

Lexicographic Data Boxes Part 3: Aspects of Data Boxes in Bilingual Dictionaries and a Perspective on Current and Future Data Boxes*

Rufus H. Gouws, *Department of Afrikaans and Dutch,
Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)*
and

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

Abstract: This article, the third in a series of three on lexicographic data boxes, firstly focuses on a number of aspects of data boxes in bilingual dictionaries with the emphasis on different approaches in bilingual dictionaries with an African language as one of the members of the treated language pair. It is not possible to provide a comprehensive discussion within the limitations of an article. Then the discussion proceeds by looking at some new ways of using data boxes in online dictionaries. It is shown that the possibilities of the new medium allow lexicographers to employ data boxes in both traditional and non-traditional ways. It is argued that data boxes are expected to fulfil a variety of purposes ranging from navigational information and the provision of salient information to giving access to relevant data in dictionary-internal and dictionary-external sources. Lexicographers of online dictionaries have introduced new ways of using data boxes that have not yet been fully discussed in metalexicographic literature. This article gives an identification and a brief discussion of some of these innovative uses of data boxes. It stresses the potential that the online environment offers lexicography. Practical and theoretical lexicographers need to be aware of these possibilities and challenges. By embarking on a more comprehensive use of data boxes dictionaries can become even better containers of knowledge and can serve their users in an optimal way.

Keywords: DICTIONARIES, DATA BOXES, POP-UP BOXES, HYPERLINKING, AFRICAN LANGUAGES, SEPEDI, ISIZULU, SEARCH DOMAIN, SEARCH UNIVERSE, DATA DISTRIBUTION

Opsomming: Leksikografiese datakassies. Deel 3. Aspekte van datakassies in tweetalige woordeboek en 'n perspektief op huidige en toekomstige data-

* This is the third in a series of three articles dealing with various aspects of lexicographic data boxes.

kassies. Hierdie artikel, die derde in 'n reeks van drie oor leksikografiese datakassies, fokus eerstens op aspekte van datakassies in tweetalige woordeboeke met die klem op verskillende benaderings in tweetalige woordeboeke met 'n Afrikataal as een van die lede van die behandelde taalpaar. Daarna gaan die bespreking voort deur te kyk na 'n paar nuwe maniere om datakassies in aanlyn woordeboeke te gebruik. Daar word aangetoon dat die moontlikhede wat die nuwe medium bied, leksikograwe in staat stel om datakassies op sowel tradisionele as nie-tradisionele maniere te gebruik. Daar word aangevoer dat die datakassies gebruik kan word om 'n verskeidenheid doeleindes te bereik, wat wissel van navigasie-inligting en die verskaffing van belangrike inligting tot toegang tot relevante data in woordeboek-interne en woordeboek-eksterne bronne. Leksikograwe van aanlyn woordeboeke het nuwe maniere bekendgestel om datakassies te gebruik wat nog nie volledig in die metaleksikografiese literatuur bespreek is nie. Hierdie artikel gee 'n identifisering en 'n kort bespreking van sommige van hierdie innoverende gebruike van datakassies. Dit beklemtoon die potensiaal wat die aanlynomgewing aan die leksikografie bied. Leksikograwe moet bewus wees van hierdie moontlikhede en uitdagings. Deur met 'n meer omvattende gebruik van datakassies te begin, kan woordeboeke selfs beter kennishouers word en hul gebruikers op 'n optimale manier dien.

Sluutelwoorde: WOORDEBOEKE, DATAKASSIES, OPSPRINGKASSIES, HIPERSKAKELS, AFRIKATALE, SEPEDI, ISIZULU, SOEKDOMEIN, SOEKUNIVERSUM, DATAVERSPREIDING

1. Introduction

The fact that data boxes are not used at all in many dictionaries and that they are often almost randomly used merely to bring together or highlight information could create the impression that data boxes have an insignificant role to play in dictionaries and should therefore only belong to the periphery of metalexigraphic discussions. In this article we wish to argue to the contrary, i.e. that data boxes are important and even essential but underutilized lexicographic components which should be used to fulfil specific needs. The user should be guided with regard to salient information which cannot typically be catered for by standard dictionary conventions and items such as those giving a paraphrase of meaning, translation equivalents and examples of use. We wish to make it clear that treatment in data boxes is not in competition with the default treatment in the article of a specific lemma. They supplement each other, default treatment is the first objective then consideration of a data box if the compiler deems it necessary to give further guidance on salient information. In Part 1 Gouws and Prinsloo (2021) and Part 2 Prinsloo and Gouws (2021) (this volume), lexicographic data boxes as text constituents in dictionaries and types and contents of data boxes were discussed with the purpose of setting the scene for the discussion of the future of data boxes and data boxes of the future in this article.

Moreover, as seen in section 3, data boxes did not lose their relevance in the transition from paper to electronic dictionaries but electronic dictionaries often employ alternative strategies for the presentation of data boxes enabled by the digital era.

Dictionaries have a genuine purpose, cf. Wiegand (1998: 299). This also applies to the different components of dictionaries, including the data boxes. Data boxes of the future should focus on what we believe is the genuine purpose, cf. Wiegand (1998: 299), of data boxes, i.e. guidance on salient data not sufficiently emphasised in the default lexicographic presentation. These include for example the contrasting of different words, aspects of the range of application, antiquation, taboos etc. However, the possibilities offered by the online environment and the innovative and dynamic options regarding the structure of dictionary articles should lead lexicographers to use data boxes in ways that include but also go beyond the mere representation of salient data. Where the first part of this article focuses on various aspects of data boxes presenting salient data, the second section moves towards new uses of data boxes. In the first section the discussion will be directed at some issues in bilingual dictionaries in which an African language is a member of the treated language pair. The second section will primarily be directed at online dictionaries in general but will also be relevant to future African language dictionaries.

2. Data boxes in bilingual dictionaries — African languages as a case in point

2.1 Different approaches to data boxes

Dictionaries for the African languages could firstly cater for the inclusion of data boxes dealing with issues not restricted to the given language pair. Secondly, they could include data boxes specific to the language family they belong to and finally data boxes dealing with unique features of individual members of the language family. With regard to these three issues data boxes can play a significant role in making the user aware of salient data. What is presented in this paragraph is a selection of a number of lemmas which should be considered for the provision of data boxes in addition to the standard treatment given in the dictionary article. The selected issues pertain to different aspects of morphology, syntax and semantics such as (a) demonstratives, (b) multiple recurring phrases as translation equivalents, (c) reference to men and women versus addressing them, (d) different constructions used for English adjectives and (e) equivalent relations. It is, however, not possible to present a comprehensive or systematic account of the full scope of required data boxes within the limitations of an article.

The African language Sepedi (a Sotho language) will be taken as example language in the following discussion with occasional reference to isiZulu (a Nguni language) both belonging to the Bantu¹ Language Family. To the knowledge of the authors the only Sepedi and isiZulu dictionaries using data boxes are the Oxford school dictionaries for Sepedi and isiZulu (henceforth ONSD and OZSD respectively).

As far as the first category is concerned it can simply be stated that data

boxes for African languages should give guidance on issues applicable to all languages such as contrasting related words, range of application, cultural considerations, etc. Secondly, attention should be given to typical characteristics of the language family such as verbal moods, nominal classes, kinship terminology, etc. Finally, data boxes should be included guiding the user on issues characteristic of the specific African language such as guidance on pronunciation, syntax, semantics, word division, etc. These issues are discussed in detail in Part 1 Gouws and Prinsloo (2021) and Part 2 Prinsloo and Gouws (2021). Consider figure 1 as an example dealing with contrast and range of application.

<p>Sepedi ** /Sepèdi/ noun 7/- ■ Pedi (language) ♦ Go tloga fao bareti ba Sepedi ba ile ba thoma go ngwala direto ka histori. <i>From there, the Pedi poets started writing poems about history.</i> In the current constitution, Sepedi is used to refer to Northern Sotho.</p>	<p>umzala * noun 1/2, 1a/2a (pl. abazala, omzala) ■ cousin ♦ Kunjalo khona mzala lixhoshwa libhekile. <i>It is like that, my cousin; it was an oversight.</i> The noun "umzala" is used for all types of cousins, except to refer to the children of the brother of your father.</p>
--	---

Figure 1: Data boxes for *Sepedi* and *umzala* in ONSD and OZSD — contrast and range respectively

The data box at *Sepedi* contrasts the names *Sepedi* versus *Northern Sotho* and informs the user that these terms refer to the same language. There is much controversy around the use of these names and what the relation or difference between these terms is, therefore guidance is required. The data box at *umzala* gives a precise indication of the range of application i.e. that it can be used to refer to cousins but not to the children of one's father's brother. A complicated system of kinship terminology exists for African languages, cf. Van Wyk and Haasbroek (1990) for Setswana and Prinsloo and Van Wyk (1992) for Sepedi. Data boxes can provide valuable guidance on kinship, e.g. on the range of application as in the case of *umzala* in figure 1 and contrasting kinship relations, e.g. relatives on father's versus mother's side, whether you address a specific relative versus speak about them. The shortage of space will always be a consideration in paper dictionaries and it is for the lexicographer to prioritise the type of information to be provided, e.g. treatment in the article by means of translation equivalents or providing a data box or both. So, for example, the compiler of OZSD and ONSD has valued the importance of text boxes at demonstratives as so important that he dedicated 40% of the page in figure 5 to text boxes.

As far as the second category is concerned, figure 2 illustrates a typical category which Sepedi and isiZulu (and probably all other members of the language family) have in common, i.e. guidance on grammatical data like nominal classes and their concords or that the English articles *a*, *an* and *the* are not translated/do not have equivalents.

