SADiLaR-II Project: Linguistic corpus enrichment for conjunctively written South African languages

**Protocol: Part-of-Speech Tagging (isiNdebele)**

2020-2021

# Aim of this protocol

The aim of this protocol is to guide annotators in the process of annotating the isiNdebele data on the part-of-speech layer for the SADiLaR-II project: *Linguistic corpus enrichment for conjunctively written South African languages*. 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”.

# Concepts and definitions

The aim of annotation on the part-of-speech layer is to identify the part-of-speech category of each token, as identified on the tokenisation layer (i.e. usually words, multi-word expressions, and punctuation). The following definitions apply:

* **Part-of-speech**: The function that a word fulfils in a sentence (i.e. in a grammatical context). It is also sometimes referred to as the lexical category, word class or lexical class of a word. The part-of speech of a word is determined by the morpho-syntactic behaviour of the word in the specific context, e.g. what happens to the words morphologically within certain syntactic contexts. In linguistics, we usually distinguish between open classes (i.e. classes to which new words can be added, and in that sense productive classes; usually nouns, verbs and adverbs), and closed classes (i.e. classes that are generally not expanded through productive, everyday processes; e.g. adjectives, pronouns, determiners, etc.).

**Example in running text**:

|  |  |  |
| --- | --- | --- |
| Umsekeli |  | N01 |
| ngeemali |  | ADV |
| kufanele |  | V |
| athole |  | V |
| imali |  | N09 |
| yokubhadela |  | POSS09 |
| amaphasela |  | N06 |
| wokudla |  | POSS06 |
| . |  | PUNC |
|  |  |  |

* **Word**: A string of characters with whitespaces on both sides, which could consist of alphanumeric characters, hyphens, apostrophes, full-stops, parentheses and/or symbols, but no other punctuation marks. In the Bantu languages, a distinction is made between linguistic words and orthographic words as these two entities do not always coincide. For disjunctively written Bantu languages, several orthographic words can correspond to one linguistic word, whereas for conjunctively written Bantu languages generally one orthographic word corresponds to one linguistic word. The preceding definition would therefore be true for orthographic words; linguistic words are linguistic units that constitute word categories. Abbreviations are considered to be words.

**Example for a disjunctive language**:

*Batho* linguistic word (noun), also an orthographic word

*ba* orthographic word

*a* orthographic word 1 linguistic word (verb)

*mo* orthographic word

*rata* orthographic word

* **Multi-word expression:** A string of words (two or more words) that is considered to be one lexical unit. Various multi-word expressions can be considered, including idioms, collocations, separable verbs, etc.
* **Part-of-speech tagging:** Generally, part-of-speech tagging takes the orthographic distinct unit as basis for tagging. For conjunctively written languages, this is straightforward. For disjunctively written languages, the term ‘part-of-speech tagging’ is extended to not only include tagging of linguistic words, but also the tagging of orthographic words. Morphological tagging is thus done parallel to word class tagging for disjunctive languages.
* **Punctuation:** “[Symbols](http://en.wikipedia.org/wiki/Symbols) which indicate the structure and organization of [written language](http://en.wikipedia.org/wiki/Writing), as well as [intonation](http://en.wikipedia.org/wiki/Intonation_(linguistics)) and pauses to be observed when reading aloud.” (<http://en.wikipedia.org/wiki/Punctuation>). Typical punctuation marks include the full-stop, comma, bracket/parenthesis, colon, semi-colon, dash, ellipsis, exclamation mark, question mark, quotation mark (single or double), slash/back-slash, and various other symbols.

# General principles

The following general principles apply:

* Annotation on the part-of-speech layer should be in excess of 98%.
* The tag set below is based on:
  + The EAGLES guidelines (<http://www.ilc.pi.cnr.it/EAGLES/home.html>);
  + An article by E .Taljard, G. Faaß, U. Heid & D.J. Prinsloo, 2008. On the development of a tagset for Northern Sotho with special reference to the issue of standardization. Literator 29(1), 1-27.
* The general principles of the Text Encoding Initiative P5 Guidelines (TEI P5; <http://www.tei-c.org/Guidelines/>) should be followed when in doubt.

