

ProcedureProtocol

**Protocol: Morphological Analysis (isiZulu)Protocol: Morphological Analysis (Siswati)**

**X.Y.Z1.0.3**

**Name of AuthorW. Pienaar**

DOCUMENT INFORMATION

|  |  |
| --- | --- |
| **Filename** | Protocol.SADiLaR.MorphologicalAnalysisSiswati.1.0.3.WP.2018-12-04.docx |
| **Type** | Protocol |
| **Title** | Protocol: Morphological Analysis (Siswati) |
| **Version** | 1.0.3 |
| **Source Author(s)** | W. Pienaar |
| **Commentator(s)** | M.J.Puttkammer |
| **Status** | Document Status |
| **Date** | 2018-12-06 |
| **Pages** | 18 |
| **Security Classification** | Public Domain Only Source Author(s) Only Distributors Only NWU |

|  |  |
| --- | --- |
| **Distribution** |  |
| **Key Words** |  |
| **Project** | SADiLaR |

DOCUMENT HISTORY

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Status** | **Author/Commentator** |
| Version of document when opened | Date of document as specified by previous author | Status of document when opened | Name of previous author(s) or commentator(s) |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

TABLE OF CONTENTS

[Introduction 6](#_Toc531698966)

[Aim of this protocol 6](#_Toc531698967)

[Concept and definitions 6](#_Toc531698968)

[CATEGORY 1: Nouns 7](#_Toc531698969)

[CATEGORY 2: Verbs and auxiliary verbs 7](#_Toc531698970)

[CATEGORY 3: Adjectives 8](#_Toc531698971)

[CATEGORY 4: Adverbs 8](#_Toc531698972)

[CATEGORY 5: Subject concords ([SC]) 9](#_Toc531698973)

[CATEGORY 6: Object concords ([OC]) 10](#_Toc531698974)

[CATEGORY 7: Demonstratives ([Dem]) 10](#_Toc531698975)

[CATEGORY 8: Quantitatives 11](#_Toc531698976)

[CATEGORY 9: Relatives 11](#_Toc531698977)

[CATEGORY 10: Pronouns 12](#_Toc531698978)

[CATEGORY 11: Copulatives 13](#_Toc531698979)

[CATEGORY 12: Particles and prepositions 13](#_Toc531698980)

[CATEGORY 13: Secondary morphemes 14](#_Toc531698981)

[CATEGORY 14: Ideophones ([Ideoph]) 14](#_Toc531698982)

[CATEGORY 15: Possessive concords ([PossConc]) 14](#_Toc531698983)

[CATEGORY 16: Interjections ([Interjec]) 14](#_Toc531698984)

[CATEGORY 17: Conjunctions ([Conj]) 14](#_Toc531698985)

[CATEGORY 18: Interrogatives 15](#_Toc531698986)

[CATEGORY 19: Abbreviations ([Abbr]) 15](#_Toc531698987)

[CATEGORY 20: Demonstrative Copulatives ([DemCop]) 15](#_Toc531698988)

[CATEGORY 21: Enumeratives 16](#_Toc531698989)

[CATEGORY 22: Tense markers 16](#_Toc531698990)

[CATEGORY 23: Foreign ([Foreign]) 16](#_Toc531698991)

[CATEGORY 24: Proper name ([ProperName]) 16](#_Toc531698992)

[General principles 17](#_Toc531698993)

[Changing the protocol 17](#_Toc531698994)

[Authoritative references 17](#_Toc531698995)

[Roles and contact details 18](#_Toc531698996)

# Introduction

## Aim of this protocol

The aim of this protocol is to guide annotators in the process of morphological analysis for the development of text resources for the SADiLaR project.

This document explicates some general principles to be followed in annotating the corpus, explains the process of annotation, and provides a detailed explanation of the tag set to be used. This is a dynamic, “living” document, and should be adapted following the procedure as explicated in the section “Changing the protocol”.

