The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries*

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, University of Ghent, Belgium, and*
D.J. Prinsloo, *Department of African Languages, University of Pretoria, South Africa*

Abstract: Good lexicographers are constantly striving to enhance the quality of their dictionaries. Since dictionaries are ultimately judged by their target users, there is an urgency to provide for the target users' needs. In order to determine such needs more accurately, it has become common practice to submit users of a dictionary to a series of tests to monitor their success in information retrieval. In most cases such feedback unfortunately comes too late so that it can at best be considered for implementation in the next or revised edition of the dictionary. In this article it is argued that feedback from the target users should be obtained *while the compilation of the dictionary is still in progress*, a process referred to as "simultaneous feedback". This concept, which offers a new methodology for compiling dictionaries, overcomes the major problem of creating and publishing entire dictionaries before feedback from target users can be obtained. By this new methodology, the release of several small-scale parallel dictionaries triggers feedback that is immediately channelled to the compilation process of a main dictionary. As such, the target users constantly guide the compilers during the entire compilation process. After a theoretical presentation of the new concept, the feasibility of simultaneous feedback is illustrated with reference to the creation of a bilingual Cilubà-Dutch learner's dictionary. It is shown how this main project has been successfully complemented by three parallel projects.

Keywords: SIMULTANEOUS FEEDBACK, NEW METHODOLOGY, MAIN DICTIONARY, PARALLEL DICTIONARIES, TARGET USERS' DESIRES, QUESTIONNAIRES, ELECTRONIC CORPORA, WORD-FREQUENCY STUDIES, CONCORDANCES, AFRICAN LANGUAGES, CILUBÀ

Opsomming: Die konsep van "gelyktydige terugvoering": Onderweg na 'n nuwe metodologie vir die samestelling van woordeboeke. Goeie leksikograwe streef voortdurend daarna om die gehalte van hul woordeboeke te verbeter. Aangesien woorde-

* An earlier version of this article was presented by G.-M. de Schryver as keynote lecture at the Fourth International Conference of the African Association for Lexicography, held at the University of Pretoria, 5-6 July 1999. This lecture was an extract from G.-M. de Schryver's M.A. dissertation (De Schryver 1999a), written under the guidance of D.J. Prinsloo and R.H. Gouws. For this version D.J. Prinsloo joined in the writing.

boeke uiteindelik deur hul teikengebruikers beoordeel word, is dit uiters noodsaaklik dat aan gebruikersbehoeftes voldoen moet word. Ten einde sodanige behoeftes noukeuriger te kan bepaal, is dit algemeen gebruiklik dat woordeboekgebruikers aan 'n reeks toetse onderwerp word ten einde hulle sukses ten opsigte van inligtingsinwinning te kan bepaal. Ongelukkig kom sodanige terugvoer in die meeste gevalle te laat deurdat dit slegs in die volgende of hersiene uitgawe van die woordeboek in berekening gebring kan word. In hierdie artikel word betoog dat terugvoering vanaf die teikengebruikers verkry moet word *terwyl die samestelling van die woordeboek nog aan die gang is*, 'n proses waarna verwys word as "gelyktydige terugvoering". Hierdie konsep wat 'n nuwe metodologie vir woordeboekmaak bied, omseil die kernprobleem van die voltooiing en publikasie van volledige woordeboeke alvorens terugvoer vanaf die teikengebruikers verkry kan word. Deur hierdie nuwe metodologie lei die vrystelling van verskillende kleiner parallelle woordeboeke tot terugvoer wat onmiddellik in die samestellingsproses van die hoofwoordeboek gekanaliseer word. Sodoende ontvang die samestellers deurlopend riglyne vanaf die teikengebruikers gedurende die volle duur van die samestellingsproses. Na 'n teoretiese uiteensetting van hierdie nuwe konsep, word die uitvoerbaarheid van gelyktydige terugvoering geïllustreer met verwysing na die voltooiing van 'n Cilubà-Nederlandse aanleerderswoordeboek. Daar word aangetoon hoe hierdie hoofprojek deur drie parallelle projekte aangevul word.

Sleutelwoorde: GELYKTYDIGE TERUGVOERING, NUWE METODOLOGIE, HOOFWOORDEBOEK, PARALLELLE WOORDEBOEKE, TEIKENGEBRUIKERS SE WENSE, VRAELYSTE, ELEKTRONIESE KORPUSSE, WOORDFREKWENSIESTUDIES, KONKORDANSIES, AFRIKATALE, CILUBÀ

1. Introduction

In modern lexicography the so-called user-perspective has emerged as an all-important criterion in the selection and lexicographical treatment of lexical items.

> Dictionaries first came into being in response to very practical needs. (Osselton 1983: 13)

> It is the function of a popular dictionary to answer the questions that the user of the dictionary asks, and dictionaries on the commercial market will be successful in proportion to the extent to which they answer these questions of the buyer. This is the basis on which the editor must determine the type of information to include. (Barnhart 1967[2] [1962]: 161)

> The user-perspective, so prevalent in modern-day metalexicography, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality. Good dictionaries do not only display a linguistically sound treatment of a specific selec-

> tion of lexical items. They are also products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are. (Gouws and Prinsloo 1998: 18)

Since the lexicographer is constantly striving to enhance the quality of a dictionary, there is an urgency to take the target users' needs, their expectations and their reference skills into account (cf. also Atkins 1998, Wiegand 1998, Van der Merwe 1999). In addition to these aspects, target users' *desires* are introduced in the present article. These desires denote a set of requests which go beyond the linguistic information required to satisfy the users' *needs*, beyond the users' *expectations* based on their perception of how dictionaries ought to be, and beyond the users' *reference skills* developed in the course of consulting so-called "standard dictionaries". Indeed, by referring to target users' desires we want to go to the core of *what* the target users would like to find in a dictionary, and *how* they would like to find it — it is the users who decide on the types of data to include in or omit from a dictionary, and it is the users who decide on the way this data is presented. For example, users might indicate that they desire an indication of frequency of use but not etymological information, or users might indicate that they desire nouns to be lemmatised in one format but adjectives in another. Information on such desires must be collected by both informal and formal means. From the moment dictionary compilers take target users' desires into account, they acknowledge that today's target users are the *privileged critics* of a dictionary.

