1. Contents

[Introduction 1](#_Toc301173553)

[Symbols and Definitions 1](#_Toc301173554)

[Xhosa 4](#_Toc301173555)

[Morphology and Allomorphy 4](#_Toc301173556)

[Basic Word Formation 5](#_Toc301173557)

[Noun Classes and Cases 5](#_Toc301173558)

[Nominative and Objective Case 5](#_Toc301173559)

[Vocative Case 6](#_Toc301173560)

[Possessive Case 6](#_Toc301173561)

[Locative Case 7](#_Toc301173562)

[Implementation 7](#_Toc301173563)

[Specific Adaptations 8](#_Toc301173564)

[Future Goals 8](#_Toc301173565)

# Introduction

This report is designed to document the creation and features of the Xhosa morphological analyzer created in the summer of 2011 at Carnegie Mellon University. This morphological analyzer was created in order to be able to tag the grammatical and content morphemes of the Xhosa language. The analyzer was created by use of the Xerox Finite State Toolkit (xfst), and runs by means of Java. The main features of the morphological analyzer will be explained in detail below, but will be outlined here. The final product can take in a file filled with Xhosa words and return an output file with #% accuracy. The output file includes a morpheme-by-morpheme tag line, as well as a root recognition tag. If the root is not contained in the established lexicon, the analyzer guesses at a probable root with #% accuracy. When using a test suite, the analyzer can add a root to the lexicon file, in order to more accurately predict words in the future. The analyzer also contains the framework for a simple dependency prediction algorithm, in order to correctly predict morpheme structure.

# Symbols and Definitions

|  |  |
| --- | --- |
| **Symbols Used in Morphological Analyzer** | |
| **1** | Class I Singular |
| **1P** | Class I Subclass, Singular |
| **1st** | First Person  First Derived Verb |
| **2** | Class I Plural |
| **2nd** | Second Person  Second Derived Verb |
| **2P** | Class I Subclass, Plural |
| **3** | Class II Singular |
| **4** | Class II Plural |
| **5** | Class III Singular |
| **6** | Class III Plural |
| **7** | Class IV Singular |
| **8** | Class IV Plural |
| **9** | Class V Singular |
| **10** | Class V Plural  Class VI Plural |
| **11** | Class VI Singular |
| **13** | Class VII Singular |
| **14** | Class VIII Singular |
| **Adj** | Adjective |
| **Adv** | Adverb |
| **AdvN** | Adverb formed from a noun |
| **Aug** | Augmentive suffix |
| **Aux** | Auxiliary particle |
| **Card** | Cardinal Number |
| **Caus** | Causative form  Causal Conjunction |
| **Cond** | Conditional Conjunction |
| **Conj** | Conjunction |
| **Coord** | Coordinative Conjunction |
| **Cop** | Copula Form of Personal Pronoun |
| **Dim** | Diminutive Suffix |
| **Fem** | Feminine Suffix |
| **Fin** | Final Conjunction |
| **Fut** | Future Tense |
| **Gen** | Gendered Word |
| **Imp** | Imperfect Tense |
| **Impf** | Imperfect Tense |
| **Ind** | Indicative Form |
| **Inf** | Infinitive Form |
| **Int** | Intensifier |
| **Irr** | Irregular |
| **Loc** | Locative |
| **Long** | Long Form |
| **Masc** | Masculine Word |
| **Mono** | Monosyllabic Root |
| **Nadj** | Adjective formed of a Noun |
| **Neg** | Negative Particle |
| **Neu** | Neutral Gender |
| **Nom** | Nominative Case of Nouns |
| **Noun** | Noun |
| **Num** | Number |
| **Obj** | Objective Case of Noun  Objective Personal Pronoun |
| **Ord** | Ordinal Number |
| **Part** | Participle |
| **Pass** | Passive Tense |
| **Past** | Past Tense |
| **Perf** | Perfect Tense |
| **Plu** | Plural |
| **Plupf** | Pluperfect Tense |
| **Poss** | Possessive Case |
| **PP** | Personal Pronoun |
| **PPrep** | Post-Prepositional Pronoun |
| **Prep** | Preposition |
| **Pres** | Present Tense |
| **Quest** | Interrogative |
| **Recip** | Reciprocal Form |
| **Ref** | Reflexive Pronoun |
| **Rel** | Relative Particle  Relative Clause |
| **Short** | Short Form |
| **Sing** | Singular |
| **Stat** | Stative Form |
| **Sub** | Subjective Pronoun |
| **Subj** | Subjunctive Form |
| **Subord** | Subordinate Conjunction |
| **Subs** | Substantive Pronoun |
| **T1** | Type I of Adjectives |
| **T2** | Type II of Adjectives |
| **Temp** | Temporal |
| **Vadj** | Adjective made of a Verb |
| **Verb** | Verb |
| **Voc** | Vocative Case |
| **Vow** | Vowel Verb |

# Xhosa

Xhosa is a Bantu language of the Nguni branch. It is an official language of South Africa, and spoken by approximately 7.9 million people. It is the dominant language in the Eastern Cape Province, and the most widespread language in South Africa, although not the most spoken. It is closely related to the other Nguni languages of Zulu, Swati, and Ndebele, to the point of mutual intelligibility. Nguni languages are members of the Southeastern Bantu language group. Xhosa is the most s

Xhosa is a tonal language, which is not recognizable in textual representation.

# Morphology and Allomorphy

Xhosa, like most Bantu languages, is highly agglutinative. Nouns, adjectives, verbs, and adverbs, are all modified by morphemes to form sentences. There are two types of morphemes: grammatical and content morphemes. Grammatical morphemes are a closed class (meaning the morpheme list is concrete), and express grammatical expressions such as tense or mood. Content morphemes are an open class (meaning new morphemes can be added) and indicate the actual meaning of the word. In Bantu languages, grammatical morphemes account for at least fifty percent of the morphemes in any given word, and usually are closer to 75% of the morphemes in any given word. Content morphemes only appear once per word, as the root of the word. Grammatical morphemes are infrequently more than one syllable and almost never more than two syllables[[1]](#footnote-1). Content morphemes, on the other hand, are nearly always more than one syllable and can have greater than five syllables, depending on the spcific language.

