A writing assistant en route to a full computational grammar for Sepedi

DJ Prinsloo* iD **& Daniel Prinsloo**

Department of African Languages, University of Pretoria, South Africa
Correspondence: danie.prinsloo@up.ac.za

A detailed user study and observations by lecturers indicate that the correct compilation of sentences in the eight verbal moods as well as for a number of similar grammatically complicated constructions in Sepedi poses a challenge in any text production situation. Feedback from target users indicated that there is a need for a computational writing assistant for the compilation and verification of correct constructions. To address this need, an extended computational sentence builder for verbal moods, adjectives and possessive constructions in Sepedi was designed and built to assist in text production and to serve as a model for other African languages. This article introduces a prototype of such an extended computational sentence builder for verbal moods, adjectives and possessive constructions in Sepedi. The emphasis throughout is on the grammatical complexity of Sepedi and how the Sepedi Helper can assist users to produce correct sentences. In contrast to typical traditional grammars of Sepedi, the tool also provides the required cognitive information through basic clickable help screens. The Sepedi Helper is a dynamic lightweight tool aimed at combining user knowledge with a text production tool, i.e. user-involved, step-by-step production of Sepedi phrases. The emphasis in the design is on simplicity.

Introduction

Writing assistants, whether designed as standalone software or linked to dictionaries or word processing software, fulfil an important role in the compilation of texts in, for example, language learning and in any text production situation such as found in a dictionary use situation. This perceived need observed by, for example, lecturers and teachers is scientifically substantiated by user studies. In the case of Sepedi, two formal user studies were conducted on the need for writing assistants in what is probably the most complicated grammatical constructions in Sepedi, i.e. verbal moods and copulatives. The need for and value of these writing assistants were clearly indicated by these user studies and are briefly discussed. Detailed discussions of such support systems are given in Prinsloo et al. (2017) and Bothma et al. (2018). The most prominent writing tools include *Interactive Language Toolbox* (Verlinde and Peeters 2012; https://ilt.kuleuven.be/inlato/), the Louvain EAP Dictionary (LEAD) (Paquot 2012; https://leaddico.uclouvain.be) and *Schrijfhulp Portaal* (https://ilt.kuleuven.be/schrijfhulp/). Typical examples for African languages include an assistant for the compilation of isiZulu possessives (Bosch and Faaß 2014), decision trees for the copulative in Sepedi (Bothma et al. 2013), guidance paths on kinship relations (Prinsloo and Bosch 2012), decision trees for colour terms in Sepedi (Taljard and Prinsloo 2013; Bothma et al. 2018). A comparison between these writing tools is regarded as future work – it falls outside the aims of the current article and cannot be accommodated in the space limitation.

This article represents an extended implementation of the Sepedi Helper (Prinsloo et al. 2015; http://sepedihelper. co.za/) as a follow-up to verbal relative constructions as presented at eLex 2011 (Prinsloo et al. 2011), eLex 2013 (Bothma et al. 2013; Prinsloo et al. 2017).

This article is also the first complete description of the Sepedi Helper. The new beta version of the Sepedi Helper (a) covers all eight of the verbal moods, (b) offers a description of pronominalisation, (c) incorporates the Sepedi adjectives, and (d) enables the building of possessive constructions. The focus is on the linguistic description of the relevant constructions but a discussion of programming aspects as well as the findings of a user study will also be presented. The compilation of the Sepedi Helper (www. sepedihelper.co.za) is a work in progress at an advanced stage of completion en route to a full and comprehensive computational grammar for Sepedi. Cross-references to relevant issues regarding the Sepedi Helper discussed in Prinsloo et al. (2017) will be given, but these issues will not be discussed in detail in this article. A concise summary of certain required aspects also dealt with in Prinsloo et al. (2017) will, however, be given as a point of departure.

The Sepedi Helper is a writing tool that users can directly access for cognitive information about moods, pronouns, adjectives and possessive constructions. It can also be linked to a word processing programme with a clickable icon or to an e-dictionary to obtain information on demand during a look-up procedure.

The compilation of a writing assistant requires close cooperation between linguist and programmer. The role of the linguist is to act as mediator between a complicated grammar and the computer programmer. The task of the linguist is to write the grammar rules in such a way that the programmer can understand them and design the computer software. We refer to such rules as linguistic algorithms. The Sepedi Helper falls within the field of study of African language linguistics and focuses on problematic Sepedi grammatical structures. It has relevancy for computational linguistics (see

the section on the description of the programming aspects) as well as computer-assisted language learning (see section on user studies). The latter disciplines, however, do not form the focus of this article.

Linguistic algorithms for computational implementation

The approach taken for the Sepedi Helper was the compilation of a linguistic algorithm that summarises the information given for selected constructions in Sepedi grammars such as those of Louwrens (1991), Poulos and Louwrens (1994), Lombard et al. (1985) and Ziervogel (1976). Grammars as such cannot be directly utilised in the compilation of an e-grammar, and grammar rules have to be presented as a computational morphological and syntactical analysis. Compare in this regard Faaß (2010) and Berg (2018) for typical approaches, in a format also understandable to the human programmer in order to write the computer software package. For the Sepedi Helper, the systems for verbal moods, nominal classes, concords, pronouns, adjectives and possessives were presented as linguistic algorithms, and these algorithms were in the end also provided in the help screens, as will be discussed below.

User studies on the Sepedi Helper and copulative decision tree

A user study was done in November 2018 at a tertiary institution on 81 first- and second-year students who do not have Sepedi as their strongest or home language. These students represent the ideal target user group since the Sepedi Helper is in principle designed to combine existing knowledge of the user with the capabilities of the Sepedi Helper. This simply means that users with little or even no knowledge of Sepedi as well as advanced users can construct sentences. Users with no knowledge of Sepedi can provide the required information in English and the Sepedi Helper will provide a full and correct Sepedi sentence. Users with a little knowledge can construct Sepedi sentences and the Sepedi Helper will be able to help them correct grammatical errors. Advanced students who manage to construct correct sentences on their own get confirmation from the Sepedi Helper that the sentences were indeed correctly compiled.