<p>bo¹ *** <i>subject concord cl. 14</i> ■ it ♦ Bogobe bo a fiša. > Bo a fiša. <i>The porridge is hot.</i> > <i>It is hot.</i></p> <p>Subject concords of noun classes, here bo, are not translated when they appear together with the subject in the sentence.</p>	<p>the *** <i>article</i> ■ - ♦ We crossed the river at the drift. <i>Sawela umfula ezibukweni.</i></p> <p>The English definite article <i>the</i>, as well as the indefinite articles <i>a</i> and <i>an</i>, have no direct translations in Zulu. In English, <i>the</i> is used before nouns.</p>
--	---

Figure 2: Data boxes at *bo* in ONSD and *the* in OZSD

As far as data boxes dealing with unique features of Sepedi and isiZulu are concerned, much guidance is required in respect of what Prinsloo (2017 and 2020) call complicated grammatical structures. Consider figures 3 and 4 as examples of unique features of prefixing and composition of specific words in Sepedi and isiZulu.

<p>ngale ** <i>locative adverb</i> C <i>LE</i>² ■ far away in that direction; far away on the other side ♦ Lowo mfana ubalekile waye washona ngale. <i>That boy ran away and eventually disappeared there, far away in that direction.</i></p> <p>The instrumental formative "nga-" can also be prefixed to the demonstrative pronoun "le", at which point the meaning of "ngale" is "with or about these/this".</p>
--

Figure 3: Data box at *ngale* in OZSD

In the treatment of *ngale* in figure 3 a cross-reference to the locative adverb *le*² is given. The data box, however, refers to the demonstrative pronoun *le*¹ which is lemmatised and treated in its appropriate alphabetical position in OZSD.

<p>nyakang <i>verb</i> ■ searching for what?; looking for what? ♦ Batho ba ba re ba nyakang? <i>What do these people say they are looking for?</i></p> <p>This form comes from the verb nyaka and the question word eng.</p>
--

Figure 4: Data box at *nyakang* in ONSD

For most African languages strong normative guidance is required since standardization is still in progress, cf. Gallardo (1980: 62). Such boxes could well be high on the list of typical data box content for these languages and OZSD and ONSD have done well in the provision of valuable information for the users in

data boxes. Consider the following six examples of data boxes for Sepedi dictionaries which could substantially enhance their value in respect of user guidance suggested as model entries for future Sepedi dictionaries.

2.2 Demonstratives

Demonstratives basically express *this* or *these* in relation to three relative distances within sight of the speaker, e.g. *monna yo* 'this man', *monna yoo* 'that man' and *monna yola* 'that man over there, yonder'. Linguists such as Louwrens (1991), Van Wyk et al. (1992), Lombard (1985) and Poulos and Louwrens (1994) distinguish three basic positions but differ in respect of the sub-positions into which demonstratives can be classified. Louwrens (1991) distinguishes between the different positions as follows:

Position 1(A) Speaker and the addressee are *close to one another*, while the object referred to is relatively *near them*

Position 1(B) Speaker and the addressee are *at a distance from each other*, while the object referred to is directly next to the speaker

Position 2(A) Speaker and the addressee are *relatively far apart*, while the object referred to is *nearer to the addressee*

Position 2(B) Refer to objects which are *very close or directly next to the addressee*

Position 3 Speaker and the addressee are *very close to one another*, while the object referred to is *far away* from them

Louwrens (1991: 106-108)

Consider table 1 as an extract from the table given in ONSD.

Table 1: Extract from the table for demonstratives in ONSD

Cl.	Pref.	Pos. 1	Pos. 2	Pos. 3	Pos. 1a	Pos. 1b	Pos. 2b
1	mo-	yo	yoo	yola	yono	yokhwi	yowe
2	ba-	ba	bao	bale	bano	bakhwi	bawe
---	---	---	---	---	---	---	---
9	(N)-	ye	yeo	yela	yeno	yekhwi	yewe
10	di-	tše	tšeo	tšela	tšeno	tšekhwi	tšewe
---	---	---	---	---	---	---	---
14	bo-	bjo	bjoo	bjola	bjono	bjokhwi	bjowe
---	---	---	---	---	---	---	---

ONSD: S18 and S19

Two issues pertaining to demonstratives are relevant to the user i.e. firstly a complete table indicating all the demonstratives of the different positions and

classes and their basic meanings, i.e. indicating three distances, 'here', 'there' and 'there (yonder)'. Secondly indication of the exact semantic relations in respect of speaker and addressee is required. The lexicographer could, for example, give the full table, e.g. as in table 1 and the basic translations as the reference address e.g. in the back matter of a paper dictionary or as a clickable pop-up window in an electronic dictionary — see the next section. The purpose is a complete illustration of the different classes and positions of demonstratives and their meanings and translation equivalents. The meanings of the different positions as described by Louwrens (1991: 106-108) above could be presented as pop-up boxes for each demonstrative or as data boxes in the central text of a paper dictionary as has successfully been done in ONSD as in figure 5.

yeng → Yuropa

254 Sesotho sa Leboa-Seisimane

yeng *iyēngi* verb = *pl. marker (ng)* & *va*¹
 o a re **yeng** = let's go; all of us (must) go • A re **yeng** go robala. *Let's go to sleep.*

yeno *iyēna* demonstrative pos. *Ia* & *va*¹ 1 (cl. 9)
 1 (cl. 9) = this (one) • O be a le tseleng ya go tla thoko **yeno**. *He was on his way to come to this side.* 2 (cl. 4) = these (ones) • Padi ye e bolela ka dišo tša mehleng **yeno**. *This novel talks about things of these times.*

This demonstrative indicates that the speaker and the addressee are relatively far apart, and that the object/object referred to is/are directly next to the speaker.

yeo *** *iyēbi* demonstrative pos. *Ii* & *va*¹ 1 (cl. 9)
 1 (cl. 9) = that (one) • Karata **yeo** ke ya go reka diaparo. *That card is for buying clothes.* 2 (cl. 4) = those (ones) • O ša kae mekotla **yeo**? *Where are you taking those bags?*

This demonstrative indicates that the speaker and the addressee are relatively far apart, and that the object/object referred to is/are closer to the addressee, but still some distance away from her/him.

yešo *iyēša* communal possessive pronoun cl. 4, cl. 9 = our; of our place/family/homestead • Re tša šomiša koloi **yešo** go go rwalela diphablo. *We will use our (family's) car to transport your possessions.*

yo *** *iyōi* demonstrative pos. *I* cl. 1 = this (one) • Monna **yo** ke malome. *This man is my uncle.*

This demonstrative indicates that the speaker and the addressee are close to each other, and the person referred to is near them.

yo² *** *iyōi* aspectual prefix = go(ing) to; go and ... • Sepela, o **yo** tšufa monna yoo wa batho. *Go and help that poor man.*

yohle * *iyōhli* quantitative pronoun 1 (cl. 9)
 = the whole • Naga **yohle** e aparetšwe ke mafula. *The whole country is flooded.* 2 (cl. 4) = all • Mehlare **yohle** e tša ripša. *All the trees will be cut down.*

yola *** *iyōla* demonstrative pos. *Iii* cl. 1 & *vo*¹ = that (one) • Mpilešše mošemane **yola**. *Call that boy over there for me.*

This demonstrative indicates that the speaker and the addressee are very close to each other, and that the person referred to is far away.

yona¹ *** *iyōna* absolute pronoun 1 (cl. 9)
 = she/he/it (in particular); her/him/it (in particular) • Ngwalolla tema ye, o tšatša dikgoba tše di lego mo go **yona**. *Rewrite this paragraph and fill in the gaps that appear in it.* 2 (cl. 4) = they (in particular); them (in particular) • Meetse a šomišwa le ge go agwa **yona** mengwako le meago ye mengwe. *Water is used when in particular houses and other buildings are built.*

yona² *** *iyōna* possessive pronoun cl. 9, cl. 4
 • [PC +] **yona** 1 (cl. 9) = her(s)/his/its • Mmusuša o kgopetše Kgoro ya Maphelo go lekola pholisi ya **yona** ka leswa. *The government has asked the Department of Health to review its policy again.* 2 (cl. 4) = their(s) • Le dire ka melao ya **yona** ka moka. *You must act in accordance with all their laws.*

yoo *** *iyōo* demonstrative pos. *Ii* cl. 1 & *vo*¹ = that (one) • Mosadi **yoo** a lego kgauswi le wena o nyaka eng? *What does that woman want?*

This demonstrative indicates that the speaker and the addressee are relatively far apart, and that the person referred to is closer to the addressee, but still some distance away from her/him.

yowe *iyōwō* demonstrative pos. *Iii* cl. 1 & *vo*¹ = that person • Monna **yowe** ga se lephodisa. *That person is not a police officer.*

This demonstrative indicates that the speaker and the addressee are some distance apart, and that the object referred to is very close to the addressee.

yunibesithi * *iyunibesithi* noun 9/10 (pl. *diyunibesithi*) = university • Baitšuti ba Yunibesithi ya Limpopo ba tša thoma ditšahlobo kgwedi ye e tšang. *Students at the University of Limpopo will start their exams next month.*

• **yunibesithing** /yunibesithing/ (pl. *diyunibesithing*) = at the university • Go na le ditšuto tša go fapana **yunibesithing**. *There are different courses at university.*

Yuropa /yuropa/ noun 9/- = Europe • Papa wa ka o šoma Yuropa bjalo ka morutši. *My father is working in Europe as a teacher.*

Figure 5: Data boxes for demonstratives in Sepedi in ONSD

In figure 5 the compiler regarded the salient information given by data boxes for demonstratives as so important that they are provided for each demonstrative and not only as a separate section in e.g. the back matter of the dictionary. Such a decision remains the prerogative of the compiler in consideration of the skills level of the target user.