## Mapping the tagset to EAGLES guidelines

Applicable EAGLES categories:

N = Noun

V = Verb

A = Adjective

P = Pronoun

B = Adverb

C = Conjunction

T = Numeral

W = Interjection

U = Unique/Unspecified

R = Residual

Z = Punctuation

The current POS categories (POS tags not applicable to isiNdebele, but applicable to other South African languages, are rendered in *italics* and in grey):

|  |  |  |
| --- | --- | --- |
| 1 | PUNC | Punctuation |
| 2 | ABBR | Abbreviation (incl. acronyms) |
| 3 | ADJ | Adjective (incl. enumerative) |
| 4 | ADV | Adverb |
| *5* | *ASP* | *Aspectual marker* |
| *6* | *AUX* | *Auxiliary stem* |
| 7 | CDEM | Class-indicating demonstrative |
| *8* | *CN* | *Class-indicating nominal prefix* |
| *9* | *CO* | *Class-indicating object concord* |
| 10 | CONJ | Conjunction |
| 11 | COP | Copulative (copulative subject concord, demonstrative copulative, copulative verb) |
| *12* | *CS* | *Class-indicating subject concord* |
| 13 | FOR | Foreign |
| 14 | IDEO | Ideophone |
| 15 | INT | Interjection |
| 16 | INTER | Question word |
| *17* | *MNEG* | *Negative morpheme* |
| 18 | N | Noun |
| 19 | NPP | Place and brand name |
| 20 | NUM | Numerative |
| *21* | *PART* | *Particle* |
| 22 | POSS | Possessive (possessive concord, possessive pronoun) |
| 23 | PROEMP | Emphatic pronoun |
| 24 | PROQUANT | Quantitative pronoun |
| 25 | REL | Relative |
| *26* | *TENS* | *Tense marker* |
| 27 | V | Verbal |
| 28 | VAUX | Auxiliary verb |

Combined / mapped Eagles and isiNdebele categories are:

|  |  |  |
| --- | --- | --- |
| **EAGLES** | **isiNdebele POS tags** |  |
| N = Noun | 18 and 19 as subcategory |  |
| V = Verb | 28 and 29 as subcategory |  |
| A = Adjective | 3 |  |
| P = Pronoun | 23, 24 and 25 |  |
| B = Adverb | 4 |  |
| K = Conjunction | 10 |  |
| T = Numerals | 20 |  |
| W = Interjection | 15 |  |
| U = Unique | 7, 11, 22, (5, 8, 9, 12, 17, 21, 27) |  |
| Z = Punctuation | 1 |  |

# Technical specifications

* All documents and annotation layers should be saved in UTF-8 format.
* Depending on the annotation software, XML encoding should be used. This will be determined by the software developers; annotators should in general not be concerned about appropriate syntax for the XML encoding.

# Authoritative references

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

* Jiyane, D.M. 1994. Aspects of isiNdebele grammar. M.A. thesis, University of Pretoria.
* Khumalo, L. 2003. An Introduction to Ndebele grammar. Cape Town: Centre for Advanced Studies of African Society.

# Changing the protocol

Since this protocol is considered 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 project manager (in consultation with the project leader) has the final say in changing this procedure document, and is the only authoritative person who might change the protocol.
* If an annotator (or other project member) wishes to change a specification or add examples to the protocol, s/he should inform the project manager by email to [Sunny.Gent@nwu.ac.za](mailto:Sunny.Gent@nwu.ac.za). The project manager considers the request, decides on the required actions and if needed 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.

# Tagset

For purposes of annotators, this tag set is by and large taken over from Taljard et al. (2008) and various documents compiled by G. Faasz and U. Heid from the IMS, Stuttgart and D.J. Prinsloo and E. Taljard, University of Pretoria. The information below refers to the current state of the tagset, but further development will probably necessitate any number of changes.

The tagset is mainly based on the lexical and morphological criteria defined by Lombard (1985) and Louwrens (1991). The logical structure of the tagset is divided into two layers of linguistic description (annotation levels):

The first annotation level (level 1) includes all mandatory, or, according to EAGLES, obligatory information, namely up to three elements: an element hinting at the word class, a second one specifying functional or syntactic properties, and a third one giving morphological specifics, cf. e.g. *PRO(noun)EMP(hatic)PERS(on)*.