## Concept and definitions

The aim of annotation on the morphological analysis layer is to identify the morphemes of each token (specifically morphologically complex words), as identified on the tokenisation layer (i.e. usually words, multi-word expressions, and punctuation). The following definitions apply:

* **Morphological analysis**: The identification and description of morphemes in complex words.
* **Morpheme**: A simplex symbolic (i.e. (grammatically) meaningful) unit in the language system. It is simplex in the sense that it does not contain smaller symbolic units as subparts; e.g. a morpheme, like the class prefix t*i-* of class 8/10.

Morphemes in Siswati include nominal prefixes (e.g. Classes 1-18), verbal suffixes (e.g. -is-, -el-, -an- ), subject and object concord, etc.

* **Root:** The verbal stem without any affixes or extension (e.g. -tseng-)
* **Stem**:
  + Nominal: nominal stem without affixes (e.g. -*kati*)
  + Verbal: root with affixes and extension.(e.g. *tsenga*, *tsengela*, *titsengela*)

Main categories for morphological analysis:

|  |
| --- |
| CATEGORY 1: Nouns |
| CATEGORY 2: Verbs and auxiliary verbs |
| CATEGORY 3: Adjectives |
| CATEGORY 4: Adverbs |
| CATEGORY 5: Subject concords |
| CATEGORY 6: Object concords |
| CATEGORY 7: Demonstratives |
| CATEGORY 8: Quantitatives |
| CATEGORY 9: Relatives |
| CATEGORY 10: Pronouns |
| CATEGORY 11: Copulatives |
| CATEGORY 12: Particles and prepositions |
| CATEGORY 13: Secondary morphemes |
| CATEGORY 14: Ideophones |
| CATEGORY 15: Possessive concords |
| CATEGORY 16: Interjections |
| CATEGORY 17: Conjunctions |
| CATEGORY 18: Interrogatives |
| CATEGORY 19: Abbreviations |
| CATEGORY 20: Demonstrative Copulatives |
| CATEGORY 21: Enumeratives |
| CATEGORY 22: Tense markers |
| CATEGORY 23: Foreign |
| CATEGORY 24: Proper name |

# CATEGORY 1: Nouns

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Class nr.** | **Example from each class** | **Class prefix**  **([BPre] / [NPre] / [NPrePre])** | **Stem**  **([NStem])** | **Suffix**  **Aug**  **([AugSuf])** | **Suffix**  **Dim**  **([DimSuf])** | **Suffix**  **Loc**  **([LocSuf])** |
| Class 1 | umfati | umu- | -fati | -kati | -ana | -e |
| Class 1a | babe | Ø | -babe |  |  |  |
| Class 2 | bafati | ba- | -fati |  |  |  |
| Class 2a | bobabe | bo- | -babe |  |  |  |
| Class 3 | umfula | umu- | -fula |  |  |  |
| Class 4 | imifula | imi- | -fula |  |  |  |
| Class 5 | licandza | li- | -candza |  |  |  |
| Class 6 | emacandza | ema- | -candza |  |  |  |
| Class 7 | sicatfulo | si- | -catfulo |  |  |  |
| Class 8 | ticatfulo | ti- | -catfulo |  |  |  |
| Class 9 | imvu | in- | -vu |  |  |  |
| Class 10 | timvu | tin- | -vu |  |  |  |
| Class 11 | lutsi | lu- | -tsi |  |  |  |
| Class 14 | busuku | bu- | -suku |  |  |  |
| Class 15 | kuhamba | ku- | -hamba |  |  |  |
| Class 16 | phasi | pha- | -si |  |  |  |
| Class 17 | kumake | ku- | -make |  |  |  |
| Class 18 | emuva | mu- | -va |  |  |  |

# CATEGORY 2: Verbs and auxiliary verbs

|  |  |  |
| --- | --- | --- |
| Verbal root | [VRoot] | e.g. -mis |
|  |  |  |
| Intransitive verbal root | [Intrans] | e.g. -etigab |
|  |  |  |
| Verb terminative (-a, -e) | [VerbTerm] | e.g. -a |

