Extending the Usage of Adjectives in the Zulu AfWN

Laurette Marais
Voice Computing
NextGen Enterprises and Institutions
CSIR
South Africa
lmarais@csir.co.za

Laurette Pretorius
Division of Computer Science,
Department of Mathematical Sciences
Stellenbosch University
South Africa
lpretorius@sun.ac.za

Abstract

The African languages Wordnet (AfWN) for Zulu (ZWN) was built using the expand approach, which relies on the translation of concepts in the Princeton WordNet (PWN), while retaining their PWN lexical categories. In this paper the focus is on the adjective as PWN lexical category. What is considered adjectival information (provided both attributively and predicatively) in English, is usually verbalised quite differently in Zulu - often as verb or copulative constructions - as may be seen by inspecting the Zulu written forms in "adjective" entries in ZWN. These written forms are not complete Zulu verb or copulative constructions and in order for them to be useful, tense, polarity and agreement have to be added. This paper presents a grammar-based approach to recover important morphosyntactic information implicit in the ZWN "adjective" written forms in order to derive a tool that would assist a user of the ZWN to render and analyse correct full forms automatically as desired by the context in which an "adjective" is used.

1 Introduction

The central role that the PWN has come to play in computational lexical semantics and meaning representation is well known and has given rise to the development of wordnets for many languages, including the African languages, with the expectation that they would play a similar role in the natural language processing (NLP) of these languages. From the outset, the developers of AfWN were confronted with, amongst others, the structural discrepancies that arise in creating "adjective" entries (Mojapelo, 2014). This resulted in "adjective" entries that are mostly based on word categories other than Zulu adjectives - categories that exhibit their own complex morphosyntactic structure. For this reason, these entries pose unique challenges for the potential users of the AfWN, creating the need for computational tools and methods for using them. We are not aware of any language other than those of the AfWN that exhibit this characteristic and has a wordnet. Indeed, dealing with this issue has not been discussed elsewhere except for Northern Sotho in (Mojapelo, 2014).

In this paper we propose a novel approach for extending the usage of "adjective" entries in the ZWN and we provide a tool that could in due course support applications of the ZWN.

2 Background

It is often stated that the expand model, by "which the source wordnet, usually the English PWN, is translated into the target language, relies on the assumption that the new language shares an underlying structure with the PWN" (Griesel and Bosch, 2020). It is also widely stated that this model is often used in cases where the target language is under-resourced. The AfWN developers adopted the expand model, mainly on the basis of the under-resourcedness of the African languages. What has received less attention is the extent to which the African languages "share an underlying structure with the PWN". The PWN has four main lexical categories: nouns, verbs, adjectives and adverbs. In this article we focus on what is known as the "adjective" (as noun qualifier) in the PWN and how noun qualification in Zulu presents unique challenges for "adjective" entries in the ZWN.

English adjectives (PWN) seldom map to adjectives in Zulu. Zulu employs qualificatives to modify nouns, namely the adjective, the descriptive possessive, the verbal relative, copulative relative and the enumerative[1] (Poulos and Msimang, 1998). The question arises how the ZWN should apply the expand model in this situation. Mojapelo (2014) notes that for Northern Sotho, "[t]he challenge ... is that while they are all meaning equivalents of the same English word category, they straddle a

[1]There are only four enumeratives in Zulu, so we disregard them here.

number of morphosyntactic categories in Northern Sotho, which nevertheless share a semantic function." This also applies to Zulu in that the written forms of the "adjective" entries in the ZWN represent a diverse set of complex mophosyntactic constructions.

Our aim in this paper is to use a computational grammar to recover these implicit constructions of the ZWN written forms and to use this information to generate and analyse all full forms for varying tense, polarity and agreement. We describe how this generation and analysis can be exposed via an end-user command line tool.[2]

The structure of the paper is as follows; In Section 3 we briefly discuss the AfWN, noting aspects of the ZWN that are applicable to our work. In Section 4 we focus on noun qualification and the "adjective" as lexical entry. More specifically, we discuss the challenges and limitations surrounding the choice of written forms to represent complex morphosyntactic constructions in Zulu. Section 5 introduces the GF Zulu resource grammar (ZRG): we explain how the verb phrase is modelled and how it facilitates the rendering of specific full forms of verb phrases. The extension procedure using the ZRG in Section 6 sets out the main contribution. Section 7 presents an evaluation as well as a discussion of the results.

3 The AfWN

Development of the AfWN started in 2008 by following the expand approach - considered the preferred approach (as opposed to the merge approach) for under-resourced languages, using the CBC.[3] Over time the need to be more African focused was recognised, which led to the use of the SIL-CAWL,[4] consisting of 1700 terms which resulted from linguistic research in Africa. Although the PWN includes four open lexical categories (noun, verb, adjective and adverb), the publications of the AfWN, more specifically the ZWN, have up to now addressed predominantly nouns and verbs (see, for example, (Griesel and Bosch, 2020)). These two lexical categories are known to allow for a mostly well-behaved mapping between the nouns and verbs of English and Zulu, provided that the concepts are lexicalised in both languages. The lexical category of adjective, however, does not allow for an equally well-behaved mapping. This has important consequences for the structure and subsequent use of ZWN entries that are labelled as "adjective", as will be explicated in subsequent sections.