It is however not easy for the compiler to determine exactly what these needs, expectations, reference skills and desires of the target users are. Many compilers simply rely on their intuition as far as these are concerned. In order to determine these needs, expectations, reference skills and desires more accurately, it has become common practice to submit users of a dictionary to a series of tests to monitor their success in information retrieval. Atkins (1998) is an excellent example of current efforts to research dictionary use. However, such research is carried out at a stage when the dictionaries are already completed and published. Feedback from target users can only be implemented in forthcoming editions of these dictionaries. Formulated differently, it means that the feedback obtained in this way comes too late. Tono (1992: 232) is correct when he remarks:

> Probably the most widely used research method with regard to the dictionary user is the survey … All variables are studied *ex post facto* that is, as they exist in the situation. No experimental variables are manipulated. Furthermore, most of the surveys on dictionary-users' study are limited to describing the *status quo.*

In this article it is argued that feedback from the target users should be obtained while the compilation of the dictionary is *still in progress*, a process for

this reason referred to as "simultaneous feedback". This concept which offers a new methodology for compiling dictionaries, overcomes the major problem of having to create and publish entire dictionaries first before one can start to evaluate them.

In a nutshell, this new methodology entails the release of several small-scale parallel dictionaries which trigger feedback that is channelled back to the compilation process of a main dictionary. Within the framework of simultaneous feedback, the variables are thus not studied *ex post facto*, as Tono notes. They are studied while they are being implemented, since the target users guide the compilers during the entire compilation process. As such, it is not a description of the *status quo*.

The concept of simultaneous feedback grew out of a dire necessity to compile a bilingual Cilubà-Dutch learner's dictionary. It quickly became evident that the main problems encountered in the lexicographical treatment of Cilubà are shared by the other African languages.[1] The solutions that will be presented therefore have implications that go beyond the Lubà language. The theoretical concept of simultaneous feedback itself goes far beyond the African languages — it applies to all languages. This is why we claim that it is a new methodology for compiling dictionaries.

2. The theoretical framework of simultaneous feedback

Compilers write dictionaries to be used by target users. Target users consult the dictionaries and in this process they come to judge those very dictionaries. Dictionaries' privileged critics are therefore the target users who were the original focus of the compilers. We have come full circle, from compilers to target users, to dictionaries and back to compilers. This full circle is not just a linguistic loop, because the loop represents indispensable feedback. However, for most dictionary projects this feedback comes too late. In fact, one would like to have feedback from the privileged critics (the target users) both from the very start and during the entire compilation process of a dictionary — hence, the chimera would be to listen to simultaneous feedback from the target users to the compilers. But is it really a chimera? With the concept of simultaneous feedback, it is *not*. A schematic representation of the theoretical framework of simultaneous feedback is shown in (1).

In (1) one recognises the three primary constituents of any dictionary compilation process, viz. target users, compilers and dictionaries. The framework itself should roughly be read from left to right, and from top to bottom — [1] through to [10]. However, as will be seen, every single component is actually interlinked with every other single component.

The compilers' central task is the compilation of a main dictionary, based on a main electronic corpus — depicted by the central arrow in (1). The first component in (1) is also the first task. That is, the formulation of a theoretically motivated model for the structure and contents of the dictionary to be com-

(1) The theoretical framework of "simultaneous feedback"

piled must be followed by a preliminary analysis of the needs, expectations and reference skills of the potential target users of the dictionary — [1]. It is of paramount importance that from this initial stage onwards, information concerning the target users' desires be gathered through informal and formal consultation with the future target users — [1]. *As such, feedback is simultaneously introduced right from the very start.*

Since the main dictionary is to derive its data from a main corpus, the compilers have to build an electronic corpus for the specific language. As a result, the compilers cannot start the compilation of the main dictionary right away and are moreover confronted with the prospect of an extremely time-consuming undertaking. In order to overcome this deadlock the main dictionary project is *complemented* by a series of small and inexpensive parallel dictionary projects. These parallel projects have the same basic structure, contents and target users as the main project — [2], [6], [10], etc. These parallel projects are to derive their data from small-scale parallel test-corpora and are to be completed in short periods of time. From the release of the first parallel dictionary onwards, informal and formal feedback is received from the parallel projects and channelled back into the time-consuming main project — [3], [7], etc. *From this instant, the compilation of the main dictionary becomes a true "work in progress" with simultaneous feedback from the target users to the compilers.*

The parallel projects are thus used as experimental tools to test a plethora of strategies in order to refine the presentation of the information in the main project under construction. Once a structured main corpus has been built, word-frequency studies can be done to assist the lexicographers in the compilation of the lemma-sign list of the main dictionary — [4]. Subsequently, concordance lines, also derived from the main corpus, supplement and confirm the compilers' intuition in the compilation of the main dictionary articles — [8]. Until completion of the main project the parallel projects continue to elicit feedback — [5], [9], etc. All these instances of simultaneous feedback ultimately enable the compilers to select the most appropriate blend of lexicographical procedures to ensure the most effective retrieval of information by the target users in the main dictionary.

Consequently, within the framework of simultaneous feedback, we cannot agree with the claim that "lexicography is not a terrain in which you can experiment from one day to the next" (Lombard 1994: 211). On the contrary, complying with the target users' desires means doing just that!

It goes without saying that this theoretical framework should be considered the backbone of any potential dictionary project. In the remainder of this article, the concept of simultaneous feedback will be applied to the compilation of a functional Cilubà-Dutch source of reference. For this project, the components [1] through to [10] shown in (1) will be analysed, with the main focus on the first few components.

[1] Formulation of a theoretical model, and Analysis of needs, expectations, reference skills and desires

(a) Formulation of a theoretically motivated model for the structure and contents of the main dictionary project

The ultimate goal of a *Woordenboek Cilubà-Nederlands* (Cilubà-Dutch Dictionary), WCN for short, is "a general and alphabetically ordered, pocket-sized unidirectional decoding learner's paper dictionary, based on frequency of usage and meant for university students". It is important to stress the fact that the promptness with which the precise target and target group were determined should not be considered as mere catch-as-catch-can choices. Indeed, before embarking on a dictionary project one must take the nature and comprehensiveness of "currently available lexica" seriously into account in order not to compile "just another general dictionary" to be used by "unspecified target users". Prior to the project being discussed here, only one general dictionary with Cilubà as the source language was readily available, viz. the *Dictionnaire Tshilubà-Français* (De Clercq and Willems 1960[3]). In this dictionary, Cilubà, French and (some "keywords" in) Dutch are covered in a one-way configuration. Unfortunately, a speaker of Dutch who is not acquainted with French is unable to use this dictionary satisfactorily. A bilingual Cilubà-Dutch dictionary is thus certainly not a redundant product.

(b) Preliminary analysis of the needs, expectations and reference skills of the target users

From the start we had a very specific target group in mind, viz. university students. Their general needs and expectations are reflected in the project's ultimate target: "a general and alphabetically ordered, pocket-sized unidirectional decoding learner's paper Cilubà-Dutch dictionary". It should be noted that these needs and expectations were only finalised through informal conversations with the future target group.