Verbal prefixes

|  |  |  |
| --- | --- | --- |
| Reflexive prefix | [ReflPre] | i- |
| Imperative prefix | [ImpPre] | yi- |
| Locative prefix | [LocPre] | ku- |
| Progressive prefix | [ProgPre] | sa- |
| Pre-locative “s” | [PreLoc-s] | s- |

Verbal Suffixes

The following tags are present in the data, use the specific tag in the table below:

|  |  |  |
| --- | --- | --- |
| Neuter | [NeutExt] | -akal- |
| Intensive | [IntensExt] | -elel- |
| Reciprocal | [RecipExt] | -an- |
| Applicative | [ApplExt] | -el- / -l- |
| Causative | [CausExt] | -is- |
| Passive | [PassExt] | -iw- / -w- |

The following tags exist in Siswati, use the specific tag in the table below if possible, otherwise use [VerbExt].

|  |  |  |
| --- | --- | --- |
| Associative | [Ass] | -anan- |
| Iterative | [Iter] | -ak- |
| Neutro active | [NAct] | -al- |
| Dispersive | [Dis] | -alal- |
| Positional | [Pos] | -am- |
| Contactive | [Con] | -ats- |
| Neutro passive | [NPas] | -ek- |
| Perfect | [Per] | -il- |
| Intensive | [Intens] | -isis- |

Auxiliary verbs ([AuxVStem])

|  |
| --- |
| -be |
| -se |

Copulative verbs ([Cop])

<to be adapted to Siswati>

|  |
| --- |
| Verb stem (e.g. le, ba) |

# CATEGORY 3: Adjectives

|  |
| --- |
| Adjective stem ([AdjStem]) |
| Adjective prefix ([AdjPref]): Same as class prefix - see Category 1 |

# CATEGORY 4: Adverbs

([Adv])

|  |
| --- |
| Basic adverbs (e.g. mbamba) |
| Derived adverbs (e.g. kabilii, kakhulu) |
| Verbs used as adverbs: See verb Category 2 |

Adverb prefixes ([AdvPre])

|  |
| --- |
| nga- |
| na- |
| ka- |

# CATEGORY 5: Subject concords ([SC])

|  |  |  |  |
| --- | --- | --- | --- |
| **Persons and classes** |  | **Std** | **Con.** |
| 1st Sing. |  | ngi e.g.  ngitsengile tincwadzi | ka e.g. angikatsengi tincwadzi |
| 1st Pl. |  | si | si |
| 2nd Sing. |  | u | u |
| 2nd Pl. |  | ni | ni |
| Class 1 | umfati | u | u |
| Class 1a | babe | (See class 1) |  |
| Class 2 | bafati | ba | ba |
| Class 2a | bobabe | (See class 2) |  |
| Class 3 | umuti | u | wa |
| Class 4 | imiti | i | ya |
| Class 5 | licandza | li | li |
| Class 6 | emacandza | a | a |
| Class 7 | sibuko | si | si |
| Class 8 | tibuko | ti | ta |
| Class 9 | imvu | I | ya |
| Class 10 | timvu | ti | ta |
| Class 11 | lutsi | lu | lwa |
| Class 14 | busuku | bu | ba |
| Class 15 | kuhamba | ku | kwa |
| Class 16 | phasi | (see class 17) |  |
| Class 17 | kudze | ku | kwa |
| Class 18 | emuva | (see class 17) |  |

Indefinite and classless

|  |  |
| --- | --- |
| Indefinite | ku (e.g. kuyashisa) |

# CATEGORY 6: Object concords ([OC])

|  |  |
| --- | --- |
| **Persons & classes** | **Object concord** |
| 1st Sing. | ngi |
| 1st Pl. | si |
| 2nd Sing. | ku |
| 2nd Pl. | ni |
| Class 1 | mu |
| Class 1a | (See class 1) |
| Class 2 | ba |
| Class 2a | (See class 2) |
| Class 3 | wu |
| Class 4 | yi |
| Class 5 | li |
| Class 6 | wa |
| Class 7 | si |
| Class 8 | ti |
| Class 9 | yi |
| Class 10 | ti |
| Class 11 | lu |
| Class 14 | bu |
| Class 15 | ku |
| Class 16 | (See class 17) |
| Class 17 | ku |
| Class 18 | (See class 17) |