## Basic Word Formation

The Xhosa alphabet contains 26 letters and uses a romanized script. The vowels in Xhosa are **a, e, i, o**  and **u.** Vowels can generally not be combined, but all vowels have a “long” pronounciation, which may be written as two of the same vowel in a row. Consonants are the letters **b, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, y**  and **z**. These consonants can be blended or combined to form new sounds. There are # different ways to combine consonants: Nasalization, … Nasalization of a consonant consists of adding an **m** or **n** before the consonant, which can be applied . In Xhosa, there are also three click sounds, as taken from the Hottentot and Bushman languages. These clicks are represented by **c, q,** and **x**. Clicks, like consonants, can be used in combinations and blends. Clicks can all be nasalized or aspirated, the former by adding an **n** before the click letter, the latter by adding an **h** after the click. Words are formed by phonemes that, at their simplest form, consist of a single consonant and a single vowel. Consonants blends and clicks may also be used to start a phoneme.

## Noun Classes and Cases

One of the unique aspects of Bantu languages is the use of a noun class system. Every noun is assigned a class, which then determines the formation of all affixes on that noun and other words in the sentence. A noun in Xhosa cannot stand on its own: even the nominative form has a class-specific prefix. Xhosa has 8 classes of nouns, and one subclass. Each class and subclass has two separate sets of prefixes for singular and plural forms, except for classes VII and VIII. By standard convention, the singular and plural form are numbered consecutively in class order, so that every class contains a singular and plural form labeled as two consecutive integers. (Ex. Class I contains the prefixes 1 (singular) and 2 (plural)). These numbers sometimes represent the class itself, seperated by a back slash.

The first class of nouns in Xhosa is the only class to also contain a subclass. The first class represents animate objects; its subclass represents proper nouns, as well as nouns representing relationships and some humanized words, such as **u-xam**, monitor lizard. The other classes of the Xhosa language generally have no unifying theme to the subject of the words, except for Class 13, which holds only abstract nouns. Many of the roots in Class 13 appear in other classes, such as **ntu**, man, which occurs as **umntu/abantu**, a man/men (Class I) and **ubuntu**, manhood (Class 13).

## Nominative and Objective Case

The nominative form of a Xhosa noun is made of a class specific prefix and a root. It simply represents the noun as a subject of a sentence or clause. It is identical to the objective case, where the noun in question is the object of a clause or sentence. The class prefixes are subject to several modifications, based on the structure of the root. The basic forms are as follows.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Subjective and Objective Prefixes*** | | | | | | | | | |
| **Class** | **I** | **I(P)** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Sing.** | um | u | um | ili | isi | in | ulu | ubu | uku |
| **Plu.** | aba | oo | imi | ama | izi | izin | izin | ------ | ------ |

The prefixes listed above actually consist of an article and a prefix proper. The article is considered to be the first vowel of the prefix. The “prefix proper” is the remaining letters. The final **‘m’** in classes I(a) and II(a) are actually truncated forms of the Zulu syllable **“mu”**, which is considered the prefix proper for those listed classes.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Articles*** | | | | | | | | | |
| **Class** | **I** | **I(P)** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Sing.** | u | u | u | i | i | i | u | u | u |
| **Plu.** | a | oo (a) | i | a | i | i | i | ------ | ------ |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Prefix Propers*** | | | | | | | | | |
| **Class** | **I** | **I(P)** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Sing.** | mu | ------ | mu | li | si | n | lu | bu | ku |
| **Plu.** | ba | ------ | mi | ma | zi | zin | zin | ------ | ------ |

The nominative and objective cases are the most frequently seen of all the noun cases. They can be modified to be augmentive or diminutive by the addition of a specific suffix, which leaves the class specific prefix unmodified. The prefixes for nominative case also appear in prepositional, possessive, vocative, and locative nouns, but in modified forms.

## Vocative Case

The vocative case expresses the act of directly addressing the root. It is formed by dropping the article of the nominative prefixes, and only using the prefix proper. In contracted forms (such as **iin** for **izin**), the uncontracted form is used, and the article removed from the uncontracted form. Vocative forms can also be followed by the suffix **ndini**, which further emphasizes the vocative nature of the word. Some examples follow: **mntu,** man! **bawondini**, father!

## Possessive Case

Possessive case takes two separate forms. If the noun in question is a member of the first subclass (i.e. a proper noun), it takes before it the preposition **ka**, with some modifications made. The first subclass of nouns do not retain their nominative prefix. All other classes take a class specific posessive prefix, which combines with the nominative prefix. The prefixes are class specific to the *object* being possessed, but are attached to the beginning of theword representing the *owner*. This means that possessive prefix and nominative prefix ***do not*** have to be of the same class. The **a** of the possessive prefix combines with the first letter of the nominative prefix: **aa** becomes **a**, **ai** becomes **e**, and **au** becomes **o**.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Possessive Prefixes*** | | | | | | | | | |
| **Class** | **I** | **IP** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | wa | ka\* | wa | la | sa | ya | lwa | ba | kwa |
| **Plural** | ba | ka\* | ya | a | za | za | za |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***\*Possessive Prefixes for Proper Nouns*** | | | | | | | | |
| **Class** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | ka | ka | lika | sika | ka | luka | buka | kuka |
| **Plural** | baka | ka | ka | zika | zika | zika |  |  |
|  |  |  |  |  |  |  |  |  |

**Examples:**

**ihashe womntu,** the man’s horse

**ihashe likaPhatho**, Pato’s horse

## Locative Case

Locative case indicates that the noun is actually the location of the action. It consists of a prefix modification and a suffix, for all classes. The nominative prefix of the noun is maintained, but the article of the prefix is changed to **e**. Then, the suffix **ni** is added to the end of the word. The last vowel sometimes undergoes a change: **a** becomes **i, o**  becomes **wo**, and **u** becomes **wi**. **E** and **i** remain the same.

**Gendered Nouns**

Very few nouns in Xhosa have true genders. The exeptions are mostly limited to the expression of male and female animals. Feminine nouns can be marked as such with the suffix **+kazi**, which has no masculine counterpart. In other situations, the masculine and feminine forms are entirely different names, such as **intombi**, a girl and **indoda**, a man. Other masculine and feminine words are formed by using a gendered word before a neutral noun. For example, **inkunzi yenkomo**, a bull and **imazi yenkomo**, a cow.

