Direct User Guidance in e-Dictionaries for Text Production and Text Reception — The Verbal Relative in Sepedi as a Case Study[*]

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, South Africa (danie.prinsloo@up.ac.za)*

Theo J.D. Bothma, *Department of Information Science, University of Pretoria, South Africa (theo.bothma@up.ac.za)*

Ulrich Heid, *Department of Information Science and Natural Language Processing, Hildesheim University, Hildesheim, Germany and Department of African Languages, University of Pretoria (heid@uni-hildesheim.de)*
and
Daniel J. Prinsloo, *Entelect, Melrose Arch, Johannesburg, South Africa and Department of African Languages, University of Pretoria, Pretoria, South Africa (dprinsloo@gmail.com)*

Abstract: This article introduces a prototype of a writing (and learning) assistant for verbal relative clauses of the African language Sepedi, accessible from within a dictionary or from a word processor. It is an example of how a user support tool for complicated grammatical structures in a scarcely resourced language can be compiled. We describe a dynamic light-weight tool aimed at combining user-knowledge with text production support, i.e., user-involved interactive text production of the complicated verbal relative in Sepedi. In this article, the focus is on access in a dictionary use situation. Although the tool is intended as a writing assistant to support users in text production; it also satisfies text reception and cognitive needs, but its focus is on solving text production issues related with the interaction between lexical items and complex grammatical structures in the African (Bantu) languages and for learning by users and/or training users in this interaction.

Keywords: WRITING TOOLS, USER-GUIDANCE, USER SUPPORT, TEXT PRODUCTION, E-DICTIONARIES, AFRICAN LANGUAGES, SEPEDI, COMPLEX GRAMMATICAL STRUC-

---

[*]   This article represents follow-up work on an initial design study for user support in complex grammatical structures presented at *Euralex 2014* (Prinsloo et al. 2014). The website *Sepedihelper.co.za* was introduced at *eLex 2015* (Prinsloo et al. 2015).

TURES, RELATIVE CONSTRUCTION

Opsomming: Direkte gebruiksleiding in e-woordeboeke vir teksproduksie en teksresepsie — die werkwoordrelatief in Sepedi as gevallestudie. Hierdie artikel stel 'n prototipe van 'n skryf- en (leer)hulpmiddel bekend vir werkwoordrelatiewe konstruksies in die Afrikataal Sepedi, wat vanuit 'n woordeboek of 'n woordverwerker toeganklik is. Dit dien as voorbeeld van hoe 'n gebruikershulpmiddel vir ingewikkelde grammatikale strukture in 'n hulpbronbeperkte taal saamgestel kan word. Ons beskryf 'n dinamiese liggewig hulpmiddel wat gemik is op die kombinering van gebruikerskennis met teksproduksie-ondersteuning, dit wil sê, gebruikersbetrokkenheid by interaktiewe teksproduksie van die ingewikkelde werkwoordrelatiewe in Sepedi. In hierdie artikel is die fokus op toegang tydens 'n woordeboekgebruiksituasie. Hoewel die werktuig bedoel is as 'n skryfhulpmiddel om gebruikers in die produksie van teks te ondersteun, voldoen dit ook aan teksresepsie- en kognitiewe behoeftes. Die fokus is egter op die oplossing van teksproduksiekwessies wat verband hou met die interaksie tussen leksikale items en komplekse grammatikale strukture in die Afrikatale asook op die aanleer van taal deur gebruikers en/of die opleiding van gebruikers in hierdie interaksie.

Sleutelwoorde: SKRYFHULPMIDDELS, GEBRUIKERSLEIDING, GEBRUIKERSONDER-STEUNING, TEKSPRODUKSIE, E-WOORDEBOEKE, AFRIKATALE, SEPEDI, KOMPLEKSE GRAMMATIKALE STRUKTURE, RELATIEFKONSTRUKSIE

1.   Introduction

Over the last ten years, several writing aid tools have been developed (see below for an overview). Their purpose is to support users who need to produce texts (e.g. in a language that is not their L1). This support can be obtained in a dictionary use situation or from a word processor. The focus in this article is on guidance in text production when using an e-dictionary for text production with verbal relatives in Sepedi. Such user support can either be obtained by checking words, sentences or paragraphs produced by the user, or by guiding him/her to adequate solutions. Most such tools focus on lexical choice (e.g. in collocations). For languages with complicated morphosyntactic structures, such tools should cover not only lexical choice, but also the interaction between lexicon and grammar. The South African African (Bantu) languages are a typical example of such languages, and we will use Sepedi as a case in point in the present article.

  We will present the prototype of *Sepedihelper*, a tool that can assist dictionary users in the construction of Sepedi verbal relative constructions, which are a typical example of the complexity that arises from the interaction between lexical choice and the grammatical system of the language. The prototype is presented in a stand-alone version, but the objective is to include it into an interactive e-dictionary, e.g. an English–Sepedi translation dictionary.

  In the remainder of this introduction, we recall the main lines of the state of the art in writing aids; in section 2, we present the concept of direct user

guidance which underlies the *Sepedihelper*. Sections 3 and 4 are devoted to the morphosyntactic properties of the relative construction and show the complexity involved in the interaction between lexical choice and the building-up of correct grammatical constructions. In sections 5 and 6 we show the principles underlying the writing support for Sepedi relatives, as well as the properties of the actual implementation of the tool. A "guided" tour from the dictionary user perspective follows in section 7 and we conclude in sections 8 and 9 with remarks on first experiences with dictionary users, as well as plans for future work. While we exemplify the principles of direct guidance for dictionary users on the Sepedi relative construction, we are convinced that more constructions from the African (Bantu) languages, as well as more generally any kind of interaction between lexical choice and grammatical (or morphosyntactic) constraints of a given language, could be dealt with along the same lines.

Writing tools have a great potential for user support in an e-dictionary, especially for text production but also for text reception of complicated grammatical structures in any language. Such tools should be designed to take the dictionary user's expertise into account in terms of the level or strategy for guidance provided.

Regarding text production, this article illustrates the working of a Builder for assisting users to write relatives in Sepedi (see Section 3). In a similar vein, the tool should be able to translate a Sepedi relative phrase into English. The nature of the support should typically also link to a user's level of knowledge of the grammatical system of L2 and should therefore take different user types, based on their knowledge of the L2 and their information needs into consideration (cf. Tarp (2008)), which can be summarised as follows:

— A user with a very limited knowledge of the language or a casual user, e.g., may prefer a machine translation option in the dictionary, with links to the grammar rules which may be consulted on demand. Cf. Bosch and Faaß (2014) as an example of direct user guidance to the correct answer in the compilation of possessive constructions in Zulu plus rule-based machine translation technology. Possessive constructions in Zulu can be regarded as complicated, requiring substantial knowledge of the nominal class system, possessive concords, exceptions to the formation rules, etc. which many inexperienced users may not have at the time of consulting the tool.