# CATEGORY 7: Demonstratives ([Dem])

|  |  |  |  |
| --- | --- | --- | --- |
| **Persons & classes** | **Demonstratives** | | |
|  | **1st Dem** | **2nd Dem** | **3rd Dem** |
| Class 1 | lo | lowo | lowa |
| Class 1a | lo | lowo | lowa |
| Class 2 | laba | labo | laba |
| Class 2a | laba | labo | laba |
| Class 3 | lo | lowo | lowa |
| Class 4 | le | leyo | leya |
| Class 5 | leli | lelo | lela |
| Class 6 | lawa | lawo | lawa |
| Class 7 | lesi | leso | lesa |
| Class 8 | leti | leto | leta |
| Class 9 | le | leyo | yela |
| Class 10 | leti | leto | leta |
| Class 11 | lolu | lolo | lolwa |
| Class 14 | lobu | lobo | loba |
| Class 15 | loku | loku | lokwa |
| Class 16 | lapha | lapho | lapha |
| Class 17 | lapha | lapho | lapha |

# CATEGORY 8: Quantitatives

Quantitative Stem ([QuantStem])

E.g. -nkhe (inclusive), -dvwa (exclusive)

Quantitative Concord ([QuantConc])

|  |  |
| --- | --- |
| **Persons & classes** | **Quantitative concord** |
| 1st Sing. | wo-/nge- |
| 1st Pl. | so- |
| 2nd Sing. | wo-/we- |
| 2nd Pl. | no- |
| Class 1 | yo- |
| Class 1a | (See class 1) |
| Class 2 | bo- |
| Class 2a | (See class 2) |
| Class 3 | wo- |
| Class 4 | yo- |
| Class 5 | lo- |
| Class 6 | o- |
| Class 7 | so- |
| Class 8 | to- |
| Class 9 | yo- |
| Class 10 | to- |
| Class 11 | lo- |
| Class 14 | bo- |
| Class 15 | ko- |
| Class 16 | (See class 17) |
| Class 17 | ko- |
| Class 18 | (See class 17) |

# CATEGORY 9: Relatives

Relative Stem ([RelStem])

E.g. -tsite

Relative Concord ([RelConc])

|  |  |
| --- | --- |
| **Persons & classes** | **Relative concord** |
| Class 1 | lo- |
| Class 1a | lo- |
| Class 2 | laba- |
| Class 2a | laba- |
| Class 3 | lo- |
| Class 4 | le- |
| Class 5 | leli- |
| Class 6 | la- |
| Class 7 | lesi- |
| Class 8 | leti- |
| Class 9 | le- |
| Class 10 | leti- |
| Class 11 | lolu- |
| Class 14 | lobu- |
| Class 15 | loku- |
| Class 17 | loku- |

Relative Suffix ([RelSuf])

|  |
| --- |
| -ko |

# CATEGORY 10: Pronouns

Absolute Pronouns ([AbsPron])

|  |  |
| --- | --- |
| **Persons & classes** | **Absolute Pronoun** |
| 1st Sing. | mine |
| 1st Pl. | tsine |
| 2nd Sing. | wena |
| 2nd Pl. | nine |
| Class 1 | yena |
| Class 1a | (See class 1) |
| Class 2 | bona |
| Class 2a | (See class 2) |
| Class 3 | wona |
| Class 4 | yona |
| Class 5 | lona |
| Class 6 | ona |
| Class 7 | sona |
| Class 8 | tona |
| Class 9 | yona |
| Class 10 | tona |
| Class 14 | bona |
| Class 15 | kona |
| Class 16 | (See class 17) |
| Class 17 | kona |