As far as the target users' reference skills are concerned, suffice it to say that as they are university students, we could presuppose that these skills would be quite well developed.

(c) Informal and formal consultations with the target users in order to determine their desires

The target users' desires were recorded in two ways. Firstly, the actual use of the *Dictionnaire Tshilubà-Français* by university students was monitored in an informal way. The gathering of informal data from actual dictionary use might seem trivial, but it definitely constituted "a battery of learners' desires" — col-

lected, it should be stressed, through natural participant observation. Secondly, information was also gathered in a more formal way via direct questions like "Imagine someone would compile a Cilubà-Dutch Dictionary. Ideally, what would *you* like to find in such a dictionary? And how would *you* like to find it?"

[2] First parallel dictionary based on a first parallel test-corpus

In order to get a dictionary project going within the framework of simultaneous feedback, one quickly needs to determine a subtarget and a subtarget group for a first pocket-sized parallel project which can be compiled in a short period of time and which is based on a first parallel test-corpus which can be built promptly — and all this within a limited budget. We projected that the first parallel project would have to result in a small unidirectional decoding learner's lexicon for Dutch-speaking learners of Cilubà at the Department of African Languages and Cultures of the University of Ghent, Belgium. For that specific subtarget and subtarget group, it was easy to foresee the series of potential Lubà lemma-signs for which the future users would be likely to take to the lexicon, namely those "heard" during the theoretical courses and practical exercises, and those "read" while translating texts from Cilubà into Dutch and studying the different syllabi. It was therefore decided to put together a first parallel test-corpus consisting of all elements encountered during the first two years of the Cilubà courses. Hence, the orthographic, semantic and grammatical information required to construct the lexicon were to be extracted primarily from those courses. The corpus itself was called the "Course Corpus", while the resulting *Lexicon Cilubà-Nederlands* (De Schryver and Kabuta 1997), henceforth LCN, was compiled in just two months and contained roughly 2 500 lemma-signs. From the sample pages shown in Appendix A, it is clear that the aim was to deviate as little as possible from so-called "standard dictionary conventions".

Instead of thoroughly discussing all the aspects of this first parallel dictionary, we will focus on two topics, nouns and verbs. On the lexicographical treatment of nouns in African languages, the dilemma has always been "How does one lemmatise nouns successfully?" (see e.g. Prinsloo and De Schryver 1999). We chose to enter nouns as singulars, followed by the gender — where the two poles of the gender are thought to enable the user to form, when applicable, the plural. Consider the nouns in (2):

(2) (a) **kààfê** [12/4] cf var **kàfê**
(b) **kaakù** [1/2a] grootouder; voorouder; **~ *mukàjì*** [ud] *grootmoeder*; **~ *mulùme*** [ud] *grootvader*
(c) **kabèji** [12/13; dim ◂ **dibèji**] papiertje; **~ *kàà bwanga*** *(dokters)voorschrift*

Here the genders 12/4, 1/2a and 12/13 are thought to enable the target users to infer the plural forms.

When it comes to verbs, the treatment of a typical verb in LCN is illustrated in (3).

(3) **-dyà** [tww; cf **spw3**, **5**] eten; **~ *kuukuta*** [ud] *eten en verzadigd z* ▸ **bidyà; cidìilu; cyàkudyà; -dìika; -dìikiibwa; -dìila; -dìisha; mudì; Mudìila-mpiku**

As can be seen in (3), a verb is lemmatised in the imperative in LCN, thus without its class prefix **ku- —** a rather traditional approach, as bringing all the verbs together under **ku/kw** would only result in an artificial overcrowding of one particular lexicon-section.

From (2) and (3) one could get the impression that articles stand in relative isolation in LCN. This view is incorrect, as a comprehensive network of cross-references can link any reference position within an article with any internal or external reference address. We will briefly look at two novelties, "verb stem node" and "noun node". In LCN one encounters two kinds of large networks, each one with its respective "reference nodes". The first network (when present) is centred around the stem of verbs. Its purpose is to use this form as a node to link all lemma-signs that are connected. However, one is only referring to lemma-signs *within* the lexicon, and nothing is claimed about other possibilities in Cilubà. The "tail slot" was created for this purpose. It is to be found at the end of an article and starts with an arrow pointing to the right followed by one or more lemma-signs connected with this single "verb stem" (the head of the article). For instance, for the verb **kudyà** in (3), one finds nine lemma-signs in the tail. Under their proper alphabetical position, all of these nine evidently contain a cross-reference back to the node, mostly through the use of an arrow pointing to the left. Such cross-references can be seen in (4) where a sample of the lemma-signs from **kudyà**'s tail slot are shown as they can be found under their proper alphabetical position:

(4) (a) **cidìilu** [7/8 ◂ app **-dyà**] **1** eetzaal; **2** kribbe
(b) **cyàkudyà** [cn sub 7/8 ◂ **-dyà**] voedsel; ***mukàndà wà byàkudyà*** *menu(kaart)*
(c) **-dìika** [iww, sta **-dyà**] eetbaar z; gegeten w; afgeknaagd w; verteren
(d) **mudì** [1/2 ◂ **-dyà**] eter; verslinder

Although one can construct tail slots for several part-of-speech types, a second important network uses nouns as nodes. The method being analogous, it suffices to consider one example of a "noun node" together with the cross-references back to the node. One such small network is shown in (5):

(5) (a) **diitu** [5/6] bos, woud
▸ **-à mwitu; kwitu; mwitu**
(b) **kwitu** [loc 17/Ø ◂ **diitu**] cf var **mwitu**
(c) **mwitu** [loc 18/Ø ◂ **diitu**; var **kwitu**; vgl **cisuku**] in 't bos/woud; a/h water
□ **-à ~** [cn adj] wild; iets dat groeit/verblijft in 't bos/woud; ***ngulube wa ~*** *wild varken/zwijn, everzwijn*

Within the framework of simultaneous feedback, we maintained that we cannot agree with the statement that one is not able to experiment in lexicography. However, one cannot really experiment without making suggestions first. This is why the first parallel project, LCN, makes suggestions. And since LCN was actually published, suggestions had to be made for every possible aspect of the lexicon.

"There are various types of financiers that a lexicographer can approach," says Alberts (1999: 4). "Private funds may, as a first option, be raised by tapping one's own … financial resources." This is exactly what was done for the first parallel project. The booklets themselves, which cost just 3US$ each to produce, were given to the target users for free. From the moment the first parallel project was distributed, feedback was instantly received.