In the user study, students first had to compile sentences in the different moods without the assistance of a writing tool, and then with the tool. Positive feedback obtained from the respondents relate to the value of the Sepedi Helper in terms of practical assistance with building up grammatically correct sentences in the different moods, indicating the changes from one mood to another and confirmation of correctness where students managed to produce the correct sentence on their own. Analysis of the data indicated an 18.5% increase in the production of correct sentences and knowledge of the moods by the Sepedi Helper. This survey and its findings are described in great detail in Prinsloo and Taljard (2019). A similar user study was done for copulative constructions (Prinsloo 2020) which indicated an even more significant increase of 27.9% in the production of correct copulatives. These studies clearly indicate the need for and value of writing assistants for grammatically complicated constructions in Sepedi.

Computational implementation of verbal moods in Sepedi

In the following sections, a brief discussion of the crucial grammatical features of verbal moods, pronominalisation, adjectives and possessives underpinning the Sepedi Helper will be given. Grammatical features covered for the moods are present, future, past; positive, negative, intransitive, transitive, double transitive; nominal forms, subject and object concords and pronouns. Figures 1 and 2 summarise the basic features of nouns and verbs required for building sentences in the Sepedi Helper.

Table 1 presents a summary of all prefixes, concords and pronouns required for the verbal moods, pronominalisation,

Table 1: Noun class system of Sepedi with prefixes, concords and pronouns utilised in the Sepedi Helper

Person/noun class	Example	Class prefix	Subject concords 1	Subject concords 2	Object concords	Demon-stratives	Possessive concords	Emphatic pronouns
1st person singular			*ke*	*ka*	*n-*			*nna* 'I'
1st person plural			*re*	*ra*	*re*			*rena* 'we'
2nd person sing.			*o*	*wa*	*go*			*wena* 'you' (singular)
2nd person plural			*le*	*la*	*le*			*lena* 'you' (plural)
Class 1	*mosadi* 'woman'	*mo*	*o/a*	*a*	*mo*	*yo*	*wa*	*yena*
Class 2	*basadi* 'women'	*ba*	*ba*	*ba*	*ba*	*ba*	*ba*	*bona*
Class 3	*molao* 'law'	*mo*	*o*	*wa*	*o*	*wo*	*wa*	*wona*
Class 4	*melao* 'laws'	*me*	*e*	*ya*	*e*	*ye*	*ya*	*yona*
Class 5	*leoto* 'foot'	*le*	*le*	*la*	*le*	*le*	*la*	*lona*
Class 6	*maoto* 'feet'	*ma*	*a*	*a*	*a*	*a*	*a*	*ona*
Class 7	*setulo* 'chair'	*se*	*se*	*sa*	*se*	*se*	*sa*	*sona*
Class 8	*ditulo* 'chairs'	*di*	*di*	*tša*	*di*	*tše*	*tša*	*tšona*
Class 9	*nko* 'nose'	*n-*	*e*	*ya*	*e*	*ye*	*ya*	*yona*
Class 10	*dinko* 'noses'	*di*	*di*	*tša*	*di*	*tše*	*tša*	*tšona*
Class 14	*bohlale* 'wisdom'	*bo*	*bo*	*bja*	*bo*	*bjo*	*bja*	*bjona*
Class 15	*go bala* 'to read/count'	*go*	*go*	*gwa*	*go*		*ga*	*gona*
Class 16	*fase* 'below'	*fa*	*go*	*gwa*	*go*	*fa*	*ga*	*gona*
Class 17	*godimo* 'above'	*go*	*go*	*gwa*	*go*	*mo*	*ga*	*gona*
Class 18	*morago* 'behind'	*mo*	*go*	*gwa*	*go*	*mo*	*ga*	*gona*

```
            pos     monna o reka puku               |
      pres.                                          |
            neg.    monna ga a reke puku             |                  | ga   ..-e
1. INDICATIVE                                        | STATEMENTS
            pos     monna o rekile puku              |
      past                                           |
            neg     monna ga se a reka puku          |                  | ga se + cons. sc.+ pres
---------------------------------------------------------------------------------------------------
            pos     ge monna a reka puku             |
      pres.                                          |
            neg.    ge monna a sa reke puku          |                  | sa   ..-e
2. SITUATIVE                                         | IF/WHILE
            pos     ge  monna a rekile puku          |
      past                                           |
            neg     ge monna a sa reka puku          |                  | sa + pres
---------------------------------------------------------------------------------------------------
            pos     monna yo a rekago puku           |
      pres.                                          |
            neg.    monna yo a sa rekego puku        |                  | sa   ..-e
3. RELATIVE                                          | THAT (IS DOING)
            pos     monna yo a rekilego puku         |
      past                                           |
            neg     monna yo a sa rekago puku        |                  | sa + pres
---------------------------------------------------------------------------------------------------
            pos     ... (gore) monna a reke puku     |
4. SUBJUNCTIVE                                       | MUST (DO SOMETHING)
            neg     ... (gore) monna a se reke puku  |                  | se    ....-e
---------------------------------------------------------------------------------------------------
            pos     ... monna a reka puku            |
5. CONSECUTIVE                                       | AND THEN ...(DO/DID SOMETHING)
            neg     ... monna a se reke puku         |                  | se    ....-e
---------------------------------------------------------------------------------------------------
            pos     ... go reka puku                 |
6. INFINITIVE                                        | TO (DO SOMETHING)
            neg     ... go se reke puku              |                  | se    ....-e
---------------------------------------------------------------------------------------------------
            pos     reka puku!                       |
7. IMPERATIVE                                        | DO (SOMETHING)!
            neg     se reke puku!                    |                  | se    ....-e
---------------------------------------------------------------------------------------------------
            pos     ... monna a reka puku            |
8. HABITUAL                                          | USUALLY (DO SOMETHING)
            neg     ... monna a se reke puku         |                  | se    ....-e
---------------------------------------------------------------------------------------------------
```

Figure 1: Sepedi verbal moods and their basic meanings and structure

adjective, possessive and relative constructions in the Sepedi Helper.

