Devices for Information Presentation in Electronic Dictionaries[*]

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

Theo Bothma, *Department of Information Science, University of Pretoria (theo.bothma@up.ac.za)*
and
Gertrud Faaß, *Department of Information Science and Natural Language Processing, Hildesheim University, Hildesheim, Germany and Department of African Languages, University of Pretoria (gertrud.faass@uni-hildesheim.de)*

Abstract: Electronic dictionaries should support dictionary users by giving them guidance in text production and text reception, alongside a user-definable offer of lexicographic data for cognitive purposes. In this article, we sketch the principles of an interactive and dynamic electronic dictionary aimed at text production and text reception guiding users in innovative ways, especially with respect to difficult, complicated or confusing issues. The lexicographer has to do a very careful analysis of the nature of the possible problems to suggest an optimal solution for a specific problem. We are of the opinion that there are numerous complex situations where users need more detailed support than currently available in e-dictionaries, enabling them to make valid and correct choices. For highly complex situations, we suggest guidance through a decision tree-like device. We assume that the solutions proposed here are not specific to one language only but can, after careful analysis, be applied to e-dictionaries in different languages across the world.

Keywords: ELECTRONIC DICTIONARIES; USER GUIDANCE; TEXT PRODUCTION; TEXT RECEPTION; DICTIONARY DESIGN, DECISION TREE STRUCTURE, COPULATIVES, KINSHIP TERMINOLOGY, INFORMATION PRESENTATION DEVICES

Opsomming: Inligtingsaanbiedingsinstrumente in elektroniese woorde-boeke. Elektroniese woordeboeke behoort woordeboekgebruikers te ondersteun deur hulle te lei ten opsigte van teksproduksie en teksresepsie volgens 'n gebruikergedefinieerde aanbod van leksi-

---

kografiese data vir kognitiewe doeleindes. In hierdie artikel skets ons die beginsels waarop 'n inter-aktiewe en dinamiese elektroniese woordeboek berus, gemik op teksproduksie en teksresepsie wat die gebruikers op innoverende wyse lei, veral ten opsigte van moeilike, gekompliseerde of verwar-rende aspekte. Die leksikograaf is genoodsaak om 'n noukeurige analise te doen van die aard van moontlike probleme ten einde 'n optimale oplossing aan te bied vir 'n spesifieke probleem. Ons is van mening dat daar verskeie komplekse gevalle bestaan waar gebruikers meer gedetailleerde ondersteuning benodig as wat tans in e-woordeboeke beskikbaar is ten einde korrekte keuses te kan maak. Vir hoogs problematiese situasies stel ons leiding deur middel van 'n keuseboom-tipe instrument voor. Ons veronderstel dat die oplossings wat hier voorgestel word nie tot 'n enkele taal beperk is nie, maar na versigtige analise op verkillende tale van die wêreld toegepas kan word.

Sleutelwoorde: ELEKTRONIESE WOORDEBOEKE, GEBRUIKERSLEIDING, TEKSPRO-DUKSIE, TEKSRESEPSIE, WOORDEBOEKONTWERP, KEUSE-BOOMSTRUKTUUR, KOPULA-TIEWE, VERWANTSKAPSTERMINOLOGIE; INLIGTINGSAANBIEDINGSINSTRUMENTE

1.   Introduction

In lexicography, the electronic era was met with great enthusiasm and expecta-tions. Early publications on electronic dictionaries were all about the potential of the new medium and the expected revolution it would bring along, thereby antiquating the paper dictionary in a decade or two. De Schryver (2009), how-ever, rightfully expresses disappointment in respect of the pace of development of electronic dictionaries. More exciting was the introduction of what could be called "true electronic features" such as pop-up boxes, audible pronunciation and sophisticated search features. Some electronic dictionaries also solve lem-matisation problems, which cannot be resolved in paper dictionaries. Typical examples are *isiZulu.net*[1] where Zulu words can be looked up without prior stem identification by simply typing in the word, or commercial products such as Amazon Kindle e-books[2] that link inflected word forms to lemmatised forms, e.g. *went* to *go* or German *ging* to *gehen*.

Electronic dictionaries of today, however, could enter a more advanced dimension in fulfilling more sophisticated needs of the users, e.g. through multiple access routes. Rundell (2009) refers to "game changing" developments that have "expanded the scope of what dictionaries can do and (in some respects) changed our view of what dictionaries are for". De Schryver (2009) calls in this context for an adaptive and intelligent dictionary (aiLEX) that will be able to "study and under-stand its user" and consequently "present itself to that user". "Intelligent" diction-aries in this sense are currently outside the scope of this article and of our approach: "intelligent" probably assumes a greater or lesser use of artificial or com-putational intelligence in the underlying programming of the dictionary, which is a realistic expansion of e-dictionaries but not one we are currently planning.

Our focus is more on supporting dictionary users by giving them guidance in text production and text reception, alongside a user-definable offer of lexicographic data for cognitive purposes. In most cases, what is currently offered in dictionaries

claiming that they give guidance in text production, is in fact still more on the level of text reception, or it leads to an overload of information, or the information provided is not appropriate for a given situation. We thus intend to sketch the principles of an interactive and dynamic electronic dictionary aimed at text production and text reception, which guides the user in innovative ways, especially with respect to difficult, complicated or confusing issues. What is at stake in the medium term are improvements in access possibilities of electronic dictionaries, including (but not discussed in this paper) access to more information presented by linking the dictionary with other online sources, cf. Heid, Prinsloo and Bothma (2012).

Our approach will result in dictionaries which are different from existing ones, and where the individualised lexical offer for a specific user in a specific usage situation and with a specific information need is paramount. In our view, it is important that a dictionary provides features that allow an individual user to get access, in the most intuitive way, to exactly the bit of information he/she needs in a given situation. This does not mean that predefined typical user needs as stated in the Function Theory of lexicography (cf. e.g. Tarp 2008a) are not used — our approach simply goes one step further. The process of individualising the lexical offer which we sketch in this article follows the principles of the Function Theory and adds at the same time the concept of dynamic interaction, as Tarp (2009a: 292) states, "in order to conceive dictionaries capable of meeting all the users' needs in specific types of situations". The improvements envisaged are not limited to adding more sophisticated search options or to the increased utilisation of pop-up windows — it will, in fact, utilise totally different concepts such as a step-by-step text production guidance and interactive selection processes.

These processes will be illustrated by means of examples ranging from fairly simple to highly complex from different languages, viz. Germanic, Romance and African, where the user can be guided to the correct text production or text reception choices by means of devices ranging from simple examples to complex decision trees.

We do not assume that the type of solutions we offer below are applicable to all lexical items in all text production or text reception information needs — the lexicographer will have to do a very careful analysis of the nature of the possible problems to suggest an optimal solution for a specific problem, and in many cases users may not need complex solutions. However, we are of the opinion that there are numerous complex situations where users need more detailed support than currently available in e-dictionaries, to make valid and correct choices. We assume that the solutions proposed here are not specific to one language only but can, after careful analysis, be applied to e-dictionaries in many different languages across the world.

There are two possible approaches to provide lexicographic support in the above-mentioned situations, viz. a "stand-alone" dictionary that will be consulted as an information tool in its own right, or the integration of the lexicographic tool into a text production environment, for example into the user's word processor. Depending on the point of departure, the user could either

find the information he/she is looking for by direct/intended dictionary-lookup or during the actual process of typing text, for example in a situation where the user can move the cursor in the text he/she is creating from word to word to get basic information.

To set up a database that could provide access to the data needed for the processes described briefly above requires a very careful planning and design of the database. The first step, however, would be to select a number of examples and to describe these in detail, deciding on the nature of the solution to be offered in each case, for example, in a text production or in a text reception situation. Once this is described in detail, it will have to be formalised according to an XML schema, a database structure or another data representation format which will need to be designed in such a way that it will make provision for different levels of complexity and detail. Careful attention will have to be paid to the granularity of the data, to enable the user to easily navigate complex steps, and in the process not to present the user with unnecessary or confusing data — only the data required at each step in the decision process are to be presented to the user at any given time. Designing such a database and the schemas associated with it is a complex process which will not be addressed in this paper. After a brief reference to the Function Theory of lexicography (section 2), we will provide a short and preliminary survey and classification of devices for information access in electronic dictionaries (section 3). We will then address new devices for information access which we conceive of as being appropriate for the purpose of providing adequate information for different use situations, viz. tables, diagrams, guidance paths and interactive decision trees. All these devices will be discussed in some depth (section 4), before a detailed example of decision support for a complex text production problem will be given (section 5). We conclude in section 6 and point to intended future work.

