example, this means parsing "tsambile" using the SRG along with the newly created Siswati lexicon. The result is shown in Figure 3, which shows the correct lexical function `be_soft_V` used to parse the string as a chunk. This confirms that the adaptation process was successful for the relevant verb root.

6 Phase 3: Evaluation/analysis

As mentioned, the third phase of the experiment focused on evaluating the accuracy of the adaptation solution by comparing the automatically generated

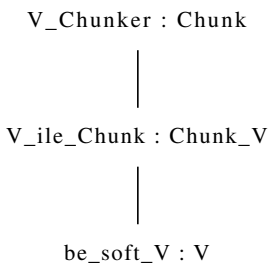

Figure 3: Abstract tree obtained by parsing the Siswati token *tsambile*

Measurement	Number of items
Total isiZulu items	1 038
Successes	488
Reparses	161
Failures	389

Table 4: Overview of bootstrapping results

Siswati terms with those manually captured from the Siswati dictionary.

The success of the boot strapping solution was evaluated by taking several factors into account. These are outlined in this section. The overall quantitative results are provided in Table 4.

The total number of isiZulu items refers to the number of lexical items consisting of a single word extracted from the isiZulu picture dictionary. As mentioned, only single word lexical items were included in the experiment. "Successes" refer to the number of isiZulu lexical items that were successfully converted to Siswati lexical items using the bootstrapping solution. A lexical item was marked as a "reparse" when the isiZulu and Siswati do not constitute the same construction. As set out in Section 5, Siswati terms failing an initial comparison were reparsed to determine whether the reason for the failure is the choice of different forms of the same root or stem by the different developers of the isiZulu and Siswati dictionaries. The reparses are therefore also considered as successfully bootstrapped, because the Siswati root or stem was accurately adapted.

Lastly, Table 4 indicates the number of failures. These were instances where the bootstrapping solution could not automatically generate the correct Siswati lexical item. In order to identify why the Siswati lexical item could not be predicted successfully, the failures were manually categorised into the subgroups shown in Table 5.

Failure classification	Number of items
Total Failures	389
Capitalisation errors	3
Class errors	21
Unique lexical item	229
Phonological differences	39
Exception errors	81
Dictionary errors	16

Table 5: Overview of bootstrapping results

6.1 Analysis of failures

Capitalisation errors: Three Siswati verbs were wrongly categorised as failures because the capitalisation of the words are misaligned. For example, the word Chinese (Eng), *iShayina* (zul) was capitalised as *liShayina* when the term in the dictionary is given as *Lishayina* (ssw). We therefore do not consider this kind of inconsistency as a true error.

Class errors: In these instances the lexical items were generated correctly but the wrong class morpheme was added. For instance *iNgisi* (zul) was used to generate *liNgisi* (ssw); however the correct form included in the Siswati picture dictionary was *Umngisi* (ssw). Only 21 class errors were made.

Unique lexical items and words sourced from other languages: Cases where the isiZulu and Siswati lexical items are not derived from the same root and the Siswati counterpart can in no way be derived from the isiZulu were classified as unique lexical items or words sourced from other languages. The isiZulu and Siswati translations for cellphone are *iselula* and *makhalekhikhini*, respectively. The isiZulu form is derived from the English whilst the Siswati form cannot be derived from either the English or isiZulu, rendering these lexical items impossible to predict using the bootstrapping solution.

Phonological differences: Failures where the Siswati deviates from the isiZulu due to a pronunciation difference that is orthographically respresented, for example language is *ulimi* in isiZulu and *lulwimi* in Siswati, were categorised as phonological differences. Many of the words with phonological differences are loan words from isiZulu in Siswati.

Exception errors: Instances where the bootstrapping solution was applied to lexical items that can be regarded as exceptions to the rules and where the changes made by the bootstrapping solution resulted in the incorrect Siswati form being generated were categorised as exception errors. In these instances the Siswati form is derived from the same root as the isiZulu form but deviates from the isiZulu in an unpredictable or unsystematic way, rendering it difficult to account for these derivations in the Siswati bootstrapping solution without causing a higher failure rate. Some of these deviations are small, for instance an <h> is added to the isiZulu word for ox *inkabi* to form *inkhabi* in Siswati, while others show a higher level of deviation, like the words for mouse *igundane* (zul) and *ligundvwane* (ssw) where the prefix could be correctly generated by the RG but the <vw> in <dvwane> could not be anticipated by or accounted for in the bootstrapping solution. In instances like *ligundvwane* (ssw) the simplicity of the rules stipulated in the bootstrapping solution prevent the correct form from being generated, however, adding layers of nuance to these rules could have rendered the correct form. The refinement of these rules to accommodate more exceptions is reserved for future research.