Verbal moods in the Sepedi Helper

The main screen of the Sepedi Helper given in Figure 2 with the Class 1 noun *monna* 'man' chosen as a noun in this case is designed as a one-stop point of departure for all cognitive information, and step-by-step guidance in the compilation of sentences in the different verbal moods, and adjective and possessive constructions.

The first clickable button at the top of the screen provides basic instructions and the indicative mood as a point of departure for constructing sentences in the present, future or past tenses. Sepedi example sentences with English translations are given. Prominent on this screen is a clickable icon to start a 60-second demo video of how to use the Sepedi Helper. The screen displays the basic meaning of the mood currently selected in capital letters, i.e. 'statements' in this case and the user can immediately get the basic meanings of the other moods by simply selecting them from the sentence type button. This is followed by a sentence type button and just below the button the meaning of the selected mood is given in capital letters, e.g. in this case 'STATEMENTS' for the 'Indicative'. Clicking on the sentence type button produces a drop-down menu giving a very brief summary of the meanings of the different moods (Figure 3).

If the user clicks on the question mark on the right-hand side, a fairly comprehensive summary of the semantic and morphological characteristics of the mood currently showing in the window is given, i.e. for 'Indicative' in this case (Figure 4).
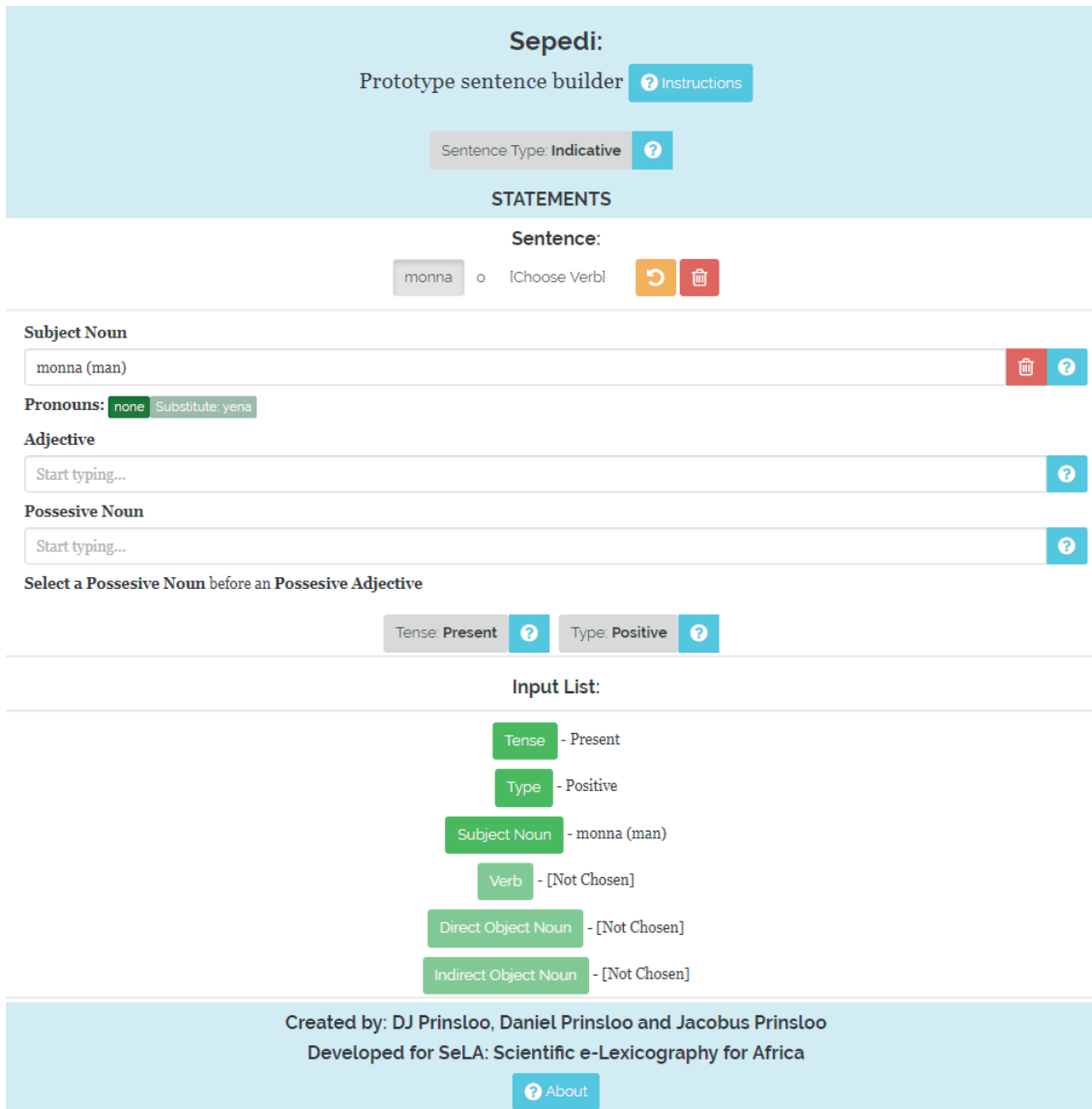
Figure 2: A section of the main screen of the Sepedi Helper (sepedihelper.co.za)