2.   Dictionaries as language information tools: the views of Function Theory

The work on information presentation in electronic dictionaries presented in this article will make use of elements of the Function Theory of lexicography as proposed, inter alia, by Bergenholtz and Tarp (cf. Bergenholtz and Tarp 2002, 2003, 2004, 2005, Tarp 2008 and 2008a, 2009 and 2009b, 2011, 2012), which claims that "dictionaries and other lexicographical work are above all *utility tools* conceived and produced with the genuine purpose of satisfying specific types of human needs, i.e. *information needs,* existing in one or several individuals in society" (Bothma and Tarp 2012: 89). The Function Theory currently works with four types of lexicographically relevant situations (cf. Tarp 2008a), viz. communicative, cognitive, operative and interpretive situations.

In this article we refer only to communicative and cognitive situations. Communicative situations may be further subdivided into a number of situations such as text production and text reception, text translation, text revision etc. Cognitive situations may also be divided into various sub-situations and refer to situations where the user may need to acquire specific knowledge to

perform a task. Operative and interpretive situations refer respectively to situations where a user needs instructions on how to perform a physical or mental action and to situations where a need exists to interpret and understand a sign, signal, symbol etc. For a brief summary of the Function Theory, see Bothma and Tarp (2012, section 2).

3.   A classification of devices for information access in electronic dictionaries

In this section, we make an attempt at provisionally classifying the devices available to lexicographers who design electronic dictionaries, for giving e-dictionary users access to lexicographic data. In our view, three types of devices need to be distinguished: contents-related devices, navigational devices and presentational devices. In section 4, we will propose a few presentational devices in addition to the standard ones briefly mentioned in section 3.3, which in our view may considerably improve the effectiveness and efficiency of electronic dictionaries, at least for certain types of phenomena.

Contrary to the situation for printed dictionaries, where both contents-related and navigational devices (i.e. items ("Angaben") and text structure markers ("Strukturanzeiger")) have been discussed in detail (Wiegand 1989), we are not aware, as yet, of an inventory or a state of the art description of the devices commonly used in electronic dictionaries. Unfortunately, we will not be able, in this paper, to even outline such a state of the art; our overview will thus have to remain sketchy. There are neither any standards nor guidelines that would suggest which devices to use for what purpose, for which users or which types of phenomena to be explained in the dictionary.

3.1   Contents-related lexicographic devices

Electronic dictionaries have inherited from printed dictionaries a number of well-known lexicographic devices such as paraphrase of meaning (definitions), examples of usage, pronunciation guidance, part of speech, etc. Such devices are commonly used in many printed and electronic dictionaries, and will not be discussed any further.

3.2   Navigation devices

In addition, electronic dictionaries make use of general design principles of web design, for example with respect to navigation devices: this includes features such as scrolling facilities, clickable headlines, site maps etc. *Wiktionary*[3], the online dictionary that belongs to *Wikipedia*, as well as the *Oxford English Dictionary Online*[4] provide examples of site maps for dictionary entries: they offer an overview of their articles in terms of clickable headlines (cf. Bothma 2011: 84). Again, these devices are also commonly in use, and will hence not be discussed any further.

3.3   Devices for information presentation

There are several devices for on-demand data provision, such as expandable or collapsible indications, cross-referencing by means of links, mouse-sensitive areas which display data when being pointed at with the mouse cursor, or pop-up boxes with lexicographic data that can be opened by pointing the mouse cursor at them and clicking. All of these provide information to the user if he/she activates them. Such devices allow for the personalisation of the lexical offer, in so far as they allow the user (at least to some extent) to decide himself or herself, how much he/she wants to see.

3.3.1   Mouse-over

An example of data provision on demand based on the mouse-over feature in an existing e-dictionary is the mouse-sensitive behaviour of the subcategorisation indications in ELDIT, the electronic learners' dictionary for German and Italian (cf. Abel 2002) realised at EURAC in Bolzano[5]: it displays subcategorisation patterns in a terminology which deliberately avoids linguistic terminology (assuming that users are not familiar with terms like "subject", "controlled infinitival" etc.) as standard practice. In figure 1, we reproduce a screen from ELDIT which shows the entry for IT *dare* ("[to] give"), with a valency formula saying "someone gives something to someone" (*qualcuno dà qualcosa a qualcuno*). If one of the fields of the formula is being pointed at with the mouse cursor (cf. figure 1), the linguistic term of the respective element of the valency formula is displayed (in figure 1, this is *complemento oggetto*), and the corresponding word or phrase in the example sentence is highlighted. The same highlighting happens if a relevant part of the example sentence is pointed to with the mouse cursor: in this case, the respective element of the valency frame and the corresponding linguistic term are indicated.
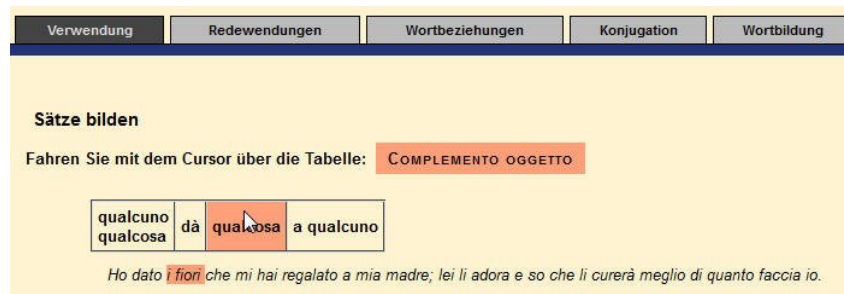

Figure 1: Screenshot from ELDIT: mouse-sensitive elements of valency frames, as an example of data provision on demand

3.3.2   Pop-up boxes

Similar to the above example, other elements of the data display can be made mouse-sensitive. The following example concerns an illustration, where lexicographic data about objects belonging to a theme or being part of a larger object are shown on demand. This device has also been prototyped, for example, in ELDIT[6]; an example is the illustration in the entry for *Haus* ("house") with its pop-up entry for *Fenster* ("window") in figure 2. The cursor changes to a hand when active items are encountered and upon clicking on the item, additional data are displayed in the pop-up box.
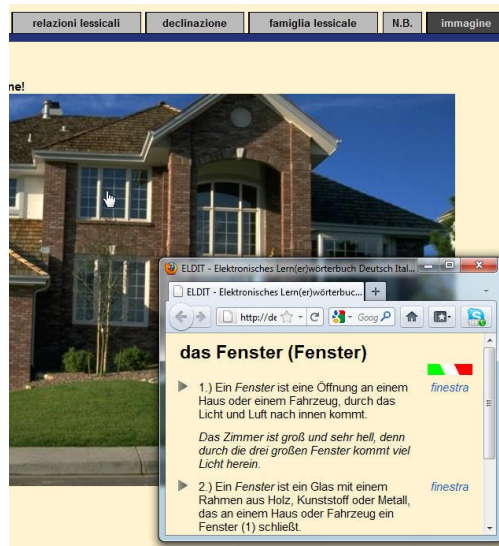

Figure 2: Screenshot from ELDIT of a mouse-sensitive illustration: a pop-up box as an example of data provision on demand (user clicking on the resp. item)

See also the *English Dictionary for South Africa* (2012, CD-ROM version) where items in illustrations are hyperlinked to the dictionary article.

This device could very well be used in technical dictionaries to illustrate terms that denote elements of complex technical objects, e.g. in explosion diagrams. To the best of our knowledge, it is, however, not much used yet in specialised dictionaries.

4.   New devices for information presentation in electronic dictionaries

This section will be devoted to a discussion of different presentational devices that have so far not often been used in electronic dictionaries; they seem, how-

ever, to be relatively effective and efficient for a number of cognitive as well as communicative dictionary functions, especially on lexical items that belong to (potentially rather complex) structured systems.

4.1   Phenomena belonging to structured systems

To act as language information tools, (general language) dictionaries should cover phenomena from all levels of linguistic description: pronunciation, morphology, syntax, semantics, and pragmatics. Many such phenomena are systematic in one way or another: for example, many of the morphological and syntactic properties of words and word combinations must be seen in the context of the grammatical system of the language or of a given subsystem. They can only (or at least better) be interpreted with reference to this grammatical (sub) system. This holds for many function words, for words belonging to inflectional paradigms, for complementation patterns or for complex words built according to specific morphological word formation principles. We say that such items belong to "structured systems".