Possessive Pronouns ([PossPron])

|  |  |
| --- | --- |
| **Persons & classes** | **Possessive Pronoun** |
| 1st Sing. | kwa |
| 1st Pl. | tsine, ketfu |
| 2nd Sing. | kwakho |
| 2nd Pl. | nine, kenu |
| Class 1 | yena |
| Class 1a | (see class 1) |
| Class 2 | bona, kabo |
| Class 2a | (see class 2) |
| Class 3 | wona |
| Class 4 | yona |
| Class 5 | lona |
| Class 6 | ona |
| Class 7 | sona |
| Class 8 | tona |
| Class 9 | yona |
| Class 10 | tona |
| Class 11 | lona |
| Class 14 | bona |
| Class 15 | kona |
| Class 16 | (see class 17) |
| Class 17 | kona |

Pronoun Stem ([PronStem])

|  |  |
| --- | --- |
| **Persons & classes** | **Pronoun Stem** |
| 1st Sing. | mi |
| 1st Pl. | tsi |
| 2nd Sing. | we |
| 2nd Pl. | ni |
| Class 1 | ye |
| Class 1a | (See class 1) |
| Class 2 | bo |
| Class 2a | (See class 2) |
| Class 3 | wo |
| Class 4 | yo |
| Class 5 | lo |
| Class 6 | wo |
| Class 7 | so |
| Class 8 | to |
| Class 9 | yo |
| Class 10 | to |
| Class 11 | lo |
| Class 14 | bo |
| Class 15 | ko |
| Class 16 | (See class 17) |
| Class 16 | ko |

Demonstrative Pronouns: See Category 7

Pronoun Suffix ([PronSuf])

|  |
| --- |
| -na |

# CATEGORY 11: Copulatives

Copulative Verbs ([Cop]): see Category 2

Copulative Particles ([CopPre]): see Category 12

Demonstrative Copulatives ([DemCop]): see Category 20

|  |
| --- |
| ngi |
| u (See subject concords) |
| ti (See subject concords) |
|  |
| Verbs |
| li |
| ba |

# CATEGORY 12: Particles and prepositions

|  |
| --- |
| Locative / temporal (e.g.ku, e) |
| Copulative (e.g. ngu, yi) |
| Copulative particles resembling subject concords: see Category 5 |

# CATEGORY 13: Secondary morphemes

|  |  |  |
| --- | --- | --- |
| Future | [Fut] | e.g tawu, ta |
| Progressive | [Pro] | e.g. sa |
| Potential | [Pot] | e.g. ka |
| Exclusive | [ExclNeg] | -ka- |
| Negative | [NegPre] | -a-, -ka-, -nga-, -nge-, -anga-, -i- |

# CATEGORY 14: Ideophones ([Ideoph])

E.g. *ekhu*, *dvu, klebhu, mbo, meku, bha, monyu, nkha, chafuchafu, gcavu, ngci, muku, hhihli, dlo, nce, nce*

# CATEGORY 15: Possessive concords ([PossConc])

|  |  |
| --- | --- |
| **Persons & classes** | **Possessive Concord** |
| Class 1 | wa |
| Class 1a | (See class 1) |
| Class 2 | ba |
| Class 2a | (See class 2) |
| Class 3 | wa |
| Class 4 | ya |
| Class 5 | la |
| Class 6 | a |
| Class 7 | sa |
| Class 8 | ta |
| Class 9 | ya |
| Class 10 | ta |
| Class 11 | lwa |
| Class 14 | ba |
| Class 15 | kwa |

Possessive prefix “ka” ([PossKA])

# CATEGORY 16: Interjections ([Interjec])

E.g. ehhe*i, hhayi*!, washa!

# CATEGORY 17: Conjunctions ([Conj])

|  |
| --- |
| ge |
| ngoba |
| kepha |
| noma |
| futsi |

# CATEGORY 18: Interrogatives

Question words ([Interrog]), e.g.

|  |
| --- |
| bani |
| ini |
| bobani |
| kuphi |
| nini |

Question particles ([InterrogSuf]), e.g.

|  |
| --- |
| -ni |

# CATEGORY 19: Abbreviations ([Abbr])

Abbreviation (e.g. Umbu. Sib. Lib)