The second level of annotation (level 2) includes recommended and optional information. This level is in most cases used for a detailed description of closed class items described in the tagger lexicon. Compare the following excerpt:

**Figure 1: Annotation levels**

|  |  |  |
| --- | --- | --- |
| **Description** | **Tag 1st level (mandatory information)** | **Tag 2nd level (optional/ recommended information)** |
| **Pronouns:** |  |  |
| emphatic personal | PROEMPPERS | 1sg,2sg,1pl,2pl |
| **Verbals:** | V | tr |
| **Morphemes:** |  |  |
| deficient | MORPH | def |

For disjunctive languages, next to all orthographic words, all linguistic words will also be tagged, resulting in two layers of POS annotation: one for all orthographic words and one for all linguistic words. For conjunctive languages, this extra layer of POS annotation is not needed.

The tagset currently distinguishes 20 categories applicable to isiNdebele and two different levels of annotation. However, only level 1 has been annotated. The first part of the tag gives a general indication of the nature of the unit in question. These are as follows:

|  |  |
| --- | --- |
| **Tag** | **Explanation** |
| PUNC | Punctuation |
| ABBR | Abbreviation (incl. acronyms) |
| ADJ | Adjective (incl. enumerative) |
| ADV | Adverb |
| CDEM | Class-indicating demonstrative |
| CONJ | Conjunction |
| COP | Copulative (copulative subject concord, demonstrative copulative, copulative verb) |
| FOR | Foreign |
| IDEO | Ideophone |
| INT | Interjection |
| INTER | Question word |
| N | Noun |
| NPP | Place and brand name |
| NUM | Numerative |
| POSS | Possessive (possessive concord, possessive pronoun) |
| PROEMP | Emphatic pronoun |
| PROQUANT | Quantitative pronoun |
| REL | Relative |
| V | Verbal |
| VAUX | Auxiliary verb |
|  |  |
|  |  |
| ***Tags not applicable to IsiNdebele*** | |
| *ASP* | *Aspectual marker* |
| *AUX* | *Auxiliary stem* |
| *CN* | *Class-indicating nominal prefix* |
| *CO* | *Class-indicating object concord* |
| *CS* | *Class-indicating subject concord* |
| *MNEG* | *Negative morpheme* |
| *PART* | *Particle* |
| *TENS* | *Tense marker* |

## PUNCTUATION

Level 1: PUNC

### ****Notes:****

* The tag PUNC is used for all punctuation marks. These include full stops, commas, colons, semi-colons, quotation marks, hyphens, exclamation marks, brackets, etc.

### Examples:

|  |  |
| --- | --- |
| ; | PUNC |
| ( | PUNC |
| ! | PUNC |
| “ | PUNC |

## ABBREVIATION

Level 1: ABBR

### ****Notes:****

* All abbreviations are tagged as ABBR. This also includes acronyms, like SONA (State of Nation Address) or NGO (non-governmental organisation).

### Examples:

|  |  |
| --- | --- |
| isib | ABBR |
| NGO | ABBR |

## ADJECTIVE

Level 1: ADJ01-11, ADJ 14-15, ADJ01a, ADJ02a, ADJLOC

### ****Notes:****

* The tag ADJ is used to tag linguistic units that consist of a noun class prefix and an adjective stem.
* The tag ADJ also includes enumeratives.

### Examples:

|  |  |
| --- | --- |
| omunye | ADJ01 |
| elikhulu | ADJ05 |
| komunye | ADJLOC |

## ADVERB

Level 1: ADV, ADVLOC

### Notes:

* Part-of-Speech tagging is about the function a word fulfils in a sentence, its grammatical context. Therefore, the tag ADV is applied if a word is used in the function of an adverb in the context of a sentence.
* Nouns belonging to the locative classes are tagged as ADVLOC, as their function is that of adverb.

### Examples:

|  |  |
| --- | --- |
| kanye | ADV |
| phambili | ADV |
| engaphasi | ADVLOC |

## [CLASS-INDICATING] DEMONSTRATIVE

Level 1: CDEM01-11, CDEM14-15, CDEMLOC

### Notes:

* Currently, no distinction is made for the different ‘positions’ or ‘distances’ of the demonstratives. Only noun class information is included in the tag.
* The demonstratives of the locative classes (16, 17, 18) are tagged as CDEMLOC.
* The demonstratives of classes 1a and 2a are tagged as CDEM01 and CDEM02 respectively.

### Examples:

|  |  |
| --- | --- |
| labo | CDEM02 |
| loyo | CDEM03 |
| lapho | CDEMLOC |

## CONJUNCTION

Level 1: CONJ

### Notes:

* Conjunctions that consist of multi-word units are not at this stage tagged as CONJ, the tokens making up the conjunctions are tagged separately.