Dictionary errors: These instances were suspected to result from incorrect dictionary entries, dialectal variants, or human inconsistencies in the Siswati picture dictionary.

Taken together, these results indicate an overall success rate of 62.5% when including capitalisation errors and failures due to the uniqueness of a lexical item - especially those sourced from other languages. If those, along with suspected dictionary errors, are excluded, the bootstrapping solution achieved a refined accuracy of 70.5%. This aligns with other findings in computational Bantu linguistics that show success rates between 65–75% for rule-based transfers between closely related languages (Marais et al., 2024), reinforcing the potential of cross-linguistic bootstrapping in under-resourced contexts.

7 Conclusion

This paper demonstrated the potential of bootstrapping a computational lexicon for Siswati using existing resources in isiZulu, leveraging their close linguistic affinity within the Nguni language cluster. Through the use of the Grammatical Framework resource grammars, we were able to evaluate the application of systematic phonological and morphological transformations to generate parallel isiZulu and Siswati terms.

Findings indicate a success rate of 70.5%, ren-

dering this approach largely viable as a means of significantly reducing the manual effort needed to develop lexicons for computational resources for Siswati. This result is comparable to, and in some cases exceeds, the success rates reported in related works that applied bootstrapping and transfer learning across closely related Bantu languages (Marais et al., 2024), and specifically isiZulu and Siswati. However, the remaining error margin underscores persistent challenges, including phonological exceptions, irregular morphological mappings, and incomplete alignment in lexical categories.

A persistent theme has been the challenge presented by words incorporated into isiZulu and Siswati from other languages, often English. Phonologically focused methods may provide a way of addressing this shortcoming of the current bootstrapping methodology.

These findings affirm the broader linguistic claim that resource sharing within language families can be productive, yet they also support earlier observations in the literature that standardisation discrepancies and dialectal variation must be addressed to optimise outcomes.

Acknowledgments

References

Krasimir Angelov and Peter Ljunglöf. 2014. Fast statistical parsing with parallel multiple context-free grammars. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 368–376.

Sonja Bosch, Laurette Pretorius, and Axel Fleisch. 2008. Experimental Bootstrapping of Morphological Analysers for Nguni Languages. *Nordic Journal of African Studies*, 17(2):23.

Marissa Griesel, Sonja Bosch, and Mampaka Lydia Mojapelo. 2019. Thinking globally, acting locally–progress in the African wordnet project. In *Proceedings of the 10th global wordnet conference*, pages 191–196.

Langa Khumalo and Dion Nkomo. 2022. The intellectualization of african languages through terminology and lexicography: Methodological reflections with special reference to lexicographic products of the university of kwazulu-natal. *Lexikos*, 32:1817–1853.

M. Lubisi, South African National Lexicography Units, and Siswati National Lexicography Unit. 2018. *Official foundation phase CAPS English-Siswati picture dictionary*. South African National Lexicography Units.

Laurette Marais and Laurette Pretorius. 2023. Extending the usage of adjectives in the Zulu AfWN. In *Proceedings of the 12th Global Wordnet Conference*, pages 303–314.

Laurette Marais, Laurette Pretorius, and Lionel Clive Posthumus. 2024. Bootstrapping syntactic resources from isiZulu to Siswati. In *Proceedings of the Fifth Workshop on Resources for African Indigenous Languages@ LREC-COLING 2024*, pages 77–85.

M.O. Mbatha, South African National Lexicography Units, and IsiZulu National Lexicography Unit. 2018. *Official foundation phase CAPS English-isiZulu picture dictionary*. South African National Lexicography Units.

Carmen Moors, Ilana Wilken, Karen Calteaux, and Tebogo Gumede. 2018. Human language technology audit 2018: analysing the development trends in resource availability in all South African languages. In *In Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists, SAICSIT '18*, page 296–304.

P.C. Taljaard and Sonja Bosch. 1988. *Handbook of IsiZulu*. J.L. Van Schaik, Johannesburg.

P.C. Taljaard, J.N. Khumalo, and Sonja Bosch. 1991. *Handbook of Siswati*. J.L. Van Schaik, Johannesburg.