**Augmentive and Diminutive Nouns**

Augmentive and diminutive forms of nouns are formed by suffixing on an appropriate suffix. To augment a noun, the suffix **+kazi** is added to the end of the word, which drops its feminine meaning. No changes are made to the root of the noun, like in **umthikazi**, a large tree. To make a noun diminutive, the suffix **+ana** or **+anyana** is added. The final vowel of the root undergoes changes to avoid consecutive vowels: **a, i** and **e** are omitted and **o** and **u** are converted to **w**. For example, **into**, a thing, becomes **intwana**, a small thing. Labial consonants that come before **+ana** are sometimes changed to their palatal equivalents, especially when the following vowel is **o** or **u**: **b** becomes **ty** or **j, p** becomes **tsh**, **m** becomes **ny**, **mp** becomes **ntsh,** and **mb** becomes **nj**. **L** also sometimes becomes **dl**. Augmentive and diminutive suffixes can also be applied to adjectives, where **+kazi** means “very” and **+ana** means “fairly”.

**Prepositional Articles**

Nouns can take a variety of prefixes with prepositional meanings. Many of these combine with the article of the noun in the same way as the possessive prefixes. These prepositional prefixes are followed by prefix proper and noun root. There are five different prepositional particles that can be used: **ka, kwa, ku, na,** and **nga. Ka** has already been covered, as the possessive form of proper nouns.

***Kwa***

**Kwa** is used before proper nouns (Class IP). It expresses the idea of “at the place of” or “in the tribe of”. **Kwa** combines like **ka** to the word it modifies, displacing the article. Some examples follow: **kwaZulu**, in Zululand and **kwaMphanda**, at Mpanda’s place.

***Ku***

**Ku** is used in the sense of “to”, “at”, or “from”. **Ku** is sometimes used in place of the locative form. In Class IP nonuns, it replaces the nominative prefix **u**. If **ku** is used before any other word, it combines with the article: if the article is **a**, it is displaced, **e** and **i** change to **w** and replace the **u** in **ku**, and **o** and **u** replace the **u** in **ku**. Examples include **kuyihlo**, to or from your father and **kwinkosi**, to or from the chief.

***Na***

**Na** means “along with” or “together with”. It also expresses the idea of possession expressed in English by “have”. It contracts with the article like **ka**. For example, **namadoda**, with the men, or **nesicaka**, with the servant.

***Nga***

**Nga** designates the idea of “by means of” or “through”. It also expresses the general ideas of “about” or “concerning”. It combines in the same way as **ka** to the noun it modifies, and takes the form **ngas** before locative form. It is often used in combination with locative case, and follows the adverbs **nge** and **nje.** It can also be followed by **kwa** or **nxa**. Examples include **(ndambetha)** **ndentonga**, (I struck him) with a stick, or **ngaphambili**, in front.

**Personal Pronouns**

Personal pronouns play a large role in Xhosa, because they serve as concords to conect the sentence. There are six different forms that personal pronouns can appear in: subjective, objective, substantive, copulaic, post-prepositional, and possessive

**Subjective Pronouns**

Subjective pronouns serve as the main pronominal concord in Xhosa. They consist of 18 different prefixes, one for each class of noun, as well as one for first and second person, singular and plural. The subjective pronouns are used before any type of predicate: verbs, adjectives, adverbs, and prepositional phrases. The subjective pronoun is placed at the front of the predicate, before all other prefixes except the negative particle **a**. When the pronoun appears before a vowel, the consonant form of the pronoun (shown in parentheses) is used instead, to prevent consecutive vowels. Some illustrations follow: **ndithetha**, I am speaking, or **silapha**, we (are) here.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Subjective Pronouns*** | | | | | | | | | | |
|  | **1st** | **2nd** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | ndi | u | u, a ,e, ka | u | li | si | i | lu | bu | ku |
|  | *(nd)* | *(w)* | *(w)* | *(w)* | *(l)* | *(s)* | *(y)* | *(lw)* | *(b)* | *(kw)* |
| **Plural** | si | ni | ba | i | a, e | zi | zi | zi |  |  |
|  | *(s)* | *(n)* | *(b)* | *(y)* | *()* | *(z)* | *(z)* | *(z)* |  |  |

**Objective Pronouns**

Objective pronouns serve as pronominal concords as well, but for the object rather than the subject of the sentence. It appears immediately before the verb stem. It can be the pronoun for either direct or indirect object. Some illustrations are **uya ndibona**, he sees me, or **ndiya lifuna ihashe**, I want the horse.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Objective Pronouns*** | | | | | | | | | | |
|  | **1st** | **2nd** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | ndi | ku | m | wu | li | si | yi | lu | bu | ku |
|  | *(nd)* | *(kw, k)* | *(m)* | *(w)* | *(l)* | *(s)* | *(y)* | *(lw, l)* | *(b)* | *(kw, k)* |
| **Plural** | si | ni | ba | yi | wa | zi | zi | zi |  |  |
|  | *(s)* | *(n)* | *(b)* | *(y)* | *(w)* | *(z)* | *(z)* | *(z)* |  |  |

**Copula**

The copulaic pronouns are used to express the ideas of “it is” or “they are”. Copulas are not required before adjectives or adverbs, but they are required before nouns and pronouns. Examples follow: **ndim**, it is I, and **nguwe**, it is you.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Copulaic Pronouns*** | | | | | | | | | | |
|  | **1st** | **2nd** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | ndi | ngu | ngu | ngu | li | si | yi | lu | bu | ku |
|  | *()* | *()* | *(ng)* | *(ng)* | *(l)* | *(s)* | *(y)* | *(l)* | *(b)* | *(k)* |
| **Plural** | si | ni | nga | yi | nga | zi | zi | zi |  |  |
|  | *()* | *()* | *(ng)* | *(y)* | *(ng)* | *(z)* | *(z)* | *(z)* |  |  |