— On the other hand, a user who has a fair knowledge of the language may require a different type of support, e.g. through decision trees, i.e., a series of basic choices made by the user. Examples have been discussed for copulatives, kinship terminology, colour terms, etc. Cf. Bothma et al. (2013), Prinsloo and Bosch (2012), Prinsloo et al. (2011), Taljard and Prinsloo (2013).

— In certain cases, a user might benefit more from a bird's eye view through well-structured guidance paths such as tables and diagrams on e.g. kinship relations, grammatical moods and meanings, etc. Cf. Prinsloo et al. (2012).

Such technologies, integrated into the dictionary, may enable the user to find the correct information at an adequate level of detail and complexity required to solve his/her information need, thereby individualising the data presented to the user in terms of his/her information need. Cf. Bothma (2011), Fuertes-Olivera and Tarp (2014), Tarp (2008, 2011, 2012), Verlinde (2011).

The purpose of such tools is to guide users to the information they are looking for, i.e., without having to first study complicated grammatical structures in order to find the required information. We use the term *user support (technologies)* as an umbrella term for all such technologies. To date, we have described only the three technologies listed above. Additional such technologies and designs exist, e.g., *Interactive Language Toolbox* (https://ilt.kuleuven.be/inlato/), *Writing assistants and automatic lexical error correction: word combinatories* (Wanner et al. 2013), *A collocation writing assistant for learners of Spanish* (Alonso Ramos et al. 2014), user driven task and problem-oriented multifunctional leximats (Verlinde et al. 2010), online data-driven lexicographic instruments on foreign language learning (Buyse and Verlinde 2013), the work of Bertels and Verlinde on lexicography and corpus analysis (Bertels and Verlinde 2011), etc. All such techniques can be embedded in an e-dictionary and are intended to give information on demand, i.e., the user has the option to consult the tool if the "standard" dictionary article does not provide sufficient data to solve the user's specific information need. It would also be possible to embed such tools in a word processor, or for the user to consult such tools as stand-alone tools; this, however, is not the focus of this article and will not be addressed further. The focus is therefore on a tool embedded in an e-dictionary; access in this case is from the e-dictionary. The user therefore consults the dictionary, in the current case about the translation of the English word "who" into Sepedi (i.e., a translation/text production information need), and upon finding that (s)he needs more help than is available in the "standard" dictionary article, accesses the *Sepedihelper* on demand.

2.   Direct user guidance as a support technique

Direct user guidance as an additional technique in the e-dictionary to provide user support for complex grammatical structures is not a solution for all user support. We regard it as a complementary technology that may be used in conjunction with other user support technologies for specific grammatical constructions, and the same way as these it should only be available to the user on demand, depending on the user's level of language knowledge, the nature of the information need and the user's choice of support tool. In Prinsloo et al. (2014), we presented a design study to show that user support through direct user guidance can provide solutions in the case of complex concordial relationships between nouns and pronouns. In terms of the Function Theory of Lexicography (Tarp (2008), Bothma and Tarp (2012), Fuertes-Olivera and Tarp (2014)), the design provides for text production, text reception and cognitive information

needs. In this article, we report on further work that has been done in this regard, viz. the development of a small-scale prototype to demonstrate the feasibility of such a tool. We describe it from the perspective of the end-user, i.e., how (s)he could go about solving his/her information need by using the prototype tool. We also briefly describe the technologies we used to develop the prototype tool and report on some observations from user-studies.

As will be clear from the discussion above and from Prinsloo et al. (2011, 2012), such techniques are made available "on demand", i.e., users are not forced to use them if they feel that their information needs have been solved by the "standard" dictionary article. In every case, the use of such a technique is therefore a conscious choice of the user to find more information or information that is easier to use, digest or apply than the information available in the dictionary, the outer text of the dictionary or other reference tools such as grammar books that the user may have available.

The importance of the user perspective as the main thrust in the compilation of modern dictionaries has been emphasized in numerous publications, e.g., Gouws and Prinsloo (2005), Tarp (2008, 2011, 2012), Fuertes-Olivera and Tarp (2014). The concept of *user-support* appropriately puts the user in focus. Compare Tarp's (2012: 253) idea of individualization when he refers to "quicker, more accurate and personalized satisfaction of the corresponding user needs". Our approach to user support furthermore does not necessarily put the user into a specific category (e.g., as a learner of the language): it is not profile-based and does not assume that the user will be interested to study a complete grammatical paradigm before being able to produce (or understand) texts. We therefore also cater for the casual, on-the-fly user, who is not interested or in a position to devote time to the in-depth learning of a foreign language, but relies on access to appropriate information from the e-dictionary and additional tools on demand.

3.   Grammatical distinctions as a problem for African language lexicographers

Constructing phrases and sentences in African languages is a complicated process resulting from the classification of nouns into different noun classes. Text production, be it through dictionary consultation or creative writing, requires a substantial amount of grammatical knowledge. Traditionally the user had to rely on paper dictionaries and grammar books. Most print dictionaries for African languages are not helpful for text production and complicated grammatical issues are dealt with in many pages of fine print in grammar books which the user has to study as a prerequisite to text production.

3.1   The notion of grammatical distinctions

A given grammatical property may be expressed in many different forms. For

example, there are different equivalents for a pronoun such as *he,* determined by the grammatical class of the noun. Nouns in African (Bantu) languages are subdivided into different noun classes, as illustrated in Table 1. These classes have their own sets of subject concords and object concords, as well as different sets of pronouns such as demonstrative, possessive, emphatic and quantitative. This means that e.g. in Sepedi, an English personal pronoun such as *he* can be expressed by up to ten different subject concords, a form like *him* by ten object concords and more than 20 pronominal forms. Consider Table 1 which distinguishes 15 different noun classes, each having their own subject concords (Sc.); object concords (Oc.); demonstratives (Dem.); possessive concords (Poss.); emphatic pronouns (Ep.) and quantitative pronouns (Qp.).