### Examples:

|  |  |
| --- | --- |
| namkha | CONJ |
| ukuba | CONJ |

## COPULATIVE

Level 1: COP

Level 2: COP\_neg, COP\_nil

### Notes:

* The tag COP includes copulative subject concord, demonstrative copulative and copulative verb occurrences.
* The tag COP is used to tag copulative verbs stems, e.g. –*le*, -*se*, *-(e)na* and *–(e)ba*

(-*be*, -*bê* and –*bilê*). For the copulative verb stem –*se* the tag COP\_neg on level 2 is used, as is the case for the verb stem –*be* (<-ba) when it is used in the negative form.

### Examples:

|  |  |
| --- | --- |
| yiKomidi | COP |
| kube | COP |

## FOREIGN

Level 1: FOR

### Notes:

* Linguistic units taken from a foreign language are tagged as FOR.

### Examples:

|  |  |
| --- | --- |
| provincial | FOR |
| systems | FOR |

## IDEOPHONE

Level 1: IDEO

### Examples:

|  |  |
| --- | --- |
| godu | IDEO |
| yeke | IDEO |

## INTERJECTION

Level 1: INT

Level 2: INT\_neg, INT\_nil

### Notes:

* On level 2, the tag INT\_neg is used to tag the negative interjection *hhayi, cha*.

### Examples:

|  |  |
| --- | --- |
| na | INT |
| nekomo | INT |

## INTERROGATIVES

Level 1: INTER

Level 2: \_man, \_time, \_loc, \_N01a, \_N02a

### Notes:

* In some grammars, the items *njani, nini* and *kuphi* are described as adverbs, but here they are tagged as question words, asking about manner (\_man), time (\_time) and place (\_loc) (locality) respectively.
* Since the interrogatives *ubani* and *obani* can generate agreement of classes 1a and 2a respectively, they are tagged as INTER\_N01a and INTER\_N02a respectively.
* Information on class numbers is included in the tag for all forms with –*fe*.
* Note that when –*ngaki* is used with the meaning of ‘how many / much’ it appears in an adjectival construction and is consequently tagged as ADJ. Only when it is used with the meaning of ‘where’ is it tagged as INTER(\_loc).
* Note that when –*njani* is used with the meaning of ‘what kind of’ it appears in an adjectival construction and is consequently tagged as ADJ. Only when it is used with the meaning of ‘how’ is it tagged as INTER\_man.

### Examples:

|  |  |
| --- | --- |
| na | INTER |
| bunjani | INTER |
| mangaki | INTER |

## NOUN

Level 1: N01-11, N14-15, N01a, N02a, NLOC, N00

Level 2: \_aug, \_dim, \_loc, \_name, \_nil

### Notes:

* Although 17 noun classes plus two subclasses (1a and 2a) are distinguished for isiNdebele, only 14 noun tags are used. All locative classes (16, 17, 18, *N*- and *kwa*-) are tagged as NLOC.
* Names of people are tagged as N01a / N02a(\_name).
* N00 is used for lexicalised nouns.
* It is not necessary to include information regarding singular and plural in the tag, since this information is contained in the class number. Generally, nouns with an uneven class number (1, 3, 5, 7, 9) are singular nouns, whereas those belonging to the evenly numbered classes (2, 4, 6, 8, 10) are plural nouns.
* The distinction singular/plural is not relevant for the locative classes.
* Apart from containing information on noun class numbers, level 2 tags for nouns also include the features augmentative (\_aug), diminutive (\_dim), locative (\_loc), name (\_name) and even a combination of these, e.g. \_dim\_loc. These appear after the underscore ‘\_’ in the tag. If none of these is present, the absence of such a feature should be indicated by \_nil.

### Examples:

|  |  |
| --- | --- |
| umuntu | N01 |
| abomma | N02a |
| imiphumela | N04 |
| iphrojekthi | N05 |
| amalunga | N06 |
| isiqhema | N07 |
| indawo | N09 |
| mayelana | N00 |
| emsebenzini | NLOC |

## PLACE AND BRAND NAME

Level 1: NPP

Level 2: NPP\_place, NPP\_brand

### Notes:

* If level 2 annotation are used, place names are tagged as NPP\_place, brand names as NPP\_brand

### Examples:

|  |  |
| --- | --- |
| KwaZulu-Natal | NPP |
| Mars | NPP |

## NUMERATIVE

Level 1: NUM

### Notes:

* For all numbers and numerals (23, 1998, iii, etc.) the tag NUM is used.