**Substantive Pronouns**

Substantive pronouns are used on their own to represent nouns. It takes a fuller form than any other pronouns. It is usually formed by adding the ending **+ona** to the end of objective pronouns. Substantive forms are used for emphasis on the noun against the verb.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Substantive Pronouns*** | | | | | | | | | | |
|  | **1st** | **2nd** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | mna | wena | yena | wona | lona | sona | yona | lona | bona | kona |
| **Plural** | thina | nina | bona | yona | wona | zona | zona | zona |  |  |

**Post-Prepositional Pronouns**

After prepositional particles, the substantive form of the pronoun is used, but usually in contracted form. Contracted form of the substantive form is formed by dropping the final **na**. This form is also used after the pronominal copula. Some examples: **Ndiza kuwe**, I come to you, and **sabethwa nini**, we were beaten by you.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Post-Prepositional Pronouns*** | | | | | | | | | | |
|  | **1st** | **2nd** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | m | we | ye | wo | lo | so | yo | lo | bo | kho |
| **Plural** | thi | ni | bo | yo | wo | zo | zo | zo |  |  |

**Possessive Pronouns**

For the most part, the pronouns used after the possessive particles are the same as the post-prepositional form. They differ in the second person. first person plural, and Class I singular. Like possessive form in nouns, the possesssive prefix agrees with the object possessed, and the pronoun agrees with the possessor. Examples: **mna neempahla zam,** I and my belongings, or **ukumkani nabantu bakhe**, the king and his people.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Possessive Pronouns*** | | | | | | | | | | |
|  | **1st** | **2nd** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** |
| **Singular** | m | kho | khe | wo | lo | so | yo | lo | bo | kho |
| **Plural** | thu | nu | bo | yo | wo | zo | zo | zo |  |  |

**Adjectives**

Like in English, adjectives in Xhosa can be used as an attribute or a predicate. Unlike English, the placement of the adjective does not matter for the distinction. Instead, the diffference between the two is demarcated by the prefix of the adjective.

There are two types of adjectives: simple and relative adjectives. Simple adjectives express the ideas of age, size, number, and quantity. They are a closed class. They have a separate set of prefix propers derived from noun prefixes. Relative adjectives convey the ideas of color, taste, and similar qualities. Relative adjectives usually originated from nouns or verbs. They have the prefixes of pronominal concords and relative articles.

There are relatively few adjectives in Xhosa, so both verbs and nouns can be adjectives. Many nouns in Xhosa are descriptive, which can serve as attributes to other nouns by means of the relative particle and copula. Abstract nouns can also serve as adjectives. They can either use the relative particle and copula or the relative concord and the prepositional particle **na**. Verbs can also serve as adjectives, if they are intransitive. They take the relative particle before and the relative clause marker **yo** at the end of the word if used in an attributive sense. As a predicate, the verb doesn’t change. Stative perfect tense is often used as an adjective. Sometimes, present tense is used to express idea that are expressed in English with adjectives, such as **ndiya godola**, I shiver/ I am cold.

**Simple Adjectives**

Simple adjectives take the same prefix as the noun they modify, after dropping the article, except Class V. When these adjectives are modifying first or second person pronouns, they take the same prefixes they do for Class I. These prefixes are demonstrated below with the adjective root **de**, long or tall.

|  |  |  |
| --- | --- | --- |
| ***Adjective “de”*** | | |
|  | **Singular** | **Plural** |
| **1st** | **ndimde** (mna) | **sibade** (thina) |
| **2nd** | **umde** (wena) | **nibade** (nina) |
| **I** | umfo **mde** | abafo **bade** |
| **II** | umthi **mde** | imithi **mide** |
| **III** | iliza **lide** | amaza **made** |
| **IV** | isilo **side** | izilo **zide** |
| **V** | imini **inde** | iimini **zinde** |
| **VI** | uluhlu **lude** | izintlu **zinde** |
| **VII** | uboya **bude** |  |
| **VIII** | ukuthetha **kude** |  |

When used as a predicate for nouns without articles, the prefixes for simple adjectives undergo a tone shift which is not represented in XhosMorph. If the noun being modified does have an article, a relative particle (**a, e,** or **i**) is used before the prefix used in attributive adjectives.

|  |  |  |
| --- | --- | --- |
| ***Adjective “shushu”*** | | |
|  | **Singular** | **Plural** |
| **I** | umfo ushushu | abafo bashushu |
| **II** | umzi ushushu | imizi ishushu |
| **III** | iliso lishushu | amehlo ashushu |
| **IV** | isilo sishushu | izilo zishushu |
| **V** | imbiza ishushu | iimbiza zishushu |
| **VI** | uluvo lushushu | izimvo zishushu |
| **VII** | ubuso bushushu |  |
| **VIII** | ukutya kushushu |  |

**Relative Adjectives**

In both predicate (modifying a noun without an article) and attributive usage, relative adjectives take the pronominal subject of the noun they modify. When they are modifying nouns with an article as a predicate, they are prefixed by the relative article, followed by the pronominal subject.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Relative Adjectives*** | | | | |
|  | **Singular** | **Plural** | **Singular** | **Plural** |
| **I** | umfo ushushu | abafo bashushu | umntu obomvu | abantu ababomvu |
| **II** | umzi ushushu | imizi ishushu | umthi obomvu | imithi ebomvu |
| **III** | iliso lishushu | amehlo ashushu | ihashe elibomvu | amahashe abomvu |
| **IV** | isilo sishushu | izilo zishushu | isitya esibomvu | izitya ezibomvu |
| **V** | imbiza ishushu | iimbiza zishushu | into ebomvu | izinto ezibomvu |
| **VI** | uluvo lushushu | izimvo zishushu | uluthi olubomvu | izinti ezibomvu |
| **VII** | ubuso bushushu |  | ubuso obubomvu |  |
| **VIII** | ukutya kushushu |  | ukutya okubomvu |  |

**Verbs**

Verbs in Xhosa contain the most number of affixes on average. Suffixes and prefixes are connected to verb roots, with up to # affixes. In a dictionary, verbal roots are shown in infinitive form, which consusts of the infinitive prefix **uku** before the root, ending in the present tense **a**. The stem of the verb consists of everything that follows the infinitive prefix, whereas the root of the verb is the stem without its final vowel. For example, the word **ukuhamba**, to go, has the stem **hamba** and the root **hamb**. XhosMorph isolates the root of words, from which the sem can be formed by adding a final **a**.