Person or noun class	Example	Sc.	Oc.	Dem.	Poss.	Ep.	Qp.
1st Person singular	nna 'I'	ke	n-			nna	
1st Person plural	rena 'we'	re	re			rena	
2nd Person sing.	wena 'you' (singular)	o	go			wena	
2nd Person plural	lena 'you' (plural)	le	le			lena	
Class 1	monna 'man'	o/a	mo	yo	wa	yena	yohle
Class 2	banna 'men'	ba	ba	ba	ba	bona	bohle
Class 3	molato 'trouble, problem'	o	o	wo	wa	wona	wohle
Class 4	melato 'problems'	e	e	ye	ya	yona	yohle
Class 5	lesogana 'young man'	le	le	le	la	lona	lohle
Class 6	masogana 'young men'	a	a	a	a	ona	ohle
Class 7	selo 'object, thing'	se	se	se	sa	sona	sohle
Class 8	dilo 'objects, things'	di	di	tše	tša	tšona	tšohle
Class 9	ntlo 'hut'	e	e	ye	ya	yona	yohle
Class 10	dintlo 'huts'	di	di	tše	tša	tšona	tšohle
Class 14	bogobe 'porridge'	bo	bo	bjo	bja	bjona	bjohle
Class 15	go reka 'to buy'	go	go		ga	gona	gohle
Class 16	fase 'below'			fa			

| Class 17 | godimo 'above' | **go** | **go** | | **ga** | **gona** | **gohle** |
| Class 18 | morago 'behind' | | | **mo** | | | |

Table 1: The noun class system of Sepedi with a few sets of concords and pro-
nouns

In Table 1 the Sepedi equivalent of the demonstrative 'this' varies depending on
the class of the noun, e.g.,

(1)   class 1: *monna **yo*** 'this man', but
        class 10: *dikgomo **tše*** 'these cattle' and
        class 14: *bogobe **bjo*** 'this porridge'

Likewise, the Sepedi equivalent of the possessive 'of' differs for each class, e.g.,

(2)   class 1: *mosadi **wa** monna* 'wife of the man', but
        class 2: *basadi **ba** monna* 'wives of the man' and
        class 7: *selepe **sa** Madika* 'Madika's axe'

Concords and pronouns representing subjects and objects also vary according
to the nominal class, e.g.:

(3) a.   O e bone 'He saw it'
         o (e.g. *monna* class 1)   ***e*** (e.g. *tau* class 9)   *bone*
         he (the man)   **it** (the lion)   saw

    b.   **o** (e.g. monna class 1)   bone   **yona** (e.g. tau class 9)
         he (the man)   saw   **it** (the (specific) lion)

    c.   Ba rekile tšohle 'They bought everything'
         ba (e.g. *banna* class 2)   *rekile*   ***tšohle*** (e.g. *dikgomo* class 10)
         they (the men)   bought   **all** (the cattle)

In 3a *o* is a subject concord and *e* is an object concord. In 3b *yona* is an emphatic
pronoun and in 3c *tšohle* is a quantitative pronoun.

3.2   Grammatical distinctions in the sentence context

If Table 1 is interpreted from a translation-based viewpoint (e.g. EN → Sepedi),
the grammatical distinctions paradigm is mono-dimensional in the sense that it
is always given for a single source language item which diverges into a single
set of equivalents. More than one instance of grammatical distinction can, how-
ever, co-occur in a single construction or phrase; thus multiple choice points
from the grammatical paradigms in Table 1 may (co-)occur in one sentence,
and below we will illustrate cases with one, two and three occurrences. In (4)

the user has to determine the correct subject concord from the paradigm *o/a/le/se/e* to complete the sentence.

Example 1: *he/they*, as the subject of a sentence (subject concords):

(4) a. **O/a/le/se/e**   thušitše mosadi
       He   helped the woman

   b. **ba/e/a/di**   ja bogobe
      They   eat porridge

In (5) the same situation prevails for the selection of the appropriate quantitative pronoun (used as an object of the verb).

Example 2: how to express *all* (quantitative pronouns):

(5)   Go bolaya   **bohle/yohle/ohle/tšohle**
         To kill   **all**

In (4) and (5) respectively the user has to deal with a single paradigm to complete the sentence. More complicated are situations where (s)he has to negotiate two, as in (6) or even more such grammatical paradigms in a single sentence, cf. (7).

    In (6) the user has to find the correct subject concord and the applicable object concords from the *two* paradigms *o/a/le/se/e* and *ba/e/a/di* to complete the sentence: *he* as a subject and *them* as an object:

(6) **O/a/le/se/e**   tlo   **ba/e/a/di**   thuša.
     He   will   them   help

The construction involves varying subjects and objects. The subject and object are in most cases not belonging to the same class as in (6) if *o* (class 1) is to represent the subject *he* and *ba* (class 2) the object *them*.

    In (7) the correct demonstrative, subject concord and object concord need to be selected from the *three* paradigms *yo/wo/le/se/ye* (demonstratives), *o/a/le/se/e* (subject concords) and *ba/e/a/di* (object concords) to complete the sentence; and again involves varying subjects and objects in terms of the correct demonstrative, subject concord and object concord:

(7) **Yo/wo/le/se/ye**   **a/wo/le/se/e**   **ba/e/a/di**   thušitšego
     He   he   them   helped
     He who helped *them*.

(7) is an example of a relative construction which can be regarded as one of the complicated structures for text production, especially for inexperienced users. It will first be described in more detail and then followed by an introduction of

the first prototype relative builder for Sepedi. The standard structure of the relative is noun + demonstrative + subject concord + verb stem with the relative suffix *-go*, as described below.

4.   The relative construction in Sepedi

The relative is described in detail in traditional Sepedi grammars such as Van Wyk et al. (1992), Lombard et al. (1985), Ziervogel (1969) and Poulos and Louwrens (1994). They agree in principle that the relative modifies a noun or pronoun and that two main types are distinguished, i.e., direct and indirect relatives. Both direct and indirect relatives typically consist of nouns, demonstratives, subject concords, object concords, verb stems, relative suffixes and pronouns. For the user to produce a who-sentence, knowledge of at least 10 pages in fine print in Poulos and Louwrens (1994) is required, and the option to use Google Translate or Microsoft/Bing Translator for a translation does not exist.

(8)   Direct relative

a. Monna yo a sepelago. 'The man who is walking.'
b. Monna yo a rekago puku. 'The man who buys/is buying the book.'
c. Monna yo a rekelago bana puku. 'The man who buys a book for the children.'

In (8a) the relative consists of a noun of class 1 (*monna*), a demonstrative of class 1 (*yo*), a subject concord *a*, an intransitive verb stem *-sepela* and a suffix *-go* indicating relative mood on the verb form. In (8b) the verb stem *-reka* is transitive and is followed by a direct object *puku*. In (8c) the verb is double transitive, indicated by the suffix *-el*, and followed by an indirect object *bana* and a direct object *puku*. Objects can be pronominalized by means of object concords or pronouns.

So, e.g. the objects in (8c) can be pronominalized as in (9).

(9)   Direct relative with a pronominalized object

a. Monna yo a **ba** rekelago puku. 'The man who buys **them** a book.'
b. Monna yo a rekelago **bona** puku. 'The man who buys **them** a book.'
c. Monna yo a **e** rekelago bana. 'The man who buys **it** for the children.'
d. Monna yo a rekelago bana **yona**. 'The man who buys **it** for the children.'

