From Corpus to Dictionary: A Hybrid Prescriptive, Descriptive and Proscriptive Undertaking

Minah Nabirye, *Department of African Languages and Cultures, Ghent University, Ghent, Belgium (minah.nabirye@UGent.be)*
and
Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Ghent, Belgium; and Xhosa Department, University of the Western Cape, Bellville, South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: Despite some heroic efforts over the past few years, Lusoga remains mostly underdeveloped. It is under continuous pressure from more prestigious languages, such as the neighbouring Luganda and especially the only official language in Uganda, English. Lusoga is undergoing rapid language shifts, with new concepts entering the language daily. Ironically, this process is taking place before Lusoga has even been properly reduced to writing. There is no single official orthography that is truly being enforced; people who do write, write as they think fit. Language data is needed for the production of reliable reference works. In the absence of a substantial body of published material in Lusoga, the researcher can resort to recording and transcribing the living language. This opens Pandora's box, in that spoken language (which is meant to be heard, and is typically less formal) is far more complex than written language (which is meant to be read, and is typically more formalised). Spoken and written variants are, by definition, different. And yet one wants to move the language forward, in a way, before the time is ripe. But then, with over two million speakers, how much longer can one wait? This article reports on the building of a new Lusoga corpus, nearly half of which consists of transcribed oral data. The writing problems encountered during the transcription effort are given detailed attention. Dealing with those writing problems in lexicography requires a multipronged approach. While most could be solved by laying down a norm, and thus through prescriptive lexicography, others need a more cautionary approach, and thus descriptive lexicography. Others still can only sensibly be solved when the lexicographer proposes certain options in defiance of existing norms and assumptions, at which point proscriptive lexicography needs to be called in.

Keywords: LUSOGA, UGANDA, ORTHOGRAPHY, SPELLING, CORPUS, ORAL, SPOKEN, TRANSCRIPTION, FULL WORDS, COMPOUNDS, MULTIPLE FORMS, LOANWORDS, BORROWINGS, FORMALITY LEVELS, CONCORDS, PRESCRIPTIVE LEXICOGRAPHY, DESCRIPTIVE LEXICOGRAPHY, PROSCRIPTIVE LEXICOGRAPHY

Obufunze: Okulondoola engeli Eitu ly'Olusoga bwe Linaatuusibwa mu Iwanika: Omutindo ogulaga Olulimi bwe luli, bwe luteekwa okuba oba bwe lube lutwalibwe. Empandiika y'Olusoga ekaali inhuma inho waile nga waliwooku obubonelo