There are a few exceptions where the verbal root does not end in **a**, but rather in **i**  or **o**. There are only three verbs that famm under these exceptions: **ukuthi**, to say or do, **ukwazi**, to know, and **ukutsho**, to say so.

Xhosa has six different moods and ten different tenses. The different moods of the verbs express direct assertion (in indicative tense), a state of being (participial), cause and effect (subjunctive), a separate timeline (temporal), a command (imperative), and the nominalizaion of the verb (infinitive). These moods are formed with prefixes, suffixes, or a lack of affixes. The tenses of Xhosa verbs are split into primary and secondary. Only the primary tenses (present, past, perfect, and future tense) can be properly morphologically analyzed by XhosMorph; the secondary tenses are generally formed by a deficient verb followed by a verb in primary tense.

**Indicative Mood**

The indicative mood of verbs expresses simple and direct assertions. It is in this mood that most tenses can be formed. Most other moods are limited to just a few tenses. Most tenses have both a short and long form, each with different usages.

Short form, present tense is formed with the stem of the verb, preceded by the pronominal concord. Other affixes, such as objective concords can also be attached to the verb stem. Short form is used when the verb is followed by an object or an adverbial phrase. Long form of present tense is essentially the same as the short form, except the auxiliary particle **ya** is inserted in between the pronominal concord and the verb stem. Long form is always used at the end of a clause, as well as when the emphasis is on the verb, even if there is an object .

**ndifuna ukuhamba**, I want to go

**sisebenza nzima**, we are working hard

Negative form, present tense only has one form, which takes after the short form of non-negative present tense indicative verbs. Negative form takes the negative particle **a** before the pronominal concord. The verbal root also takes the negative suffix **i** at the end, rather than **a**.

**Andithandi**, I do not love

Perfect tense, short form is formed by changing the final **a** of short form, present tense, to an **e**. Long form changes that final **a** to **ile**. The different forms are used in the same way as short and long form of present tense. Perfect tense corresponds partly with the English perfect and past tenses. If this tense is used as past tense, it refers to the recent past or some other definite point in time. Perfect tense has many different allomorphies, depending on the structure of the verb stem. Stems ending in **ala**, **ana**, or **atha** become **ele**, **ene**, and **ethe**, respectively. Many verbs ending in **ela** simply become **ele** in long form. Other specific verb stems take irregular perfect forms: **hlala** becomes **hleli**, **hlutha** becomes **hluthi**, **ma** becomes **mi** or **mile**, **mitha** becomes **mithi**, and **tsho** becomes **tshilo**. A few verbs only have short form of the perfect tense: **thi** only has the perfect form tense of **the**, and **ba** becomes **be**.

The negative form of perfect tense, like that of present tense, only has one form. The negative prefix **a** starts the word, and the perfect **e** at the end of the word is replaced with the negative particle **nga**.

**Andilimbanga**, I have not tried.

Past tense only has a short form. Because perfect tense can be used to describe things that happened in the recent past, past tense also only describes long past, or indefinite past. Past tense is formed by inserting an **a** after the pronominal subject. The pronominal subject takes its consonantal form in past tense. **Wasinceda wena**, you helped us, and **Indoda yamlahla** **umkakhe**, the man deserted his wife.

The negative form of past tense is technically formed by adding the negative particle **a** to the beginning of the word. However, more often than not, the negative form of perfect tense is used instead of the proper past tense negative.

**Andahamba**, I did not go

**Andihambanga,** I did not go

Future tense, short form is created in a similar manner as past tense, where **o** is inserted after the consonantal form of the pronominal subject. Long form of future tense is actually split into two separate words, the first being the pronominal concord followed by the auxiliary verb **ya** (*to go*) and the second being the verb stem, prefixed by **ku**. The difference between short and long form of future tense is very different than present tense. Long form expresses simple future tense, such as **ndiya kuhamba**, I will go. Short form has sense of conditionalism, such as **ndongena ke**, I will enter then (in the circumstances you mention). Short form also has an implied reference to the person being spoken to. It corresponds most with an emphatic future in English.

**Ndiya kuhlala apha**, I shall stay here

**Uya kuhamba nale ndoda?**, Will you go with this man?

Future negative tense, unlike all other primary tenses, does have separate forms for short and long negative forms. Long form of future tense is formed by changing the first word, the auxiliary **ya**, by adding the negative prefix **a** and changing the final **a** to **i**. The second word does not undergo any change. Short form, negative completely changes the non-negative form. It is formed with the negative prefix **a**, followed by the pronominal concord in normal form, the shortened infinitive prefix **ku**, then the verb stem.

**Andiyi kuhlala apha**, I shall not stay here

**Abantu abakubona inkosi yabo**, we shall return soon

The secondary tenses of the indicative tense are organized into two sets, recent past and distant past. The secondary tenses consist of two words: an auxiliary verb followed by the real verb root.

The secondary tenses of the indicative are largely formed using an auxiliary verb followed by a primary tense of the verb root. There are two sets of secondary tenses: recent past and long past. Recent past uses the auxiliary verb **ukuba** in its perfect tense followed by either the present tense, perfect tense, or future tense. Distant past uses the auxiliary verb **ukuya** in past tense, also followed by present, perfect, or future tense. The root in present tense is called imperfect tense, perfect tense is called pluperfect tense, future tense is called future imperfect tense. Long form of both recent and distant past are formed in the same way: the respective auxiliary verb with a pronominal concord is used in perfect tense, followed by a normal formation of the appropriate primary tense. Short form for each recent and distant past is different for each set, as well as each tense associated with it.

For recent past, the secondary tense generally contracts the auxiliary form to **be** , which is then prefixed to the beginning of the primary tense word. In the future imperfect tense, **be** contracts with the first word of the long form of future tense. This means that in long form, future imperfect tense consists of three words, two auxiliary verbs followed by the verb root. Short form still consists of two words.