In (9a) and (9b) *bana* is pronominalized by its object concord *ba* and emphatic pronoun *bona* respectively. Likewise, in (9c) and (9d) *puku* is pronominalized by its object concord *e* and emphatic pronoun *yona* respectively. All of these constructions can occur in the present, future and past tense, in the positive or negative. Consider, e.g., (8a) in the three tenses present, future and past in the positive and negative in Table 2:

A	Direct Relative intransitive		
1	Present tense positive	Monna yo a sepelago	The man who is walking
2	Present tense negative	Monna yo a sa sepelego	The man who is not walking
3	Future tense positive	Monna yo a tlogo sepela	The man who will walk
4	Future tense negative	Monna yo a ka se sepelego	The man who will not walk
5	Past tense positive	Monna yo a sepetšego	The man who walked
6	Past tense negative	Monna yo a sa sepelago	The man who did not walk

Table 2:   Present, past and future tense, positive and negative

All of the constructions in (8), (9) and Table 2 also apply for the indirect relative which differs from the direct relative in the use of an additional nominal before the verb as in (10):

(10)   Object concord as pronominalized object

Monna yo mosadi a mo rekelago puku. 'The man for whom the woman buys a book.'

In this case the demonstrative belongs to *monna* but the subject concord to *mosadi*. A detailed discussion of the indirect relative is given in the traditional grammar books cited above.

Several hundreds if not thousands of possible relative constructions could be formed for relatives through the combination and permutation of the following possibilities:

— direct/indirect relative
— present tense/future tense/past tense
— positive/negative
— intransitive/transitive/double transitive
— object concord/object pronoun
— 18 noun classes, etc.

The 18 most typical types for the direct relative are the following:

1.   Direct relative intransitive positive
2.   Direct relative intransitive negative
3.   Direct relative transitive positive with object noun
4.   Direct relative transitive negative with object noun
5.   Direct relative transitive positive with object noun pronominalized with a concord

6. Direct relative transitive negative with object noun pronominalized with a concord

7. Direct relative transitive positive with object noun pronominalized with a pronoun

8. Direct relative transitive negative with object noun pronominalized with a pronoun

9. Direct relative double transitive positive with indirect and direct object nouns

10. Direct relative double transitive negative with indirect and direct object nouns

11. Direct relative double transitive positive with indirect object pronominalized with a concord and direct object

12. Direct relative double transitive negative with indirect object pronominalized with a concord and direct object

13. Direct relative double transitive positive with indirect object pronominalized with a pronoun and direct object

14. Direct relative double transitive negative with indirect object pronominalized with a pronoun and direct object

15. Direct relative double transitive positive with indirect object and direct object pronominalized with a concord

16. Direct relative double transitive negative with indirect object and direct object pronominalized with a concord

17. Direct relative double transitive positive with indirect object and direct object pronominalized with a pronoun

18. Direct relative double transitive negative with indirect object and direct object pronominalized with a pronoun

Consider, e.g., the amount of knowledge presupposed from the dictionary user if (s)he wants to produce a fairly simple English single transitive sentence such as the *man who buys her a book* in Sepedi. (S)he has to know

(a) the Sepedi word for *man*, i.e., *monna*,

(b) to which of the possible 15 noun classes it belongs, i.e., class 1, in order to

(c) select the correct demonstrative from 15 possibilities, i.e., class 1 *yo*,

(d) the subject concord for class 1,

(e) that an irregular relative concord is used for this noun class, i.e., *a* and not *o*,

(f) the Sepedi word for *buy*, i.e., *reka*,

(g) the Sepedi word for *book*, i.e., puku,

(h)   the Sepedi word to which *her* refers, e.g. mosadi 'woman', mosetsana 'girl', etc.,

(i)   to which of the noun classes the object belongs in order to select the correct object concord from 15 possibiliies, i.e., *mo*,

(j)   what the relative suffix is, i.e., *go*

(k)   that the object concord is used pre-verbally.

in order to construct *monna yo a mo rekelago puku*.

5.   Direct guidance for relative constructions

In principle, in respect of the relative, guidance can be given by means of three possible types of access depending on the user's need in terms of text production and his/her knowledge of the language.

The on-the-fly user will benefit most from assistance resembling machine translation for both text production and text reception purposes. The typical situation could be where the user simply wants to know how to say an English sentence such as *the man who bought the car* in Sepedi or how to translate the equivalent Sepedi sentence *monna yo a rekilego mmotoro* into English. As remarked by Prinsloo et al. (2014: 820), they simply need an on-the-spot solution and might not even be interested in learning Sepedi or English.

The focus of this article is on the user with limited knowledge of Sepedi who needs help — to a greater, or lesser extent — to create relative sentences in Sepedi. The user might even be someone like a second or third language speaker of Sepedi who is quite proficient in the language but requires confirmation as to the correctness of the sentences produced, e.g. in the case of relative constructions with irregular nouns and verbs. The relative builder, accessible via the article for "who" in the e-dictionary, attempts to cater for these different proficiency levels in a natural way by offering the user the opportunity to take shortcuts, e.g. if (s)he knows the Sepedi words or a longer route to the Sepedi words through dictionary lookup. No attempt was made to cater for formal user proficiency levels, e.g. users requested to indicate their level of expertise (e.g. Bothma 2011 and De Schryver 2003).

In building the relative construction, the system performs the following steps for the sentence *the children who love the food*:

(i)   *children*: the tool provides the correct equivalent from the e-dictionary, i.e., *bana* tagged for part of speech as N02 (noun of class 2, cf. Table 1);

(ii) *who*: keeping the agreement constraint from the sentence formation rule (noun + demonstrative + subject concord + verb + relative suffix (*-go*), cf. (7) and (8)), the tool extracts the demonstrative for class 2 from the closed-class list of demonstratives, i.e., *ba*;

(iii) (subject concord): The insertion of the SC is coded in the rule for relatives: it requires, in addition to the demonstrative in (ii), the subject concord for the noun in (i). As in (ii), the tool proposes *ba*, i.e., the subject concord for class 2.

(iv) *food*: the tool provides the correct equivalent from the dictionary, i.e., *dijo* tagged for part of speech as N08 (noun of class 8, cf. Table 1);

(v) *love*: as for (i), the system selects the correct Sepedi equivalent: *rata*, plus adding the relative suffix. The adding of *-go* is built into the relative construction rule.

Result: Bana ba ba ratago dijo. *The children who love the food.*

The processes (i) to (v) are the same for automated text production support. The user enters the entire English phrase, e.g. *the man who bought the car* and the system applies (i) to (v) to construct the Sepedi sentence *monna yo a rekilego mmotoro*. It is also possible to type the full Sepedi sentence, in which case the process is reversed to produce the English translation *the man who bought the car.*

In these specific cases, no user knowledge (neither lexical nor grammatical knowledge) is required, and the process is fully automated. However, the tool can also be used interactively, which requires the user to make specific choices in the construction of the relative, as discussed in section 7 below.