4 Noun Qualification

4.1 English

While noun modification in English is achieved through a variety of word categories and constructions,[5] our focus is on the adjective as it occurs in the PWN. Broadly speaking, an adjective in English is a word that defines, qualifies or modifies the meaning of a noun.[6] The use of an adjective in English is either attributive or predicative. The attributive use is the most common use with the adjective almost always coming before the noun. Adjectives are said to be predicative when they are used as the complement of the verb *to be*, or other similar verbs such as *get*, *become*, *grow*, etc. In Tables 1 and 2 we see that no matter how the adjective is used (attributively or predicatively in any tense and polarity), its form (*accessible*, *blind*, etc.) remains basically unchanged. This is, however, not the case when modifying nouns in Zulu and the other languages of the AfWN.

4.2 Zulu

In Zulu, noun qualification is essentially achieved by means of not only the adjective, but also the descriptive possessive, the verbal relative and the copulative relative (see, for example PoulosMsimang). These so-called qualificatives differ in terms of morphosyntactic structure, which raises the question of how they should be handled in the ZWN given the expand model.

Mojapelo (2014) notes for Northern Sotho that "[t]he immediate issue, first of all, is the absence of a one-to-one correspondence between the adjective in English and that in Northern Sotho ... The issue is that a lexicalised equivalent of the sense expressed by an English adjective cannot be ignored on the grounds that it is not an adjective, nor can it be categorized as an adjective while it not". She

[2] Available at: https://github.com/LauretteM/gf-afwn

[3] http://globalwordnet.org/resources/gwa-base-concepts/

[4] https://www.sil.org/resources/archives/7882

[5] For example, the adjective, adjectival phrase, noun, genitive, participle, and even adverbs and sentences.

[6] In English an attributive noun functioning as an adjective qualifying a noun is often used instead of the genitive case or the dative case as in many other languages.

concludes: "The proposal is that while it is understandable that only stems be considered, invariant parts that are separate from the stem but that will help to disambiguate it be retained". It is clear that this approach was also followed by the developers of the ZWN, although the conjunctive orthography of Zulu presents challenges with regards to isolating invariant parts.

For the purpose of this paper, we refer to qualificatives that are employed in Zulu in contexts where English employs an adjective, as adjective-like qualificatives.

Some adjective-like qualificatives, namely the verbal and copulative constructions occur in various tenses and polarities. It is also important to note that an adjective in English, stated positively, is often lexicalised in Zulu by means of a negative predicate-based construction (see Section 4.2.2). Polarity is therefore often an inherent aspect of the lexicalisation of an English adjective as a Zulu qualificative. Descriptive possessives, on the other hand, do not exhibit tense and polarity.

While an exposition of Zulu linguistics, including the nominal classification and concordial agreement systems, falls outside the scope of this article (see, for example, (Poulos and Msimang, 1998) and (Taljaard and Bosch, 1988)), we give a short overview of the adjective-like qualificative constructions found in the ZWN.

4.2.1 Constructions Based on Adjective Stems and Primitive Relative Stems

There are a limited number of so-called adjective stem and primitive relative stems in Zulu.[7] For example, in Zulu, 'the big house' and 'the house that is big' are expressed using the same construction, namely the relative descriptive copulative in the present, positive: *indlu enkulu*. In fact, strictly speaking Zulu does not have an attributive form of the adjective. Rather, nouns are either modified by adjectives in relative clauses or in main clauses. This usage roughly corresponds to what is typically meant by attributive and predicative, and so for the purpose of this paper, we will refer to any relative construction as the attributive form of a qualificative, and to the main clause predicate construction as the predicative form.

[7]For a list of the most common ones, see for example (Poulos and Msimang, 1998, pp.142)

4.2.2 Verbal Constructions

Almost any verb stem can be used to form a qualificative in the form of a verbal relative.

The so-called direct verbal relative[8] represents the attributive use of the English PWN adjective to which it is mapped. Table 1 lists some forms of the verb root *-ngenek-*, which means 'to be accessible'. The ZWN written form for 'accessible' is *ngenekayo*, which is clearly derived from the present positive relative form of the verb. Each entry in the table shows in bold the part of the verb form that also appears in the written form. We see that the full written form appears only once, with *ngeneka* appearing more often and in some cases only *ngenek* is found in the table entry.

The relative suffix *-yo* that is often used in the long form of the verb, is optional, and this can be seen from the Zulu written forms *mangalisa* for 'fabulous', *mangalisayo* for 'amazing'. The choice to include the *-yo* is purely stylistic and the two English senses are in fact lexicalised by the same Zulu construction. However, in some contexts the *-yo* may not be used, as seen in the written form *mangalisa kakhulu* for 'thundering'. Similarly, long and short forms exist for the present and past positive predicative forms. For example, in the presence of an adverb, the short form is typically required, as can be seen in the table.

Sometimes, the English adjective is lexicalised in the ZWN in the negative, such as *ngaboni* for 'blind', which is clearly derived from the present negative relative form of the verb *-bon-*, which means 'to see'. In such cases, however, the written form is even less clearly related to the various other forms of the verb, as can be seen in Table 2. The table lists some forms of the verb root *-bon-*, but the polarity of the Zulu verb is flipped. To be specific, in Zulu, 'to be blind' is lexicalised as 'not to see'. The negation of the verb in the ZWN written form is essential to communicate the correct meaning, but the negative morpheme *nga* is only found in relative constructions, i.e. the attributive forms. Different negative morphemes appear in the predicative forms.

From these examples it is clear that the written forms in the ZWN do not readily allow for the generation of all forms of the Zulu qualificatives that they represent. Not only must additional morphemes be supplied to express the correct agree-

[8]The indirect relative construction falls outside the scope of this article.

ment, tense and polarity, but as we have seen, it is necessary to have knowledge of the internal structure of the written form in order to know how it, or substrings of it, can be used.