For distant past, the secondary tense uses the past tense of the auxiliary ukuya, wherein the last **a** is softened to **e**. In short form, this auxiliary is contracted to just the pronominal concord in consonant form, followed by **a**. This shortened form is combined with the primary tense word, in a similar way as the recent past.

The negative tenses of the secondary tenses of indicative mood form negatives in the same way as the respective primary tense used in the second word of long form. In short form, the part of the contracted form that corresponds with the long form second word is changed to the respective negative primary tense.

**ndaye ndingahambi**, I was not going

**ndandingahambi,** I was not going

XhosMorph treats secondary tenses as separate words rather than in total. The auxiliary verb is tagged as an auxiliary form, and the primary verb is tagged in the primary tense in which it appears.

When the indicative mood makes an assertion, it is often followed by a special mood called participial. Participial mood describes something acting or being acted upon. It thus functions as an adjective and a verb. Participial mood is virtually indistinguishable from the primary tenses of indicative form, except in past tense, which ends in **yo**. Some classes change pronominal concords in every tense except past: Class I singular changes from **u** to **e**, Class I plural changes from **ba** to **be**, and Class III plural changes from **a** to **e**.

**Subjunctive Mood**

The subjunctive mood describes an action which is a sequence or a result of another action. It thus follows another verb¸which is in the indicative, imperative, or infinitive mood. In present tense, subjunctive are preceded by the final conjunctions, **ukuba**, that, **ukuze**, in order that, **hleze**, lest. Subjunctives can also appear in past tense. With Class I singular, **u** becomes **a**. Present tense is formed in the same way as perfect tense in the indicative mood. Past tense is formed in the same way in subjunctive as indicative mood.

Subjunctive form becomes negative in a different way for each tense. Present tense takes the negative particle **nga** after the pronominal concord. It also ends in **i** instead of **a**. Past tense takes the negative prefix **a** before the normal subjunctive past tense. The ending does not change.

**Temporal Mood**

Temporal mood is used in dependent clauses that describe “when” an action happened. It is similar to a contracted form of future indicative tense. It takes the consonantal form, followed by **a**, like in past tense, and the abbreviated infinitive form, **ku**. Class I singular becoms **a**. Temporal mood can be in present, past, or perfect tense. Past and perfect tense are both formed by using the temporal mood of the deficient verb **ukuba,** to be, followed by the perfect tense of the participial form.

|  |  |  |
| --- | --- | --- |
| **Temporal Mood** of the verb **ukubuya**, to return | | |
|  | **Singular** | **Plural** |
| **1st** | ndakubuya | sakubuya |
| **2nd** | wakubuya | nakubuya |
| **I** | akubuya | bakubuya |
| **II** | wakubuya | yakubuya |
| **III** | lakubuya | akubuya |
| **IV** | sakubuya | zakubuya |
| **V** | yakubuya | zakubuya |
| **VI** | lwakubuya | zakubuya |
| **VII** | bakubuya |  |
| **VIII** | kwakubuya |  |

The temporal form of the verb **ukubona**, to \_\_\_\_\_\_, followed by **ukuba** and a present tense verb, expresses the idea of “while” an action is happening. **Bakubona ukuba baya thetha**, while they were speaking, or, **akubon’ ukuba uyemka**, as he was departing. The introductory verb **ukuthi**, to do thus, is also be used before temporal mood. **Ndithe ndakuvuka ndahamba indlela yam**, (I did thus) when I awoke I went on my way.

Temporal form does not take a negative form.

**Imperative Mood**

Imperative mood expresses a command or a request. It only occurs in present tense and only in second person. The singular form is simply the verb stem, without any prefix. Plural form is the same as singular, with the suffix **ni** added to the end of the stem. If the imperative is needed for any other voice than second, the deficient verb **ukuma** can be used in imperative singular form (**ma**) and followed by any verb in present subjunctive. **Hamba**, [you] go, and **vulani amehlo enu**, open your eyes (pl.).

The negative form of the imperative mood is formed in two separate ways. The first way is to use the negative form of the past subjunctive. The second is to use the deficient verb **musa** (do not, or stop doing) followed by the infinitive form. **Musu ukuthetha**, do not speak, or **ungahambi**, do not go.

**Infinitive Mood**

Infinitive mood names the verbal action. It is also classified as a Class VIII noun, and therefore govern other words as a verb while being governed by verbs itself. It is equivalent to the English infinitive, gerund, and verbal noun. It only appears in present tense, where it is prefixed by **uku**. **Ukufika kwakhe kwalindelwa**, his arrival was expected.

Negative form of the infinitive mood is formed by inserting the negative particle **nga** after the infinitive prefix. The negative ending **i** is also used for infinitive form. **Ukungahambi**, not going.

**Verbal Voices**

Verbs can be expressed from several viewpoints beyond simple subject agreement. These different viewpoints are called “voices”. **Active voice** is when the subject is the performer of an action. Examples have been covered extensively in the moods and tenses already outlined. The other main voice in Xhosa is **passive voice**. This voice is used when the subject is the person or thing acted upon. Passive voice is formed by inserting a **w** before the final vowel of the verb. Thus, the ending **a** becomes **wa**, **e** becomes **we**, and **ile** becomes **iwe** (contracted from **ilwe**). **Siya thandwa**, we are loved, or **inqwelo yatsalwa ngamahashe**, the wagon was drawn by horses.

Some allomorphic changes occur when making a verb passive. Labial sounds before the passive **w** are changed to the corresponding palatal sound.

|  |  |  |
| --- | --- | --- |
| **Labial** | **Palatal** | **Example** |
| p | tsh | **bopha**, to bind, becomes **botshwa** |
| b | ty | **hlaba**, to stab, becomes **hlatywa** |
| b | j | **bubisa**, to destroy, becomes **bujiswa** |
| mb | nj | **bamba**, to seize, becomes **banjwa** |
| m | ny | **luma**, to bite, becomes **lunywa** |
| m | nyu | **tyumza**, to crush, becomes **tyunyuzwa** |

If m occurs before a consonant other than a labial, a **u** is added before the passive **w**. This **u** was originally present in active form, but is now contracted to just **m**.