obulaga enkola y'obuzila mu kugizimba. Olusoga lukaawagamiile wansi w'ennimi nga Oluganda n'olulimi olw'eiwanga, Olungeleza. Ensonga endala ku dhikaalemeisa Olusoga okwetengelela n'okuba nti lulina ebigambo ebilwingila buli olukeile. Eky'embi, enkyukakyuka ebigambo bino ebiyaaka by'eleetawo eli kwidha mu kiseela nga Olusoga lwene lukaali kufunilwaku mpandiika ntongole. Waile walifu dh'Olusoga buti dhiliwo kamaala, wazila ndala ku dho eli kugobelelwa mu kuwandiika kubanga abantu bakaawandiika nga bwe babona. Ebiwandiiko ebili ku Lusoga oba mu lulimi Olusoga byetaagibwa okulaga Olusoga bwe lulina okuba ela n'ebilina okwebuuzibwaku. Eibula ly'ebiwandiiko oti ni bino litegeeza nti omunoonheleza alina okwefunila entambi dh'Olusoga olwogelwa ela yeewanulila Olusoga oluli mu ntambi edho okusobola okutegeela engeli olulimi olwo bwe luli mu kiseela ekyo. Kino kileeseewo obuzibu obundi nti Olusoga olwogelwa tilutongoze ate lulimu emigote kamaala egyandibaile gilondoolwa okusinziila ku mutindo omusengeke singa lubailem mu buwandiike. Ekika ky'Olusoga oluli mu mbeela eyogelwa kya ndhawulo ku kili mu mbeela y'obuwandiike. Okuwanula Olusoga okuva mu ntambi kw'aba nti kugiililiile kwagayaga na kunoga kibala kikaali kwenga bukalamu. Aye engeli ye kili nti Olusoga lulina aboogezi abaswika mu bukaile obubili, abanoonheleza ku Lusoga baalisaine kutandiika li? Olupapula luno lulambulula enzimba y'eitu ly'Olusoga nga ekitundutundu ky'eitu lino kiviile mu Lusoga olwawanulwa okuva mu mbeela eyogelwa. Obuzibu obwayagaanibwa mu mpandiika y'ebigambo ebyawanulwa n'obusimbiibwaku eisila. Okulaga engeli y'okuzigula obuzibu bw'empandiika mu isomo ly'amawanika kwetaagisa enkola eteekubila inho ku nsonga ndala aye enoonheleza engeli esinga kugasa omutendela gw'obuzibu obulondoolebwa. Waile nga obuzibu obundi busobola okugondhoolwa okugiila ku mutendela oguteebwawo, gwalaga olulimi nga bwe luteekwa okuba, obuzibu obundi bwetaagamu okwegendeleza nga wano mubaamu enkola y'okulaga nga olulimi bwe luli. Ate bwo obuzibu obundi bwandyetaagisa abawandiisi b'amawanika okuwaayo obudhulizi obusinziilwaku endowooza dhaibwe edhitagobelela mitendela giliwo nga balaga olulimi bwe lube lutwalibwe waile nga kino kyandiba nga kikontana n'amateeka g'olulimi agaliwo.

Ebigambo ebikulu: OLUSOGA, UGANDA, WALIFU, EMPANDIIKA, EITU LY'OLUSOGA, ENDHOGELA, OKUWANULA OLUSOGA, EBIGAMBO EBILAMBA, EBIGAMBO EBIGAITILILE, ENNAMULA KU BIGAMBO EBYAWUKANA, EBIGAMBO EBYEYAZIKE, EMITENDELA GY'OBUTONGOLE, ENNHUNGA YA NALIINA, OLULIMI BWE LUTEEKWA OKUBA, OLULIMI BWE LULI, OLULIMI BWE LUBE LUTWALIBWE

1.   The Ugandan mother-tongue education policy

In September 2005, the Ugandan Parliament passed the teaching of nine regional indigenous languages in primary schools. Implementation officially started in February 2007. The languages concerned are Runyoro-Rutooro, Runyankore-Rukiga, Luganda, Lusoga, Rukonjo, Lugbara, Acholi, Ateso, and Karamojong (see NCDC 2006 and 2006a).

Elementary considerations required to prepare the introduction of all these languages into the mother-tongue education system are, however, pending. This is mostly detrimental for languages like Lusoga which are proposed as a medium of instruction for the very first time. The most crucial requirements for the successful implementation of this policy are the training of

mother-tongue educators and the production of reference works for the target audience.

Nabirye and De Schryver (2010) identified two target audiences for Lusoga: the primary-school teachers and their pupils. For these two audiences, Lusoga is supposed to be used both as the medium of instruction and as a subject in primary 1-3, but only as a subject in primary 4-7. Primary teachers do not learn Lusoga throughout their education and they are not Basoga trained to teach Lusoga as a subject either. As teachers they are nonetheless expected to conduct lessons in and on Lusoga, and to grade the pupil's Lusoga exercises.