4.2.3 Identifiying and associative copulative constructions

The identifying copulative in Zulu is translated in English with the verb 'to be', for example *umuntu nguthisha* ('the person is a teacher'). It is used in the ZWN to lexicalise a small number of adjectives. For example, the written form for 'false' is given as *ngamanga*, which is derived from the relative clause that means 'which is a lie', as in *impendulo engamanga* ('the answer that is a lie').

The associative copulative in Zulu is translated in English with the verb 'to have', for example *umuntu unemoto* ('the person has a car'). It is often used in the ZWN to lexicalise adjectives. For example, the written form for 'believable' is given as *nokukholwa*, which is derived from the relative clause that means 'which has belief', as in *impendulo enokukholwa* ('the answer that has belief').

These copulative constructions are used attributively and predicatively in the various tenses and polarities, as can be seen in the example in Table 3.

4.2.4 Descriptive possessives

Descriptive possessives are inherently attributive in nature and meaning equivalent predicates cannot be derived in a predictable way. This is in contrast to verbal and copulative (predicate-based) constructions that can express the same meaning in the attributive and predicative use by means of relative clauses and predicates in main clauses, respectively. Our focus in this paper is therefore on the predicate-based qualificatives found in the ZWN.

4.3 Problem Statement

A valuable contribution, namely that of mapping English adjectives to Zulu qualificatives, has been achieved in the ZWN. This was done by implicitly capturing the morphosyntactic constructions that represent the lexicalisations of the English senses in Zulu. However, it is clear from the discussion above that there is a gap between what the ZWN provides in its written forms and what would be required by language processing applications. A sophisticated computational solution is required to effectively deal with the complexity of the adjective-like qualificatives in the ZWN.

5 The GF Zulu Resource Grammar

Grammatical Framework is a computational grammar framework and programming language for writing multilingual grammars. A GF multilingual grammar has an abstract syntax as interlingua and one or more concrete syntaxes, one for each language. The abstract syntax defines categories and functions which are implemented in the concrete syntaxes as linearisation categories and linearisation functions. By defining the linearisation categories and functions of a language, the GF runtime is enabled to linearise abstract syntax trees into natural language strings or to parse natural language strings into abstract syntax trees (Ranta, 2011).

A central project of Grammatical Framework has been the development of a Resource Grammar Library (RGL), the core of which is a common abstract syntax that defines linguistic structures found in most languages. For example, it includes categories for nouns and verbs and functions for predication and noun modification.

The original intent of the RGL was to serve as a linguistic software library to enable rapid development of application specific grammars (Ranta, 2009). Implementing the RGL categories and functions for a language would once and for all capture the general morphology and syntax of the language, to be reused by grammars aimed at a specific use case or application. More recently, however, attempts have been made to employ the general use grammars of the RGL towards wide-coverage parsing (Ranta et al., 2020). The RGL supports close to 40 languages and has, for example, been used to develop a parallel Swedish and Bulgarian Wordnet resource (Angelov, 2020).

The Zulu Resource Grammar (ZRG) models the morphology and syntax of Zulu. This is achieved by the implementation of a deliberate selection of functions from the common abstract syntax, in addition to a set of extra language specific abstract functions.[9] A large lexicon and an extension that defines chunks have been developed to enable the use of the ZRG as a wide-coverage Zulu parser.

In the GF RGL, as in the ZRG, the `VP` category is used to model generalised predicates for which tense, polarity and agreement is not yet fixed. `VP`s are used in two main ways, namely to supply the predicate in main clauses and to construct relative

[9] See the README at https://github.com/GrammaticalFramework/gf-rgl/blob/master/src/zulu/README.md

Use	Tense	Pol.	Zulu	English
Attr.	Pres.	Pos.	*indlu* ***engeneka(yo)***	the house that is accessible
		Neg.	*indlu enga****ngeneki***	the house that is not accessible
	Past	Pos.	*indlu* ***engeneki****le(****yo****)*	the house that was accessible
		Neg.	*indlu enga****ngeneka****nga*	the house that was not accessible
	Fut.	Pos.	*indlu ezo****ngeneka***	the house that will be accessible
		Neg.	*indlu engazuku****ngeneka***	the house that will not be accessible
Pred.	Pres.	Pos.	*indlu iya****ngeneka***	the house is accessible
	Pres.	Pos.	*indlu* ***ingeneka*** *kakhulu*	the house is very accessible
		Neg.	*indlu ayi****ngeneki***	the house is not accessible
	Past	Pos.	*indlu* ***ingeneki****le*	the house was accessible
	Past	Pos.	*indlu* ***ingeneke*** *kakhulu*	the house was very accessible
		Neg.	*indlu ayi****ngeneka****nga*	the house was not accessible
	Fut.	Pos.	*indlu izo****ngeneka***	the house will be accessible
		Neg.	*indlu ayizuku****ngeneka***	the house will not be accessible

Table 1: Examples of Zulu qualificatives derived from the Zulu written form *ngenekayo* ('accessible')

Use	Tense	Pol.	Zulu	English
Attr.	Pres.	Pos.	*umuntu o****ngaboni***	the person who is blind
		Neg.	*umuntu o****bona****(yo)*	the person who is not blind
	Past	Pos.	*umuntu o****ngabona****nga*	the person who was blind
		Neg.	*umuntu o****boni****le(yo)*	the person who was not blind
	Fut.	Pos.	*umuntu o****nga****zuku****bona***	the person who will be blind
		Neg.	*umuntu ozo****bona***	the person who will not be blind
Pred.	Pres.	Pos.	*umuntu aka****boni***	the person is blind
		Neg.	*umuntu uya****bona***	the person is not blind
	Past	Pos.	*umuntu aka****bona****nga*	the person was blind
		Neg.	*umuntu u****bon****(il)e*	the person was not blind
	Fut.	Pos.	*umuntu akazuku****bona***	the person will be blind
		Neg.	*umuntu uzo****bona***	the person will not be blind

Table 2: Examples of Zulu qualificatives derived from the Zulu written form *ngaboni* ('blind')

clauses. In the former case, the agreement is fixed by the subject noun phrase, while in the latter case, it is fixed by the noun phrase being modified.