[3] Informal and formal feedback on the first parallel project

(a) Informal Files

On the day LCN was released, a series of Informal Files were opened to track every remark concerning the lexicon and its use. As far as dictionary users are concerned, we are convinced that these files represent a rare first-hand insight into target users' (initial) opinions regarding a dictionary's release and its subsequent use. It should be stressed that all the opinions were recorded in our capacity as natural participant observers. A different kind of feedback, namely academics' informal feedback, was received through discussions with scholars and following the reading of a paper at a Lexicography Seminar in Tervuren, Belgium (De Schryver 1998). The feedback in the Informal Files ranges from

plain suggestions to improve the layout, to unclear articles, codes and abbreviations, and valuable recommendations for different lemmatisations.

(b) The quest for formal feedback

Besides informal feedback, its formal counterpart was also collected. From the manifold standardised methods that are available to obtain formal feedback we opted for questionnaires through a mail survey. In order to view the different ways of receiving feedback from a somewhat broader perspective, one can make use of Galtung's methodological typology for differentiating between various forms of social research. According to him, there are nine "basic ways" of collecting data, as listed in (6).

(6) Basic ways of data-collection (Galtung 1967: 109-121)

We see that the data collected in the Informal Files can be assigned to two types: the target users' remarks can be assigned to type 1, while all the scholars' remarks can be assigned to type 4. The type of data-collection to be expounded on now mainly belongs to type 9, and to a lesser extent also to type 8. Regarding the target users, it becomes apparent that we moved from one extreme (type 1) to the other (type 9). In this way we hope to have balanced the type of feedback collected.

Once one starts browsing the literature, one realises that the advantages of mailed questionnaires are legion. One huge drawback of mailed questionnaires, however, is the large percentage of nonresponses. Over the years, different procedures for reducing this large percentage of nonresponses have been suggested, the most tempting probably being Dillman's "Total Design Method" (1978), for short TDM. Dillman claims that his method boosts the response from 30 to 70%, a percentage comparable to face-to-face interviews. Basically, TDM is constructed around three keywords:

(a) "personal character": one should try to give the mailed questionnaire a personal touch;
(b) "trade tricks": one should use all the latest trade tricks, in that the questionnaire itself should be well-balanced and specifically designed for the specific target respondents; and

(c) "follow-ups": one should incorporate three follow-up mailings.

TDM being a *total* design method, (a), (b) and (c) should *all* be implemented in order to attain the 70% response. With the variant we used, 69% of the questionnaires were eventually returned.[2]

(c) LCN Questionnaire

The Questionnaire itself, given in full as Appendix B, was four pages long and was packed with fifty questions covering all the various aspects of LCN, as well as some potential future endeavours. In spite of this large number of questions — with many of them even subdivided into subquestions — nearly all respondents tried to reply to every single question. It is a golden rule in an anonymous questionnaire not to place "personal questions" at the beginning, since the respondents might otherwise doubt whether the questionnaire is indeed intended to be anonymous. This is why the three questions with a personal touch were placed at the end (Q46-48). In addition, the questions were grouped together according to subject: Generalities (Q1-5), Layout (Q6-12), Abbreviations and Symbols (Q13-16), Grammar "within" the *Lexicon* (Q17-22), Nouns (Q23-29), Adjectives (Q30), Verbs (Q31-34), Cross-references (Q35-36), Outlook (Q37-40) and a Round-up (Q41-50). The way these different groups were laid out had to please the eye (cf. Dillman's keyword (b) above).

All questions were kept very short and with as simple a wording as possible. This is why, e.g., *inbrengen* (enter) was used instead of lemmatise, or *trefwoord* (headword) instead of lemma-sign. The use of easy phraseology and examples simplified communication between compilers and target users.

Closed questions ("yes/no" or "multiple-choice", where one simply had to tick off one or more alternatives) were the most common type. Other questions were more open-ended, in that different possibilities had to be rated using numerals. Lastly, only one tenth of the questions were true open questions (in which the respondents were asked to write their answers down).

(d) Analysing the formal target users' feedback on LCN

The analysis of the target users' formal feedback on LCN can be classified as follows:

(i) the target users' needs and expectations;
(ii) the target users' reference skills; and
(iii) the target users' desires.

In this tripartite structure one instantly recognises the theoretical framework — [1] in (1), and the preliminary analysis. One or more representative questions

from the Questionnaire will now be used briefly to illustrate each of these classifications.

(i) With Q3 we went to the core of the target users' needs: We wanted to know for which type of data they consulted LCN. The results are shown in (7).[3]

(7) Why take to LCN?

It is clear that every competence level (whether elementary, intermediate or advanced) used LCN primarily for semantic purposes (27%, 32% and 27% respectively). In addition, all levels rank the tonal dimension in the second place, orthography in the third, and tables in the fourth.

(8) Slot order

One cannot claim that this is also what the target users actually *need*. Since the target users were asked for which purposes they consulted LCN, the chosen data types were restricted to those available in LCN. To determine the target users' *expectations* in a lexicon of this kind, one can consider Q39. In Q39 we wanted to learn what particular kind of information the target users would like to find first in a future Cilubà-Dutch Dictionary. The results are shown in (8).

All three levels expect a translation/description first. The answers to Q39 thus confirm the conclusion drawn from Q3; that the target users' first and foremost need is semantic information.

(ii) As for the target users' *reference skills*, the Questionnaire dealt with both general reference skills and reference skills in a Lubà context. Regarding the general skills, the results confirm the preliminary analysis that these are rather well developed. As for the skills in a Lubà context, the answers to one particular question are very revealing. In Q15 the understanding of the "place-holder symbol", also called "repetition symbol" (cf. e.g. Hartmann and James 1998: 119), was tested in a Lubà context. This nontypographical structural marker, the tilde (~) in LCN, simply replaces the lemma-sign within a particular article. However, considering the variety of answers to Q15 shown in (9), one is surprised.

(9) Target users' reference skills in a Lubà context

In order to understand the confusion raised by Q15, one needs to take account of the following Lubà rule. A word-final high tone becomes low when the preceding syllable contains a falling tone or a low tone *and* when it is followed by a pronominal prefix carrying a low tone. With this rule, the above becomes **bânà bààbò**. Although this simple rule is mentioned within the first weeks of the Cilubà courses, the answers to Q15 show that the rule is *not* applied by the majority of target users. Even if one makes provision for this potential change in tone in the front matter of the dictionary by pointing out that one should take "floating tones" into account — as was done in LCN on p. xv, § 4. — this is not sufficient. The answers show that the place-holder symbol cannot be used whenever a word-final high tone changes into a low tone, as only one out of ten target users possesses this one "reference skill in a Lubà context".