Typically, the particularised presentation of lexical data in semasiological dictionaries, i.e. the individualised access to each lemma entry, does not bring the systematic nature of such phenomena to the fore, but rather obscures it by distributing the members of the set across the whole macrostructure. For some dictionary use situations, this is not a major issue, and some lexicographers counterbalance this effect by including systematic morphological or syntactic overview tables (inflection paradigms, inventories of closed class items, subcategorisation tables, etc.) into their dictionaries, for example as outer texts, in an appendix or in a dictionary grammar (cf. Gouws 2009, 2010).

The same property of being part of a structured system is also present in certain lexical semantic phenomena, especially in those which can be structured in terms of taxonomies (e.g. animals, plants), of ordered lists (e.g. names of the days of the week or military ranks) or in terms of relational networks (e.g. kinship terms, cf. Prinsloo and Bosch 2012). In specialised languages, often the meaning of a term is not independent from a given system of related terms; well-known examples are technical taxonomies or juridical terms which denote concepts that are part of a legal system.

The above examples all concern lexical items that are related with other lexical items in one way or another. For a user, knowing about such relationships may in some cases just be part of his or her grammatical knowledge, in other cases it may be a necessary precondition for successful use of the respective items in communication. A prominent example of the latter case is the translation of legal terms which belong to legal systems that are not isomorphic between source and target language (cf. Mayer 1998).

To present items belonging to a structured system, we suggest the use of one of the different presentational devices which we intend to discuss in the following: tables, diagrams, guidance paths and decision trees. While the first two are mainly oriented towards cognitive functions, the latter two are primar-

ily relevant for communicative functions.

4.2   Devices for presenting structured systems: tables

In the first place, presenting the structured system which a given item belongs to, only serves a cognitive purpose: the user of the dictionary may learn about the wider context of the item he or she is looking up. In the case of translation, for example of legal documentation, it may, however, also have a communicative function.

Tables, being two-dimensional, tend to provide good possibilities for presenting items that denote objects which belong to lists or taxonomies, i.e. to systems with up to two classification criteria. A table is thus a recommended lexicographic device for restoring contextualisation where the relation between a list of items that have a certain number of characteristics in common has to be illustrated. Tables may be linked (in the sense of data on demand) to each individual entry they contain, and they can thus also be made accessible from each item.

Table 1 is an example of a table that summarises morphosemantically related lexical items. It contains a summary of eight moods as distinguished by Lombard (1985) for Sepedi (also known as Sesotho sa Leboa or Northern Sotho). This table gives a bird's eye view of the forms used for the different moods and distinguishes between imperfect (present tense) and perfect (past tense) forms, moods (1-3 versus 4-8) and gives examples of positive and negated forms in each case. Finally, a concise summary of the negation strategies and a simplistic indication of the meaning of each mood is presented in the rightmost column.

1. INDICATIVE	pres	pos	*monna o reka puku* the man buys a book	*…-a*
		neg	*monna ga a reke puku* the man does not buy a book	*ga ...-e* STATE-MENTS
	past	pos	*monna o rekile puku* the man bought a book	*ga se*+cons SC+pres
		neg	*monna ga se a reka puku* the man did not buy a book	
2. SITUATIVE	pres	pos	*ge monna a reka puku* if the man buys a book	
		neg	*ge monna a sa reke puku* if the man does not buy a book	*sa ...-e* IF/WHILE
	past	pos	*ge monna a rekile puku* if the man bought a book	
		neg	*ge monna a sa reka puku* if the man did not buy a book	*sa+pres*

3. RELATIVE	pres	pos	*monna yo a rekago puku* the man who is buying a book	*…ago*
		neg	*monna yo a sa rekego puku* the man who is not buying a book	*sa...-ego* THAT (IS DOING)
	past	pos	*monna yo a rekilego puku* the man who bought a book	
		neg	*monna yo a sa rekago puku* the man who did not buy a book	*sa…-ago*
4. SUBJUNCTIVE		pos	*... (gore) monna a reke puku* … (so that) the man buys a book	MUST (DO SOMETHING)
		neg	*... (gore) monna a se reke puku* … (so that) the man does not buy a book	*se ...-e*
5. CONSECUTIVE		pos	*... monna a reka puku* … (then) the man bought a book	AND THEN (DO/DID SOMETHING)
		neg	*... monna a se reke puku* … (then) the man did not buy a book	*se ...-e*
6. INFINITIVE		pos	*... go reka puku* … to buy a book	*go …-a* TO (DO SOMETHING)
		neg	*... go se reke puku* … not to buy a book	*go+se ...-e*
7. IMPERATIVE		pos	*reka puku!* buy a book!	DO (SOME-THING)!
		neg	*se reke puku!* do not buy a book!	*se …-e*
8. HABITUAL		pos	*... monna a reka puku* the man (usually) buys a book	*…-e* USUALLY (DO SOME-THING)
		neg	*... monna a se reke puku* the man (usually) does not buy a book	*se ...-e*

Table 1: The modal system in Sepedi: table presentation

We will come back to possibilities of making use of table 1 for communicative purposes below, in section 4.4.1.

4.3   Devices for presenting structured systems with multiple relations: diagrams

As is the case with tables, diagrams also fulfil a contextualisation function, but in addition they answer to the requirements of visually presented relations. A diagram of the wiring system of a car not only brings together all wires, globes and fuses, but also presents them schematically in relation to each other. The same is true, for example, for kinship terminology, where not only a set of terms that belong together is brought together, but also the exact relations in terms of e.g. older and younger generations, male and female, husbands and wives, are illustrated (cf. the diagram extracted from the *Macmillan English Dictionary for Advanced Learners* (2007: 502), figure 3). Diagrams allow the lexicographer to keep track of more than two types of ordering criteria, thus being adequate for presenting items that denote objects related through several different types of relations. Kinship terminology is a good example of this type of sets of lexical items; this is illustrated in figure 3.
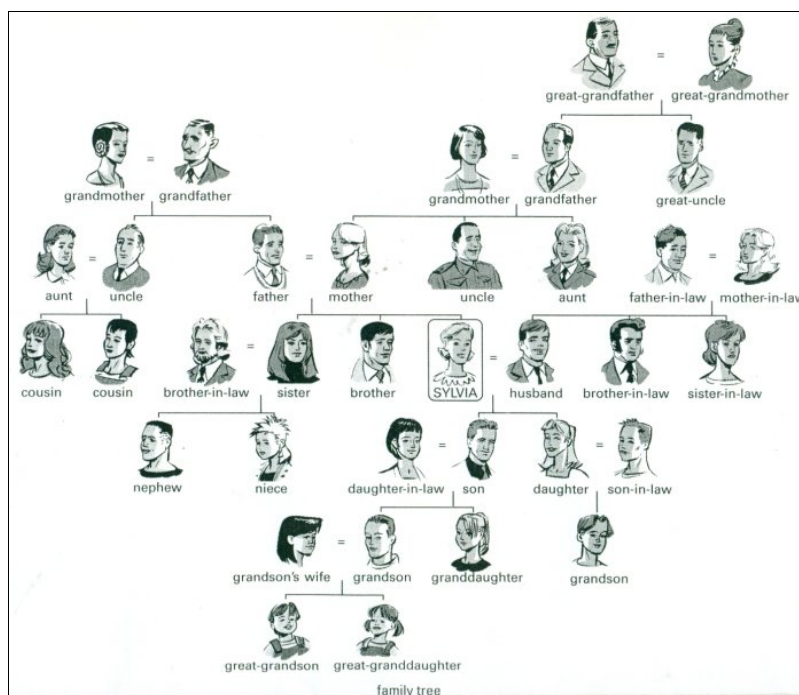

Figure 3: Family tree in MED (2007: 502)

4.4   Devices for guiding users through structured systems: guidance paths

Tables as well as diagrams provide an overview of a structured system, a sort of bird's eye view. Both do not necessarily presuppose any reading direction or any specific entry point. In this sense, these devices rather serve cognitive, overview-related functions, much more than communicative ones.

In our view, situations where the user needs to precisely search for a given item from a structured system are best served by what we call "guidance paths", i.e. paths through tables or diagrams that can be activated by the user.