In the ZRG, the `VP` linearisation category contains a table with all full forms of the predicate as it appears in the main clause and the relative clause, for every tense, polarity and agreement value, and also, if applicable, distinguishing between a long and a short form. Figure 1 shows a snippet of the code defining the `VP` linearisation category, along with the parameters that define the dimensions of this table. For example, in the `VP` for the verb `bon_V` ('to see'), we can obtain the indicative, present, positive by selecting the values `MainCl`, `Third C1_2 Sg`, `Pos`, `PresTense` and `True`, which will yield the form *uyabona* ('sees'). Implementing a function that takes a `VP` as argument therefore involves making the appropriate selections based on the context in which the `VP` is used.

6 Extending the Usage of ZWN Adjectives

The ZWN is under active development and a second release is expected soon. For this publication, our work was based on preliminary data acquired from the developers ahead of the new release. Due to the status of development at the time, the data dump did not include links to the senses of PWN 3.1 or the Zulu usage examples. However, inspection of the data showed that a significant number of adjective entries had remained essentially in tact from the first version (1338 out of 1590). We therefore decided to focus on the adjective entries in the preliminary data of the new release that also appeared in the first release. In this way, we could ensure that the ZWN written forms in our dataset were as current as possible, while their English senses could be obtained via the first release's links to PWN 2.0.

Our contribution is three-fold: we recover implicit morphosyntactic constructions from the written forms by parsing them using the ZRG; we provide functionality to generate and analyse full forms of these constructions; we do this via a mostly automatic process, which can be reused for future versions of the ZWN, and for the other languages in the AfWN once resource grammars for these languages are available.

The notion of using a fully fledged syntax parser for parsing mostly single token written forms of a wordnet seems incongruous at first glance. However, as an agglutinating language with a conjunc-

Use	Tense	Pol.	Zulu	English
Pred.	Pres.	Pos.	*impendulo* ***ingamanga***	the answer is a lie
		Neg.	*impendulo ayi****ngamanga***	the answer is not a lie
	Past	Pos.	*impendulo ibi****ngamanga***	the answer was a lie
		Neg.	*impendulo ibinge****ngamanga***	the answer was not a lie
	Fut.	Pos.	*impendulo izoba* ***ngamanga***	the answer will be a lie
		Neg.	*impendulo ayizukuba* ***ngamanga***	the answer will not be a lie
Pred.	Pres.	Pos.	*impendulo i****nokukholwa***	the answer has belief
		Neg.	*impendulo ayina****kukholwa***	the answer does not have belief
	Past	Pos.	*impendulo ibi****nokukholwa***	the answer had belief
		Neg.	*impendulo ibingena****kukholwa***	the answer did not have belief
	Fut.	Pos.	*impendulo izoba* ***nokukholwa***	the answer will have belief
		Neg.	*impendulo ayizukuba na****kukholwa***	the answer will not have belief

Table 3: Examples of Zulu qualificatives derived from the Zulu written forms *ngamanga* ('false') and *nokukholwa* ('believable')

```
param
 CType = MainCl | RelCl ;
 Agr = First Number | Second Number | Third ClassGender Number ;
 Polarity = Pos | Neg ;
 BasicTense = PresTense | FutTense | PastTense | RemFutTense | RemPastTense ;

VP = {
 s :  CType => Agr => Polarity => BasicTense => Bool => Str ;
 ...
}
```

Figure 1: Code snippet of `VP` linearisation category with the field `s` as a table of full form strings

tive orthography, single tokens in Zulu may represent full sentences or clauses. The morphosyntactic discrepancies between English adjectives and Zulu qualificatives, in fact, has resulted in such clauses being included routinely as written forms of lemmas, as discussed in Section 4.2. In order to benefit from the lexical semantic contribution of the ZWN, a sufficiently powerful method for identifying and manipulating the relevant constructions is needed. Our contention is that a syntax parser and lineariser, such as provided by GF, is a minimum requirement for taking full advantage of the ZWN.

6.1 Preparing to Parse

The ZulMorph[10] morphological analyser (Pretorius and Bosch, 2003) was used to perform a first pass through the written forms, since it is the state-of-the-art morphological analyser for Zulu and contains a large lexicon (Bosch, 2020). It was found that 501 of the 1338 written forms contained at least one token that could not be analysed by ZulMorph, and these written forms were consequently not considered. An inspection of the failures showed that the majority of tokens that failed to analyse contained an error, although in a number of cases the absence of the relevant root or stem in the ZulMorph lexicon caused the failure. This left 837 written forms to be parsed.

The lexicon for parsing was also prepared with the help of ZulMorph. The morphological analysis is done per token, and the analyser provides all possible analyses. All these analyses of the tokens in the written forms under consideration were used to identify roots and stems for inclusion in a GF lexicon module. No attempt was made to select the applicable analyses from among the various possibilities – all roots and stems were included, leaving the disambiguation step to the parser.