(iii) The bulk of the Questionnaire dealt with the target users' *desires*. The Questionnaire was rather exhaustive in this respect, as it dealt with dictionary aspects ranging from grammar in the lexicon to nouns, adjectives, verbs, cross-references, layout, and a final round-up. Since it is impossible to review all this feedback here, we will restrict our discussion to two specific issues: the lem-

matisation of nouns on the macrostructural level and the "node networks" centred around verb stems.

In the Questionnaire, seven questions dealt specifically with nouns, of which four dealt with the lemmatisation of nouns on the macrostructural level. The latter four are shown in (10).

(10) Nouns: macrostructure

In Q23, after reminding target users of the lemmatisation approach utilised in LCN, we sought to learn how they would like to look up nouns. For only 58% of the elementary learners, LCN's approach corresponded to the way they would like to look up nouns. For intermediate and advanced learners, LCN is on the right track, as 93% and 91% of them respectively mark LCN's approach as conforming to their desires. Since 25% of the elementary learners opted for both singulars and plurals, whilst *none* of the more advanced learners did, this suggests a certain unfamiliarity with the noun class system among beginners. In addition, one target user rightly remarked that entering both singulars and plurals would be handy but uneconomical because of the space it would take up in the dictionary.[5]

The target users' desire is thus: *For intermediate and advanced learners one can safely lemmatise nouns under their singulars as long as provision is made for*

irregular plurals, but for elementary learners additional guidance on the noun class system will have to be provided for if this is to be a successful approach. Alternatively, one might consider lemmatising both singulars and plurals for inexperienced users.

In Q24 we wanted to know where the target users would like to find irregular plural forms. When one considers the large number of respondents who added "under both", a certain logic appears. Moreover, one target user (rightly) observed that cross-references should link plurals and singulars.

Hence, the target users' desire is: *Irregular plural nouns must be mentioned both within the article of their singular forms and be lemmatised under their proper alphabetical position, whilst cross-references should link both forms.*

In Q25 we sought to learn whether target users need gender information, and if so, in which format. What is extremely interesting here is the option not one single target user chose, namely that class numbers as well as nominal prefixes are redundant. Though some respondents suggested including both class numbers and nominal prefixes, roughly half of the elementary and intermediate learners and nearly all the advanced learners opted for class numbers.

The target users' desire is thus: *Noun gender information is crucial and should, for advanced learners, preferably be coded using class numbers, whilst elementary and intermediate learners might desire to receive some additional guidance.*

In Q28 we wanted to know the target users' opinions about the "stem tradition". 67% of the elementary, 87% of the intermediate and no fewer than 91% of the advanced learners consider this tradition to be much too complicated. If one looks at the respondents' additions, they suggested that it might reveal lexical relations, that it might be useful to add the stem within each noun article, that it might be better to include both stems and nouns, and that the lexicon might become too large with the inclusion of stems. Also, one target user (rightly) suggested that this approach would simply require too much look-up time.

Hence, the target users' desire is: *Under no circumstances should nouns be lemmatised according to the "stem tradition".*

The true power of a questionnaire is well illustrated through these four questions. As a second topic to illustrate this power, one can consider the "node networks" centred around verb stems. Without going into the underlying theoretical details of the need for harmonising lumping and splitting, Q35 asked the target users whether they thought the node networks were useful. The results are shown in (11).

Even though the answers to Q35 are somewhat dispersed around different "useful aspects", especially for the intermediate level, the target users' judgement stands out clearly: the node networks *are* useful. The favoured aspect is that one is able to see "connections".

(11) Cross-references

It is hoped that these examples illustrate that a carefully designed questionnaire enables the lexicographer to track, evaluate and adjust all possible aspects of the dictionary process.

When simultaneous feedback was described at the start of this article, it was not only claimed that a dictionary project could be launched within this framework, but also that it should run concurrently with the development of electronic corpora. The best time to start compiling the main electronic corpus is between the release of the first parallel project and the stage when the feedback on that first parallel project is analysed. In doing so, the progress of the project is not delayed. In truth, simultaneous feedback from the parallel projects, combined with the wealth of information derived from the electronic corpus, constitute the basic building blocks for the main dictionary. The creation of an electronic Lubà corpus is precisely the subject of the next section.

[4] Main electronic corpus: Word-frequency studies, and The lemma-sign list

The *Collins COBUILD English Language Dictionary* (Sinclair 1987), known as COBUILD1, was derived from the earliest electronic megacorpus, the 7,3-million-word COBUILD Main Corpus (Renouf 1987). As COBUILD1 contained roughly 70 000 "references", this meant that approximately 100 running words in the corpus were required for each reference in the dictionary. Our objective was to apply this *same ratio* to the second parallel project. Increasing LCN's lemma-sign list by 20%, from 2 500 to 3 000 lemma-signs, seemed feasible. Applying the ratio derived from COBUILD1 meant that we had to create an electronic corpus of 300 000 words.

Furthermore, from the various corpora types discussed in the literature on corpus linguistics (cf. e.g. Kennedy 1998: 2, 20, 52, 62, Summers 1993: 186, 190)

we retained the notion of a "structured corpus", and decided to follow a combined genre/topic stratification.

The resulting corpus, which has been named "Recall's Cilubà Corpus", henceforth RCC, consists of seven subcorpora (i.e. magazines, traditional stories, informal literature, textbooks, scientific works, religious works and miscellaneous sources) of circa 40 000 words each and one subcorpus (i.e. poetry and proverbs) of circa 20 000 words.[6] The outline of this "structured Lubà corpus of 300 000 words" is thus straightforward. Each genre/topic area is given equal weight, except for the block poetry and proverbs, which is only half the size of the others. Moreover, the different spoken sections in RCC amount to 17,4%.

A recent article by Kilgarriff (1997) is titled "Putting Frequencies in the Dictionary". As the title suggests, frequencies are *put in* the dictionary. This, of course, might be a nice addition to an existing dictionary. In our understanding of the utilisation of word-frequency studies, we would rather use frequency data to help in deciding *what* to include *how* in a dictionary. As such, a lemmatised frequency list is a crucial aspect of simultaneous feedback. If one wishes to compile, say, a parallel dictionary containing 2 000 lemma-signs, one can start by selecting the 2 000 most frequent lemmata in the corpus. If, however, the aim is to compile a parallel dictionary containing 5 000 lemma-signs, one can select the top 5 000 lemmata in the corpus, etc.