4.4.1   Guidance paths through tables

Table 2 contains a small extract from table 1 (cf. section 4.1), for the lexical material used in Sepedi to express the situative. It is similar in layout to table 1, but converted into a set of choice options, represented by the arrows. By following the appropriate path, the user will be guided to a valid expression, e.g. by following the path "situative → past → positive" to *-ile*.
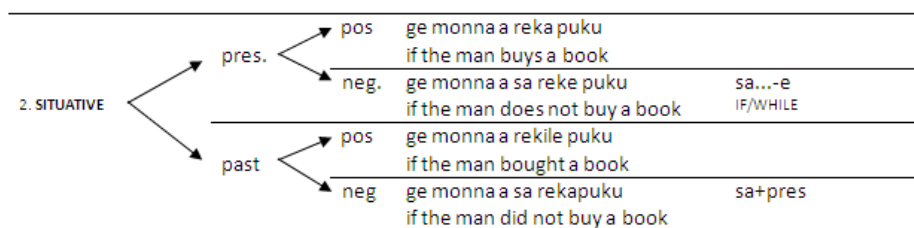

		pos	ge monna a reka puku	
	pres.		if the man buys a book	
		neg.	ge monna a sa reke puku	sa...-e
2. SITUATIVE			if the man does not buy a book	IF/WHILE
		pos	ge monna a rekile puku	
	past		if the man bought a book	
		neg	ge monna a sa rekapuku	sa+pres
			if the man did not buy a book	

Table 2: Extract from table A (modal system of Sepedi), with choice options for text production

Table 3 reflects another extract from the modal system of Sepedi (cf. table 1) and suggests a guidance path for expressing the negative form of the moods that do not distinguish time. The path links the relevant moods with the negative forms, and the vertical bars and framed boxes on the right hand side emphasise the fact that a single negation strategy is used in all cases. This layout is aimed at production guidance in the formation of negated sentences.

4. SUBJUNCTIVE	pos	... *(gore) monna a reke puku*			MUST (DO SOMETHING)
	neg	... *(gore) monna a se reke puku*	*se...-e*	MUST NOT DO SOMETHING	
5. CONSECUTIVE	pos	... *monna a reka puku*			AND THEN ...(DO/DID SOMETHING)
	neg	... *monna a se reke puku*	*se...-e*	AND THEN DID NOT DO SOMETHING	
6. INFINITIVE	pos	... *go reka puku*			TO (DO SOMETHING)
	neg	... *go se reke puku*	*se...-e*	NOT TO DO SOMETHING	
7. IMPERATIVE	pos	*reka puku!*			DO (SOMETHING)!
	neg	*se reke puku!*	*se...-e*	DON'T DO SOMETHING!	
8. HABITUAL	pos	... *monna a reke puku*			USUALLY (DO SOMETHING)
	neg	... *monna a se reke puku*	*se...-e*	USUALLY NOT DO SOMETHING	

Table 3: A guidance path for expressing negated forms of moods in Sepedi

Table 4 is a guidance path highlighting the tripartite relation between (i) the grammatical term, used to denote the mood, (ii) a basic semantic indication of its meaning, given in context with a positive and negated example and a simplistic summary of the negation strategy used. The emphasis here is on reception guidance in terms of the meaning of a given modal expression (cf. the arrows in the schema), to be read from the middle either to the right hand side (paraphrase) or to the left hand side (grammatical terminology). The schema can also be used for text production, either starting from meaning paraphrases (rightmost column), or from the grammatical terminology (leftmost column), with the guidance path leading in either case to the lexical expressions in the middle column.

Table 4: Possible guidance paths for highlighting the relation between a Sepedi mood and its meaning

4.4.2   Guidance paths through diagrams

If we take up the example of kinship terminology discussed above, in section 4.3, we may recall that the table shown in figure 3 serves mainly cognitive purposes, as it allows the user to understand which names of family members exist in English. In a text reception situation, the user who learns English as a foreign language may wish to understand what a term like *cousin* means: ideally, he or she will just have to enter the item into a search box of the dictionary and get all those relatives highlighted and linked to the "central" person in the diagram ("Sylvia") who can be called *cousin*. So, the dictionary proposes a path through the relational diagram. Along this path, the distinctive properties of the items related (in this case the distinctive properties of *cousin*) can be collected, cf. figure 4.
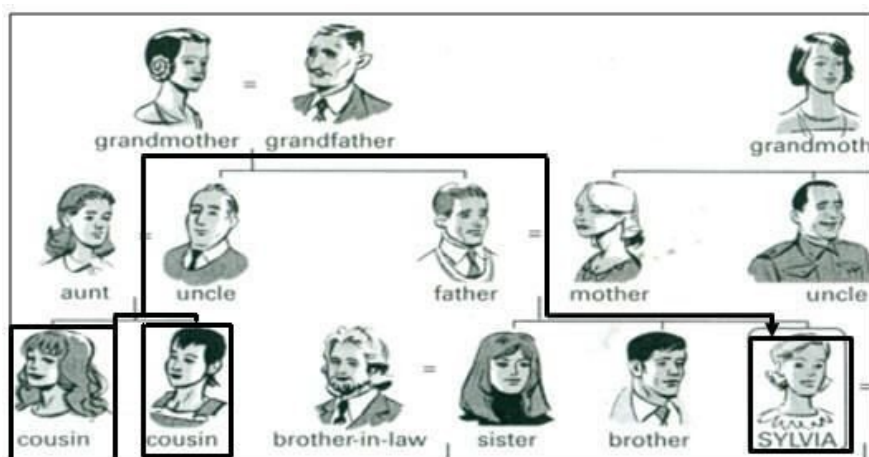
Figure 4: Guidance path for text reception, within (an extract from) the kinship diagram from figure 3

In a text production situation, likewise, a path through the diagram should be constructed, but this time step by step, on the basis of the user's decisions and moves, e.g. from "Sylvia" to her father (step 1 in figure 5), further on to the parents of her father (step 2), to the uncle (step 3), and finally to the children of the uncle (step 4). When the path stops, the dictionary should provide the lexical items searched for. Obviously, the paths for text reception and text production are the same, but used in different ways. For a more complicated system of kinship terms than that of English, namely that of isiZulu and Sepedi, see Prinsloo and Bosch 2012 and our discussion in section 4.5.3, below.
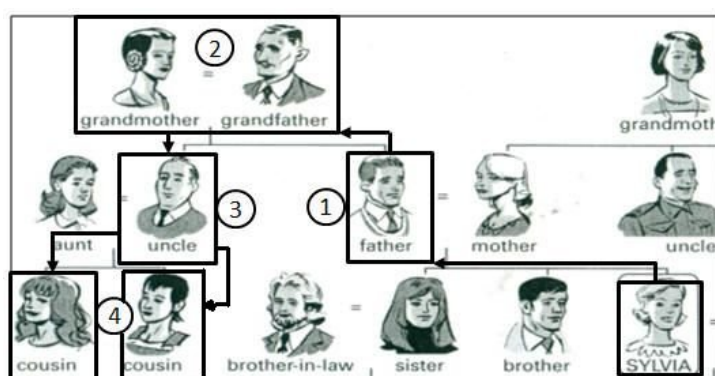

Figure 5: Guidance path for text production, within the kinship diagram from figure 3

The procedures sketched above provide guidance for communicative situations to the user, for text understanding or for text production. They could be implemented by means of tagged static data, including overlays to diagrams or tables.

4.5   Interactive decision trees

The example of kinship terminology shows the guidance an electronic diction-ary could give to the user by means of simple highlighting of paths that lead through a structured system. For issues of low and medium complexity, within (morpho) syntax, ontologies and lexical semantic networks, the decisions to be taken at each step in a text production scenario, as well as the respective meaning components derived from the full path in a text reception scenario, may be relatively intuitive and easy to follow for a user.

An alternative, and maybe even more efficient way of supporting the user, especially in text production and translation towards a foreign language, is what we call *interactive decision trees*.

This device is an attempt to model lexical choice, for production-oriented communicative functions, as a decision process. Text production decisions are made, under constraints from different levels of linguistic description. At each choice point, a few options are open, and the full decision algorithm thus reminds of a tree where the nodes are the choice points and the arcs are the options. This is why we speak of a decision tree, as it is customary in computer science.

Where the interaction of the constraints leading to the right lexical expres-sion is rather complex, the user may thus profit more from stepwise guidance through the tree than from a full table-based or diagram-based overview of the underlying system.