6.2 Parsing the Written Forms

The focus in this paper is on predicate-based qualificatives, and as shown in Section 4.2, they have typically been captured in the ZWN as incomplete relative constructions. Using the GF runtime, it is possible to restrict parsing to a certain syntax category. Our parsing strategy consisted of making several attempts on each written form, each time with a different category restriction. This included relative clauses, verb phrases, noun phrases, adverbs and locative nouns. We also used a fallback strategy for relative clauses, where if parsing failed on the written form as is, we attempted to parse it again after prefixing a relative agreement morpheme. The GF runtime returns an iterator through which all possible parses can be accessed. Our three-step heuristic for selecting from these

[10]Available at: https://portal.sadilar.org/FiniteState/demo/zulmorph/

a single parse for each written form was to select present tense relative clause parses, then to favour long verb roots where applicable,[11] and finally to revert to the tree with fewest nodes.

This automatic parsing and selection strategy, which admittedly involves quite a bit of guesswork, is an attempt to recover implicit linguistic information from the written forms alone, with no reference to the corresponding usage examples. When these become available in the new release, the accuracy of parsing and selecting will improve due to the additional available context.

As it is, however, of the 837 written forms, we were able to obtain at least one parse for 783. We further excluded written forms for which the selected parse included possibly spurious object agreement morphemes. Consequently, we were able to select a present tense relative clause parse for 628 written forms. Of the remaining written form parses, 104 were direct parses of noun phrases (including those in locative forms), adverbs or locative nouns. We are therefore relatively confident that these written forms do not represent relative constructions and hence fall outside the scope of this work. In total, therefore, our success rate at obtaining a plausible parse for the ZWN "adjective" written forms can be estimated as $(628 + 104)/837 = 0.879$.

6.3 An Adjective Application Grammar

The purpose of the adjective application grammar is to simplify the manipulation of Zulu qualificatives by providing a mapping between English adjective senses and the ZRG functions that define them.

In the GF RGL, a technical distinction is made between a relative clause (`RCl`) and a relative sentence (`RS`): the former is not fixed with regards to tense and polarity, while the latter is. When parsing the written forms, we used the `RS` category in order to capture tense and polarity information. Our selected parses all reflect the present tense, but differing inherent polarity (see the discussion in Section 4.2.2). It is this inherent polarity and the description of the predicate as a `VP` that can be re-used to construct ZRG trees to express all full forms of the ZWN written form.

The 628 parsed written forms together with their linked English senses, constitute 881 unique (English sense, Zulu written form) pairs. Table 4 gives some examples, showing how each pair (columns 1 and 2) gives rise to a function name, a ZRG `VP` and an inherent polarity value (columns 3, 4 and 5). This forms the basis of the adjective grammar.

Table 4 gives some examples of mapping between English adjective senses and the Zulu syntax elements that have been recovered from the ZWN written forms.

The function names in the abstract syntax of the adjective grammar are derived from the English senses, while the linearisation functions of the Zulu concrete syntax make use of the associated `VP` and inherent polarity value. In the code snippet in Figure 2, we show the linearisation category of `ZWN_APred`, as well as an example of a function definition for obtaining a `ZWN_APred`, and its corresponding linearisation function definition.

In the adjective grammar, agreement information is manipulated via the `ZWN_Pron` category, which encapsulates pronouns modeled in the ZRG. This is convenient because Zulu is a pro-drop language, which means that the relative and main clauses can be linearised alongside pro-dropped pronouns (which are linearised as empty strings) to obtain only the qualificative strings.

Figure 3 in Appendix A shows an attributive example of an adjective grammar tree, which effectively constructs a ZRG tree that contains a relative clause, shown in Figure 4. Figures 5 and 6 give the corresponding predicative case, which involves main clause predication. The adjective grammar trees are simple, while exhibiting the full morphosyntactic behaviour of relative and main clauses in Zulu by making use of the ZRG.

6.4 A Grammar-based Tool for Extending Use of the ZWN

Our command line tool shows how the adjective grammar can be used to generate and analyse the adjective-like qualificatives of the ZWN. The tool is presented as an end-user tool, but the core functionality could just as easily be embedded into an NLP pipeline.

By making use of the linearisation functionality in the GF runtime, it allows a user to specify the English adjective, along with the required tense, polarity and agreement information in order to obtain the correct form of the corresponding Zulu qualificative(s). Each of these command line parameters map in a straight forward way to a function in the adjective grammar, which is used to construct

[11] For the purposes of this work, verb root extensions were considered as part of the root.

English sense	ZWN Written form	Function	ZRG VP	Inherent polarity
articulate	cacile	articulate_1_A	UseVStative cac_V	Pos
decided	cacile	decided_1_A	UseVStative cac_V	Pos
accessible	ngenekayo	accessible_1_A	UseV ngenek_V	Pos
accessible	finyelelekayo	accessible_2_A	UseV finyelelek_V	Pos
amazing	mangalisayo	amazing_1_A	UseV mangalis_V	Pos
fabulous	mangalisa	fabulous_1_A	UseV mangalis_V	Pos

Table 4: Examples of mappings between English adjective senses and Zulu qualificative constructions

```
- linearisation category
ZWN_APred = { vp :  VP ; pol :  ZPol } ;

- function
thundering_1_A : ZWN_APred ;

- linearisation function
thundering_1_A = { vp = AdvVP (UseV mangalis_V) kakhulu_Adv ; pol = ZPos } ;
```

Figure 2: Implementing an English adjective as a Zulu qualificative

the correct tree. The tool also allows the use of wild cards, in which case the linearisations for all possible values (and combinations of values) are given. Example output is shown in Figure 7 in Appendix A.

Conversely, the tool allows the user to provide a Zulu qualificative string in order to obtain its corresponding English adjective(s), along with the tense, polarity and agreement information. The input is parsed to obtain a tree, whose nodes contain the required information. Example output is shown in Figure 8 in Appendix A.