2.   The available Lusoga data

To date, no more than a handful of textbooks, spelling guides, and dictionaries exist for Lusoga. The available literature is limited to a few dozen booklets with stories only, as well as some religious publications. Unfortunately, in all this material, a variety of spellings is found. Moreover, given very few people currently write in Lusoga, there are numerous word forms in speech which have not been codified in any of the existing reference works. Word forms that lack a full conceptualization to merit a specific writing format are considered to be especially problematic. This state of affairs thus poses an additional problem when the aim is to uplift the status and use of Lusoga. Merely building a corpus of the language, where that corpus is internally consistent, proves to be a challenge in itself. In today's linguistic world which is driven by the use of electronic text corpora — with which a range of pedagogically sound reference works could and should be compiled — this poses problems for the corpus builder.

Should the corpus builder unify the various orthographies? Or is it better to reflect and thus keep the various, original spellings? If a single orthography is to be chosen, which one, or rather, whose orthography? Conversely, if a variety of orthographies is (or even if *all* are) kept, which one(s) should be used when transcribing new (i.e. oral) material? Answering these questions is not as trivial as it may seem. For example, while linguists and lexicographers may want to standardize the spelling, they run the risk to 'overcorrect' their original primary sources, making changes which end up masking the true processes at hand. But then, can one really handle a corpus which contains a variety of spellings for what are basically the same words?

In De Schryver and Nabirye (2010) it was shown how such a heterogeneous corpus can indeed be queried, and how, with it, one can undeniably acquire greater insights into the structure of the language. Teachers will — in contrast to linguists — need more definite guidance, however, and even though one of the orthographies (viz. Namyalo et al. 2008) has now been approved as the standard for writing Lusoga, being aware of the range of outstanding writing problems is a first step towards solving them (assuming they are all solvable, which, as we will see, is not always the case).

Over the past few years, an organic Lusoga corpus has been built,[1] the composition of which is currently (July 2011) as shown in Table 1.

Table 1: Composition of the Lusoga Corpus (July 2011)

Year	Genre	Breakdown of the data	Tokens	%
1998	Religious texts	The New Testament, …	199 853	18.20
1998-1999	Newspapers (now defunct)	*Kodh'eyo*, *Ndiwulira*	187 393	17.07
1998-2003	Short stories	*Kintu*, *Ababita Ababiri*, …, *Ebindi kw'idembe ery'Obw'omuntu mu nsi Yoonayoona*	150 560	13.71
2008-2009	Reports	Busoga clan leadership, Private sector, Academia, …	24 166	2.20
2008-2011	Transcribed recordings	Songs, Traditional ceremonies, Speeches, Sermons, Interviews, Radio broadcasts, …	413 827	37.69
2009-2010	Translations out of English	PEAP (*Poverty Eradication Action Plan*), ICEE (*International Centre for Eye Education*), FIDA/PLAN (inheritance laws), Home Structure, Public Health (questionnaire for baseline survey on institutional deliveries), JLOS (village imprisonment laws), Private Sector Uganda, Commercial advertisements, Message addressed to the Busoga clan governance	19 814	1.81
2010-2011	Electronic texts	E-mails (Personal and Mailing lists), Facebook postings, Material in electronic form, Various abstracts, …	102 365	9.32
			1 097 978	**100.00**

As may be seen from Table 1, the total size of the Lusoga corpus now stands at 1.1 million running words or tokens. With well over 400 thousand tokens, the oral component of this Lusoga corpus is as high as 38%. Although all the transcriptions in the oral component have been made by the same person[2] — dili-

gently transcribing and adding material on a daily basis over a period of four years now — it has proven near impossible to be 'consistent'. This is mainly so because the transcription effort brought new and unusual constructions to the fore, which have not been documented in any of the existing orthography booklets (Byandala 1963, Kajolya 1990, LULANDA and CRC 2004, Namyalo et al. 2008), nor in any of the existing vocabulary lists or dictionaries (Korse 1999a, Gonza 2007, Nabirye 2009). Future writing guides and dictionaries (which ought to include extra-matter sections covering the orthography in depth) will thus do well to take cognizance of the problematic cases noted.