2.3 Multiple recurring phrases as translation equivalents

These are cases where an English word can be translated by means of a grammatical pattern determined by the different noun classes, — *every* as a typical case in figure 6.

every *** adjective 1 ■ ka moka ♦ Not every person has registered for the elections. *Ga se batho ka moka bao ba ingwadišeditšego dikgetho.* 2 ■ (cl. 1, cl. 3) ofe goba ofe ■ (cl. 5) le lengwe le le lengwe; lefe goba lefe ■ (cl. 9) efe goba efe; efe le efe

Figure 6: *Every* in ONSD

The duplication of the adjective construction, as in figure 6, i.e. Class 5 *le lengwe le le lengwe* (*le lengwe*: a certain/other (day) + *le*: and + *le lengwe*: another one) reflects a single instance of the recurring pattern for all other classes. This example is fine but it is important to inform the user by means of a data box that this can be done for all classes keeping in mind that the concords used have to match the nominal class to which the noun belongs. A data box as in figure 7 which gives examples from more noun classes is recommended at the article for *every*. Such a box should preferably have a cross-reference to the back matter where a full table with a description of the strategy, i.e. that the concept 'every' is expressed by means of the duplication of 'another' (e.g. *monna yo mongwe* 'another/certain man' *le* 'and' *yo mongwe* 'another one') is given.

Every

Can be expressed for all classes by duplication of the adjective construction:

Class. 1: Monna yo mongwe le yo mongwe 'every man'

Class. 3: Monwana wo mongwe le wo mongwe 'every finger'

Class. 5: Lesogana le lengwe le le lengwe 'every young man'

Class. 7: Selepe se sengwe le se sengwe 'every axe'

Class. 9: Puku ye nngwe le ye nngwe 'every book'

Figure 7: Proposed data box for *every* in Sepedi dictionaries

Prinsloo and Gouws (2006) describe this type of repetition of a phrase across the different classes as in figure 7 as grammatical divergence and all such occurrences belonging to different classes, e.g. *this man/finger/axe*, etc. or *he/she/him/her*, etc. could be treated in data boxes with great success.

2.4 Reference to men and women versus addressing them

Groot Noord-Sotho-woordeboek (GNSW) gives the following translation equivalent paradigm for *mohumagadi*: "queen, king's wife, chief's wife, chieftainess, lady, Mrs [a term of courtesy applied to any married woman]", *mohumagatšana*: "miss, queen (of cards)" and for *mosadi*: "woman, married woman, wife". Although all three words refer to a woman/adult female person, the user should be warned that it is inappropriate to address a woman as *mosadi*. The same holds true for a man/adult male person *monna* 'a man' versus *morena* 'Mr.' Consider the suggested data boxes for *women* and *men* in figure 8. This is pragmatic data, a function of data boxes. It is for the lexicographer to decide whether it should be emphasized by inclusion in a data box.

<p>Although <i>mosadi</i>, <i>mohumagadi</i> and <i>mohumagatšana</i> all refer to women, do not address a woman as <i>mosadi</i> e.g. when greeting her, use <i>mohumagadi</i> / <i>mohumagatšana</i>: <i>Dumela mohumagadi!</i> Good day madam!</p>	<p>Although <i>monna</i> and <i>morena</i> both refer to men, do not address a man as <i>monna</i> e.g. when greeting him, use <i>Morena</i>: <i>Dumela morena!</i> Good day sir!</p>
---	---

Figure 8: Proposed data boxes for *addressing females and males* in Sepedi dictionaries

The data box in figure 8 or the applicable sections thereof could be given at the articles of *man* and *woman*.

2.5 Hair

Compilers of Sepedi dictionaries should give clear guidance on the correct meanings and use of Sepedi words dealing with different kinds of hair. ONSD lemmatised *hair* and gives a translation equivalent *moriri*. This is the singular form, i.e. a/one hair. It would be better to give *meriri* 'hair (plural)' as transla-

tion equivalent since in most cases reference to the plural is made. No data box is suggested here, only different treatment of the lemma.

No mention is made of, e.g. the hair of an animal or guidance that *hair* is normally used in the plural form in Sepedi, i.e., *meriri*. GNSW gives the following translation paradigm for *boya*: "hair of an animal, wool, hair of human body (but not of head)" and translates *meriri* as human hair and *mariri* as mane (of a lion). ONSD translates *boya* as wool, animal hair, fur and *mariri* as mane and adding "of a lion" in brackets. *Meriri* is lemmatised but not treated and cross-referred to the singular *moriri* which is simply translated as hair — it should have been translated as human hair. Stronger guidance is required in respect of *hair*, *animal hair*, *mane*, *boya* and *mariri* in order to prevent the user from, e.g. incorrectly using *meriri* to refer to wool or animal hair or *boya* to refer to hair on the head of a person. A data box contrasting *boya*, *meriri* and *mariri* as in figure 9 is recommended at the article of *hair*.

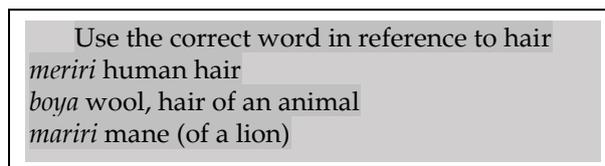


Figure 9: Proposed data box for *boya*, *meriri* and *mariri*

2.6 Different constructions used for English adjectives

A number of English adjectives such as *last*, *own*, *naughty*, etc. are not expressed as adjectives in Sepedi but through different constructions. So, for example, is there no single-word adjective for *naughty* in Sepedi — it is expressed by either a full sentence in the relative mood or by means of a possessive construction as in example (1).

(1)

a. Verbal relative

Mošemane yo a selekago (*mošemane* noun class 1 'a boy' + *yo* demonstrative class 1 + *a* subject concord class 1 + *seleka* verb stem 'be naughty' + *go* relative suffix) 'A naughty boy'

b. Possessive construction

Mošemane wa go seleka (*mošemane* noun class 1 'a boy'+ *wa* possessive concord class 1 + *go* infinitive class prefix class 15 + *seleka* verb stem 'be naughty' 'A naughty boy')

A data box, e.g. as in figure 10 presented at the article of *naughty* will provide the required guidance to the user provided that the target users should have

basic grammatical knowledge of Sepedi. If not, grammatical terms such as *verbal relative* and *possessive construction* should be briefly described in terms of their meaning, i.e. "who is doing something, something of something else respectively". Both options can even be given with the semantic one in brackets, i.e. Verbal relative (who is doing something) and Possessive construction (something of something else).

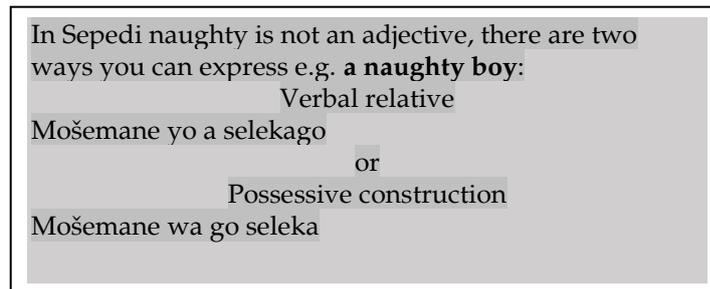


Figure 10: Proposed data box for *naughty*

2.7 Equivalent relations

As a final example consider instances of semantic divergence where a polysemous source language word has more than one translation equivalent (Gouws and Prinsloo 2005). A single Sepedi word *bala* has different translation equivalents, namely *read*, *count* and *study*. Sepedi has two homonyms *-tala*. The one member of the homonym pair has *old* as its translation equivalent whereas the second homonym has both *green* and *blue* as equivalents. ONSD translates the homonyms *-tala* correctly as respectively *old* and *green*, *blue*. The user will be well-guided if alerted by means of a data box such as figure 11 because being able to distinguish between *green* and *blue* could be vital in text production situations. A typical situation could be where it is crucial to distinguish between different specific functions performed by e.g. green versus blue buttons on a control panel.

The *Sesotho sa Leboa / English Pukuntšu Dictionary* (SEPD) could mislead the user because only *blue* is given as translation equivalent for *-tala*. The proposed data box in figure 11 could be placed at the article of *-tala* in the Sepedi to English side as well as at the articles for *blue* and *green* in the English to Sepedi side of the dictionary to warn the user that additional clarification might be required in text production situations. It is for the compiler to decide whether sufficient guidance in respect of *-tala* translated as *green* or *blue* was given in the default treatment of the lemma or whether a data box is desired to focus the attention of the user on the different senses. So, for example, the compiler of SEPD will be well-advised to firstly give *green* also as translation equivalent for *-tala*, illustrated by typical examples for each equivalent and further supported by a data box.

-tala: *blue versus green*
Legodimo le letala 'blue sky'
Bjang bjo botala 'green grass'

Figure 11: Proposed data box for *tala*

The same holds true for a data box such as figure 12 for *-bala* where the lemma *-bala* has *read*, *count* and *study* as translation equivalents. Although examples will help the user, real success in the treatment of *-tala* and *-bala* would at best be achieved by means of a data box that displays these salient semantic issues.

Bala can mean read
OR count OR study e.g.
Ke bala dikuranta 'I read
the newspapers' or 'I count
the newspapers' Ba bala
tikrii ya BA. 'They study for
a BA degree'

Figure 12: Proposed data box for *bala*

In contrast to one Sepedi word having more than one English equivalent as in figures 11 and 12, a single English word can also have more than one Sepedi equivalent. Sepedi has two words for *ask*, i.e. *botšiša* 'ask, e.g. a question' and *kgopela* 'ask for something'. Consider figure 13 as a data box giving the required guidance on the range of application for *botšiša* and *kgopela*. Such a data box is especially required in the English to Sepedi side of dictionaries such as SEPD where *ask* is simply translated as *botšiša*, *kgopela* without any indication of the range of application.