6.   A software implementation: underlying technology and essential components of the relative builder

Essential grammatical components are a machine readable English/Sepedi dictionary with part of speech markup. The syntax and components of the relative construction are hard-coded. The functioning prototype was developed using AngularJS and Bootstrap on the front-end. The back-end was developed in PHP that uses a SqlLite database. The current prototype is hosted at www. sepedihelper.co.za. The application is written using best practices to clearly differentiate between logic/content and interface/display, as well as between input and output, to allow maximum flexibility. These characteristics allow for easier improvement, maintenance and extension of the application. Due to the differentiation between front-end and back-end, both can be replaced with other technologies, if necessary. Integrating this application with Microsoft Word or with the Open-Source alternatives like Libre-Office and Open-Office will require a rewrite of the software for each. Integration with e-dictionaries, on the other hand, should be less complex. The writing tool would need minimal improvements and refactoring to allow it to be used as a component inside e-dictionaries.

For the prototype, a limited subset of the data of an existing e-dictionary was used, copying only the relevant data fields of the e-dictionary to a new e-dictionary database. We therefore foresee that a new e-dictionary database will not be required for a full implementation of this tool — the tool will simply access the database of the dictionary. In a full version, a standard bilingual Sepedi/English e-dictionary will be used. The e-dictionary database of this standard dictionary will, however, have to be modified to make provision for the fields that are required by the proposed tool.

The main form of input data required are tagged wordlists. Nominal word lists require the noun itself, a translation equivalent paradigm as well as a noun class indication, e.g. "*badiredi* (employees, workers), N02" is a full database entry represented in the database in three fields. 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, e.g., the database entry for the present tense verb, *reka*, is as follows: "*reka* (buy, buying, buys), present, transitive". The past tense entry for *reka* i.e., *bought* would be indicated in the database as "*rekile* (bought), past, transitive, *reka*". The *reka* at the end is given to enable the transformation rule applicable to verbs.

Lastly, it is worth mentioning that great emphasis was placed on performance. The SqlLite database duplicates the data in a special full-text-search (FTS) table that is extremely fast to query. Users start typing and the entire wordlist is searched for a partial match. This would take unacceptably long to complete with normal database "LIKE" operators. 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, however, not constantly modified, so updating the wordlists during maintenance periods is an acceptable trade-off and constraint for a cost-free and performant database. Wordlists that remain static as part of the grammar rules of Sepedi (demonstratives, subject concords, etc.) 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 scale easily to accommodate much larger wordlists or a full e-dictionary (modified to contain the required database fields).

7.   Using the relative builder

A prototype of the relative sentence builder is available at www.sepedihelper. co.za and is briefly described in the following section.

The user will be able to access the tool directly from the web address or in future from an e-dictionary or word processing software, as illustrated in Figure 1:
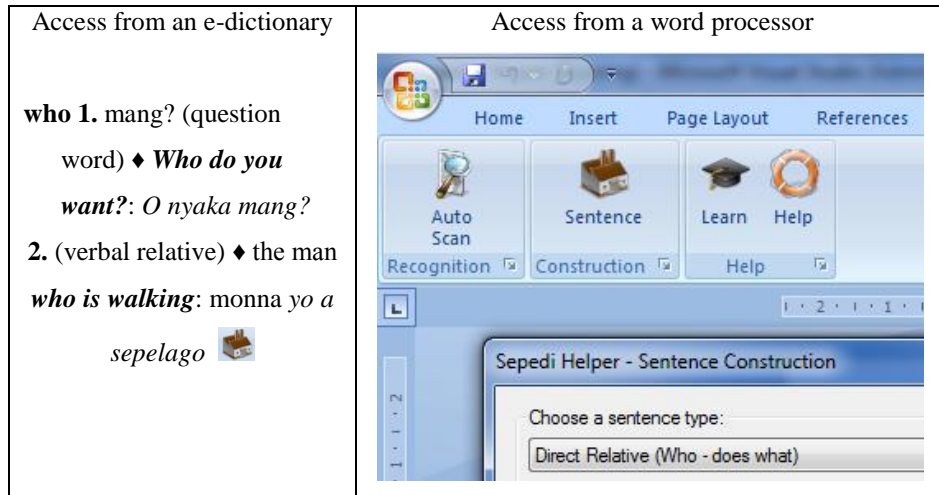
Access from an e-dictionary	Access from a word processor
who 1. mang? (question word) ♦ ***Who do you want?***: *O nyaka mang?* **2.** (verbal relative) ♦ the man ***who is walking***: monna *yo a sepelago*	

Figure 1:   Accessing the relative construction tool

In the dictionary, an additional icon (picture of a factory) allows the user to launch the tool, which opens the tool in in a pop-up window or in the user's word processor. When the user is already working in the word processor, an icon on the toolbar allows him/her to launch the tool. The Relative Builder can therefore equally well be used as a dictionary component or as a writing assistant. The dictionary user will be offered guidance from within the dictionary article of all English and Sepedi lemmas which are relevant to the relative construction as in Figure 1, left column, e.g. *who*, *what*, *which* and all of the Sepedi demonstratives *yo*, *ba*, *wo*, etc. as well as the relative suffix *-go*. The user who requires assistance to build a relative construction from within a word processor can click on the factory icon in the taskbar as in Figure 1, right column.

The tool currently offers assistance for all 18 types of relatives listed in section 3.2 above.