Intransitive verbs can also appear in passive tense, with the impersonal pronoun of Class VIII as theiir subject, in contracted form. **Kuya hanjwa**, it is being travelled, or **kuya zondwana apha**, there is mutual hatred here.

Some examples of the Passive Voice

**Siya qeqeshwa ngabo**, we are being trained by them.

**Amaza aqhutywa ngumoya (qhuba)**, the waves are being driven by the wind

**Kwashunyalyelwa (shumayela) kuloo ndawo**, preaching was going on at that place

**Other Allomorphies**

Xhosa verbs also have separate allomorphic transformations for verbs whose stem is either a single consonant or with a root that begins with a vowel.

Monosyllabic verbs sometimes have their short forms lengthened by an insertion of an extra syllable. The present participle and compounds insert **si** before the stem of the verb, like in **ndisiva**, I hearing, or  **be ndisiva**, I was hearing. The imperative form of the verb adds **yi** to the beginning of the stem (**yiza**, come). The passive voice inserts **i** before the passive ending, like in **ukupha** becoming **ukuphiwa.**

Vowel verbs have many similarities to the peculiarities of monosyllabic . The present participle inserts the letter **s** before the stem, for example **be ndisenza**, I was doing. Imperative prefixes **y** before the stem (**yaakha**, build), and in passive verbs, an **i** is inserted before the passive ending (**ukwakhiwa**, to be built). Vowels that come before the stem are either omitted or converted to their semi-vowels. For example, **sioyika** becomes **soyika**, and **ukuamkela** becomes **ukwamkela**.

A few monosyllabic verbs were originally vowel verbs starting with the letter **i**. This is evidenced by the fact that the pronominal concord beore this vowel changes from **a** to **e**. The vowels that fall under this category are **ukuuba, ukuhla, ukuma, ukuza, ukupha, ukusa,ukuva**. Some other verbs take **e** instead of **a** in the long form of present tense and long past, like **ukusuka,ukunyuka,ukumba,** and **ukumka**.

**Derivative Forms**

**Directive Form**

The directive form indicates that the action of the verb is directed towards some person, place or thing. It expresses purpose and direction, and is used before names of persons and things for which something is done and words indicating the place towards motion is directed. It is formed by changing the final **a** of simple form and changing it to **ela**.

Directive form frequently has two objects, direct and indirect. The former is the object of the simple verb, the latter is the object of the applied idea. **Inombi imthungela unina** ingubo, the girl is sweing a dress for her mother. The reflexive article can be used in the directive form to express the idea of “by oneself” or “of oneself”. **Ukuzihlalela**, to sit by oneself. Directive form sometimes occurs with verbs that do no have corresponding simple forms, lie **ukukhangela**, look at**, ukusondela**, draw near, and **khanyela**, deny.

When these verbs are to be treated as locative, they are followed by an adverb or noun in locative case. **Hambela entabini**, go towards the mountain. The infinitive form of the directive form is used to express location. **Indlu yokudlela**, a room to eat in. Passive form is also frequently used with directive verbs. **Ndakhelwa indlu**, a house was built for me.

**Causative Form**

Causative form indicates the action of the verb was a result of a secondary agent “forcing” or “making: the subject do something. It is generally formed by the suffix **isa** instead of **a**. It changes intransitive verbs to transitive. Several causative verbs have a short and long form, usually ones ending in **ala**. The long form is formed by adding the causative suffix **isa** instead of the normal **a**. Short form drops the **ala** at the end of the word, and replaces it with **eza**. Other words that end in **la** or **ka** replace it with **za** consistently.

Some verbs that appear in causative tense were derived from nouns or adjectives. **Duludulusa,** struggle with difficulties. Transitive verbs also become “doubly transitive”, where they can take two objects. **Ndababonisa indlela**, I showed them the way. Causitive form can also be used to mean “help to do”, like in **bophisa**, help to bind.

**Uya ndiqumbisa**, you make me angry.

**Bahlanganise abantu**, assemble the people.

**Reciprocal Form**

Reciprocal form indicates that the action is performed by two or more mutual actors. It is formed by changing the final **a** to **ana**. For example, **siya** **thandana**, we love one another. It is found sometimes with verbs that do not have a simple form, such **hlangana**, meet, or **fumana**, find. Recirprocal form is usually used in the plural, such as **baya fana**, they resemble each other, but it can be used with singular subjects if the prepositional prefix **na** follows, like in  **ingwe ifana nekati**, the leopard resembles the cat. The verb **ukufana**, resemble, does not undergo any changes to become reciprocal. For example, **ufana nam**, he resembles me.

**Stative Form**

Stative form expresses the state of being resulting from subjection to an action of the simple verb. It differs from passive tense in that the agent of the action is not explicitly mentioned. It changes the final **a** to **eka**. Using stative form makes intransitive verbs transitive. **Ndizilahlile iigusha zam**, I have lost my sheep. Stative form is also used, especially in negative sentences, to denote fitness or capacity for performing an action, like in **intlabo iya hlabeka**, the awl pierces, fitness or capacity for undergoing an action, like in **umhlabaawulimeki**, the ground does not plough well. In these cases, it corresponds to the English suffix **–able**. Normally, however, the stative form corresponds to “be”, “become”, or “get”, such as **lumeka**, get bitten.

Perfect tense of the stative form is used frequently to form adjectives, such as **lahlekile**, lost (from **lahla**) or **umntu othembekileyo**, a trusted or trustworthy person. It can also be followed by an objective of general reference, such as **sihlaziyekile amandla**, we were refreshed (in) strength.

If the final **a** is replaced with **akala**, it is considered “intensive stative form”. Some verbs have both forms of stative form, such as **boneka**, be seen, or **bonakala**, appear. Allomorphies occur whenever the simple verb would end in **la**: **ahlula** becomes **ahluka** and **dlathula** becomes **dlathuka**. Sometimes, these transitions have dropped their original **la** ending word.

**Deficient Verbs**

Deficient verbs are used as auxiliary verbs to describe concepts that affixes alone cannot express in Xhosa.