In Sepedi there is a difference
between asking a question and
asking for something:
Go botšiša potšišo 'to ask a question'
but
Go kgopela tšhelete 'to ask money'
Go kgopela tshwarelo 'to ask for
forgiveness'

Figure 13: Proposed data box for *ask*

Data boxes are also required at the articles of *wear*, e.g. *apara/apere* 'wear clothes' versus *rwala* 'wear a hat' and the many Sepedi equivalents for *close*, e.g. *tswalela* 'close a gate/door' versus *khurumetša* 'close a container, e.g. the lid of a bottle' versus *khupetša* 'conceal' versus *moma* 'close(d) mouth', etc.

All these data boxes convey salient information which should be presented to users in a way that draws their attention. The use of data boxes is an ideal presentation method to enable such a transfer of information.

3. Salient data and more than salient data

Data boxes can be regarded as important and even essential but often underutilized dictionary components which should be used to fulfil a specific need, i.e. typically guiding the user towards carriers of salient data. From the preceding sections it should be clear that data boxes can contribute in a systematic way to assist in the presentation of data that cannot be sufficiently accommodated in the default search positions of dictionary articles or article stretches. The system prevailing in the decision to use data boxes is based on the salience of the specific data. The lexicographic method of using data boxes should not be performed in a haphazard way or as a form of lexicographic face-lifting, cf. Wiegand and Gouws (2011: 238). Lexicographers should have a clear understanding of the reasons why these boxes are used. The presentation of salient lexicographic data can be regarded as one of the major motivations for the use of data boxes in printed dictionaries. Irrespective of what happens in the development of online dictionaries printed dictionaries should preferably continue to use data boxes and even to increase their use. Innovative strategies could complement the traditional way of using these dictionary components. Printed dictionaries of the future could employ data boxes in various ways to respond to new lexicographic challenges.

Data boxes also have an important role to play in online dictionaries. The examples and discussion of data boxes in the following sections of this article should not be regarded as of a language-specific nature but rather relevant to all languages, including the African languages.

Online dictionaries, especially those that were originally planned and published as printed dictionaries, often use data boxes in the same way as found in printed dictionaries. Figure 14 shows the use of a data box as found in the article of the lemma sign *underground* in the OALD to present a usage note: (a comparable usage note is also presented in the articles of the lemmata *metro*, *subway* and *tube*):

underground NOUN



'ʌndəgraʊnd  BrE ; 'ʌndəgraʊnd  NAmE

1 often the Underground (BRITISH ENGLISH) (NORTH AMERICAN ENGLISH **subway**) [SINGULAR] an underground railway/railroad system in a city

- *underground stations*
- *the London Underground*
- *I always travel by underground.*

► **compare** METRO, TUBE

2 the underground [SINGULAR + SINGULAR OR PLURAL VERB] a secret political organization, usually working against the government of a country

 **Usage note:** underground / subway / metro / tube

- A city's underground railway/railroad system is usually called the **underground** (often **the Underground**) in *British English* and the **subway** in *North American English*. Speakers of *British English* also use **subway** for systems in American cities and **metro** for systems in other European countries. **The Metro** is the name for the systems in Paris and Washington, D.C. London's system is often called **the Tube**.

Figure 14: Data box in the article of the lemma *underground* in the OALD

In addition, online dictionaries also display new ways of utilising data boxes. This was already alluded to in Prinsloo and Gouws (2021). This can be seen in figure 15 where the *Merriam-Webster* uses data boxes for navigation in the article of the lemma sign *dull*:

- ↓ Other Words from *dull*
- ↓ Synonyms & Antonyms
- ↓ Choose the Right Synonym
- ↓ More Example Sentences
- ↓ Learn More about *dull*

Figure 15: Data box in the article of *dull* for navigation in *Merriam-Webster* for navigation to other search positions

A click on these navigation boxes guides the user to the relevant addresses as seen in figure 16, the address of the link to *Synonyms & Antonyms*:

Synonyms & Antonyms for *dull*

Synonyms: Adjective

blunt, blunted, dulled, obtuse

Synonyms: Verb

benumb, blunt, cauterize, damp, dampen, deaden, numb

Antonyms: Adjective

cutting, edged, edgy, ground, honed, keen, pointed, sharp, sharpened, whetted

Antonyms: Verb

sharpen, whet

[Visit the Thesaurus for More](#) 

Figure 16: The address of the link in one of the data boxes in the article of *dull*

The boxes shown in figure 15 do not only have a navigational purpose. They give users the opportunity to unlock textual venues that accommodate additional lexicographic data, as seen in figure 16.

4. Innovative uses of data boxes in online dictionaries

The transition from printed to online dictionaries can rightly be regarded as extremely important with radical and far-reaching consequences. This can be seen in many aspects of online dictionaries, for example, as indicated by Heuberger (2020: 404), with regard to accessibility of data, multimedia functions, customization, hybridization, user input and storage space. The transition to online lexicography has also had a huge impact on research in the field of metalexicography. Theories of lexicography were primarily developed for the printed environment. The online environment demands a re-assessment of all aspects of these theories, including the various dictionary structures. Some structures of printed dictionaries, for example the article structure, will also prevail in online dictionaries although certain adaptations are needed; some structures, for example the frame structure, are not maintained in online dictionaries. In addition, online dictionaries can also display structures that do not occur in printed dictionaries. An example of such a structure is the screenshot structure, cf. Gouws (2014: 165). When using an online dictionary the user is confronted by various screenshots that are populated by dictionary articles and partial articles. These screenshots display innovative uses of data boxes. With regard to the use of data boxes online dictionaries show that the lexicographic practice has embarked on procedures not yet adequately described or discussed in metalexicographic publications. In the subsequent sections of this article a few occurrences of data boxes in online dictionaries will be identified and briefly discussed in order to show the need for a comprehensive look at data boxes of the future.

4.1 Highlighting data types

Online dictionaries often have dynamic article structures and even multi-layered dynamic article structures (Gouws 2014: 165). The internal access structures provide the user with access routes to the required data in its specific search zone and article layer. This is seen in *ellexiko* where the opening screenshot of an article contains data indicators that help the user to move to a next layer of the article and then perhaps to a further layer. When reaching a specific search zone the data indicator as well as the search zone is boxed by means of a thin frame. This frame helps the user to identify the boxed items as a destination of the search route. The following screenshots show this process. Figure 17 is the opening screenshot of the article of the lemma *Arm* (arm) in *ellexiko*:

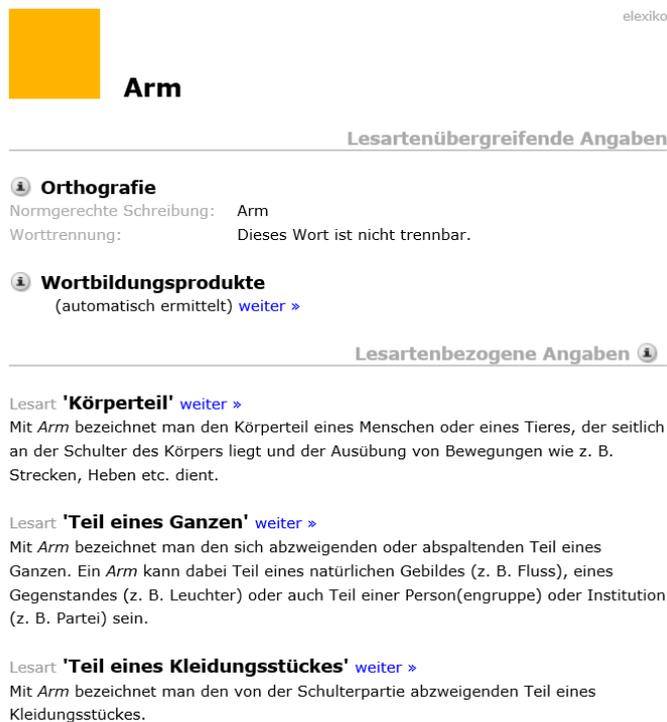


Figure 17: The opening screenshot of the article of the lemma *Arm* in *ellexiko*

A user looking for grammatical data regarding the sense of this word referring to a body part finds the data indicator *Körperteil* (Body part) and clicks on the entry *weiter* (=further) next to it. This click moves the user to the partial article presented in figure 18:

The screenshot shows the 'Arm' lemma page in the 'elexiko' database. The page title is 'Arm' with the sub-label 'Lesart: 'Körperteil''. Below the title, there are navigation links: 'zur Übersichtsseite' and 'Lesarten im Überblick'. A horizontal menu contains several categories: 'Bedeutungserklärung', 'Kollokationen', 'Konstruktionen', 'Sinnverwandte Wörter', 'Gebrauchsbesonderheiten', and 'Grammatik'. The 'Bedeutungserklärung' category is selected, and its content is displayed in a box. The content includes a title 'Erläuterung der Bedeutung / Funktion', a paragraph explaining the meaning of 'Arm', and links for 'Belege anzeigen' and 'Illustrationen anzeigen'. Below the paragraph, it states 'Wortklasse: Partitivum'. A 'Drucken' button is located at the bottom right of the page.

Figure 18: First comment on semantics in the article of the lemma *Arm* in *elexiko*

This screenshot shows the paraphrase of meaning of the specific sense of *Arm* with a thin line putting the paraphrase of meaning, its appropriate data indicator (*Bedeutungserklärung*) (=explanation of meaning) as well as links to example sentences (*Belege*) and illustrations (*Illustrationen*) in a data box. To the right of the data indicator bar the user can find the indicator *Grammatik* (=Grammar) and a click on that indicator opens the next layer, as seen in figure 19:

The screenshot shows the 'Arm' lemma page in the 'elexiko' database, with the 'Grammatik' category selected. The 'Grammatik' section is displayed in a box. It contains the following information:

Wortart:	Nomen (Maskulinum)
Nominativ Singular:	(der/ein) Arm
Genitiv Singular:	(des/eines) Armes (meist)
	(des/eines) Arms (oft)
Nominativ Plural:	(die) Arme

Figure 19: The search zone for grammatical data in the article of *Arm* in *elexiko*

This screenshot shows the grammatical data along with the relevant data indicator appearing in a thinly framed data box.