The core of the Sepedi Helper lies in the following step headed by 'Sentence', i.e. 'Choose subject noun' and 'Choose Verb'. At this point, the default settings are for building an indicative sentence in the present tense positive. See Prinsloo et al. (2017) for a similar basic build-up for the relative mood. The user clicks on 'Choose Subject Noun'. A 'Subject Noun' box appears prompting the user to type in a noun. There is an indication that the pronoun setting is 'none' (see below). The user clicks on 'Start typing' and types in a word in English or in Sepedi if (s)he knows the Sepedi word. Say, for example, the user types in *man*, the system offers some assistance in a drop-down menu for words which include *man*. The user chooses *monna* and the sentence constructor shows the first step of the build-up, i.e. *monna* and its correct subject concord *o*. The same process is repeated for 'Choose Verb'. The system inserts the verb and 'knows' that – because at this point it is a sentence in the indicative without an object – the present tense marker *a* has to be inserted, thus the built-up result (Figure 5).
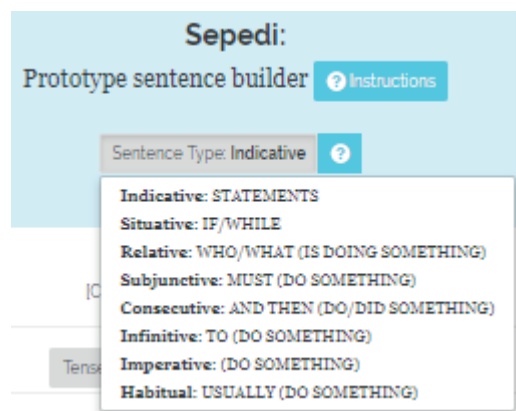
Figure 3: Brief indication of the meaning/semantic connotation of the verbal moods

4

Indicative

Indicative expresses: **statements**
Present

Most common structure: Noun + subject concord + verb + object
Example: **Positive:** Monna o reka puku. 'The man buys a book'
  Negative: Monna **ga** a rek**e** puku. 'The man does not buy a book'
  Negation strategy: **ga** … verb ends in **-e**

If nothing follows the verb: insert present tense marker "**a**":
  Monna o **a** reka 'The man buys'
In negative subject concord class 1 changes from o to **a**:
  Monna ga **a** reke puku
Can occur with direct and indirect object:
  Monna o rekela bana puku
   indirect object direct object
  'The man buys for the children a book.
Nouns can be replaced by **pronouns**:
Example (**bana**): Monna o rekela **bona** puku
Object nouns can also be replaced by object concords:
Example: Monna o **ba** rekela puku

Future
Most common structure: Noun + subject concord + fut. +verb + object
Example: **Positive:** Monna o tlo reka puku.
  'The man will buy a book'
  Negative: Monna a **ka se** rek**e** puku. 'The man will not buy a book'
  Negation strategy: **ka se** … verb ends in **-e**
In negative subject concord class 1 changes from o to **a**
The combination **ke** = I + **ka se** becomes **nka se**
  Nka se reke puku 'I shall not buy the book'
Past
Most common structure: Noun + subject concord + verb + object
Example: **Positive:** Monna o rekile puku. 'The man bought a book'
  Negative: Monna **ga se a reka** puku. 'The man did not buy a book'
  Negation strategy: **ga se** + **alternative concord** + **verb back to the present tense**
(Alt. concords: 1ps: ka, 1pp: ra, 2ps: wa, 2pp: la, cl.1: a, cl.2: ba, cl.3: wa, cl.4: ya, cl.5: la, cl.6: a, cl.7: sa, cl.8: tša, cl.9: ya, cl.10: tša, cl.14: bja, cl.15-18 and locatives: gwa)

Figure 4: Brief summary of the structure and meaning/semantic connotation of the indicative

Figure 5: Sentence in the present tense with the present tense marker *a*

The user is prompted to add an object, e.g. *Sepedi*. The system removes the present tense marker, and the result is an indicative sentence in the present tense, positive (Figure 6).

The system offers pronominalisation options for the subject noun and object noun and, in case of double transitive verbs, pronominalisation options for both the indirect object

Figure 6: Result of the build-up for an indicative sentence in the present tense, positive

noun and direct object noun – see detailed discussion on pronominalisation below.

The sentence in Figure 6 can now be changed to any of the other moods, in any tense, positive or negative by clicking and selecting items from the three grey-shaded buttons 'Sentence Type', 'Tense' and 'Type'. Consider for example the results of the selection of the situative, past tense negative in Figure 7.

The system provides the user with the conjunction *ge* which typically introduces the situative, but it is given in brackets to indicate that it can be omitted in some situative sentences. The system 'knows' that (a) the subject concord of class 1 changes from *o* in the indicative to *a* in the situative, (b) that the negative morpheme for the situative is *sa*, and (c) that the past tense form of the verb reverts back to the present tense form.

Pronominalisation

All aspects of pronominalisation by means of emphatic pronouns and object concords, including all instances of sound strengthening (Ziervogel 1976), also referred to as nasal strengthening, are handled by the Sepedi Helper. The sentence built by the Sepedi Helper in Figure 8 contains a
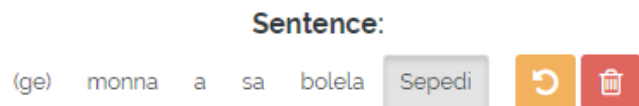
Sentence:

| (ge) | monna | a | sa | bolela | Sepedi | ↺ | 🗑 |

Figure 7: Result of instructions to change *monna o bolela Sepedi* to the situative, past tense, negative

Sentence:

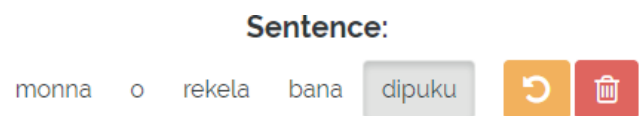| monna | o | rekela | bana | dipuku | ↺ | 🗑 |

Figure 8: A double transitive Sepedi sentence

subject noun (*monna* 'man') and two object nouns (*bana* 'children' and *dipuku* 'books'): *monna o rekela bana dipuku* 'the man buys books for the children/buys the children books'.