### Examples:

|  |  |
| --- | --- |
| 2.2 | NUM |
| 2005 | NUM |
| 74(a) | NUM |

## POSSESSIVE

Level 1: POSS01-11, POSS14-15, POSSLOC, POSSPERS, POSSKA

Level 2: POSSPERS\_1pl, POSSPERS\_2pl

### Notes:

* The tag POSS includes possessive concord and possessive pronoun.
* Even though the possessive pronouns of the noun classes of isiNdebele are morphologically identical to the emphatic pronouns, their status as possessive pronouns is signalled by the possessive concord preceding them.
* Although the status of forms such as *ekhethu*, *ekhabo* and *ekhenu* is uncertain, these forms are tagged as PROPOSSPERS\_1pl, PROPOSS02 and PROPOSSPERS\_2pl respectively, since they are almost always preceded by a possessive concord

### Examples:

|  |  |
| --- | --- |
| wephrojekhti | POSS01 |
| sokutlama | POSS07 |
| kamasipala | POSSKA |

## EMPHATIC PRONOUN

Level 1: PROEMP01-11, PROEMP14-15, PROEMPLOC, PROEMPPERS

Level 2: PROEMPPERS\_1sg, PROEMPPERS\_1pl, PROEMPPERS\_2sg, PROEMPPERS\_2pl

### Notes:

* Emphatic pronouns are also called absolute pronouns in some grammars.
* Even when these pronouns appear in collocation with their corresponding nouns, they are tagged as PROEMP, even though in such cases they have strictly speaking, no pronominal function.

### Examples:

|  |  |
| --- | --- |
| bona | PROEMP02 |
| kizo | PROEMPLOC |
| khona | PROEMP15 |

## QUANTITATIVE PRONOUN

Level 1: PROQUANT01-11, PROQUANT14-15, PROQUANTLOC

### Notes:

* Even when these pronouns appear in collocation with their corresponding nouns, they are tagged as PROQUANT, even though in such cases they have strictly speaking, no pronominal function.

### Examples:

|  |  |
| --- | --- |
| boke | PROQUANT02 |
| zoke | PROQUANT10 |
| koke | PROQUANT15 |

## RELATIVE

Level 1: REL

### Notes:

* This POS tag can occur with a relative stem or with a relative prefix.

### Examples:

|  |  |
| --- | --- |
| angeze | REL |
| elibanzi | REL |
| esingaba | REL |

## VERBAL

Level 1: V

Level 2: V\_tr, V\_itr, V\_dtr

### Notes:

* The tag V is used to tag what is generally termed verb stems.
* Prefixes (reflexive prefix, object concord first person singular, object concord class 1 prefixed to verbs stems with initial consonant *b*-, stabilizing prefix *e*- with monosyllabic stems) and suffixes that are written conjunctively to the verb root or stem are not at this stage tagged separately.
* On level 2, the degree of transitivity of verbs is indicated as V\_tr (transitive), V\_itr (intransitive), V\_dtr (double transitive).
* Categories such as tense, mood and aspect are not indicated, since these features cannot be assigned automatically by computational means.

### Examples:

|  |  |
| --- | --- |
| babe | V |
| ukwakha | V |
| inikelwe | V |

## AUXILIARY VERB

Level 1: VAUX

Level 2: VAUX\_tr, VAUX-itr, VAUX\_dtr

### Notes:

* The tag VAUX is used to tag auxiliary verbs – in contrast with auxiliary verb stems for disjunctive languages.

### Examples:

|  |  |
| --- | --- |
| ibe | VAUX |
| ukungabi | VAUX |

# Appendixes

## Roles and contact details

The following roles in this project can be identified:

* 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. Also responsible for the protocol document. If you have any questions regarding language specific requirements and specification, contact the project manager.
  + Sunny Gent: [Sunny.Gent@nwu.ac.za](mailto:Sunny.Gent@nwu.ac.za), 081 424 9715.
* Annotator: Responsible for annotating the corpus, according to the above procedures and principles.

# Tags not applicable to isiNdebele

NB: All tags following are not used in the POS tagging of isiNdebele!