To enable the compilation of the lemma-sign list for the second parallel dictionary using data from word-frequency studies, we needed a "lemmatised frequency list" and information on the "distribution" of those lemma-signs across the eight subcorpora. Statistical software tools were thus required with which RCC could be analysed. We singled out WordSmith Tools, for short WST (cf. URL WordSmith), for its versatility in handling corpus queries.[7]

In order to generate a lemmatised frequency list for the most frequent types we simply went through the top 1 000 items of the corpus and lemmatised them "by hand". For nouns this meant that when we encountered a singular form, we added the frequency of the plural form (or vice versa), where applicable. For verbs this meant that we kept track of those verbs we had already encountered and added the frequency of every single "conjugated form" we encountered subsequently. In addition to this "true lemmatisation" we joined divergent orthographies. The latter was done for all possible parts of speech. In this way we obtained a new list in which well over the first 600 types had been lemmatised.

The inclusion or omission of particular lemmatised forms cannot be based solely upon overall frequencies. This is best illustrated by the counts in (12), which show the distribution of all the unlemmatised constituents of *one* particular lemma across the eight subcorpora.

Whether the singular noun **munsantu** or the plural form, or whether the word-final **-u** or the word-final **-o** variant, the distribution is in all instances extremely uneven, with a clear bias towards religious works. Therefore, lem-

(12) Distribution of one lemma across the eight subcorpora

mata of which all constituents show such an uneven distribution certainly do not qualify for inclusion among the "truly frequent" lemma-signs.

This procedure enabled us to determine the top 600 lemmata in the 300 000-word RCC. This data provides us with the means to mark the 600 most frequent lemma-signs in the second parallel dictionary project. While work on the electronic corpus was progressing, a last type of feedback on the first parallel dictionary project was received, namely formal academics' feedback.

[5] Reanalysis of the target users' desires, and Formal academics' feedback

Regarding this form of simultaneous feedback, the idea is simply to contrast the target users' desires with formal academics' feedback. Since a number of lemmatisation strategies proposed in LCN are a kind of package deal, for example the node networks discussed above, one cannot implement *all* the target users' desires *at once* without taking the formal academics' critical reviews into account. On the node networks suggested in LCN, Gouws and Prinsloo (1998: 17, 31-32) wrote:

> The mediostructure, that is the system of cross-referencing, is a lexicographic device that can be used to establish relations among different components of a dictionary … An excellent example in African language lexicography where mediostructure has been employed as a powerful access structure is the *Lexicon Cilubà-Nederlands* (LCN) compiled by De Schryver and Kabuta. This dictionary is highly successful in interconnecting the knowledge elements represented in different sectors of the dictionary on several levels of lexicographic description to form a network … The compilers of LCN are aware of the benefits of "keeping together what semantically and grammatically belong together" but also of the need (a) to avoid extremely long entries and (b) to ensure proper treatment of each derivation in terms of grammatical, tonal and lexical information … The compilers of LCN thus succeeded in harmonising lumping and splitting, capturing the advantages of both these approaches. It can, of course, be argued that the listing of the different derivations [in the tail] occupies precious space in the dictionary. However, by substantially reducing the font size, this redundancy is diminished.

Simultaneous informal and formal target users' feedback, interlinked with

simultaneous informal and formal academics' feedback, in combination with the active use of electronic corpora, should ultimately produce a "good dictionary answering all potential questions". The second parallel dictionary was such an attempt.

[6] Second parallel dictionary based on a second parallel test-corpus

The second parallel dictionary, the *Beknopt woordenboek Cilubà-Nederlands* (De Schryver and Kabuta 1998), henceforth BCN, is probably best introduced with some sample pages. Two pages can be found in Appendix C. Notwithstanding our conviction that the prime function of a lemmatised frequency list is to be the basis for the compilation of the lemma-sign list of a dictionary, we can obviously also use this data to *indicate* frequencies in the dictionary. In terms of Kilgarriff (1997) it means "putting frequencies in the dictionary". On BCN's page 23 for example, shown in Appendix C, one sees that **dîba** is preceded by a ←, which means that it belongs to the 200 most frequent lemmata; **dibòko** is preceded by a ↑, which means that it belongs to the 400 most frequent ones; and **dib**. is preceded by a →, which means that it belongs to the 600 most frequent ones. Together with all the lemmata that have not been given a symbol (80% of them), this illustrates the four-way categorisation utilised in BCN. It should be noted that this procedure is a variant of the approach found in the *Longman Dictionary of Contemporary English* (Summers 1995[3]), known as LDOCE3. As such, this is probably the first African language dictionary in which frequencies are indicated.

Of the other new features introduced in BCN, we will now focus on one which illustrates well how the target users' feedback can effectively be taken into account, whilst not jeopardising the academics' feedback. It will be recalled that we needed a way to provide additional guidance on the noun class system for elementary learners (and to a lesser extent also for intermediate learners). From the sample pages in Appendix C a page on the left-hand side of BCN, thus an even page, displays an inserted text which lists the most important affixes. In addition, a page on the right-hand side of BCN, thus an odd page, aims to display the most important abbreviations and symbols (also meeting a target users' need). Since these two "running footers" are repeated throughout the central lemma-list, we propose calling them the even-repetitive inserted text, and the odd-repetitive inserted text, respectively ERIT and ORIT.

The affixes tabulated in ERIT are straightforward. With the canonical form of the nominal prefixes for both participants and classes as a point of departure, ERIT then lists the subject concords, followed by the object concords and the pronominal prefixes, and ends with the morphemes used to form negative tenses. In other words, *the 100 items tabulated in ERIT synthesise the core of the Cilubà concordance system at a single glance.*

The abbreviations and symbols tabulated in ORIT were carefully chosen. Firstly, some guidance with respect to the frequency-annotation symbols ←, ↑,

and → is given. Secondly, in order to decide which abbreviations to include in ORIT and which not, the frequencies of all the abbreviations in BCN were counted. Finally, the most important symbols used throughout BCN are listed in ORIT's last column.

With these two repetitive inserted texts, it is truly hoped that *all* the target users will more easily find their way, not only through the noun class system, but also through the microstructure of the articles and the complex network of cross-references. Just as was the case with the first parallel dictionary, the release of this second parallel dictionary immediately provided feedback.