In contrast to Angelov (2020), where full forms were included as tables in the wordnet resource, our decision to instead provide a computational tool is based on the sheer number of full forms of the Zulu qualificatives, which could be as many as 384.

7 Evaluation and Discussion

As noted in Section 6.2, when considering those written forms for which a morphological analysis could be found for each token, a plausible parse for 87.9% of written forms could be obtained. This was used the basis for an application grammar and wrapper tool that could generate and analyse full forms of adjectives with different tense, polarity and agreement values, as well as form, whether attributive or predicative.

This is a novel contribution. The Zulu resource grammar along with the GF runtime system is the foundation of this generation and analysis capability. While some work has been done to develop GF resource grammars for other African languages (Ng'ang'a, 2012; Kituku et al., 2021), these grammars have not yet been demonstrated to support the kind of application grammar development presented here. As such, there is no baseline to compare our work to, which presents a challenge for evaluation.

The only comparable computational tool, in terms of accuracy and scope, is the ZulMorph morphological analyser, which is an FST that can be applied to surface forms in order to obtain full morphological analyses and vice versa. While it cannot disambiguate analyses for multitoken expressions and hence is not suitable for the generation and analysis task presented here, it can be utilised in the evaluation of the output of the Zulu resource grammar, as mediated by the adjective application grammar.

7.1 Evaluation

The following methodology was implemented to evaluate our system:

1. From the mappings (see examples in Table 4), randomly select 50 entries in order to obtain ZRG VPs.
2. For each ZRG VP, randomly select a value for tense, polarity, agreement, form and length (long form or short form), and construct the relevant full abstract syntax tree as done in the application grammar.
3. For each tree, linearise the tree into a string, and obtain ZulMorph analyses for each token in the string.
4. Using the information given by the abstract syntax tree and the (possibly multiple) morphological analyses per token, attempt a selection of ulMorph analyses that correspond to the abstract syntax tree.

Qualificative property	% of Total
Verb	61.5%
`UseV`, `UseVStative`, `ComplV2`	
Associative copulative	18.6%
`CopNPAssoc`	
Locative copulative	2.2%
`CopLocative`	
Adjective/primitive relative	2.0%
`CopAP`	
Identifying copulative	1.6%
`CopNP`	
Negative	9.3%
`PNeg`	

Table 5: Properties of adjective-like qualificative constructions in the "adjective" entries of the ZWN

5. If and only if such a selection is possible for all tokens in the generated form, generation is accepted as correct.

Figure 9 in Appendix A shows a snippet of the simple web based tool that was used to visualise and select ZulMorph analyses given a ZRG abstract syntax tree. It was found that ZulMorph lacked sufficient coverage of, for example, contracted past tense forms. Hence, in 9 out of the 50 entries, the tokens generated by the ZRG could not be analysed. These were analysed manually to confirm their correctness. Out of the 50 entries selected, only 2 were determined not to have been generated correctly. This was, however, not due to errors made by the ZRG, but due to incorrect parses obtained for the ZWN written forms initially. Consequently, we estimate that our tool has an accuracy of $48/50 = 0.96$.

7.2 Discussion

This high degree of accuracy allows us to make a few quantitative observations about adjective-like qualificatives in Zulu, especially with regards to relative constructions. Such an analysis is, to our knowledge, in itself a novel contribution.

Of the adjective-like Zulu qualificatives for which a parse could be selected, 628 (85.8%) represent relative, or predicate-based, constructions, while the remaining 104 (14.2%) represent descriptive possessives or adverbs. This confirms the importance of being able to process the predicate-based qualificatives effectively.

The summary in Table 5 shows the representation of certain properties of predicate-based qualificatives in the data. The properties correspond to functions in the ZRG parses, used to arrive at the percentages. We see, for example, that verbal constructions constitute a large majority of adjective-like qualificatives, namely 61.5%. The second largest group are the associative copulatives (18.6%), while constructions based on adjective and primitive relative stems make up only 2.0% of the total. We also see that 9.3% of adjective-like qualificatives inherently exhibit negative polarity.

8 Conclusion

The analysis in the previous section illustrates the morphosyntactic diversity of the adjective-like Zulu qualificatives. We have shown how a computational grammar-based approach can overcome the challenge this poses in order to take full advantage of the ZWN by facilitating its potential use in NLP applications for Zulu.

The process we have developed could be repeated whenever new versions of the ZWN are released. Moreover, as shown in the previous section, the adjective-like qualificatives in the ZWN typically represent constructions based on verbs and nouns. Future work will include developing functionality to similarly generate and analyse full forms of verb and noun entries of the ZWN, as well as replicating the work for other languages in the AfWN once resource grammars for them are developed.

Acknowledgements

This work has been funded by the South African Centre for Digital Language Resources (SADiLaR).

References

Krasimir Angelov. 2020. A parallel WordNet for English, Swedish and Bulgarian. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3008–3015, Marseille, France. European Language Resources Association.

Sonja Bosch. 2020. Computational morphology systems for Zulu–a comparison. *Nordic Journal of African Studies*, 29(3):28–28.

Marissa Griesel and Sonja Bosch. 2020. Navigating challenges of multilingual resource development for under-resourced languages: The case of the African Wordnet project. In *Proceedings of the first workshop on Resources for African Indigenous Languages*, pages 45–50, Marseille, France. European Language Resources Association (ELRA).

Benson Kituku, Wanjiku Nganga, and Lawrence Muchemi. 2021. Grammar engineering for the ekegusii language in grammatical framework.

Mampaka Lydia Mojapelo. 2014. Morphosyntactic discrepancies in representing the adjective equivalent in African WordNet with reference to Northern Sotho. In *Proceedings of the Seventh Global Wordnet Conference*, pages 355–362, Tartu, Estonia. University of Tartu Press.