4.5.1   A simple example of decision trees

A trivial example of a decision tree, which we discuss here to show the logic of the device rather than its linguistic contents, is the translation of English pos-sessive determiners to French. The morphosyntactic systems of the two lan-guages are not isomorphic. While English possessives have for some persons different forms depending on the natural gender of the possessor (*his* (masc.) vs. *her* (fem.)), French possessives agree with the grammatical gender and number of the possessed object (*mon livre* (masc. object) vs. *ma bouteille* (fem. object) vs. *mes livres*, *mes bouteilles* (plural object)), but don't mark the natural gender of the possessor.

To correctly translate English possessives to French, the user has to decide on a number of properties of the source and target items involved. Obviously, the relevant linguistic attributes (e.g. number, gender, …) and their values (e.g.

singular, plural, …) are indicated in tables that give the full picture of the possessive determiners in the two languages. But for the actual task of translating, say, *your bottle* to French, the user only needs to be aware of a limited number of decisions to be taken: the first choice point concerns the number of the English item (cf. step 1 in figure 3); let us assume the user selects "singular"; the next decision has to do with the number of the possessed object (step 2, user decides again for "singular"); this leads to the third choice point, concerning the grammatical gender of French *bouteille*, namely "feminine" (step 3). The decision tree ends at the French form *ta*, which is appropriate in this context. Figure 6 schematises the process: solid lines indicate the choices made by the user, and broken lines point to alternative options at each choice point.
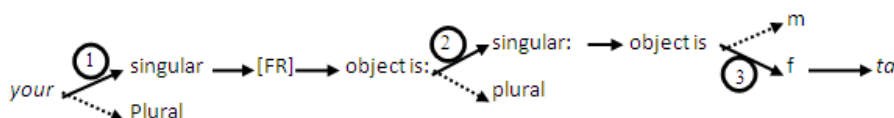

Figure 6: Schema of a decision tree for the translation of English *your* to French, in the context *your bottle*

Obviously, the above made-up example is oversimplified and redundant with most learners' grammatical knowledge. It was merely used to show the logic and the type of guidance we intend to provide through interactive decision trees.

There are, however, other cases that are more complex and may thus indeed require the use of decision trees to provide users with exactly the information needed in a given situation. Sepedi copulatives are a prime example of such a case; we will thus elaborate on the facts pertaining to this construction in the remainder of this section, giving a fuller picture of possible user guidance with respect to this phenomenon in section 5.

4.5.2   Decision trees for complex phenomena: the case of Sepedi copulatives

The system of Sepedi copulatives could be presented in a bird's eye view as a table (cf. Appendix 1) but by virtue of its sheer size, this solution would be suboptimal for both text production and text reception. Standard particularised dictionary entries would lose the information about the systematicity of the phenomena; we thus see the interactive decision trees as the most effective and efficient device to convey exactly the data a user would need with respect to these phenomena in a given type of communicative situation.

In a text production situation, the decision algorithm for the selection of copulatives entails distinguishing between an *identifying* vs. a *descriptive* vs. an

associative relation existing between the subject and its complement, see also sections 1 to 3 in Appendix 1:

(1)
   is
   [identifying. copulative], ke lengwalo (it *is* a letter)
   [descriptive. copulative], mosadi o bohlale (the woman *is* clever)
   [associative copulative], Satsope o na le Sara (Satsope *is* with Sara)

Learners of Sepedi who want to use copulatives in speech or text production have at best to do an intensive study of the copulatives from dictionaries and grammar books. Dictionaries typically provide basic and sometimes even inadequate information (cf. figure 7, below). Grammar books such as Poulos and Louwrens (1994) on the other hand, provide an overload of grammatical information (37 pages), in an effort to cover all the relevant and possible copulatives. Such details may be useful in a cognitive situation where the user would like to learn everything about the copulative, but they are hardly useful in a text production situation where the user simply wants guidance on which form to use. Such information overload could easily lead to "information death" (cf. Bergenholtz and Bothma (2011)).

Dictionaries, and especially electronic dictionaries, fail to give even basic communicative guidance or to treat all three main copulative relations shown in (1). Consider the article for the lemma *is* in the *Sesotho sa Leboa (Northern Sotho)–English Dictionary* (2003) in figure 7.

Figure 7: The lemma is in the *Sesotho sa Leboa (Northern Sotho)–English Dictionary* (2003)

In this example, two of the three copulative categories, i.e., the identifying and the associative copulatives, have not been treated, not to mention giving proper receptive or productive guidance. Paper dictionaries for Sepedi reflect the same deficiencies.

In the e-environment it is, however, possible to provide the user with the required guidance on which form is the correct one for a given situation, and to provide exactly the amount of information that is needed for each of the possible choices. Compared with a table, a decision tree will in such a case reduce the amount of presented information considerably, and the user can, at any stage, decide that his/her information need has been met and return to his/her primary task, namely to write a text.

For example, when the user wants to write *the woman is clever* in Sepedi he/she should be guided to *mosadi o bohlale* and be guarded from the typical error **mosadi ke bohlale*. The user can then be guided to subsequent levels of decisions, e.g. concerning person and noun class of the subject, tenses and moods, as well as a number of lexicalised exceptions, cf. Appendix 1 for a tabular overview of the full set of items to be considered, and section 5 for a detailed example of how decision-tree-based guidance can be conceived of.

4.5.3   Presenting decision trees to the dictionary user

The phenomena sketched above may usefully be presented to the user in terms of subsequent choices that could take the form of check boxes or other selection devices, in a graphical user interface (GUI). The visual appearance of the interface should make clear that the selections are the result of a decision process involving several steps. Instead of complex tables giving all options, a path through sub-tables should be shown, but together with links to synoptic tables which indeed allow the user to see the full picture if he/she wishes to. For a set of function words of the same category, the basic decision tree is constant. Users will only follow different paths through this tree, depending on their actual needs.

The choice points and options may equally be presented in the form of interlinked questions of a questionnaire. These are based on exactly the same logic and internal representation and thus formally equivalent to the decision trees. An example of this device has been proposed by Prinsloo and Bosch (2012), for Zulu and Sepedi kinship terminology, the system of which is considerably more complex (i.e. contains more attributes) than that of English discussed above. Without going into the details of this system, we reproduce, in figure 8, the questionnaire provided by Prinsloo and Bosch (2012), to exemplify this option of presentation. The authors use five levels of selection boxes in their questionnaire; these correspond to five choice points, cf. figure 8.

The example in figure 8 gives step-by-step guidance to correctly address a male speaker's father's older sister supplemented by additional information on the target term.

Level 1

I am a boy ☑	I am a girl ☐

Level 2

I want to speak **to** my aunt ☐	I want to speak **about** my aunt ☑

Level 3

My aunt is:	
My father's:	**My mother's:**
Older sister ☑	Older sister ☐
Younger sister ☐	Younger sister ☐
Older brother's wife ☐	Older brother's wife ☐
Younger brother's wife ☐	Younger brother's wife ☐

Level 4

MY FATHER'S OLDER SISTER
Rakgadi
Rakgadi yo mogolo
Click here for:
Additional information on a man's father's brothers and sisters ☐
Additional information on a man's wife's father's brothers and sisters ☐
Additional information: on kinship relation trees ☐
Other relations using *Rakgadi* ☑

Level 5

OTHER RELATIONS USING RAKGADI
My father's older brother's **daughter**:
Rakgadi e ka ba morwedi wa <u>ramogolo</u>
"Rakgadi can be the daughter of my father's elder brother"
My father's younger brother's **daughter**:
Rakgadi e ka ba morwedi wa <u>rangwane</u>
"Rakgadi can be the daughter of my father's younger brother"

Figure 8: Section of decision tree guidance for Sepedi kinship terms, presented as questionnaire, from Prinsloo and Bosch (2012)

4.5.4   Technical options for realising decision trees

The internal representation of lexical and grammatical data should be adapted to the decision-tree-like access to the data. For this, not only synoptic tables of

the respective words, but also a representation of the selection rules is needed.

We list in the following a number of interface solutions, by order of decreasing specialisation, which should be considered:

— Solve the problem, suggest the correct solution, and give a visual presentation and link to *Read mor*e sections such as FAQs or outer texts.

— Supply a link to *Read more* information where distinctions on a cognitive level are made.

— Supply a link to guidance on the basis of e.g. *Frequently made errors*.

— Give good, typical examples of use throughout.

All envisaged forms of presentation should be based upon a grammatical description of the construction to be tackled. One could argue that most of these issues have been sufficiently described in standard grammars. However, one should not assume that the format of these descriptions is such that they are ready to use for building decision trees. A reorganisation of the data and a fine-grained structuring (e.g. through markup) will be necessary.