**UKUBA, to be, to become, to suppose**

In perfect tense, followed by the participle of a present tense, ukuba forms the first set of secondary indicative tenses, like in **ndibe ndithetha**, I was speaking. In future tense, followed by a present or perfect tense, it expresses a [rogressive or completed action, such as **siya kuba siya hlamba**, we shall be watching. As remote past tense, it can be used in replacement of the remote past tense of **ukuya**, before intransitive participles, like in **sabe sifuneka**, we were sought for.

When used on its own, **ukuba** carries the sense of “to be” in all tenses and moods, except present indicative, where the copula replaces it. It can be used in present inicative tense if it follows **nga** or **sa**. In negative past subjunctive, it expresses an urgent wish. **Akwaba wawukho**, would that you were present! **Ukuba** can also be used in the sense of “to become”, especially in past, present, and future tenses, like in **ma sibone ukuba amaphupha akhe oba njani na**, let us see what will become of his dreams. If it is followed by **nga**, **ukuba**is used in the sense of “as soon as”, like in **liba lingafuna ukukhwaza ixhego**, as soon as the old man tried to shout.

**UKUYA, to go**

By itself, **ukuya** means “to go”. It is alo used to make the long form of the secondary distant past tenses, and long form of future indicative tense. Remote past tense softens the final **a** to **e**. **Ndaye ndihamba**, I was travelling. **Ukuya** can be added to other verbs of motion to give them the idea of “going”, like in **uye wafika**, he arrived. When this form is followed by an infinitive, a contracted form is used in informal speech, such as **baleka uyokubona**, run and see, which replaces **baleka uye ukubona**. In past subjunctive tense, **ukuya** is used to connect verbs referring to different points of time, like in **ndivule ivenkile, ndaye ndithemba ukufumana inkxaso**, I have opened a shop and hope to obtain support.

**UKUMA, to stand**

**Ukuma** is used, as mentioned before, in the sense of “let” or “must”. It is followed by the present subjunctive of the complementary verb. In Class I singular and Class III plural, a **k** is inserted after the m, to prevent the tow vowels from touching. Used with the second person, **ma** expresses an urgent imperative, as in **ma nibuye**, you must return.

There are several other deficient verbs, like **ukunga**,

**Interrogatives**

Questions are indicated in Xhosa with the suffix **na**. It comes at the end of a clause, or stands as a clause on its own. Question words are derivatives of either **nina**, what?, or **phina**, where?. There are many derivatives of each, including **nanina**, through what?, or **ngaphina**, how many? The most common question words are incorporated into XhosMorph as both suffixes and separate words.

**Adverbs**

Simple adverbs are not very numerous in Xhosa, since so many nouns and verbs can serve as adverbs. Locative nouns can be used as adverbs, as can nouns or pronouns with the prepositional prefix **nga**. Adjectives can becomes adverbs by prefixing **ka**, as in **kakhulu**, greatly. XhosMorph covers a list of approximately 50 common adverbs, but cannot only tag further adverbs that appear in the **ka**+adjective form.

**Conjunctions**

Conjunctions in Xhosa are also relatively few, but can be formed from verbal forms or pronouns. Conjunctions are divided into subordinative and coordinative conjunctions. Coordinative conjunctions join sentences of the same rank, like **ke**, and, but, so, then. The five classes of subordinative are declarative, conditional, relative, causal, and final. Declarative conjunctions join a dependent assertion to a principal clause; the first clause is always past tense, and the second is always present tense, like **ukuba**, that. Conditional clauses follow hypothetical or conditional clauses. Relative conjunctions mark clauses expressing time, place, and manner, like **xa**, when. They also function as adverbs. Causal conjunctions are used after clauses expressing cause and motive, as in **kuba**, because. Final conjunctions are used after clauses expressing purpose, like the conjunction  **ukuze**, that, in order that. XhosMorph can handle around 30 conjunctions, but has no way of tagging conjunctions that have not been manually added to the lexicon.

# Implementation

The morphological analyzer was implemented using the Xerox Finite State Toolkit, as developed by Xerox. The xfst files involved in XhosMorph consist of a lexicon file (xhos.lex), which details all of the basic forms of acceptable words, a xfst rule file (xhRules.fst), which refines the lexicon to only include properly formed words, a root guesser (xhGuess.fst), which contains an algorithm to guess a properly formed root, and a strategy file (xhStrat.fst) which allows both known roots and guessed roots to be included in the final results.

All of the xfst files are compiled into the Java executable XhosMorph.jar. This wrapper class allows the user to select the file they would like to have morphologically analyzed, runs it through the lexicon, rules, guesser, and strategy files to come up with a list of possible morphologies for wach word inputted, from which it picks the most probable morphological analysis, based on simple dependency algorithm. XhosMorph correctly analyzes #% of words that consist of a root already contained in the lexicon, and #% of all words that contain an unknown root.

All grammar concepts outlined in the morphology and allomorphy section were implemented in XhosMorph. The first layer of analysis consists of a lexicon file. The lexicon file outlined the basic morphology of Xhosa, but did not delve into any allomorphy.

# Specific Adaptations

XhosMorph contains several adaptations that make it ideal to be expanded to other Bantu languages. The most notable feature of XhosMorph is that of the test suite kit. When a test suite is run through XhosMorph, the researcher can choose the correct analysis from a list of all possible morphological anayses. This selection allows XhosMorph to store that selection in an internal probability system, which can then recognize morphemes that are more likely to occur given certain contexts. XhosMorph will also add any unknown roots to its lexicon file automatically, once a correct analysis is selected. This adaptation of the lexicon file allows XhosMorph to generate more accurate results after every test suite.

XhosMorph also has a basic dependency calculation feature. This allows the analyzer to look at entire sentences to provide information about in what context a certain prefix is being used. For Bantu languages, this is especially important, as the subject determines the pronominal concord prefixes for all following adjectives, adverbs, and verbs. XhosMorph can analyze the information gien by the entire system, recognize singular, plural, and concord forms. This information is then used by the system to generate more accurate results. Without this dependency calculation, XhosMorph’s reliability is as #%. With the dependency, the reliability is #%.

# Future Goals

XhosMorph contains the framework to expand to other Bantu languages. The test suite toolkit allows for easy additions into the

1. Exceptions occur in the locative nouns and prepositional standalone words [↑](#footnote-ref-1)