[7] Informal and formal feedback on the second parallel dictionary

At present, feedback on the second parallel dictionary, BCN, is informal. A second series of Informal Files is being assembled in which this feedback is documented. The picture looks very positive. Nonetheless, it is still too early to analyse them as was done for LCN's Informal Files. As for academics' informal feedback, a number of very encouraging e-mails from various scholars were received in which BCN has been labelled a "super book" or even a "chef d'oeuvre". These scholars point out "many interesting features", of which the two repetitive inserted texts and the experimentation with explicit frequency information seem to be favoured. One scholar even claimed that Cilubà "is currently better served, corpus-wise, than most of the big US dictionary publishers (who remain doggedly in the precorpus era)".[8]

[8] Main electronic corpus: Concordance lines and Dictionary articles

Regarding *macrostructural* planning, the second parallel project, BCN, has reached a point where virtually all target users are satisfied with the general dictionary structure and layout, and especially with the means utilised to lemmatise the different parts of speech. As such, simultaneous feedback in combination with the counts derived from the electronic corpus proved to be of enormous value in solving some of the long-standing lemmatisation problems for African languages.

On the *microstructural* level the electronic corpus also proves to be invaluable. Indeed, concordance lines derived from the corpus constitute the main tool for enhancing the dictionary articles, in respect of examples of usage, typical collocations, ordering of senses, etc.

[9] Continuous monitoring of the target users' desires, and Recruiting of additional staff

Van der Merwe (1999: 4) claims: "It is not easy to determine the needs of dictionary users. In the literature a variety of methods used by researchers are

mentioned, for example questionnaires, interviews, tests and observation. Not one of these methods has been successful up till now in doing a complete and efficient needs analysis." She is entirely in the right when she maintains that "not one of these methods" can have the intended result, and this is precisely why, within the framework suggested here, various methods are used and contrasted with one another. Moreover, the continuous monitoring of the target users' desires is built into the framework of simultaneous feedback. This ensures that the target users' desires transcend the target users' needs, expectations and reference skills.

The features provided by WordSmith Tools (WST) with which an electronic corpus can be analysed, are impressive. Yet, at the same time, the corpus-query options of this software are so numerous that the amount of data generated becomes too much to cope with by any small team of lexicographers. Viewed differently, while momentum has increased after each and every feedback batch, the staff has remained the same size. If future parallel projects are to extract the full potential out of concordance information to enable the creation of corpus-based dictionary articles, it is high time to enlarge the team with *several* mother-tongue speakers of Cilubà. Any dictionary project which has gone through two parallel projects will feel the need for recruiting of additional staff while the project grows. Hence the question is how one should read the last component of the theoretical framework in this light — [10] *more* parallel dictionaries?

[10] More parallel dictionaries used as experimental tools in order to refine the main dictionary

Taking on additional staff members requires, in one way or another, a substantial amount of funding. As long as this is not available, the entire project seems bound to remain a small dictionary project. However, this does not mean that *more* parallel projects cannot be launched in order to plan aspects of the main dictionary project.

Potential additional parallel projects could revolve around word-frequency studies. This led us to introduce a new approach to the phonetic description of a language's lexicon, the findings of which were published as *Cilubà Phonetics: Proposals for a "Corpus-based Phonetics from Below"-approach* (De Schryver 1999). Of particular interest to the main dictionary project, is that this study resulted in two short wordlists, a *Phonetic Frequency-Lexicon Cilubà-English* consisting of 350 lemma-signs (De Schryver 1999: 55-68), and a converted *Phonetic Frequency-Lexicon English-Cilubà* (De Schryver 1999: 69-87), both for short PFL. Provided that the target users know the conventions of the International Phonetic Alphabet (IPA), these two wordlists enable them to "retrieve", "learn" and "pronounce" the 350 most frequent *words* from the Lubà language.

(13) The concept of simultaneous feedback as implemented in the current dictionary project

Round-up of [1] - [10]

The theoretical plan of the concept of simultaneous feedback as presented in (1) might have seemed a bold statement. The feasibility of this concept, however, has been illustrated from a very practical perspective through a small-scale Cilubà-Dutch dictionary project. The way the concept has been implemented thus far in this undertaking is summarised in (13).

In conclusion

In this article it was argued that the feedback from target users, indispensable in the compilation of any modern dictionary, should be obtained while the compilation of the dictionary is still in progress. This is achieved by launching several small-scale parallel dictionaries from which informal as well as formal feedback from users and academics is immediately channelled back to the compilation process of a main dictionary. Such simultaneous feedback is not only invaluable in the compilation of the macrostructure but also provides useful information for improvement on the microstructural level. Furthermore, the ever-increasing role of the corpus as a lexicographic aid has been emphasised and it was indicated how corpus creation should be planned to fit into the concept of simultaneous feedback by introducing subcorpora.

The parallel dictionaries fulfil the urgent need that the target users can quickly be provided with lexica while awaiting an unabridged electronic corpus-based main dictionary. Moreover, channelling feedback from parallel projects into the main project enables the compilers to integrate all criticism into the compilation methodology itself. We are consequently convinced that the concept of simultaneous feedback successfully provides present-day lexicographers with an inexpensive head start and a tool for producing dictionaries compiled according to the latest trends and most modern approaches in lexicography.

It was emphasised throughout this article that the concept of simultaneous feedback can in principle be applied to any language. It is expected, however, that it will make a significant impact on dictionary compilation for the *African languages.* We therefore conclude with the words of Gouws (1998):

> One of the characteristic features of the Renaissance was the emergence of dictionaries and their spreading of the light of learning so desperately needed in the aftermath of the Dark Middle Ages. One of the characteristic features of the African Renaissance could be the emergence of dictionaries for all the African languages and productive and fruitful metalexicographic co-operation between all the stakeholders.

It is hoped that the concept of simultaneous feedback be part of that African Renaissance.

Notes

1. Since this article is being submitted for publication in a South African journal, sensitivity with regard to the term *Bantu* languages is exercised in our choice to use the term *African* languages instead. However, keep in mind that the latter includes more than just the Bantu language family.
2. This response compares favourably with similar projects by other experts. For example, of the 1 600 questionnaires distributed in the EURALEX- and AILA-sponsored *Research Project into Dictionary Use* (Atkins and Varantola 1998: 24), 1 140 responses (71,25%) were eventually received.
3. In order to turn the respondents' rankings into percentages, a formula was used whereby different weights were given to each number, with number "1" the highest weight and "…" (the absence of a number) the lowest weight. This same formula was used to quantify all answers in which the respondents were asked to give a ranking.
4. This is a reference to LCN's front matter where four types of "singulars with a cross-reference to their plural form" are discussed.
5. Single, lengthy additional remarks by respondents, like this one, have not been tabulated throughout the analysis of the Questionnaire.
6. The publication of LCN in December 1997 marked the inauguration of the Research Centre of African Languages and Literatures (Recall), based at the Department of African Languages and Cultures of the University of Ghent, Belgium. *Recall's Cilubà Corpus* (RCC) was the first electronic corpus built at Recall.
7. This suite of lexical analysis tools was presented by M. Rundell (1998: 16/3.19) during the Afrilex-Salex'98 Tutorial.
8. Since this quote originates from informal communication, the author will not be named.