In a text production or a text reception situation, a user can consult the (stand-alone) dictionary to obtain the required information. However, it is also possible that the support tool is integrated into a word processor the user is using to construct his/her text. In such a case, the user may require feedback on his/her own text production efforts based on his/her grammatical knowledge without specifically consulting the dictionary. The e-dictionary should then be integrated into the word processor as a grammar checker, similar to the features currently available in popular word processing software.

The following example is intended to show how we envisage practical text production work with a version of the decision tree dictionary integrated with the user's text processing environment.

Let us depart from a most common error scenario in Sepedi, for example, the user typing **selepe ke bogale* in order to express *the axe is sharp*. Learners usually know that *ke* means *it is* and that no distinction is made between *he is, she is, it is* and *they are* in Sepedi: all convert to *it is*, e.g. *(mosadi) ke mooki* ″she is (it is) a nurse″ so they incorrectly use *ke* with *bogale* "sharp". As a second example consider **mosadi o mooki* instead of *mosadi ke mooki* "the woman is a nurse″. Learners are accustomed to using the subject concord *o* with class 1 nouns in sentence construction and it is the correct form in two out of the three copulative relations (descriptive and associative copulatives; so attempting to use it also in the identifying copulative is a common error).

The student types **selepe ke bogale* in a word processor linked to the electronic dictionary and either *ke* alone or all three words together are flagged as incorrect. A quick solution is offered by means of a suggestion box, in this case offering three possibilities, namely *se, se le* "is" and *se lego* "who/what is". The user who has basic knowledge of the modal system will know which one to

select.

Most users, however, would need further guidance, and this is offered by a decision process guiding him/her through the three possible moods (see Appendix 1, sections 1–3) and the Indicative *se*, Situative *se le* or Relative s*e lego* of the decision tree examples below in section 5 for the descriptive copulative with sub-decisions. The process for **mosadi o mooki* is similar, i.e. a decision process guiding the user through the three possible moods (Indicative *ke*, Situative *e le* or Relative *e lego*) of the decision tree for the identifying copulative respectively, with sub-decisions.

4.5.5 Lexicographic processes towards the preparation of decision trees

The process to produce a dictionary article that provides a decision tree requires at least three sequential steps, building on one another:

— Step 1 would be to acquire comprehensive and accurate data for the set of rules etc. to be described. This includes the grammatical rules as well as relevant examples, common errors, etc.

— In Step 2 the lexicographer, in collaboration with a database expert, needs to reorganise the data so that it will be possible for a programmer to implement a decision tree. This requires at least two sub-processes:

— The logic of the decision process needs to be worked out very carefully, i.e., what is the logical sequence of the decisions, how much information is required to make and/or support the decisions, when are what type of examples needed, when are links to outer texts required, etc.

— The data need to be marked up in such a way that each of the data elements defined in the analysis of a specific complex problem can be identified at the required level of granularity. This implies that the database should make provision for such extensions, either by using an extensible XML schema or additional tables and fields in a relational database, depending on the original design of the system (cf. Bothma (2011)).

— In Step 3 the programmer takes the flow diagram of the decision tree together with all the explanations, examples and linked data, and implements this. The programmer should also design a "user-friendly" interface that is intuitive for the average user and guides him/her to follow the correct trail through the decision tree for the given information need.

It may be feasible to use multiple devices for the same phenomenon, as is clear from the preceding examples. For example, in the case of the kinship terms, one can provide a diagram only (see section 4.3 above), or a diagram with guidance

paths by means of different overlays (see section 4.4.2 above), or even a decision tree (see section 4.5.3 above). As indicated earlier, not all articles in a dictionary will necessarily be enhanced by means of any of these devices. Similarly, if one of these devices is applicable in a given case, it would evidently not be optimal to use multiple devices for the same phenomenon in the same article. The lexicographer will therefore have to do a very careful analysis of all lemmas to identify those lemmas and structured systems that can optimally benefit from any enhancement through the use of one of the above-mentioned information presentation devices, and he or she will subsequently have to do an equally careful analysis to decide which of the proposed technologies will be the best option in any given situation.

5.   Exemplification: complex cases of copulative selection

In this section, we come back to the structured system of the Sepedi copulative, which was mentioned in section 4.5.2, above. We now explain possible guidance scenarios in detail.

5.1   Different levels of user guidance

Figure 9 provides a schematic illustration of a pop-up guidance screen sequence for *selepe ke bogale*.
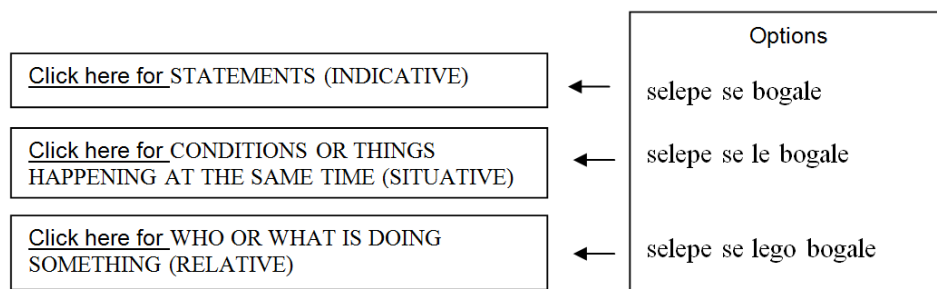
Figure 9: Dictionary feedback for *selepe ke bogale*

The user is informed that the only three correct options are *selepe se bogale*, *selepe se le bogale* and *selepe se lego bogale*. In many circumstances, observing the three correct options will enable the student from his/her knowledge of the language to select the correct one. If more guidance in respect of the descriptive relations in the Indicative, Situative and Relative is required, the user can click the buttons on the left hand side in figure 9 to display the information given in Figures 10, 11 and 12. The first option *selepe **se** bogale* is in the Indicative Mood, *selepe **se** le bogale* in the Situative Mood and *selepe **se** lego bogale* in the Relative Mood.

These grammatical terms are, however, avoided on the first level of guidance, and even on the second level, the distinction in meaning, i.e. statements, versus conditions etc. is given priority to the grammatical terms Indicative, Situative and Relative, which are given in brackets. These terms could be perceived as user-unfriendly by those users who do not know the grammar. In figures 3–5 they are, however, mentioned for the benefit of users who do have knowledge of the grammar.
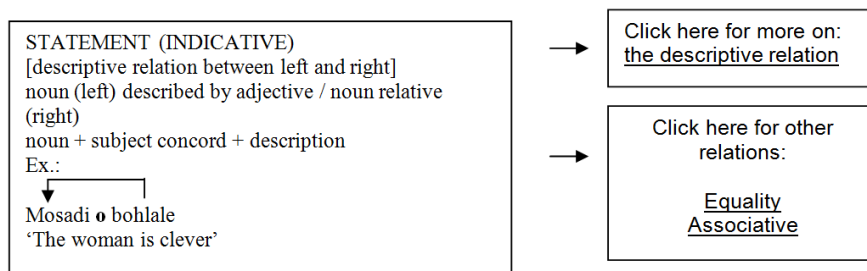
STATEMENT (INDICATIVE)
[descriptive relation between left and right]
noun (left) described by adjective / noun relative (right)
noun + subject concord + description
Ex.:

Mosadi **o** bohlale
'The woman is clever'

Click here for more on: the descriptive relation

Click here for other relations:

Equality
Associative

Figure 10: Pop-up information boxes for *selepe se bogale*

CONDITION OR THINGS HAPPENING AT THE SAME TIME (SITUATIVE)
[descriptive relation between left and right]
noun (left) described by adjective / noun relative (right)
noun + subject concord + le + description
Ex.:

(Ge) mosadi a **le** bohlale
'(If) the woman is clever'

Click here for more on: the descriptive relation

Click here for other relations:

Equality
Associative
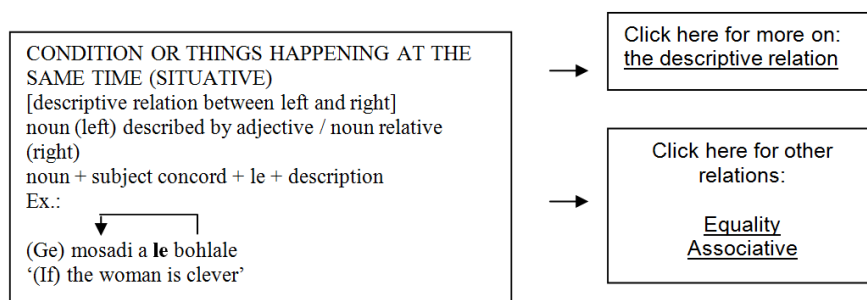
Figure 11: Pop-up information boxes for *selepe se le bogale*