Reformulated: the transcriptions in the Lusoga corpus tapped into the most vibrant part of the language, namely the oral part, and enabled a realization of the unusual constructions that now need attention. The problem cases noted during the transcription exercise thus constitute the spelling issues to be addressed. Results from this study will patch gaps in the existing reference works (both writing guides and dictionaries), and will enable Lusoga primary teachers to improve their knowledge of writing Lusoga; a factor that is required to enable the judgement of the right and wrong usage of Lusoga in the Lusoga lesson exercises. As will be shown below, however, there are also cases where there is no right or wrong, at which point the lexicographer can resort to *pro*scriptive lexicography (i.e. '*pro*posing' that a certain approach to writing Lusoga be used, rather than insisting on either 'prescribing' a norm, or merely 'describing' everything seen in the language; cf. Bergenholtz 2003, and Bergenholtz and Gouws 2010).

3.   Writing problems noted

In this section each group of examples chosen for discussion represents a different type of problem noted during the transcription of the oral data. All examples are authentic, in that they have been taken from the transcriptions. They are also representative, in that many more similar cases have been encountered. They are called 'problematic' because none of the existing reference works includes guidance on how to reduce the spoken to the written form in these cases.

3.1   Full words

The existing orthographic specifications are for example challenged by speech forms that have not yet fully been conceptualized and passed on into the writing format. Moreover, results from the dictionary testing carried out in Nabirye (2008) showed that all respondents failed to demarcate word boundaries. The conceptualization and formalization of undocumented oral constructions are thus a challenge. Guidance is for instance needed in order to be able to distinguish between forms like 'today' and 'the day of today', as in (1a-b). In our tran-

scriptions, we differentiated between the two by writing the first conjunctively, the second disjunctively. The *leelo* 'today' in example (1b) is a variable formative and can be replaced with other words in the construction, such as 'Sunday' in (1c), whose lexicalization is independent from (1a).

(1)	(a)	today	*(o)lwaleelo*	(full construction written conjunctively)
	(b)	the day of today	*(o)lwa leelo*	(separate constructions written disjunctively)
	(c)	the day of Sunday	*(o)lwa Saabbiiti*	(as previous, with capitalization)

In the lexicalization of the Lusoga names of the months, as in (2a), we find that *Ogwokusatu* 'March' is a compounded form which does not require any further information to fulfil its function. It is a full, self-standing construction and a proper name which is capitalized. *Ogw'okusatu* 'a third' as in (2b), however, requires qualification to specify the subject of 'a third'. In other words, *omwezi ogw'okusatu* 'a third month' is not the same as 'March'. Or still, the subject of third as in for example *omulundi ogw'okusatu* 'a third time' is a dependant construction and requires contextual analysis to arrive at its full meaning.

| (2) | (a) | March | *(O)gwokusatu* | (full construction, capitalized, no abbreviation or complement) |
| | (b) | a third | *ogw'okusatu* | (not capitalized, needs a noun complement) |

A distinction between structures such as (2a) and (2b) may be achieved through the use of capitalization and abbreviation. This is further evident in (3a) where *Gwakubili* 'February' is specified by the object (the underlined part). Without a specifier, as in (3b), we lack assurance that the fifth day also refers to 'Friday'. Here, it is an indefinite fifth day. In (3c), however, though the time sequence lacks an object, the context in which the term is used is sufficient to come to the conclusion that it is not referring to the beginning of just any sequence but specifically to the book of Genesis. Cases (3a) and (3c) are proper names which are by convention capitalized; cases such as (3b) are not.

(3)	(a)	*lit*. month **second time** <u>2001</u>	*mwezi Gwakubili <u>nkumi ibili ikumi na mulala</u>*
	(b)	*lit*. day **fifth**	*olunaku lwakutaanu*
	(c)	*lit*. before we go in **first**	*nga tukaali gya mu Lubelyebelye*

When forms such as (3c) are preceded by a possessive concord, as in (4a), the resulting forms may denote an independent concept. A similar example is shown in (4b). Even though such forms have not been entered or described in

any of the existing reference works, it is clear that they should and that they should be written as one full word, seeing that they refer to a single concept.