All three nouns in this sentence can be pronominalised simultaneously, e.g. *monna* and *dipuku* by means of an emphatic pronoun and *bana* by its object concord. This is achieved by selecting the appropriate options 'Substitute: yena', 'Insert ba', 'Substitute: bona', 'Insert: di' and 'Substitute: tšona' offered by the Sepedi Helper as in Figure 9. The result is given in Figure 10.

Use of the object concord of class 1 *mo-* and especially the first-person singular *n-* results in numerous instances of sound strengthening all correctly handled by the Sepedi Helper, see an extract of these rules given in Example (1).

(1a) Examples not reflecting any sound changes:
Madika o tseba nna 'Madika knows me' > *Madika o a ntseba* 'Madika knows me' (*n+tseba*)
Tau e tšhaba monna 'The lion is afraid of the man' > *Tau e a mo tšhaba* 'The lion is afraid of him'

(1b) Examples in which phonological changes occurred:
Madika o bona nna 'Madika sees me' > *Madika o a mpona* 'Madika sees me'
Tau e bona monna 'The lion sees the man' > *Tau e a mmona* 'The lion sees him'

(1c) Sound strengthening resulting from the use of the object concord of the first-person singular:
n+araba 'me+answer' > *nkaraba* (*a > ka*)
n+bona 'me+see' > *mpona* (*b > p* and *nb > mp*)
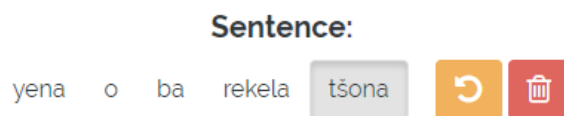*n+direla* 'me+work for' > *ntirela* (*d > t*)

Sentence:

| yena | o | ba | rekela | tšona | ↺ | 🗑 |

Figure 10: All three nouns pronominalised in a double transitive Sepedi sentence

Subject Noun

| monna (man) | 🗑 | ❓ |

Pronouns: none Substitute: yena

Indirect Object Noun

| bana (children) | 🗑 | ❓ |

Pronouns: none Insert: ba Substitute: bona

Direct Object Noun

| dipuku (books) | 🗑 | ❓ |

Pronouns: none Insert: di Substitute: tšona

Figure 9: Pronominalisation options offered for all three nouns in the Sepedi sentence

n+etela 'me+visit' > *nketela* (*e* > *ke*)
n+fora 'me+deceive' > *mphora* (*f* > *ph* and *nf* > *mph*)
n+gopotša 'me+remind' > *nkgopotša* (*g* > *kg*)
n+humiša 'me+enrich' > *nkhumiša* (*h* > *kh*)
n+hlompha 'me+respect' > *ntlhompha* (*hl* > *tlh*)
n+iša 'me+take' > *nkiša* (*i* > *ki*), etc.

The handling and presentation of adjectives

The construction of adjectives in Sepedi requires a four-part formula for classes 1–7 and 14–18. Consider Example (2) for classes 1–7 and 14:

(2) Formula: 1   2   3   4
           noun + demonstrative + class + adjective
                               prefix   stem
     Class 1   *monna yo mogolo* 'a big man'
     Class 2   *banna ba bagolo* 'big men'
     Class 3   *mmotoro wo mogolo* 'a big car'
     Class 4   *mebotoro ye megolo* 'big cars'
     Class 5   *lesogana le legolo* 'a big young man'
     Class 6   *masogana a magolo* 'big young men'*
     Class 7   *selepe se segolo* 'a big axe'
     Class 14 *bogobe bjo bošweu* 'white porridge'

Adjective constructions for classes 8 to 10 are subject to sound strengthening and follow a three-part formula as given in Examples (3) and (4).

(3) Sound strengthening
     -bedi > *-pedi* 'two' (*b* > *p*)   *-golo* > *-kgolo*
                                                 'big' (*g* > *kg*)
     -hubedu > *-khubedu* 'red' (*h* > *kh*) but *-hlano* > *-tlhano*
                                                 'five' (*hl* > *tlh*)
     -ne > *-nne* 'four' (*n* > *nn*)   *-raro* > *-tharo*
                                                 'three' (*r* > *th*)
     -so > *-ntsho* 'black' (*s* > *tsh*)   *-šweu* > *-tšhweu*
                                                 'white' (*š* > *tšh*)

(4) Formula: noun + demonstrative + strengthened stem
     Class 8   *dilepe tše kgolo* 'big axes'
     Class 9   *nku ye kgolo* 'big sheep'
     Class 10 *dinku tše kgolo* 'big sheep'

The step-by-step build-up of adjectives in the Sepedi Helper is similar to the construction of sentences described above. The user types a noun in English or Sepedi and types in an adjective in English or in Sepedi, e.g. *big*, and the result is given in Figure 11.

In the sentence builder line, the adjective *monna yo mogolo* is correctly constructed and the system also provides the correct subject concord for class 1, prompting the user to insert a verb if (s)he wishes to continue with the construction of a complete sentence (Figure 12).

Object nouns can also be extended in the same way to include adjectives and all nouns can be pronominalised simultaneously if the user would like to construct such a
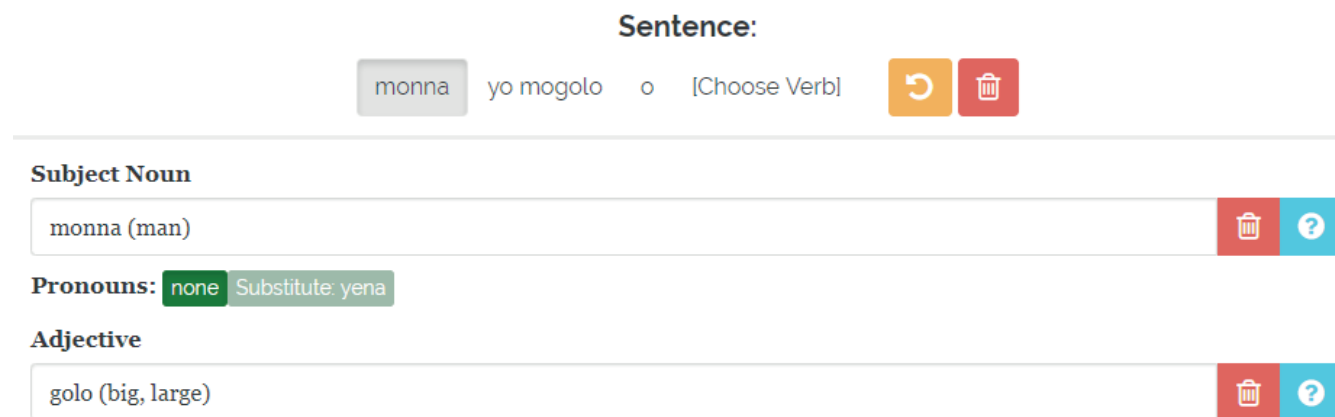
Sentence:

monna   yo mogolo   o   [Choose Verb]   ↺   🗑

Subject Noun

monna (man)   🗑   ❓

Pronouns: none   Substitute: yena

Adjective

golo (big, large)   🗑   ❓

Figure 11: Providing the basic noun and adjective to the Sepedi Helper

Sentence:

monna   yo mogolo   o   reka   puku   ↺   🗑

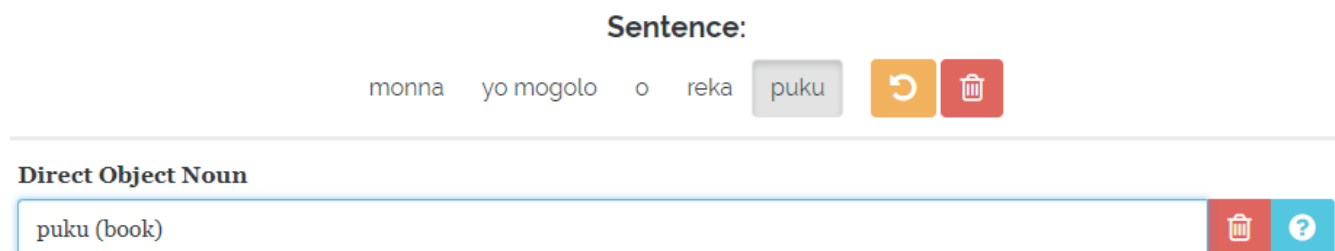**Direct Object Noun**

puku (book)   🗑   ❓

Figure 12: An adjective construction as the subject of a Sepedi sentence

complicated sentence. Consider, for example, *monna yo mogolo o rekela bana ba banyane dilo tše tšhweu* 'the big man buys white objects for the little children', pronominalised as *yena yo mogolo o ba rekela tšona tše tšhweu* 'the big one buys it for them' (Figure 13).

Possessive constructions

Possessive constructions are characterised by a three-part formula, as in Example (5) with the possessive concord generated by the first noun.

(5)   Formula for the possessive construction

1	2	3
noun +	possessive concord +	noun
lerato	*la*	*Modimo* 'love of God'
dipuku	*tša*	*ngwana* 'books of the child (the child's books)'
bohlale	*bja*	*mosadi* 'the wisdom of the woman'

The full set of possessive concords is given in Table 1.

The build-up process is similar to the one for the adjective. The user is prompted to provide the possession and the possessor, and the system selects the correct possessive concord. The possessive construction is given in the build-up line with the possibility of extending it to a full sentence, e.g. *ngaka ya kgoši e alafa balwetši* 'the king's doctor heals the patients' (Figure 14).

To see the power and accuracy of the Sepedi Helper, consider the creation of a complex double transitive Sepedi sentence in the indicative past tense negative with three nouns extended to include two adjective and two possessive constructions. The sentence *monna yo mogolo wa mošate ga se a balela banamedi molao wo motelele wa mmušo* 'the big man from the capital city did not read the lengthy governmental regulation to the passengers' can also be pronominalised for three nouns, which can then be translated as 'the big one from the capital city did not read it to them' (Figure 15).

The capabilities of the Sepedi Helper can be summarised in terms of Figure 15 as follows:
- The sentence can have an intransitive, transitive or double transitive verb with up to two objects;
- It can be presented in any of the eight verbal moods;
- It can be presented in any tense (present, future or past);
- Positive or negative forms can be constructed;
- All nouns can be pronominalised simultaneously, subject and direct object by means of the emphatic pronoun and the indirect object by means of object concord or
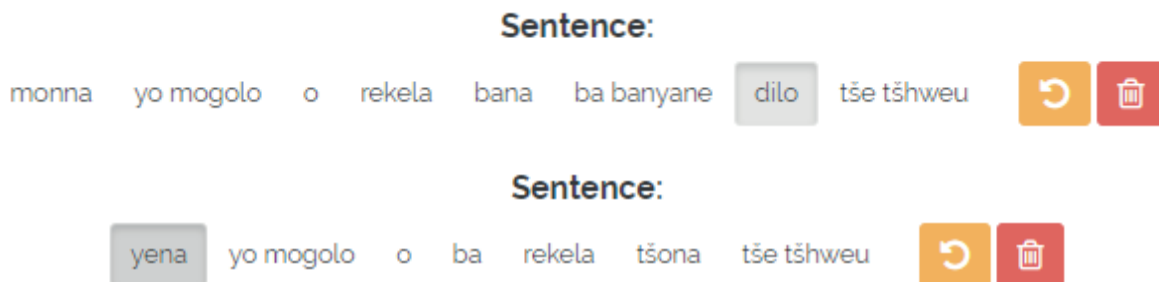

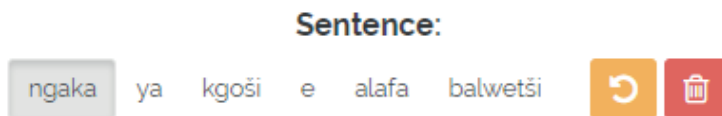**Figure 13:** Subject and object nouns occurring in adjective constructions