Both figure 18 and figure 19 display data that are part of the default treatment of nouns in this dictionary. The data box is not used to distinguish salient from less salient data with regard to the presentation in the article as a whole but it does highlight the data salient for the specific consultation — the destination unlocked by the preceding click of a data indicator. This use of data boxes is done in a consistent and systematic way in *lexiko*. It highlights the identification of specific items and enhances the retrieval of the required information. In addition, the type of data box in figure 18 and figure 19 also contributes to improve the layout of the screenshot. This approach is made possible by the dynamic nature of articles in online dictionaries and is in sharp contrast to the limitations due to the static nature of articles in printed dictionaries.

This use of data boxes can also be seen in the following partial article of the lemma sign *koekje* (figure 20) in the Dutch dictionary ANW (*Algemeen Nederlands Woordenboek*). Having navigated from the opening screenshot of the article to the screenshot presenting the partial article in which the sense of *koekje* "small cake" is treated, the user finds a typical partial article layout with three sections presented in columnlike way.

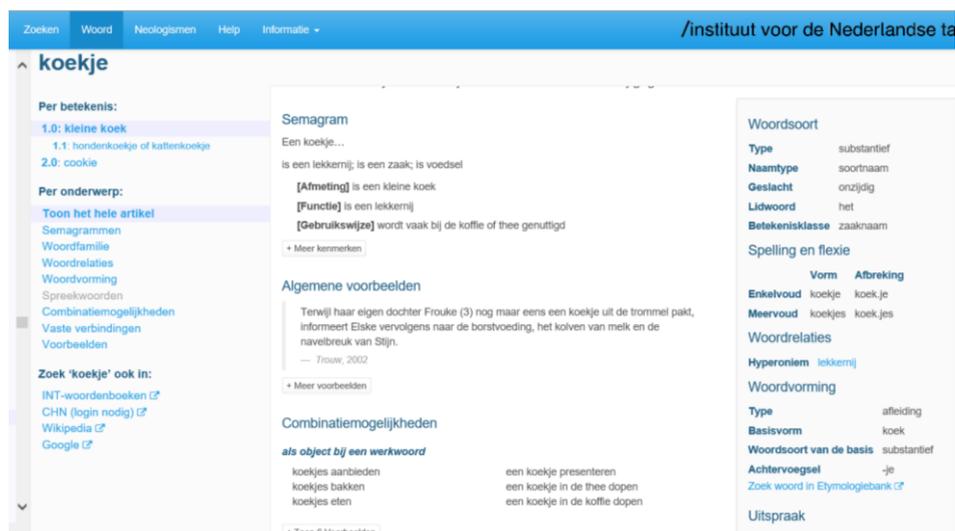


Figure 20: Partial article of the lemma *koekje* in the ANW

The right-hand section contains a data box, unfortunately not as clearly visible in the figure, that accommodates items giving part of speech, spelling and inflection, word relations and pronunciation. In all articles of this dictionary items giving these data types are presented in a data box, situated in the same

position in the screenshot. A similar data box is also seen in figures 21 and 22, giving screenshots with partial articles for two senses of the lemma *muisk* (mouse):

The screenshot shows the ANW partial article for the lemma *muisk* (mouse), sense 1.0. The article is structured as follows:

- Per betekenis:**
 - 1.0: bekend klein knaagdier
 - 1.1: *muisk* in dierkundige zin
 - 1.2: voorstelling van een muisk
 - 2.0: verheven stuk handpalm
 - 3.0: accessoire bij een computer
 - 4.0: geslachtsorgaan van een vrou
- Per onderwerp:**
 - Toon het hele artikel
 - Semagrammen
 - Woordfamilie
 - Woordrelaties
 - Woordvorming
 - Spreekwoorden
 - Combinatiemogelijkheden
 - Vaste verbindingen
 - Voorbeelden
- Zoek 'muisk' ook in:**
 - INT-woordenboeken
 - CHN (login nodig)
 - Wikipedia
 - Google
- muisk 1.0**
 - 

(Publiek domein)
 - Semagram**
 - Een muisk...
 - is een zoogdier; is een dier
 - [Afmeting]** wordt ongeveer 3 tot 10 cm lang, de staart niet meegerekend
 - [Geluid]** piept; maakt een piepend geluid
 - [Geur]** heeft een muffe geur
 - + Meer kenmerken
 - Wetenschappelijke naam:** *Mus musculus*
 - + Toon details
 - Algemene voorbeelden**
 - Het leek op het spel van de kat van de bureu met een muisk; erop af springen, met een poot slaan, even wachten en weer een tik uitdelen tot het
- Woordsoort**
 - Type** substantief
 - Naamtype** soortnaam
 - Geslacht** mannelijk of vrouwelijk
 - Lidwoord** de
 - Betekenisklasse** diernaam
 - Spelling en flexie**

	Vorm	Afbreking
Enkelvoud	muisk	muisk
Meervoud	muisk	muisk
Verkleinvorm	muiskje	muiskje
 - Woordrelaties**
 - Hyperoniem: [voed](#)

Figure 21: Partial article of the lemma *muisk* in the ANW

The screenshot shows the ANW partial article for the lemma *muisk* (mouse), sense 3.0. The article is structured as follows:

- Per betekenis:**
 - 1.0: bekend klein knaagdier
 - 1.1: *muisk* in dierkundige zin
 - 1.2: voorstelling van een muisk
 - 2.0: verheven stuk handpalm
 - 3.0: accessoire bij een computer
 - 4.0: geslachtsorgaan van een vrou
- Per onderwerp:**
 - Toon het hele artikel
 - Semagrammen
 - Woordfamilie
 - Woordrelaties
 - Woordvorming
 - Spreekwoorden
 - Combinatiemogelijkheden
 - Vaste verbindingen
 - Voorbeelden
- Zoek 'muisk' ook in:**
 - INT-woordenboeken
 - CHN (login nodig)
 - Wikipedia
 - Google
- muisk 3.0**
 - 

Bron: Darkone (CC licensed)
 - Betekenisbetrekking**
 - metafoor
 - Betrokken betekenissen** 1.0 : 3.0
 - Semagram**
 - Een muisk...
 - is een apparaat
 - [Afmeting]** past onder een handpalm
 - [Vorm]** heeft een ellipsvormige, platte onderkant en een bolle bovenkant
 - [Attribuut]** heeft aan de voorzijde twee grote drukknoppen en vaak een scrollwielje
 - + Meer kenmerken
 - Algemene voorbeelden**
 - Het Web combineert als eerste alle bestaande informatievormen op zo'n manier dat de gebruiker niet veel meer hoeft te doen dan met het pijltje van de muisk keuzes te maken.
- Woordsoort**
 - Type** substantief
 - Naamtype** soortnaam
 - Geslacht** mannelijk of vrouwelijk
 - Lidwoord** de
 - Betekenisklasse** zaaknaam
 - Spelling en flexie**

	Vorm	Afbreking
Enkelvoud	muisk	muisk
Meervoud	muisk	muisk
Verkleinvorm	muiskje	muiskje
 - Woordrelaties**
 - Hyperoniem** [apparaat](#)
 - Synoniem** [computermuisk](#)
 - Woordvorming**

Figure 22: Partial article of the lemma *muisk* in the ANW

In figures 20–22 data boxes continue their assignment as containers of lexicographic data but they play an additional role, namely to improve the article

layout and make data easier accessible to the users due to a conspicuous way of presentation. This is a function of data boxes that still needs further exploration. It demands dedicated future work which falls beyond the scope of this article.

4.2 Adding data

4.2.1 Lexicographic data

In online dictionaries data boxes are also used to highlight the access to additional dictionary-internal data that the lexicographer regards as relevant to the word treated in the specific article. Some articles in *Merriam-Webster* have a section "From the editors of *Merriam-Webster*." Below this heading a data box is given in which different types of data can be found. The data are usually of a lexicographic nature and help to fulfil a cognitive function. The data could be a reference to other articles in the dictionary that contain words in the same semantic field as the lemma or it can focus on a discussion of certain related aspect. In figure 23, a screenshot of a partial article of the lemma *bicycle*, shows this data box with its data indicator "10 words every true cyclist will know." A click on this data indicator in the box guides the user to a list of ten articles. This list includes articles with lemmata like *penny-farthing*, *peloton*, *velocipede* and *tandem bicycle*.

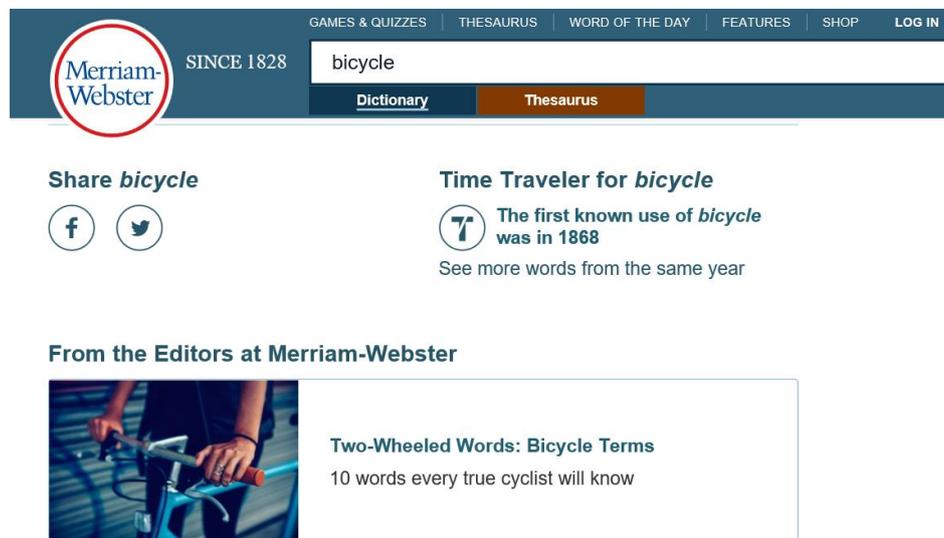


Figure 23: Additional data in *Merriam-Webster*

In this list the article of the lemma *velocipede* has an extensive treatment — with the paraphrase of meaning comparable to that given in the article of the lemma *velocipede*, as seen in figure 24:

Definition of *velocipede*

: a lightweight wheeled vehicle propelled by the rider: such as

a : TRICYCLE

b : a 3-wheeled railroad handcar

c *archaic* : BICYCLE

Figure 24: Paraphrase of meaning of *velocipede* in the MW

The treatment of *velocipede* in the article list is as seen in figure 25:

"Velocipede



Cyclist on Michaux velocipede, 1868

Definition: a lightweight wheeled vehicle propelled by the rider

The predecessor of the pedal bicycle was a two-wheeled vehicle that was propelled with the feet while seated. It was patented in 1818 in Germany, and its name, *Laufmaschine*, ("running machine") was trans-

lated into French as *vélipède* (from the Latin roots meaning "swift" and "foot"). In English, this contraption was known as the *dandy horse* or the *hobbyhorse*.