WHO OR WHAT IS DOING SOMETHING (RELATIVE)
[descriptive relation between left and right]
noun (left) described by adjective / noun relative (right)
noun + (demonstrative) + subject concord + lego + description
Ex.:

Mosadi yo a **le**go bohlale
'The woman who is clever'

Click here for more on: the descriptive relation

Click here for other relations:

Equality
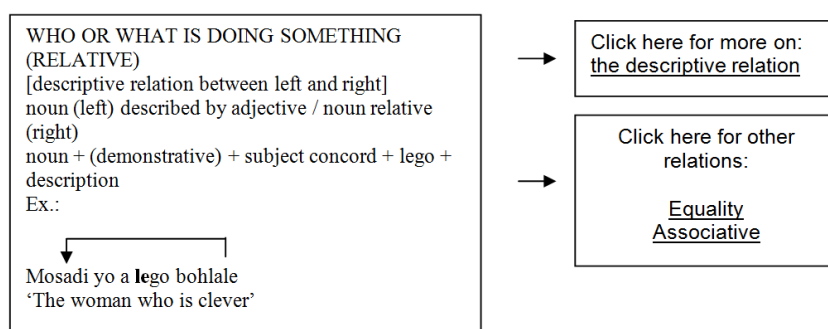Associative

Figure 12: Pop-up information boxes for *selepe se lego bogale*

In each case, the panel given in the left part of the mock-up provides the information needed for text production. Users with more (cognitive) needs can access a fuller picture via the buttons on the right hand side. Clicking the top buttons on the right hand side in figures 10–12 results in the presentation of more detailed information on the descriptive copulative, as shown on the left hand side of these figures. The buttons at the bottom on the right hand side of these figures result in pop-up boxes giving more information on the other two relations, i.e. equality and associative for each specific mood. So, for example, clicking the "Equality" option in figure 10 will provide more information on equality relations in the Indicative mood, the "Associative" option in figure 11 gives more information on associative relations in the Situative mood, etc.

5.2   From text production guidance to full grammatical guidance

Pop-up boxes giving more information and typical examples of the descriptive relations can be provided on a third level. These pop-up boxes will cover the entire sections 1–3 in the appendix. Compare, for example, guidance towards section 2, as in figure 13.

STATEMENT (INDICATIVE)	
Positive	Negative
(Nna) *ke* bohlale	(Nna) *ga ke* bohlale
'I am clever'	'I am not clever'
(Rena) *re* bohlale	(Rena) *ga re* bohlale
'We are clever'	'We are not clever'
Monna *o* bohlale	Monna ga *a* bohlale
'The man is clever'	'The man is not clever'

Figure 13: Pop-up information box for persons and classes in the Indicative Mood of the Descriptive Copulative

A second scenario is where comprehensive guidance is required, e.g. when the user wants to know how to say *is* in Sepedi. In this case a combination of decision processes is required. These processes are enriched with data from processed corpora linked with the dictionary.

6.   Conclusions

The project described in this article is driven by two underlying motivations, namely the urge to compile electronic dictionaries that can do better than current ones through maximal utilisation of advanced information technologies,

and the need for dynamic dictionaries guiding the user in innovative ways. We believe that step-by-step guidance, mainly through sequences of choices, the provision of additional relevant information on request as well as protection against incorrect conclusions users may draw from the data available to them, are the cornerstones of the design of such dynamic dictionaries.

As indicated in section 2, this project makes use of certain aspects of the Function Theory of lexicography, viz. we discuss the information needs from the perspectives of text reception, text production and cognitive situations. The examples we have discussed show that the same basic data can be used for providing the required level of guidance for each of the envisaged situations. Through the use of presentation devices of the kind illustrated in this article, the data are filtered to show only those items that are relevant for the specific information need. It is therefore not necessary to create different underlying data sets to guide the user in different situations. It is only necessary to provide the appropriate filters and presentation devices to extract precisely the data needed to solve the user's specific information need.

Careful analysis of the data needed to solve the user's information need in any given situation is therefore required: if the user has a text reception information need, data required for text production should not be included in addition in the material he/she gets presented, and *vice versa*. This selection task should be performed by a lexicographer with an excellent understanding of the language structures of the specific language, as well as of pitfalls of text reception and text production in the language. However, it also requires very serious input from an expert in database technologies and/or knowledge representation languages to organise and code the data in such a way that all relevant and indeed only relevant data required in terms of the specific type of information need can be extracted from the database. Data markup and data granularity are therefore extremely important issues to be considered (cf. also Bothma 2011: 90, 100).

Our current work only describes different devices that can be used to present information to users. How this can be implemented in principle is shown in the examples that we have provided. Practical implementation issues have clearly not yet been addressed in this article. This constitutes future work we intend to do, which includes a careful study of different database technologies and knowledge representation languages, to decide how practical implementation can best be done. Part of such a study is to do rapid prototyping of the same phenomenon by means of different technologies to find out what would, in practice, be the most effective and efficient way of realising the theory implicit in our discussions. And, obviously, once this is done, user evaluation would be required in terms of both the functionalities that are provided and also the specific interfaces that can be designed for each type of device.

As a matter of principle, data will be provided to the user only on demand. Therefore we don't envisage that each and every user will automatically be confronted with, for example, decision tree guidance in the case of

looking up the Sepedi copulatives in a text production situation. If the user finds sufficient data in the "standard" description of the copulative for text production and his/her information need is satisfied, the search ends and the user can carry on with his/her work task. If, however, the user needs further guidance because his/her knowledge of the language is insufficient to interpret the "standard" description correctly, additional information would be available on demand by means of the decision tree. The user is therefore in control of the information searching process and the amount of information displayed will not lead to information overload for the user who does not need extra information. Inherently such a choice of being presented with more (or less) information results in adaptive electronic dictionaries, as envisaged by Tarp (e.g. Tarp 2011), Bothma (2011) and many others. Such presentation devices will evidently not be added for all lemmas in a dictionary since not all of them are equally complex. Again, the lexicographer will be required to make a very careful analysis to decide which lemmas require such an extensive additional treatment.

We believe that continued research along the lines we have indicated in this article, together with further enhancements such as linking to corpora and other external data (cf. e.g. Heid, Prinsloo and Bothma 2012) and enhanced adaptive and interactive features will indeed lead to innovative electronic dictionaries that can better address the specific, individualised information needs of users in different situations.

7.   Endnotes

1.   http://www.isizulu.net
2.   http://www.amazon.com
3.   http://www.wiktionary.org
4.   http://www.oed.com
5.   http://dev.eurac.edu:8081/MakeEldit1/Eldit.html
6.   http://dev.eurac.edu:8081/MakeEldit1/Eldit.html

8.   References

Abel, Andrea. 2002. Darstellung der Verbvalenz in einem elektronischen Lernerwörterbuch Deutsch–Italienisch (ELDIT). Neue Medien, neue Ansätze. Braasch, A. and A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 413-418. Copenhagen: Center for Sprogteknologi, University of Copenhagen.

Amazon Kindle. [Online.] Available http://www.amazon.com.

Bergenholtz, Henning and Theo J.D. Bothma. 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.

Bergenholtz, Henning and Sven Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica. International Annual for Lexicography* 18: 253-263.

Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31:171-196.

Bergenholtz, Henning and Sven Tarp. 2004. The Concept of "Dictionary Usage". Dollerup, Cai (Ed.). 2004. *Worlds of Words. A Tribute to Arne Zettersten. Nordic Journal of English Studies* 3(1): 23-36.

Bergenholtz, Henning and Sven Tarp. 2005. Wörterbuchfunktionen. Barz, Irmhild, Henning Bergenholtz and Jarmo Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 11-25. Frankfurt/Bern/New York/Paris: Peter Lang.

Bothma, Theo J.D. 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.

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

De Schryver, Gilles-Maurice. 2009. State-of-the-Art Software to Support Intelligent Lexicography. Zhu, R. (Ed.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary & Lexicology:* 565-580. Beijing: Beijing Normal University. Also: http://www.hcxf.cn/read.asp?id=570.

ELDIT. http://dev.eurac.edu:8081/MakeEldit1/Eldit.html.

English Dictionary for South Africa. 2012. Cape Town: Pharos and KDictionaries.