Figure 14: A possessive construction as the subject of a Sepedi sentence
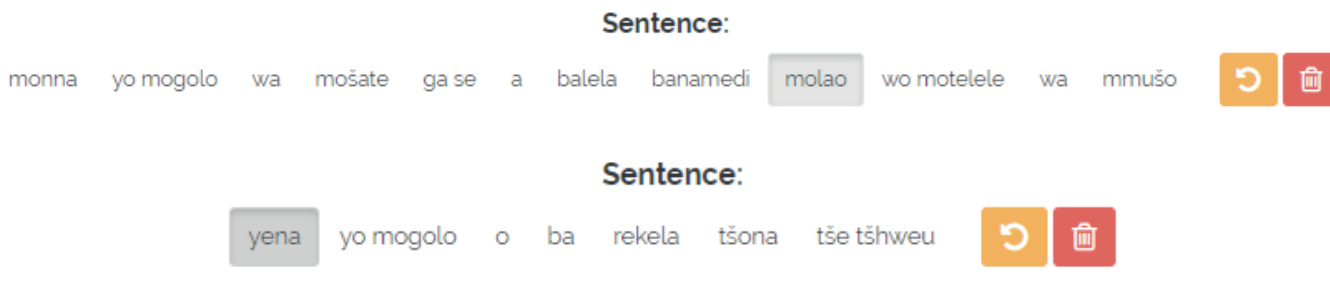

Figure 15: A complex Sepedi sentence compiled by the Sepedi Helper

emphatic pronoun;
- Subject and direct objects can be extended to adjective constructions; and
- Subject and direct objects can appear in possessive constructions.

Programming aspects

In layman's terms, this web application is meant as a tool or component to be used in e-dictionaries, word processors or as a standalone tool. The user can be an English speaker or a Sepedi speaker, and all Sepedi words are accompanied by English translations. The interface itself is in English. Guidance is given in English and includes video tutorials to guide the users.

The essential philosophy followed with the software implementation is in a sense like building a bridge from its opposite ends at the same time. The one end is a strong focus on user experience usability of the system as well as user-friendliness in making sure that the user is not overwhelmed by an enormous set of inputs. The language grammar is broken down into minimal decisions that a user has to make and expands as the user gives more information. This input information is sent as a request to the server which in turn sends back something like a 'deep' sentence output. The sentence has links back to its input and helps the user understand how a change in one word could affect many different parts of the sentence.

The server is the other end of the bridge and contains the pure linguistic functionality. The server uses word lists and linked information per word type and passes it through grammar rules as per the users' input. This is quite standard, yet the novelty and the success of the system relies on mapping the users' input optimised for user understanding to a format optimised for grammar rule processing and back.

The users' interface uses AngularJS v1 along with Bootstrap 3. The server uses PHP (a general-purpose scripting language) with nouns, verbs and adjectives stored in SqLite database. Other sentence structures and language constructs are defined directly in the server source code as this does not need to change, e.g. a list of subject concords per noun class.

The current version of the application should be able to integrate quite well with any online dictionaries and similar web applications as the web front-end was written as a large web component built from smaller components. However, at this time it cannot be integrated with Microsoft Word or the rest of the Office suite as well as with open-source alternatives like Libre-Office or any similar applications. The time and cost to do this versus the reward is still being investigated.

As mentioned, the main form of input data required are tagged wordlists. Nouns have an accompanying noun class that is critical in Sepedi linguistics and the grammar rules that the app uses. Additionally, as mentioned, all word list entries are accompanied by an English translation equivalent paradigm. The following is an example of such an entry: *batho* 'people, humans', N02. Verbs require more information, e.g. the verb itself, its translation equivalents, tense, transitivity and lastly, if past tense, the verb entry indicating the present tense form of the verb. The latter is required when converting to and from certain rules and from present to past tense. So, for example, the verb entry: *bala* 'read, reading, reads', present, transitive. The past tense entry for *bala*, i.e. 'read', would be indicated as *badile* 'read', past, transitive, *bala*. The *bala* at the end is given to enable the transformation rule applicable to verbs. The user also has the full ability to make nouns appear as pronouns and have the input of the sentences changed accordingly for all the different types of sentence structures.

This application is still only regarded as a functional prototype that has seen a few years of use in a university classroom. It contains a limited yet significant subset of data, which is significant enough to serve as verification of the system implementation covering all usage cases. For example, nouns and verbs chosen are not exhaustive but cover most language rules and all noun classes. In a full version, a much larger input set of verbs, nouns and adjectives would be needed.

The web app already tries to cater for such scaling by using a SqLite full text searching feature called FTS5 tables. This special table mostly eliminates the cost of 'LIKE' queries with text. In this context, 'LIKE' is meant to mean the database query operation that partially matches a set of letters typed in by the user to a list of possible complete words. Words matched in their entirety are relatively quick to look up. This performance consideration is critical to ensure the system remains usable with almost instant word suggestions.

The FTS5 also has another useful feature, and it can rank the relevancy of its textual match. The small set of user-searching input will show the words that have the closest match with the database entries. This ranking function is called the bm25 ranking function (http://www.staff.city.ac.uk/~sb317/papers/foundations_bm25_review.pdf). A negative constraint for using SqlLite and FTS tables is that database performance degrades with writing operations (changes, inserts and deletions) due to locking tables during updates. Wordlists are not constantly changing dynamic data, so updating the wordlists during maintenance periods is an acceptable trade-off and constraint for a cost-free and performant database.

All other wordlists such as demonstratives and subject concords are hardcoded in PHP. Such lists require manual updating but allow performance gains that are well worth the trade-off. The development choices described above should allow the application to easily scale to accommodate much larger wordlists.