# CATEGORY 20: Demonstrative Copulatives ([DemCop])

|  |  |  |  |
| --- | --- | --- | --- |
| **Persons & classes** | **1** | **2** | **3** |
| Class 1 | nangu | nango | nanguya |
| Class 1a | (see class 1) |  |  |
| Class 2 | naba | nabo | nabaya |
| Class 2a | (see class 2) |  |  |
| Class 3 | nankhu | nankho | nankhuya |
| Class 4 | nayi | nayo | nayiya |
| Class 5 | nali | nalo | naliya |
| Class 6 | (see class 3) |  |  |
| Class 7 | nasi | naso | nasiya |
| Class 8 | nati | nato | natiya |
| Class 9 | (see class 4) |  |  |
| Class 10 | (see class 8) |  |  |
| Class 11 | nalu | nalo | naluya |
| Class 14 | nabu | nabo | nabuya |
| Class 15 | naku | nako | nakuya |
| Class 16 | (see class 15) |  |  |
| Class 17 | (see class 15) |  |  |

# CATEGORY 21: Enumeratives

Enumerative stem ([EnumStem]):

|  |
| --- |
| -nye |
| -phi |

Enumerative Concord ([EnumConc]):

|  |  |
| --- | --- |
| Class 1 | mu- |
| Class 1a | mu- |
| Class 2 | ba- |
| Class 2a | ba- |
| Class 3 | mu- |
| Class 4 | mi |
| Class 5 | li- |
| Class 6 | ma- |
| Class 7 | si- |
| Class 8 | zi- |
| Class 9 | yi- |
| Class 10 | zi- |
| Class 11 | lu- |
| Class 14 | bu- |
| Class 15 | ku- |
| Class 17 | ku- |

# CATEGORY 22: Tense markers

|  |  |
| --- | --- |
| Present tense | [Pres] |
| Perfect | [Perf] |
| Future tense  (See also Category 13) | [Fut] |
| Negative present | [NegPre] |
| Negative future | [NegFut] |

# CATEGORY 23: Foreign ([Foreign])

Tag used for any foreign words e.g. finance, management, service.

# CATEGORY 24: Proper name ([ProperName])

Tag used for proper names.

# General principles

The following general principles apply:

* Annotation on the morphological analysis layer should be in excess of 98%
* The general principles of the Text Encoding Initiative P5 Guidelines (TEI P5; <http://www.tei-c.org/Guidelines/>) should be followed when in doubt.

# Changing the protocol

Since this protocol is considered to be a dynamic document, it is subject to change from time to time. For purposes of good governance, the following principles and procedures should be followed:

* The language expert (in consultation with the project leader and project manager) has the final say in changing this procedure document, and is the only authoritative person who might make final changes to the protocol.

If an annotator (or other project member) wishes change a specification, s/he should inform the linguistic manager by email: Wikus.Pienaar@nwu.ac.za. The linguistic manager in liaison with the project leader considers the request and adds the change to the protocol document. The protocol document is saved with a new version number. All other annotators are then informed of the change, and requested to change previous annotations, where applicable.

# Authoritative references

The following language-specific works of reference should be considered authoritative when in doubt:

* ZIERVOGEL, D. 1952. A grammar of Swazi (siSwati). Witwatersrand University Press.
* TALJARD, PC ET AL 1991. Handbook of Siswati. Pretoria: J L Van Schaik.

*Please add more references used.*

# Roles and contact details

The following roles in this project can be identified:

**Linguistic manager**: Responsible for overall linguistic quality in liaison with the language experts and the project leader.

*Wikus Pienaar*

*Wikus.Pienaar@nwu.ac.za*

**Project leader**: Responsible for successful technical delivery of project.

*Martin Puttkammer*

*Martin.Puttkammer@nwu.ac.za*

**Project manager**: Responsible for timeous delivery of project, within budget. If you suspect that you will not be able to deliver on a set milestone, contact the project manager.

*Sunny Gent*

*Sunny.Gent@nwu.ac.za*

**Language expert**: Responsible for language correctness.

**Annotator**: Responsible for annotating the corpus, according to the above procedures and principles.