Wanjiku Ng'ang'a. 2012. Building swahili resource grammars for the grammatical framework. In *Shall We Play the Festschrift Game?*, pages 215–226. Springer.

George Poulos and Christian T. Msimang. 1998. *A Linguistic Analysis of isiZulu*. Via Afrika, Cape Town, South Africa.

Laurette Pretorius and Sonja E Bosch. 2003. Finite-state computational morphology: An analyzer prototype for Zulu. *Machine Translation*, 18(3):195–216.

Aarne Ranta. 2009. The GF Resource Grammar Library. *Linguistic Issues in Language Technology*, 2.

Aarne Ranta. 2011. *Grammatical framework: Programming with multilingual grammars*, volume 173. CSLI Publications, Center for the Study of Language and Information Stanford.

Aarne Ranta, Krasimir Angelov, Normunds Gruzitis, and Prasanth Kolachina. 2020. Abstract Syntax as Interlingua: Scaling Up the Grammatical Framework from Controlled Languages to Robust Pipelines. *Computational Linguistics*, 46(2):425–486.

P. C. Taljaard and S. E. Bosch. 1988. *Handbook of IsiZulu*. J. L. Van Schaik, Pretoria, South Africa.

A Additional Examples

A.1 Obtaining Full Forms via an Adjective Grammar

Figures 3 and 5 show application grammar trees for expressing 'thundering' in a specified syntactic context. Figures 4 and 6 show how the same full forms are represented as trees in the ZRG.

A.2 Generating Full Forms via the Command Tool

Figure 7 shows an example of output from the command line tool. This was obtained by the following request: `python3 afwn_adjectives.py generate ? Pos 2 ? blind`

Figure 8 shows an example of output from the command line tool. This was obtained by the following request: `python3 afwn_adjectives.py analyze awubonanga`

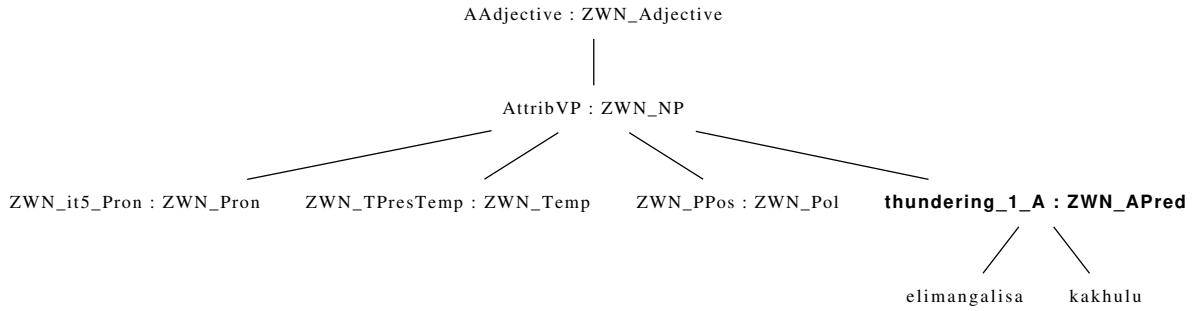

Figure 3: The adjective grammar tree for expressing 'thundering' in the attributive form in the present, positive and modifying the pronoun 'it' of class 5

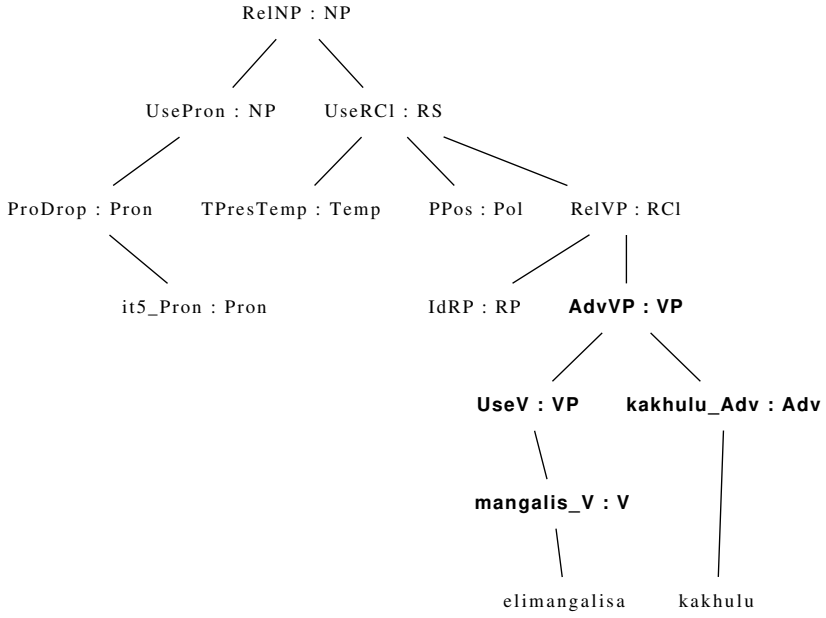

Figure 4: The resource grammar tree for expressing the concept of 'thundering' attributively as a present, positive relative clause modifying the pro-dropped pronoun 'it' of class 5.