The user's existing knowledge at any given point is taken into account by offering them choices e.g. to enter Sepedi words directly or to go via English. In the build-up process given below, the user would like to express in Sepedi the sentence *the children who love/like it*, where *it* refers to "food". The user departs from the dictionary article for *who* by clicking on the factory icon in the article for *who* in the e-dictionary (or consults the helper by clicking on the factory icon if the helper is accessed from a word processor). In both cases (s)he is presented with the *Sepedihelper* screen in Figure 2.
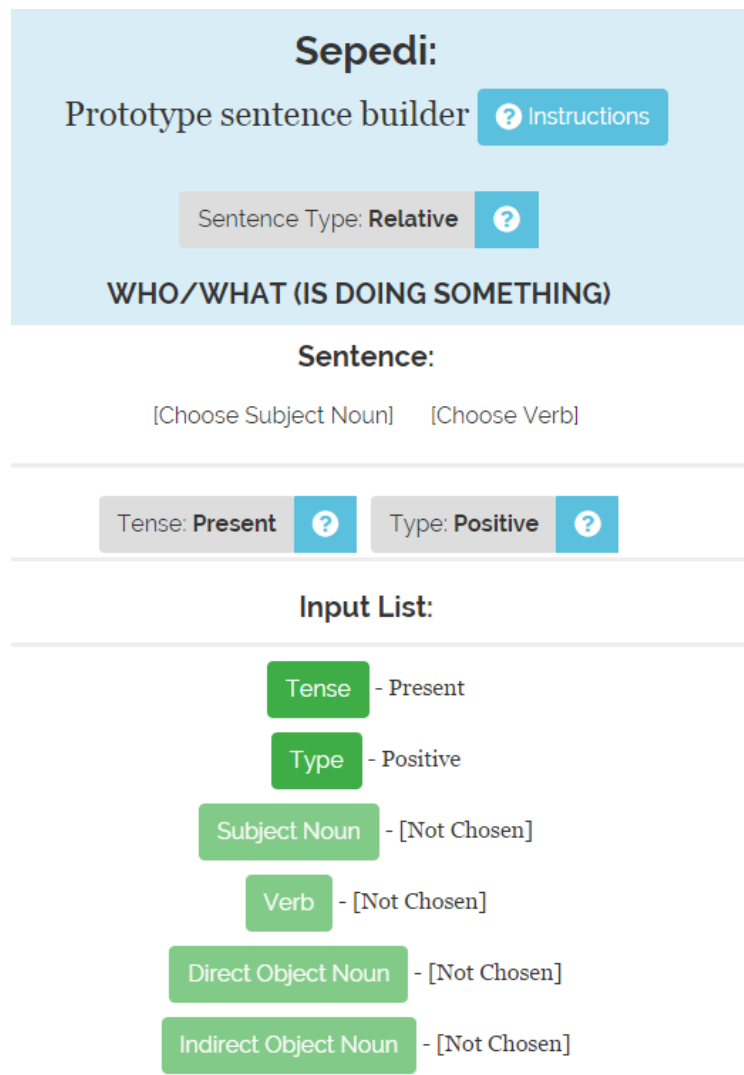
Figure 2:   The Sepedi relative builder main screen

The user is informed that cognitive information at different levels can be obtained by clicking on the question mark "?" icons. The inexperienced or first-time user can obtain cognitive information on direct relatives and direct relatives with an object concord. Typical examples are given and a few suggestions for building relative constructions are also presented, cf. Figure 3.

Relative mood:

Relative expresses: **who/what is doing something**

Present

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

Can occur with direct and indirect object:

	Monna yo a rekelago	bana puku
		indirect object direct object
	'The man who buys	for the children a book.

Nouns can be replaced by **pronouns**:

Example (**bana**):   Monna yo a rekelago **bona** puku

Object nouns can also be replaced by object concords:

Example: Monna yo a **ba** rekelago puku

Future

Most common structure:		Noun + demonstrative + subject concord + fut. + go +verb + object
Example:	**Positive**:	Monna yo a tlogo reka puku.
		'The man who will buy a book'
	Negative:	Monna yo a **ka se** rek**e**go puku.
		'The man who will not buy a book'
	Negation strategy:	**ka se** … verb ends in **-e**

The combination ke = I + ka se becomes **nka** se

                                Nka se reke puku 'I shall not buy the book'

Past

Most common structure:		Noun + demonstrative + subject concord + verb + go + object
Example:	**Positive**:	Monna yo a rekilego puku.
		'The man who bought a book'
	Negative:	Monna yo a **sa reka**go puku.
		'The man who did not buy a book'
	Negation strategy:	**sa** + **verb back to the present tense**

Figure 3:   Help screen explaining the basics of relative sentence construction

The gradual build-up and eventual completed sentence will be displayed in a horizontal line under the heading **Sentence**, as in Figure 9. Initially this line reflects all the required and optional elements of the relative construction, e.g. [Choose Subject Noun] and [Choose Verb] to be replaced step-by-step with real words and concords in the build-up process.

For step 1 the instruction is to choose a noun. If the user knows the Sepedi word (s)he can type it in, as in Figure 4.
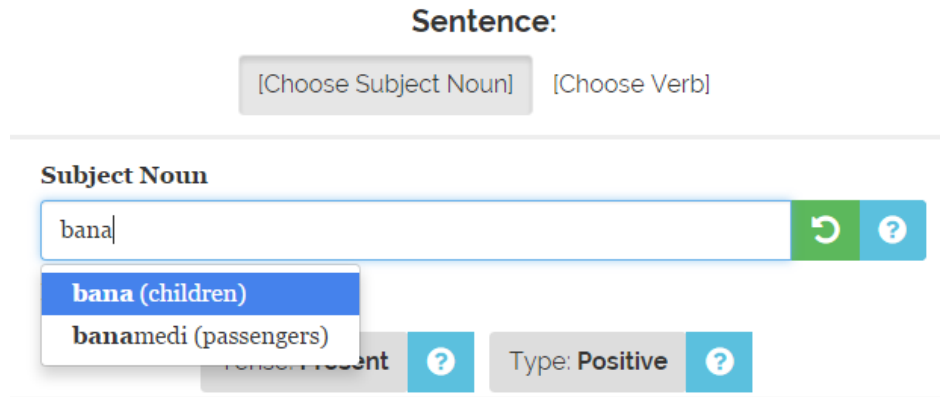
Figure 4: Inserting the Sepedi subject noun in Step 1 of the relative builder

The system presents the word with its translation equivalent(s) (currently from a limited database). In a full implementation a more comprehensive set of entries with direct translations will be provided. Words as well as navigation links that would open up viewing it as a dictionary entry for more comprehensive help with a word will be offered. The translation equivalents help the user to ascertain that (s)he is dealing with the right Sepedi noun. If the user does not know the Sepedi word, (s)he simply types the English word to find the Sepedi translation and selects *bana*, the required Sepedi item, as in Figure 5:
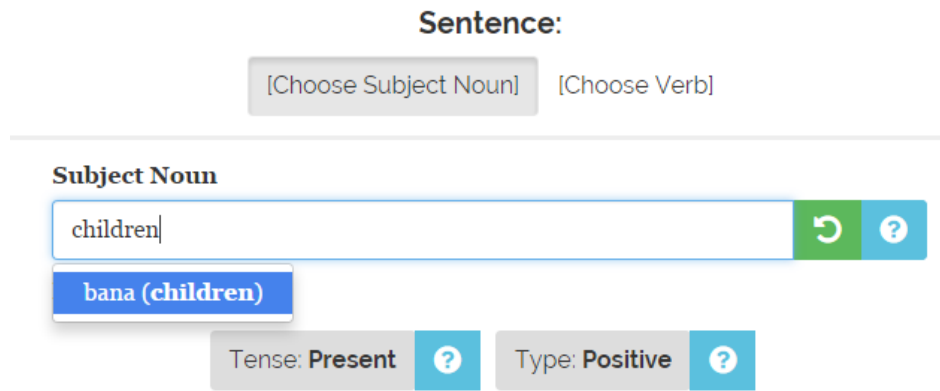
Figure 5: Inserting the Sepedi subject noun by typing the English noun in Step 1 of the relative builder