Faaß, Gertrud. 2010. *A Morphosyntactic Description of Northern Sotho as a Basis for an Automated Translation from Northern Sotho to English.* PhD Dissertation, University of Pretoria. http://upetd.up.ac.za/thesis/available/etd-10092010-134539.

Heid, Ulrich, Danie J. Prinsloo and Theo J.D. Bothma. 2012. Dictionary and Corpus Data in a Common Portal: State of the Art and Requirements for the Future. *Lexicographica* 28. To appear.

isiZulu.net. [Online.] Available http://www.isizulu.net.

Lombard, Daniel P. 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L. van Schaik.

Macmillan English Dictionary for Advanced Learners. Rundell, M. and G. Fox. (Eds.). 2002. *Macmillan English Dictionary for Advanced Learners.* London: Macmillan.

Mayer, Felix. 1998. Eintragsmodelle für terminologische Datenbanken: ein Beitrag zur übersetzungsorientierten Terminographie. [Forum für Fachsprachen-Forschung, Bd. 44.] Tübingen: Narr.

Oxford English Dictionary Online (OED). [Online.] Available http://www.oed.com.

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

Prinsloo, Danie J. and Sonja 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, Danie J., Ulrich Heid, Theo J.D. Bothma and Gertrud 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. Bled: Trojína.

Rundell, Mike. 2009. *The Road to Automated Lexicography: First Banish the Drudgery ... Then the Drudges?* Keynote speech at eLex, Louvain-la-Neuve, 22–24 October 2009. http://uclouvain.be/en-271028.html.

Sesotho sa Leboa (Northern Sotho)–English Dictionary. http://africanlanguages.com/sdp/.

Tarp, Sven. 2008. Revival of a Dusty Old Profession. *Hermes. Journal of Linguistics* 41: 175-188.

Tarp, Sven. 2008a. Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. Tübingen: Niemeyer.

Tarp, Sven. 2009. Reflections on Data Access in Lexicographic Works. Nielsen, Sandro and Sven Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz:* 43-62. Amsterdam/Philadelphia: John Benjamins.

Tarp, Sven. 2009a. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.

Tarp, Sven. 2009b. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, Henning, Sandro Nielsen and Sven Tarp (Eds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow.* Linguistic Insights. Studies in Language and Communication 90: 17-32. Bern: Peter Lang.

Tarp, Sven. 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. e-*Lexicography: The Internet, Digital Initiatives and Lexicography*: 55-70. London/New York: Continuum.

Tarp, Sven. 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica 28.* To appear.

Wiegand, Herbert E. 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires.* Volume 1: 409-462 Berlin: De Gruyter.

Wiktionary. [Online.] Available http://www.wiktionary.org.

APPENDIX: THE COPULATIVE IN SEPEDI

Overview of the copulative constellations (Faaß 2010: Table 3.30, p.128)

Copulative	*Identifying (1)*		*Descriptive (2)*		*Associative (3)*	
Category	stative	dynamic	stative	dynamic	stative	dynamic
Tense						
Present	X	X	X	X	X	X
Perfect	X	X	X	X	X	X
Future		X		X		X
Mood						
Indicative (pos/neg)	X	X	X	X	X	X
Situative (pos/neg)	X	X	X	X	X	X
Relative (pos/neg)	X	X	X	X	X	X
Consecutive (pos/neg)		X		X		X
Subjunctive (pos/neg)		X		X		X
Habitual (pos/neg)		X		X		X
Infinitive (pos/neg)		X		X		X
Imperative (pos/neg)		X		X		X

STATIVE COPULATIVES

Section (1): Identifying Copulative (Equality between noun and complement)

Indicative *The man* *is a teacher*			*Situative* *If/while the man* *is a teacher*		*Relative* *The man who* *is a teacher*	
Pers/Num/Cl.	pos	neg	pos	neg	pos	neg
1PS	*ke*	*ga ke*	*ke le*	*ke se*	*ke lego*	*ke sego*
1PP	*re*	*ga re*	*re le*	*re se*	*re lego*	*re sego*
2PS	*o*	*ga o*	*o le*	*o se*	*o lego*	*o sego*
2PP	*le*	*ga le*	*le le*	*le se*	*le lego*	*le sego*
CL1 – CL18	*ke*	*ga se*	*e le*	*e se*	*e lego*	*e sego*

Section (2): Descriptive Copulative (Complement describes noun)

	(1) Indicative *The man* *is clever*		*(2) Situative* *If/while the man* *is clever*		*(3) Relative* *The man who* *is clever*	
Pers/Num/Cl.	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**
1PS	*ke*	*ga ke*	*ke le*	*ke se*	*ke lego*	*ke sego*
1PP	*re*	*ga re*	*re le*	*re se*	*re lego*	*re sego*
2PS	*o*	*ga o*	*o le*	*o se*	*o lego*	*o sego*
2PP	*le*	*ga le*	*le le*	*le se*	*le lego*	*le sego*
CL1	*o*	*ga a*	*a le*	*a se*	*a lego*	*a sego*
CL2	*ba*	*ga ba*	*ba le*	*ba se*	*ba lego*	*ba sego*
CL 3	*o*	*ga o*	*o le*	*o se*	*o lego*	*o sego*
CL 4	*e*	*ga e*	*e le*	*e se*	*e lego*	*e sego*
CL 5	*le*	*ga le*	*le le*	*le se*	*le lego*	*le sego*
CL 6	*a*	*ga a*	*a le*	*a se*	*a lego*	*a sego*
CL 7	*se*	*ga se*	*se le*	*se se*	*se lego*	*se sego*
CL 8	*di*	*ga di*	*di le*	*di se*	*di lego*	*di sego*
CL 9	*e*	*ga e*	*e le*	*e se*	*e lego*	*e sego*
CL 10	*di*	*ga di*	*di le*	*di se*	*di lego*	*di sego*
CL 14	*bo*	*ga bo*	*bo le*	*bo se*	*bo lego*	*bo sego*
CL 15-18	*go*	*ga go*	*go le*	*go se*	*go lego*	*go sego*

Section (3): Associative Copulative (Complement and noun associated)

	(1) Indicative *The man is* *with his dog*		*(2) Situative* *If/while the man is* *with his dog*		*(3) Relative* *The man who is* *with his dog*	
Pers/Num/ Class	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**
1PS	*ke na le*	*ga ke na (le)*	*ke na le*	*ke se na (le)*	*ke nago le*	*ke se nago (le)*
1PP	*re na le*	*ga re na (le)*	*re na le*	*re se na (le)*	*re nago le*	*re se nago (le)*
2 PS	*o na le*	*ga o na (le)*	*o na le*	*o se na (le)*	*o nago le*	*o se nago (le)*
2PP	*le na le*	*ga le na (le)*	*le na le*	*le se na (le)*	*le nago le*	*le se nago (le)*
CL1	*o na le*	*ga a na (le)*	*a na le*	*a se na (le)*	*a nago le*	*a se nago (le)*
CL2	*ba na le*	*ga ba na (le)*	*ba na le*	*ba se na (le)*	*ba nago le*	*ba se nago (le)*
CL 3	*o na le*	*ga o na (le)*	*o na le*	*o se na (le)*	*o nago le*	*o se nago (le)*
CL 4	*e na le*	*ga e na (le)*	*e na le*	*e se na (le)*	*e nago le*	*e se nago (le)*
CL 5	*le na le*	*ga le na (le)*	*le na le*	*le se na (le)*	*le nago le*	*le se nago (le)*
CL 6	*a na le*	*ga a na (le)*	*a na le*	*a se na (le)*	*a nago le*	*a se nago (le)*
CL 7	*se na le*	*ga se na (le)*	*se na le*	*se se na (le)*	*se nago le*	*se se nago (le)*
CL 8	*di na le*	*ga di na (le)*	*di na le*	*di se na (le)*	*di nago le*	*di se nago (le)*
CL 9	*e na le*	*ga e na (le)*	*e na le*	*e se na (le)*	*e nago le*	*e se nago (le)*
CL 10	*di na le*	*ga di na (le)*	*di na le*	*di se na (le)*	*di nago le*	*di se nago (le)*
CL 14	*bo na le*	*ga bo na (le)*	*bo na le*	*bo se na (le)*	*bo nago le*	*bo se nago (le)*
CL 15-18	*go na le*	*ga go na (le)*	*go na le*	*go se na (le)*	*go nago le*	*go se nago (le)*