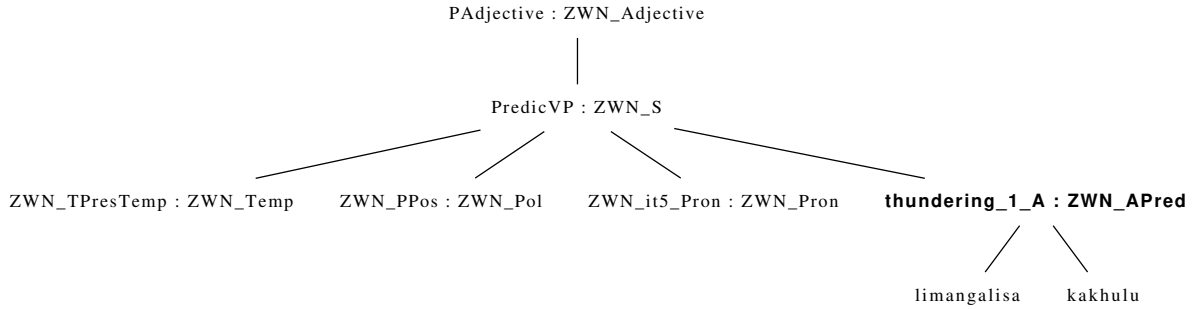

Figure 5: The adjective grammar tree for expressing 'thundering' in the predicative form in the present, positive with the pronoun 'it' of class 5 as subject

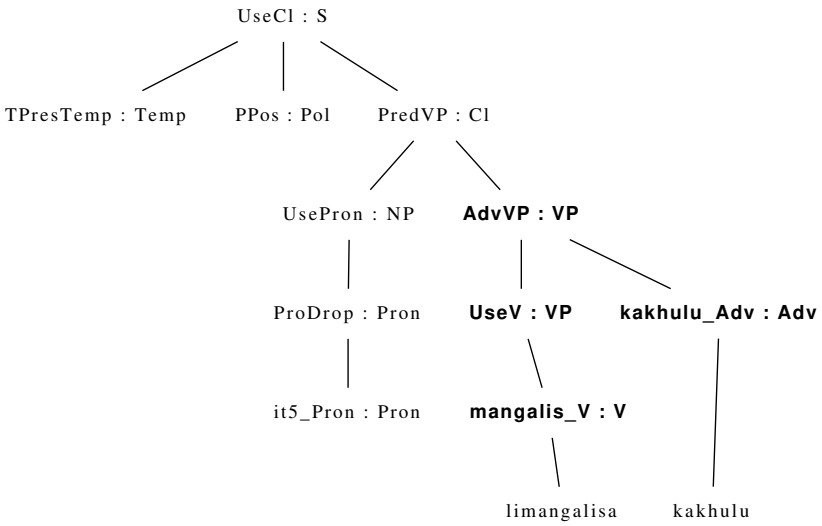

Figure 6: The resource grammar tree for expressing the concept of 'thundering' predicatively as a present, positive clause with pro-dropped pronoun 'it' of class 5 as subject.

Tense	Polarity	Class	Form	Long/short	Qualificative
Fut	Pos	2	Attr		abangazukubona
Fut	Pos	2	Pred		abazukubona
Past	Pos	2	Attr		abangabonanga
Past	Pos	2	Pred		ababonanga
Pres	Pos	2	Attr		abangaboni
Pres	Pos	2	Pred		ababoni
RemFut	Pos	2	Attr		abangayukubona
RemFut	Pos	2	Pred		abayukubona
RemPast	Pos	2	Attr		abangabonanga
RemPast	Pos	2	Pred		ababonanga

Figure 7: Output of the command line tool when requesting all positive forms of 'blind' when modifying a plural noun of class 2, such as *abafundi* ('pupils').

Tense	Polarity	Class	Form	Long/short	Adjective
Past	Neg	3	Pred		conscious
Past	Pos	3	Pred		blind
RemPast	Neg	3	Pred		conscious
RemPast	Pos	3	Pred		blind

Figure 8: Output of the command line tool when requesting an analysis of *awubonanga*

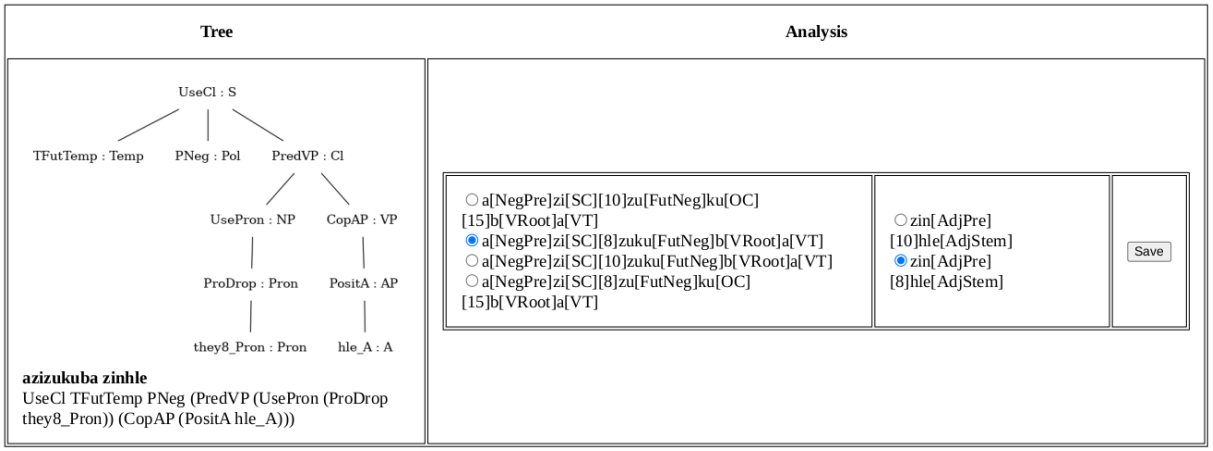

Figure 9: Example of selecting ZulMorph analyses to correspond with the ZRG abstract syntax tree