Acknowledgement

Financial assistance from the National Research Foundation (NRF) towards the presentation of "The Concept of 'Simultaneous Feedback'" as a paper at Euralex 2000 is hereby acknowledged.

Cited references

URL (Universal Resource Locator)

WordSmith http://www.liv.ac.uk/~ms2928/wordsmith/screenshots

References

Alberts, Mariëtta. 1999. The Importance of a Business Plan when Planning a Lexicographic Project. Paper presented at the Fourth International Conference of the African Association for Lexicography, held at the University of Pretoria, Pretoria, 5-6 July 1999. *Lexikos* 9: 188-197.

Atkins, B.T. Sue (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators.* Tübingen: Max Niemeyer.

Atkins, B.T. Sue and Krista Varantola. 1998. Language Learners Using Dictionaries: The Final Report on the EURALEX / AILA Research Project on Dictionary Use. Atkins, B.T. Sue (Ed.). 1998: 21-81.

Barnhart, Clarence L. 1967[2] [1962]. Problems in Editing Commercial Monolingual Dictionaries. Householder, Fred W. and Sol Saporta (Eds.). 1967[2] [1962]. *Problems in Lexicography*: 161-81. Bloomington: Indiana University.

De Clercq, Auguste and Emile Willems. 1960[3]. *Dictionnaire Tshilubà-Français*. Leopoldville: Imprimerie de la Société Missionnaire de St. Paul.

De Schryver, Gilles-Maurice and Ngo S. Kabuta. 1997. *Lexicon Cilubà-Nederlands. Een circa 2500-lemma's-tellend strikt alfabetisch geordend vertalend aanleerderslexicon met decodeer-functie ten behoeve van studenten Afrikaanse Talen en Culturen aan de Universiteit Gent*. Ghent: Recall.

De Schryver, Gilles-Maurice and Ngo S. Kabuta. 1998. *Beknopt woordenboek Cilubà-Nederlands & Kalombodi-mfùndilu kàà Cilubà (Spellingsgids Cilubà). Een op gebruiksfrequentie gebaseerd vertalend aanleerderslexicon met decodeerfunctie bestaande uit circa 3.000 strikt alfabetisch geordende lemma's & Mfùndilu wa myakù ìdì ìtàmba kumwèneka (De orthografie van de meest gangbare woorden)*. Ghent: Recall.

De Schryver, Gilles-Maurice. 1998. Cilubà Lexicography: A First Approach. Unpublished paper presented at a Lexicography Seminar, Linguistic Department of the Musée royal de l'Afrique centrale, Tervuren, 20 February 1998.

De Schryver, Gilles-Maurice. 1999. *Cilubà Phonetics: Proposals for a "Corpus-based Phonetics from Below"-approach.* Ghent: Recall.

De Schryver, Gilles-Maurice. 1999a. *Bantu Lexicography and the Concept of* Simultaneous Feedback*. Some Preliminary Observations on the Introduction of a New Methodology for the Compilation of Dictionaries with Special Reference to a Bilingual Learner's Dictionary Cilubà-Dutch.* Unpublished M.A. dissertation. Ghent: University of Ghent.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

Galtung, Johan. 1967. *Theory and Methods of Social Research*. Oslo: University of Oslo.

Gouws, Rufus H. 1998. Chairperson's Annual Report: 1997-1998. Unpublished report read at the Third International Conference of the African Association for Lexicography, Potchefstroom University for Christian Higher Education, Potchefstroom, 29-30 June 1998.

Gouws, Rufus H. and D.J. Prinsloo. 1998. Cross-referencing as a Lexicographic Device. *Lexikos* 8: 17-36.

Hartmann, Reinhard R.K. and Gregory James. 1998. *Dictionary of Lexicography*. London: Routledge.

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.

Kilgarriff, Adam. 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.

Lombard, F.J. 1994. Lexicographer, Linguist and Dictionary User: An Uneasy Triangle? *Lexikos* 4: 204-214.

Osselton, Noel E. 1983. On the History of Dictionaries. The History of English-language Dictionaries. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 13-21. London: Academic Press.

Prinsloo, D.J. and Gilles-Maurice de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.

Renouf, A. 1987. Corpus Development. Sinclair, John M. (Ed.). 1987. *Looking Up: An Account of the Cobuild Project in Lexical Computing*: 1-40. London: Collins ELT.

Rundell, Michael. 1998. Corpus Exploration: Extracting Lexicographic Data from a Corpus. Atkins, B.T. Sue, Michael Rundell and Rufus H. Gouws. 1998. *Afrilex-Salex'98. A Training Course in the Compilation of Bilingual Dictionaries*: 16/3.1-19. Unpublished course material of a tutorial held at the University of Pretoria, Pretoria, 7-18 September 1998.

Sinclair, John M. (Ed.). 1987. *Collins COBUILD English Language Dictionary*. London: HarperCollins Publishers.

Summers, Della. 1993. Longman/Lancaster English Language Corpus — Criteria and Design. *International Journal of Lexicography* 6(3): 181-208.

Summers, Della (Director). 1995[3]. *Longman Dictionary of Contemporary English*. Harlow: Longman Dictionaries.

Tono, Yukio. 1992. The Effect of Menus on EFL Learners' Look-up Processes. *Lexikos* 2: 230-253.

Van der Merwe, Michele. 1999. The User Perspective in South African Lexicography. Unpublished paper presented at the Fourth International Conference of the African Association for Lexicography, held at the University of Pretoria, Pretoria, 5-6 July 1999.

Wiegand, Herbert E. 1998. *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Berlin: Walter de Gruyter.

Abbreviations used in reference to dictionaries and lexica

BCN	*Beknopt woordenboek Cilubà-Nederlands.* De Schryver and Kabuta. 1998.
COBUILD1	*Collins COBUILD English Language Dictionary.* Sinclair (Ed.). 1987.
LCN	*Lexicon Cilubà-Nederlands.* De Schryver and Kabuta. 1997.
LDOCE3	*Longman Dictionary of Contemporary English.* Summers (Director). 1995[3].
PFL	*Phonetic Frequency-Lexicon Cilubà-English-Cilubà.* De Schryver. 1999: 55-87.
WCN	*Woordenboek Cilubà-Nederlands*. In progress.

Appendix A

Appendix B

Appendix C