Velocipede became a term for any wheeled vehicle propelled by the rider in English, including early bicycles. In French, it became the term used for the improved pedal-powered version designed in the mid-1800s. This was shortened to *vélo* to become their modern word for "bicycle."

The pedal velocipedes were nicknamed *boneshakers* because they were made with wooden wheels and iron frames. After the introduction of rubber tires, *boneshaker* became a slang term for an uncomfortable or outmoded bicycle.

The *veloc-* of *velocipede* is also the root of *velocity*."

Figure 25: The treatment in the article list of the lemma *velocipede* in the MW

This use of data boxes like that in the article of the lemma *bicycle* shows a significant change in the way in which lexicographers employ this article component — data boxes present a departure slot from where the user can depart to article-external but dictionary-internal data venues. By including these isolated thematically-bound article stretches the lexicographer increases the extent of the dictionary as a search region and the relevant data boxes ensure access to these new venues in the search region.

4.2.2 Non-lexicographic data

Online dictionaries contain typical lexicographic data. Data boxes participate in accommodating the lexicographic data. However, the online environment opens possibilities for dictionaries to become containers of more than just traditional lexicographic data. As components of dictionary articles data boxes in online dictionaries can contain data that even go beyond a display of lexicographic data relevant to the treatment of the word represented by the lemma sign of the specific article. The data distribution structure of these dictionaries can also make provision for the satisfaction of more general cognitive needs. Irrespective of the lemma functioning as guiding element of an article the articles in *lexico.com* contain a data box in which the "word of the day" is given and another box displaying the most recent "word of the year". This is seen in figure 26 with the word of the day in an orange coloured box and the word of the year in a green coloured box:

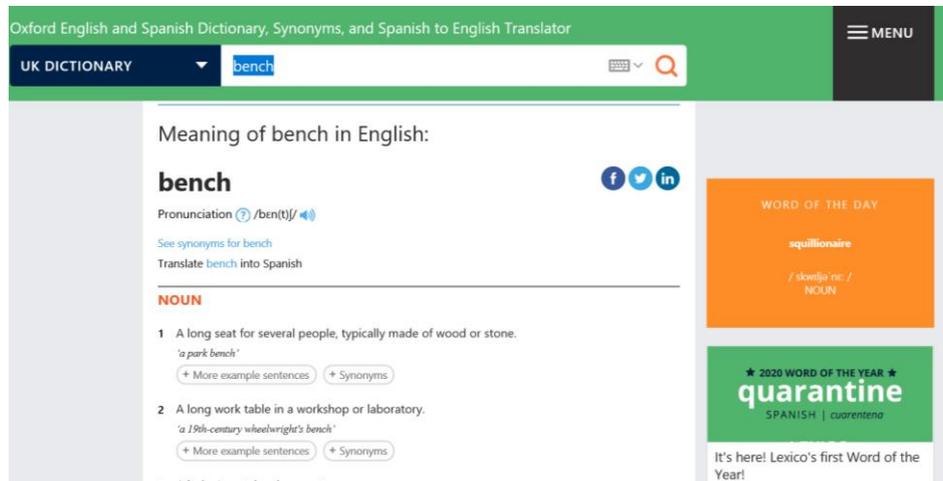


Figure 26: Word of the day in *lexico.com*

Because these boxes are presented in every article, knowledgeable users of this dictionary will know that they can retrieve this information from the dictionary and know where to find it. For a user consulting the dictionary for the first time or consulting it to find other data in a dictionary article these data boxes offer a data bonus and additional consultation success.

Lexicographers of online dictionaries also use the lesser space restrictions to include data boxes with non-lexicographic data that could be seen as a type of *lexicotainment*, where *lexicotainment* could refer to the presentation of data that do not contribute to achieving the genuine purpose of the dictionary, but enable the retrieval of information that might not be lexicographically relevant but may enrich the consultation procedure. Schierholz (2015: 340) also refers to "reading dictionaries for entertainment or to kill time (which is called 'lexicotainment'". The following screenshot of a partial article of the lemma *bench* in *lexico.com* shows a data box that contains a brief quiz of which the topic is not related to the lemma of the article accommodating this data box. On any given day this quiz will not be the same in all articles. However, the subsequent data box with "trending words" is the same in all articles. The data in this latter data box are not actually a form of *lexicotainment* because this box given in figure 27 rather adds to the fulfilment of a cognitive function of the dictionary and therefore this data fall within the scope of the genuine purpose of the dictionary.

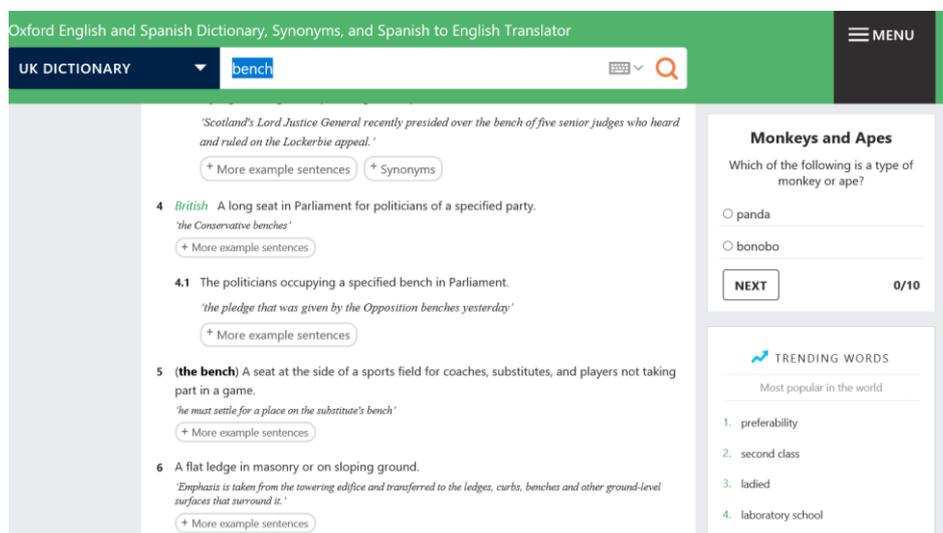


Figure 27: A quiz and "Trending words" in *lexico.com*

Lexicographers can also employ data boxes to respond to questions from their dictionary users. Articles in the *Merriam-Webster* dictionary contain a data box "Ask the editors". This data box given in figure 28 contains separate boxes with the response of the lexicographer to questions put by the users:

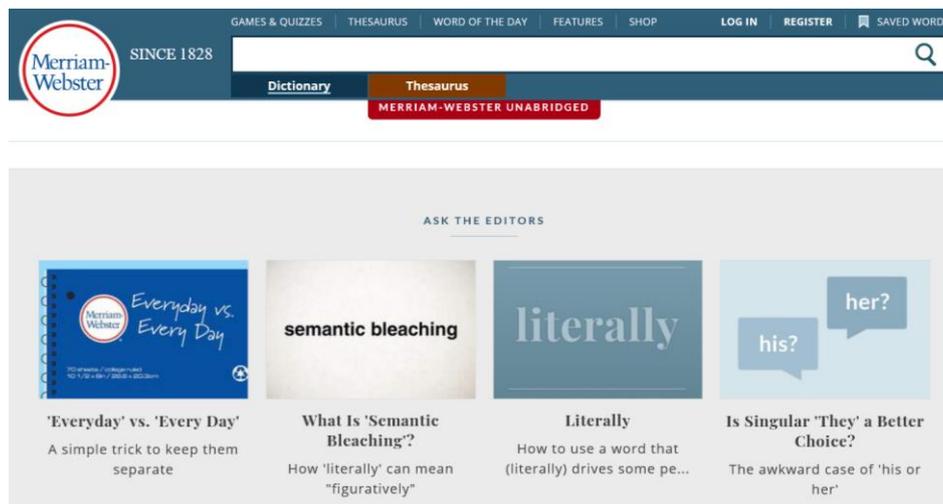


Figure 28: Response of editors to user questions

This data box is used to introduce an innovative communication opportunity between dictionary maker and dictionary user. This use of data boxes and the further possibilities that could arise demand more comprehensive attention from the field of metalexigraphy.