## ASPECTUAL MARKER

Level 1: ASP *Not applicable for isiNdebele*

Level 2: ASP\_pot, ASP\_prog, ASP\_nil

### Notes:

* The deficient verbs forms, also called deficient auxiliary verb forms, *-ya, -no, -dio, -yo* and -z*o* are tagged as ASP.
* On the second level of annotation the progressive morpheme *sa* and the potential morpheme *nga* are tagged as ASP\_prog and ASP\_pot respectively.

## AUXILIARY

Level 1: AUX *Not applicable for isiNdebele – see VAUX*

### Notes:

* The tag AUX is used to tag what is generally termed auxiliary verb stems.
* See tag 29 VAUX for full auxiliary verbs.

## [CLASS-INDICATING] NOMINAL PREFIX

Level 1: CN15 *Not applicable for isiNdebele*

### Notes:

* Currently, this tag is only used for tagging the noun class prefix of class 15, which is written disjunctively from the stem following it.

## [CLASS-INDICATING] OBJECT CONCORD

Level 1: CO01-10, CO14-15, COLOC, COPERS *Not applicable for isiNdebele*

Level 2: COPERS\_1pl, COPERS\_2pl, COPERS\_2sg

### Notes:

* Noun class information is included in the tag. Tags for first and second person distinguish between singular and plural.
* Object concords that are written conjunctively to the verb stem (e.g. object concord 1st person singular, object concord of class 1 *mo*- preceding a verb stem with initial consonant b-) are not at this stage tagged.
* The object concords of classes 1a and 2a are tagged as CO01 and CO02 respectively.
* Object concords of the locative classes are tagged as COLOC.

## [CLASS-INDICATING] SUBJECT CONCORD

Level 1: CS01-10, CS14-15, CSLOC, CSINDEF, CSNEUT, CSPERS *Not applicable for isiNdebele*

Level 2: CSPERS\_1sg, CSPERS\_1pl, CSPERS\_2sg, CSPERS\_2pl

### Notes:

* No distinction is currently made between the unmarked form of the subject concord, and the co-called ‘consecutive form’ thereof.
* The tag CSINDEF is used to tag the indefinite subject concord *go / gwa*
* The tag CSNEUT is used to tag the neutral subject concord *e / ya*, which is inter alia often found in identifying copulatives.
* Subject concords of the locative classes are tagged CSLOC.
* Subject concord of nouns in classes 1a and 2a are tagged as CN01 and CN02 respectively.

## NEGATIVE MORPHEME

Level 1: MNEG *Not applicable for isiNdebele*

### Notes:

* Negative morphemes *akusi* is tagged as MNEG.

## PARTICLE

Level 1: PART *Not applicable for isiNdebele*

Level 2: PART\_cop, PART\_agen, PART\_hort, PART\_loc, PRT\_que, PART\_temp, PART\_ins, PART\_con

### Notes:

* The category ‘particle’ is a contentious one and is not recognized by all grammarians. The term ‘prefix’ is often used to refer to members of this category.
* The category ‘particle’ is a closed category.
* The tag PART\_cop is used for the copulative particle *ngi* ‘it/he/she is, they are’. The negative form of this copula *akasi* is tagged as @@.
* The agentive particle *a* found following after passive verbs is tagged as PART\_agen.
* The hortative particle *a* is tagged as PART\_hort.
* Five locative particles (PART\_loc) are distinguished for Northern Sotho, i.e. *ka*, *lapha*, *ko*, *le* and *kwa*.
* The question particles (PART\_que) are *alo, nini, ini, yi*.
* The temporal particle (PART\_temp) is used to indicate a point in time.
* The instrumental particle (PART\_ins) is *nga*.
* The connective particle *ni* which is typically found following after the copulative verb stem *–na* is tagged PART\_con

## TENSE MARKER

Level 1: TENSE *Not applicable for isiNdebele*

Level 2: TENSE\_fut, TENSE\_pres, TENSE\_past, TENSE\_neg

### Notes:

* The future tense morphemes –*zoku* and *-yoku* are tagged as TENSE\_fut.
* When the morpheme –*a* … *ka* appears together with the negative morpheme *se* it functions as a marker of the future tense, and in such cases is marked as TENSE\_fut.
* The past tense morpheme –*a* only occurs in a specific negative form of the so-called past tense and is the only past tense morpheme in Northern Sotho. It is tagged as TENSE\_past.
* The so-called present tense morpheme –*ya* which appears only in (some) present tense verbs is marked as TENSE\_pres.