The web app attempts to use best practices with software implementation. The technologies used are sufficient for their current target audience, but some technology replacements would be required to scale to thousands of users, for instance using Elastic-Search instead of SqLite, introducing clusters of servers with load balances and so forth. Maintenance and extension are substantially supported by the componentised architecture and would also allow for changing or swapping out of components as needed to allow for new requirements.

Developments that are still in early testing are the addition of adjectives, possessive nouns and adjectives for possessive nouns. This proves how extendible the current underlying system is and shows that the prototype is viable for Sepedi, and should be for similar languages. The system can produce a large percentage of all possible sentence types and with

additional vocabulary would enable a user to use it as a fully-fledged writing tool. As more functionality is added, it will remain a challenge not to erode the user experience. The principal way to uphold both is to keep concerns separated. This means that each component/module should only have one responsibility; this enables it to be replaceable, maintainable and improvable without requiring compounding changes in other components/modules; the changes remain isolated per component/module.

Conclusion and future work

In this article, the need for writing assistants in Sepedi was indicated mainly on the basis of user studies conducted on verbal moods and copulatives, and the aim was formulated as the description of an extended computational sentence builder for verbal moods, adjectives and possessive constructions. A full description of the linguistic complexity of the Sepedi Helper was attempted, with special attention being paid to pronominalisation, adjective and possessive constructions. It was shown that the Sepedi Helper is a powerful text production tool that combines the knowledge of the user with text production support requiring only basic inputs from the user. For the inexperienced user, it enables production of correct Sepedi sentences and phrases, and for advanced users, confirmation of correct text production. It can be concluded that The Sepedi Helper, in its current stage of development, covers the crucial and most frequently used constructions in Sepedi and is well en route to the eventual goal of compiling a complete e-grammar for Sepedi. Future work includes the addition of more constructions such as enumeratives, passives, verbal suffixes and copulative constructions as well as a detailed comparison of the Sepedi Helper with other writing tools mentioned in this article.

ORCID

DJ Prinsloo: https://orcid.org/0000-0003-0054-4676

References

Berg AS. 2018. Computational syntactic analysis of Setswana. Unpublished Doctoral Thesis. North-West University, South Africa.

Bosch SE, Faaß G. 2014. Towards an integrated e-dictionary application –The case of an English to Zulu dictionary of possessives. In: *Proceedings of the XVI EURALEX International Congress: The User in Focus.* http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202014/euralex_2014_056_p_739.pdf

Bothma T, Heid U, Prinsloo DJ. 2013. Implementing decision support for text production in e-dictionaries. *eLex 2013. Electronic lexicography in the 21st century: thinking outside the paper.* Institute of the Estonian Language and Trojina, Institute for Applied Slovene Studies. 17–19 October, Tallinn, Estonia.

Bothma T, Prinsloo DJ, Heid U. 2018. A taxonomy of user guidance devices for e-lexicography. *Lexicographica* 33(1): 391–422. https://doi.org/10.1515/lexi-2017-0018

Faaß G. 2010. A morphosyntactic description of Northern Sotho as a basis for an automated translation from Northern Sotho into English. Unpublished doctoral thesis, University of Pretoria, South Africa.

Lombard DP, Van Wyk, EB, Mokgokong PC. 1985. *Introduction to the Grammar of Northern Sotho.* Pretoria: Van Schaik.

Louwrens LJ. 1991. *Aspects of Northern Sotho Grammar.* Pretoria: Via Afrika.

Paquot M. 2012. The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool. In: Granger S, Paquot M (eds), *Electronic Lexicography.* Oxford: Oxford University Press. pp. 163–186. https://doi.org/10.1093/acprof:oso/9780199654864.003.0009

Poulos G, Louwrens LJ. 1994. *A Linguistic Analysis of Northern Sotho.* Pretoria: Via Afrika.

Prinsloo DJ. 2020. User studies on the Sepedi Copulative Decision Tree. *Southern African Linguistics and Applied Language Studies* 38(4): 323–335. https://doi.org/10.2989/16073614.2020.1858124

Prinsloo DJ, Bosch SE. 2012. Kinship terminology in English-Zulu/Northern Sotho dictionaries – a challenge for the Bantu lexicographer. In: Fjeld RV, Torjusen JM (eds), *Proceedings of the 15th Euralex International Congress*, 7–11 August, Oslo, Norway. pp. 296–303.

Prinsloo DJ, Bothma TJD, Heid U, Prinsloo Daniel J. 2017. User guidance in e-dictionaries for text production and text reception – The verbal relative in Sepedi as a case study. *Lexikos* 27(1): 403–426. https://doi.org/10.5788/27-1-1408

Prinsloo DJ, Heid U, Bothma TJD, Faaß G. 2011. Interactive, dynamic electronic dictionaries for text production. In: Kosem I, Kosem K (eds), *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011.* 10–12 November, Bled: Trojína. pp. 215–220. http://www.trojina.si/elex2011/elex2011_proceedings.pdf

Prinsloo DJ, Prinsloo Daniel, Prinsloo JV. 2015. A writing assistant as support tool for electronic dictionaries. Paper presented at *eLex 2015: Electronic Lexicography in the 21st century: Linking lexical data in the digital age*, 11–13 August, Herstmonceux Castle, UK.

Prinsloo DJ, Taljard E. 2019. User studies on the Sepedi Helper writing assistant. *Language Matters* 50(2): 73–99. https://doi.org/10.1080/10228195.2019.1627390

Taljard E, Prinsloo DJ. 2013. Lexicographic treatment of so-called cattle colour terms in Northern Sotho. *AFRILEX 2013*, 2–4 July, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa.

Verlinde S, Peeters G. 2012. Data access revisited: The Interactive Language Toolbox. In: Granger S, Paquot M (eds), *Electronic Lexicography.* Oxford: Oxford University Press. pp. 147–162. https://doi.org/10.1093/acprof:oso/9780199654864.003.0008

Ziervogel D. 1976. *Handboek van Noord-Sotho.* Pretoria: Van Schaik.