4.3 Information and reference tools

As utility tools dictionaries, whether in printed or online format, are carriers of data from which users can retrieve information. Online dictionaries are no longer only regarded as isolated tools but they are part of a larger family of reference tools. Besides presenting lexicographic data to their users online dictionaries often also guide the users to dictionary-external sources — either in the same search domain, a dictionary portal, or in the search universe where other lexicographic and non-lexicographic sources can be targeted. Although the mediostructure of printed dictionaries also makes provision for cross-reference positions accommodated by cross-reference items with a dictionary-external address these cross-references typically are embedded within the dictionary article — either within a search zone complementing another item or in a search zone dedicated to dictionary-external cross-references, as seen in figure 29, the article of the lemma *Benutzungsgrund* in the *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research* (WLWF: Wiegand et al. 2010)

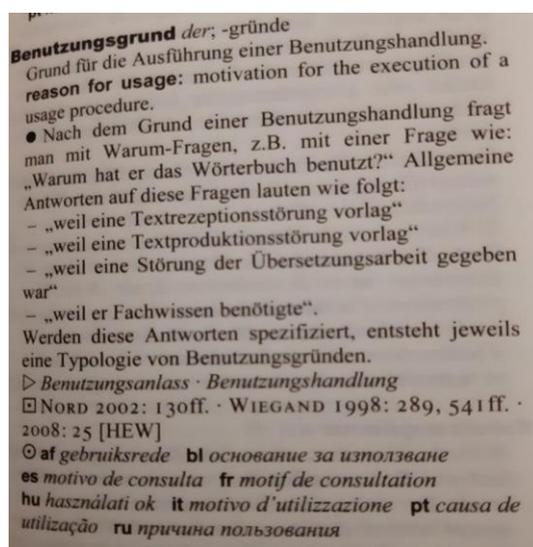


Figure 29: Article of the lemma *Benutzungsgrund* in the WLWF

In this article the typographical structural indicator "☐" identifies the search zone populated by items giving dictionary-external cross-reference addresses, as seen in figure 30:

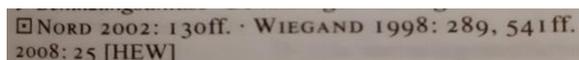


Figure 30: Dictionary-external cross-reference addresses in the WLWF

Online dictionaries can cross-refer users in a much better way to a specific reference address, for example by including a link with an unambiguous data marker, as seen in figure 15. It can also be done by directing users to sources in either the same search domain or in the search universe (Gouws 2021: 15; 2021a). Unlike presenting these sources as items in a search zone populating the obligatory microstructure of the dictionary, as seen in figure 29, the lexicographer can use a data box that contains, among others, a reference to different sources from which the user can retrieve additional information. This is seen in figure 31, the article of the lemma *Zug* in *dict.cc* where a click on the information icon in the left and the right margins of the article activates a pop-up data box, seen in figure 31 in the bottom right-hand corner of the article.

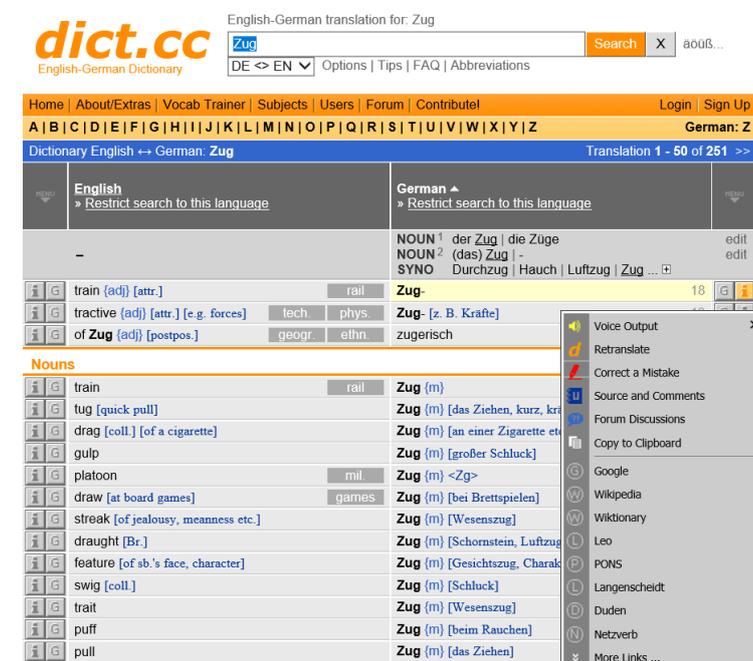


Figure 31: Article of the lemma *Zug* in *dict.cc*

The lower section of this data box is used to convey another type of salient data, namely the titles of dictionary-external sources. A click on any of these sources guides the user to the treatment of the item of which the information icon was clicked in the source given in the data box. Here the data box assists the user in a way not typically found in printed dictionaries.

This is another innovative use of data boxes. The dictionary introduces an occurrence of this type of text constituent that gives access to data relevant to the lemma and it also gives access to other sources where additional relevant information could be retrieved.

5. The future

Data boxes have made a significant contribution in ensuring a more comprehensive and diverse transfer of lexicographic data. Lexicographers of both printed and online dictionaries have been innovative in introducing different ways of best employing data boxes. Certain procedures, for example the procedure of boxing salient data, became established in printed dictionaries. This tradition has been continued in some online dictionaries. The reality of lesser space restrictions but also dynamic article structures, new layout possibilities and easier linking of items in a dictionary article to either dictionary-internal or dictionary-external addresses have resulted in new ways of using data boxes in online dictionaries. Many of these ways have not been sufficiently discussed in metalexigraphic literature and this paper emphasises the need for such a discussion. Not only lexicographic data but also relevant non-lexicographic data can be accommodated in data boxes. This offers numerous opportunities to lexicographers when devising the data distribution structures of their dictionaries. Much more attention can now be given to the possibility of a stronger focus on the cognitive function of dictionaries.

Electronic dictionaries of the future are expected to continue the tradition of the paper and current electronic dictionary to present data as part of the treatment of the lemma. So, for example in figure 14 the data box is presented directly following the treatment of the second sense in the article of *underground*. In this way the users have no option whether they want to see the data box or not. Presentation of databoxes in this way can add to information overload and increase text density. Hyperlinking could be a better or alternative approach to the presentation of data boxes in future electronic dictionaries. Electronic dictionaries employ hyperlinking and pop-up boxes to such an extent that almost every item in a dictionary article is hyperlinked to a pop-up box. Such pop-up boxes provide the user with information on various issues ranging from convention explanation, phonetic and grammatical information and translation equivalents or individual words used in paraphrase of meaning; thus, a complicated cross-referencing system. This system is designed on the basis of two approaches namely hovering and clicking. In the case of hovering no deliberate action from the user is required but an opportunity is

offered to them to obtain more information through a deliberate clicking action. Consider an inventory of pop-up boxes obtained through hovering and clicking for *mosadi* compiled by Prinsloo and Van Graan (2021: 54) in figure 32.

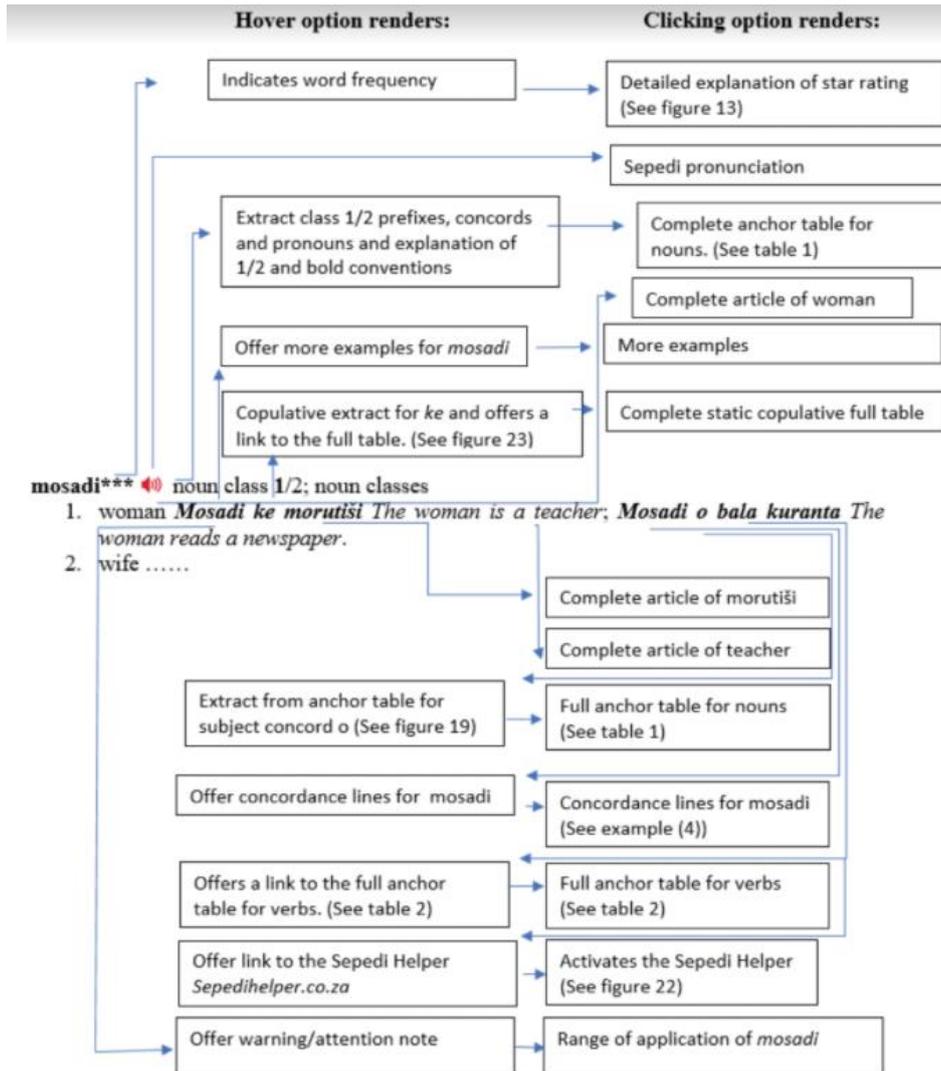


Figure 32: Model article for *mosadi* with arrows indicating hovering and clicking options

The first 20 pop-up boxes deal with a variety of issues such as frequency indication, pronunciation, grammatical guidance, additional examples, complete

articles of words used in the translation equivalent paradigm, etc. The final two pop-up boxes are data boxes offering salient information pertaining to the range of application of *mosadi*. A hierarchy exists between these two data boxes. Through hovering over the warning/attention note as is the case in the top left box on frequency, the user is informed about a box that can be obtained through clicking. Two considerations are at stake here. Firstly, the issue of information overload and secondly, a hierarchical drill-down strategy. Prinsloo and Bothma (2020: 87) say in this regard:

A user does not need such an information overload to solve a very specific information need in a given situation – the user typically prefers to be provided with exactly the required amount of information to solve his/her information need in the given situation. In an e-environment, this information overload can easily be circumvented by initially providing only basic information that builds upon existing knowledge, but then providing, through drill-down options on demand, either more basic or more in-depth information about the problem at hand.