The system automatically performs a lookup for the part of speech of *bana*, i.e., a noun from Class 2, and generates the demonstrative *ba* and the subject concord *ba* in the next two fields, i.e., the generated demonstrative and the generated subject concord as well as displaying the current stage in the build-up process in the Generated Complete Sentence line, i.e., *bana ba ba* prompting the user to enter the remaining required component, i.e., the verb, Figure 6.
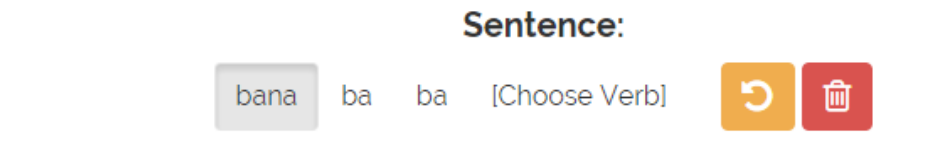
Sentence:

| bana | ba | ba | [Choose Verb] | ↺ | 🗑 |

Figure 6: Generating the demonstrative and subject concord in the Sepedi relative builder

The user repeats the same process as in Step 1 to enter the verb *rata* in Step 2. The system automatically adds the required relative suffix *-go* to the verb and prompts the user to add a direct object, Figure 7.
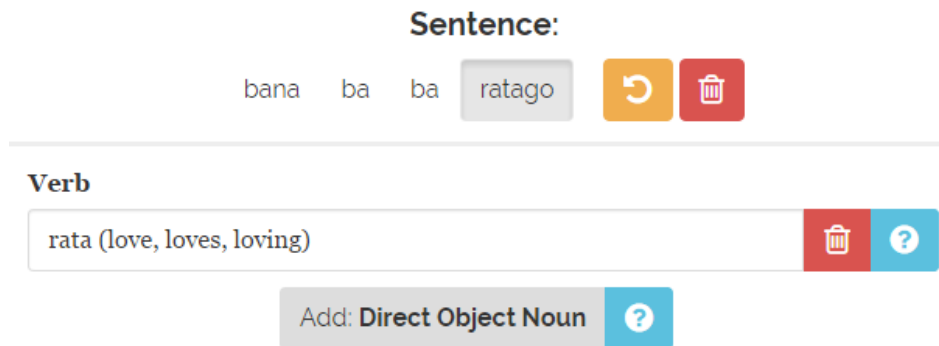
Sentence:

| bana | ba | ba | ratago | ↺ | 🗑 |

Verb

| rata (love, loves, loving) | 🗑 | ❓ |

Add: **Direct Object Noun** ❓

Figure 7: The Sepedi relative phrase for *the children who love*

The user enters the object noun *dijo* "food", the builder completes the relative and asks the user if (s)he would rather prefer to pronominalize the object noun, i.e., the children who love *it*. At this point the system has already determined the part of speech of *dijo*, i.e., a noun in class 8. A choice is offered between insertion of the object concord *di* or substitution of *dijo* by its pronoun *tšona*.
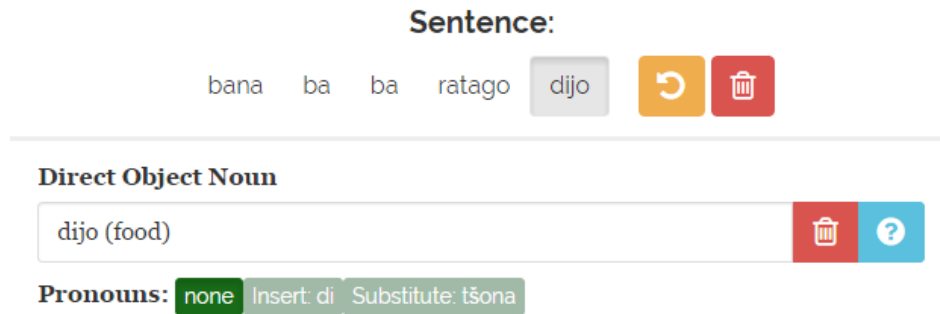
Sentence:

bana   ba   ba   ratago   dijo

Direct Object Noun

dijo (food)

Pronouns: none Insert: di Substitute: tšona

Figure 8:   The Sepedi relative phrase for *the children who love the food*

If the user chooses the object concord, the system inserts it in the correct syntactic order which is pre-verbal.
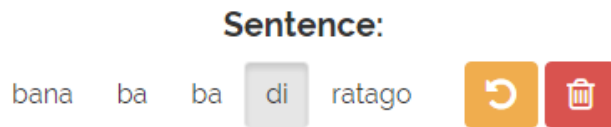
Sentence:

bana   ba   ba   di   ratago

Figure 9:   The Sepedi relative phrase for *the children who love it*

Consider the same situation as in Figures 8 and 9 for a noun from a different nominal class, as in Figures 10 and 11.

Sentence:

bana   ba   ba   ratago   modiredi

Direct Object Noun

modiredi (employee)

Figure 10: The Sepedi relative phrase for *the children who love the employee*

Sentence:

bana   ba   ba   [mo]   ratago   [↺]   [🗑]

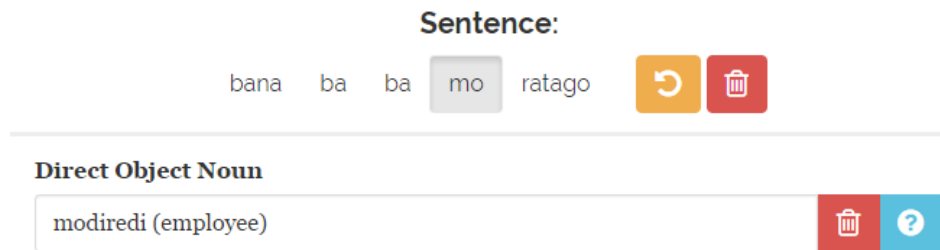**Direct Object Noun**

modiredi (employee)   [🗑] [❓]

Figure 11: The Sepedi relative phrase for *the children who love him*

In Figure 10, the noun is *modiredi* (employee) which belongs to class 1 and the object concord of class 1 is *mo*. Thus the system correctly generated the object concords for *dijo* and *modiredi* as *di* versus *mo* respectively.

8.   User-studies

Reflecting on user-studies is not the aim of this article — user feedback will be dealt with in more detail in a forthcoming publication. Recently, two user studies totaling 109 users were conducted on the Sepedihelper and the relative construction in particular. These studies, of which some were performed in class by students on their cellphones, indicate that most users fail to produce "who"-phrases correctly without external help and benefitted much from the guidance by the writing assistant. In these studies the user had to attempt compilation of the phrase first and then to use the writing assistant. The value of the assistant was clear. The most typical problems in respect of producing a "who"-sentence are summarized as follows:

— Wrong mood

— Wrong sentence position for the subject of the sentence

— Both demonstrative and/or subject concord left out

— Wrong negation morpheme or negation morpheme in wrong syntactic position

— Did not change the concord of class 1 to *a*

— Omission of the relative suffix

— Incorrect spelling / word division

— Got it right — no guidance but confirmation from the tool

Users generally regarded the tool as user-friendly and easy to use.

9.   Conclusion and future work

User support through direct guidance (and other support mechanisms) for complex grammatical structures allows the user to navigate via the shortest route to the information (s)he is looking for in an e-dictionary without having to work through long and often complicated grammar-type representations of complex grammatical structures. Such guidance is always available on demand, i.e., the user is not forced to work through any such support mechanisms if (s)he finds that the "standard" data in the e-dictionary are sufficient to solve his/her information need in a given situation. However, if more information is needed or if the standard presentation of the information (be this in the e-dictionary, in outer texts or in reference tools) is too difficult or complex to be easily understood, the user would have an alternative mechanism (or alternative mechanisms), accessible on demand from within the e-dictionary (or from within a word processor or as a stand-alone application) to obtain the relevant information.

The proposed direct guidance functions also successfully combat information overload and fulfil the needs of not only the learner of the language but also of the casual on-the-fly-user of the dictionary; its flexibility is intended to provide a step towards individualization. Different access points are available to the user depending on his/her pre-existing knowledge. It is therefore not a profile-based tool.

We envisage that such mechanisms be implemented as "plug-in modules" in entries of specific lemmas of an e-dictionary, i.e., an additional link/button is shown to the user on screen which (s)he can follow on demand. Such tools could therefore be used as writing tools integrated in a word processor, again activated by the user on demand, if (s)he requires to check the correct formulation of a complex grammatical construction, i.e., checking whether his/her own original construction is grammatically correct. Such additional functionality could be part of our future work — this is different from constructing a sentence in a word processor (or dictionary-linked tool), and rather similar to spelling and grammar checkers that currently occur in popular word processing software.

Future work includes the full-scale implementation of user support for complex structures proposed in this paper as a module within e-dictionaries. Identifying and categorising additional support techniques and developing prototypes and the full-scale implementation of such additional support techniques are also envisaged, as well as identifying further complex grammatical structures for which additional user support techniques may need to be developed. We will also investigate the possibility of the reuse of all such modules for user support in word processors and writing tools, as well as for language instruction and computer-assisted language learning (CALL).

Acknowledgement

References

Alonso Ramos, M., M. Garcìa Salido and O. Vincze. 2014. Towards a Collocation Writing Assistant for Learners of Spanish. Faaß, G. and J. Ruppenhofer (Eds.). 2014. *Workshop Proceedings of the 12th Edition of the KONVENS Conference, Hildesheim, Germany, October 8–10, 2014*: 77-88. Hildesheim: Universitätsverlag Hildesheim.

Bertels, A. and S. Verlinde. 2011. La lexicographie et l'analyse de corpus: nouvelles perspectives. *Meta: Journal des Traducteurs* 56(2): 247-265.

Bosch, S.E. and G. Faaß. 2014. Towards an Integrated e-Dictionary Application — The Case of an English to Zulu Dictionary of Possessives. Abel, Andrea, Chiara Vettori and Natascia Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*: 739-747. Bolzano/Bozen: EURAC Research.

Bothma, T.J.D. 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 71-102.

Bothma, T.J.D., U. Heid and D.J. Prinsloo. 2013. Implementing Decision Support for Text Production in e-Dictionaries. *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 2013, Tallinn, Estonia.

Bothma, T.J.D. and S. Tarp. 2012. Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.

Buyse, K. and S. Verlinde. 2013. Possible Effects of Free On Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of Linguee and the Interactive Language Toolbox. *Procedia: Social and Behavioral Sciences* 95: 507-512.

De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.

Fuertes-Olivera, P.A. and S. Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Lexicographica. Series Maior 146. Berlin/Boston: De Gruyter Mouton.

Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Lombard, D.P., E.B. van Wyk and P.C. Mokgokong. 1985. *Introduction to the Grammar of Northern Sotho.* Pretoria: Van Schaik.

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

Prinsloo, D.J. and S.E. Bosch. 2012. Kinship Terminology in English–Zulu/Northern Sotho Dictionaries — A Challenge for the Bantu Lexicographer. Fjeld, Ruth Vatvedt and Julie Matilde Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012*: 296-303. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Prinsloo, D.J., T.J.D. Bothma and U. Heid. 2014. User Support in e-Dictionaries for Complex Grammatical Structures in the Bantu Languages. Abel, Andrea, Chiara Vettori and Natascia Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*: 819-827. Bolzano/Bozen: EURAC Research.

Prinsloo, D.J., U. Heid, T.J.D. Bothma and G. Faaß. 2011. Interactive, Dynamic Electronic Dictionaries for Text Production. Kosem, I. and K. Kosem. 2011. *Electronic Lexicography in the 21st Century. New Applications for New Users*. *Proceedings of eLex 2011, Bled, 10–12 November 2011*: 215-220. Ljubljana: Trojina, Institute for Applied Slovene Studies. http://www.trojina.si/elex2011/elex2011_proceedings.pdf.

Prinsloo, D.J., U. Heid, T.J.D. Bothma and G. Faaß. 2012. Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22: 290-320.

Prinsloo, D.J., Daniel Prinsloo and J.V. Prinsloo. 2015. A Writing Tool for Sepedi. *E-Lex 2015*. Herstmonceux Castle, United Kingdom, 11–13 August 2015.

Taljard, E. and D.J. Prinsloo. 2013. *Lexicographic Treatment of So-called Cattle Colour Terms in Northern Sotho.* Paper delivered at AFRILEX 2013 — 18th Annual International Conference of the African Association for Lexicography, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa, 2–5 July 2013.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica Series Maior 134. Tübingen: Niemeyer.

Tarp, S. 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 54-70.

Tarp, S. 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28: 253-267.

Van Wyk, E.B., P.S. Groenewald, D.J. Prinsloo, J.H.M. Kock and E. Taljard. 1992. *Northern Sotho for First-Years*. Pretoria: Van Schaik.

Verlinde, S. 2011. Modelling Interactive Reading, Translation and Writing Assistants. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 275-286.

Verlinde, S., P. Leroyer, and J. Binon. 2010. Search and You Will Find. From Stand-alone Lexicographic Tools to User Driven Task and Problem-oriented Multifunctional Leximats. *International Journal of Lexicography* 23(1): 1-17.

Wanner, L, S. Verlinde and M. Alonso Ramos. 2013. Writing Assistants and Automatic Lexical Error Correction: Word Combinatorics. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia*: 472-487. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.

Ziervogel, D. 1969. *Handboek van Noord-Sotho*. Pretoria: Van Schaik.