Drill-down actions through clicking in figure 32 lead to another (deeper) level of information. In the case of frequency indication the drill-down action renders a pop-up box with detailed information on the star rated convention used for frequency indication in the dictionary. In the *Macmillan Dictionary* (MED) such drill-down actions result in the provision of a wealth of information for the user. Hovering over the frequency star convention in any dictionary article guides the way to several levels of drill-down options. The first level is detailed information "RED WORDS AND STARS". Second levels obtained through further clicking are clicking on a video entitled "Smart learning with Red Words and Stars" and a clicking option to download a "Red Words & Stars pack". The same holds true for the data box "Range of application of *mosadi*" in figure 32 where the drilling-down action results in data boxes such as the one designed for *mosadi* in figure 8 above.

One of the exciting possibilities in online lexicography is the use of data pulling procedures (Gouws 2018; 2021). The successful employment of data pulling procedures can be enhanced by a clear indication of the information retrieval structure of the specific dictionary. In this regard it is important that users need be made aware of the dictionary-external sources functioning in the relevant search domain as well as the search universe. Data boxes can make a huge contribution in presenting a position in a dictionary article where the menu of dictionary-external sources can be given — as seen in figure 31.

In further metalexigraphic research into data distribution options in dictionaries as well as into the enhanced use of data pulling procedures an increased use of data boxes should be negotiated. This is a dictionary component that could continue to play a significant role in future dictionaries.

6. Conclusion

This article as well as the preceding two articles in this trio have focused on a variety of aspects related to data boxes in printed and online dictionaries. The first article (Gouws and Prinsloo 2021) gave a metalexigraphic perspective with a focus primarily on the occurrence of lexicographic data boxes as text constituents in dictionaries. In the second article (Prinsloo and Gouws 2021) the types and contents of data boxes were discussed. This third contribution put the emphasis on data boxes in bilingual dictionaries with an African language as one of the treated languages. It also looked at new ways in which existing online dictionaries have used data boxes.

The current use of data boxes in both printed and online dictionaries can form an important point of departure for the future use of this type of text constituents. **Accommodating salient data should remain a significant assignment to data boxes.** In addition, the use of data boxes to ensure an improved article layout and data distribution gives future lexicographers numerous options to enhance the quality of their dictionaries. As dynamic utility tools dictionaries can also use data boxes as text constituents that form a bridge between dictionary-internal and dictionary-external consultation procedures.

Data boxes have played a significant role in the lexicographic practice. This role should be maintained and increased in future dictionaries. Better collaboration between metalexigraphers and practical lexicographers can ensure an exciting use of data boxes when fully exploiting the potential of the online environment.

Endnote

1. The term 'Bantu' got stigmatized during the Apartheid Era in South Africa. Therefore, the term 'African' is preferred in South Africa even in reference to what is internationally referred to as 'Bantu languages'. The discussion in this article is, however, focused on the Bantu language family and most of the issues described cannot necessarily be generalized to be applicable to other languages on the continent of Africa. To respect the view of those opposed to the term 'Bantu', it will only be used in cases where a distinction between African languages (languages spoken in Africa) versus a member of the Bantu language family is essential.

Acknowledgement

This research is supported in part by the South African Centre for Digital Language Resources (SADiLaR). Findings and conclusions are those of the authors.

References

Dictionaries

- ANW. *Algemeen Nederlands Woordenboek*. Available at <http://anw.inl.nl/>. (Date of access: June 2021.)
- dict.cc. *Online-Wörterbuch Englisch-Deutsch/Deutsch-Englisch*. Available at <https://www.dict.cc/>. (Date of access: June 2021.)
- elexiko. *Online-Wörterbuch zur deutschen Gegenwartssprache*. Available at <https://www.owid.de/docs/elex/start.jsp>. (Date of access: June 2021.)
- GNWS = Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho-woordeboek / Comprehensive Northern Sotho Dictionary / Pukuntšū ya Sesotho sa Leboa*. Pretoria: J.L. van Schaik.
- lexico.com. *Oxford English and Spanish Dictionary, Thesaurus, and Spanish to English Translator*. <https://www.lexico.com/>.
- MED = *Macmillan Dictionary*. Available at <https://www.macmillandictionary.com/>. (Date of access: August 2021.)
- Merriam-Webster. Available at <https://www.merriam-webster.com/dictionary/>. (Date of access: June 2021.)
- OALD = *Oxford Advanced Learner's Dictionary*. Available at http://oald8.oxfordlearnersdictionaries.com/dictionary/underground_3. (Date of access: June 2021.)
- ONSD = De Schryver, G.-M. (Ed.). 2007 *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšū ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.
- OZSD = De Schryver, G.-M. (Ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgesi, Esishicilelwe abakwa-Oxford*. Cape Town: Oxford University Press Southern Africa.
- SEPD = Mojela, M.V., M.C. Mphahlele, M.P. Mogodi and M.R. Selokela. 2006. *Sesotho sa Leboa / English Pukuntšū Dictionary*. Cape Town: Phumelela.
- Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer, and W. Wolski (Eds.). 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. Volume 1. Berlin: De Gruyter.

Other references

- Gallardo, A. 1980. Dictionaries and the Standardization Process. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography*: 59-69. Columbia: Hornbeam Press.
- Gouws, R.H. 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Gouws, R.H. 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke *Lexikos* 28: 177-195.
- Gouws, R.H. 2021. Expanding the Use of Corpora in the Lexicographic Process of Online Dictionaries. Piosik, M., J. Taborek und M. Woźnicka (Eds.). 10. *Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven*: 1-19. Berlin/Boston: De Gruyter.

- Gouws, R.H.** 2021a. Lexicography and Documentation in a Multilingual Environment. Unpublished paper delivered at the 14th International Conference of the Asian Association for Lexicography, ASIALEX 2021, 12–14 June 2021.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. and D.J. Prinsloo.** 2021. Lexicographic Data Boxes. Part 1: Lexicographic Data Boxes as Text Constituents in Dictionaries. *Lexikos* 31: 330-373.
- Heuberger, R.** 2020. Monolingual Online Dictionaries for Learners of English and the Opportunities of the Electronic Medium: A Critical Survey. *International Journal of Lexicography* 33(4): 404-416.
- Lombard, D.P.** 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L. van Schaik.
- Louwrens, L.J.** 1991. *Aspects of Northern Sotho Grammar*. Pretoria: Via Afrika.
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.
- Prinsloo, D.J.** 2017. Africa's Response to the Corpus Revolution. Xu, Hai (Ed.). 2017. *Proceedings of the 11th International Conference of the Asian Association for Lexicography, ASIALEX 2017, 10–12 June 2017, Guangzhou, China: Lexicography in Asia: Challenges, Innovations and Prospects*: 20-31. Guangzhou, China: ASIALEX. Available at <http://asialex.org/pdf/Asialex-Proceedings-2017.pdf>. [Consulted January 2021.]
- Prinsloo, D.J.** 2020. Detection and Lexicographic Treatment of Salient Features in e-Dictionaries for African Languages. *International Journal of Lexicography* 33(3): 269-287.
- Prinsloo, D.J. and T.J.D. Bothma.** 2020. A Copulative Decision Tree as a Writing Tool for Sepedi. *South African Journal of African Languages* 40(1): 85-97.
- Prinsloo, D.J. and R.H. Gouws.** 2006. Lexicographic Presentation of Grammatical Divergence in Sesotho sa Leboa. *South African Journal of African Languages*. 26(4) 184-197.
- Prinsloo, D.J. and R.H. Gouws.** 2021. Lexicographic Data Boxes. Part 2: Types and Contents of Data Boxes with Particular Focus on Dictionaries for English and African Languages. *Lexikos* 31: 374-401.
- Prinsloo, D.J. and N.D. van Graan.** 2021. Principles and Practice of Cross-referencing in Paper and Electronic Dictionaries with Specific Reference to African Languages. *Lexicography, Journal of Asialex* 8(1): 32-58.
- Prinsloo, D.J. and J.J. van Wyk.** 1992. Verwantskapsterminologie van die Noord-Sotho. *South African Journal of Ethnology* 15(2): 43-58.
- Schierholz, Stefan J.** 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25: 323-352.
- Van Wyk, E.B., P.S. Groenewald, D.J. Prinsloo, J.H.M. Kock and E. Taljard.** 1992. *Northern Sotho for First-Years*. Pretoria: J.L. van Schaik.
- Van Wyk, J.J. and F.T. Haasbroek.** 1990. Verwantskapsterminologie van die Batswana. *South African Journal of Ethnology* 13(4): 159-179.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Volume 1*. Berlin/New York: Walter de Gruyter.
- Wiegand, H.E. and R.H. Gouws.** 2011. Theoriebedingte Wörterbuchformprobleme und wörterbuchformbedingte Benutzerprobleme I. Ein Beitrag zur Wörterbuchkritik und zur Erweiterung der Theorie der Wörterbuchform. *Lexikos* 21: 232-297.