(4) (a)   *lit*. (something in cl. 7) of **first**   >   primary 1   *ekyolubelyebelye*
    (b)   *lit*. (something in cl. 7) of **daylight**   >   lunch   *ekyomusana*

The examples in (5) are cases of homography and reduplication. In our transcriptions (5b) is only a full construction when it is written conjunctively, while in (5c) the form can be repeated as many times as the emphasis allows, say as *boona boona boona boona*, but the meaning is not the same as in (5a). Distinguishing between the different lexical forms in (5) is dependent on establishing the function each form is intended to perform.

(5) (a)   also   *boona*   (single lexical construction)
    (b)   all   *boonaboona*   (single lexical construction,
                                                written conjunctively)
    (c)   all (+ emphasis) *boona boona …*   (compounded construction,
                                                written disjunctively)

Some constructions are difficult to specify because they sound twisted and are not only a puzzle in speech but also in writing. The writing problems are founded on similarities which make it difficult to establish boundaries in the entire construction. An example is given in (6c), which is best approached in successive steps, as in (6a) and (6b), leading to (6c).

(6) (a)   grandparents   *badhaadha*
    (b)   grandparents of the grandparent   *badhaadha ba dhaadha*
    (c)   grandparents of the grandparent's   *badhaadha ba dhaadha ba dhaadha*
          grandparent

A thorough understanding of the Lusoga grammar is required to crack the puzzle in such constructions. The current absence of a proper grammatical description of Lusoga calls for extra caution during the transcription exercise to appreciate the different parts in the construction.

3.2   Compounds

Compounding is not problematic in speech where word boundary considerations are unconscious, but it may be problematic in writing where the process needs to be applied consciously. The specification of possible compounding procedures is thus required.

3.2.1 Compounding with prefixes

For example, the *Eiwanika* (i.e. the monolingual Lusoga dictionary, Nabirye 2009) defines *muna-* as a prefix used to link a subject to the object intended. A person of the journalist trade will for example be called *munamawulile*. At the time of compiling the *Eiwanika*, the usage of this prefix was not fully ascertained and seemed to crop up once or twice only, so that only the most dominant usages were given as examples. There are however new ways the same prefix is being exploited in the corpus, still serving the purpose defined but linking to a wider spectrum of objects not earlier conceived possible, as in (7a).

(7) (a)   someone from Bilumba   *Munabilumba*
    (b)   people from Mayuge   *Banamayuge*
    (c)   someone from Masese   *Munamasese*

Such compounds clearly still need to be fully conceptualized as single concepts. In the e-mail and Facebook sections of the corpus, for example, (8i) is also found as (8ii-iii).[3]

(8)   Ugandan   (i)   *(O)munauganda*
       also found as: (ii)   **(o)muna Uganda*
       also found as: (iii) **(o)munaUganda*

An analogous prefix is also being exploited in more elaborate ways than those primarily intended. The prefix *(o)mwise-* (sg.)/*(a)baise-* (pl.) usually denotes the belonging to a restricted context of a clan as in (9). The context restriction is however being relaxed and made to cover contexts other than the clan, as in (10).

(9) (a)   of the Menha clan   *Omwisemenha*
    (b)   of the Igaga clan   *Abaiseigaga*
    (c)   of the Mususwa clan   *Abaisemususwa*

(10)   supporter of the NRM   *mwiseNRM*

This usage innovation puts both the prefix *(o)muna-* (sg.)/*(a)bana-* (pl.) and the prefix *(o)mwise-* (sg.)/*(a)baise-* (pl.) on the same footing with regard to their function. It is therefore essential to realize that all the cases noted in (7), (8i), (9) and (10) are correct forms, and where the derived compounds are proper names they should be capitalized, except for (10) where the subject referred to is not definite but could be any supporter of the political party NRM. Given all the examples represent a single concept in the construction, they should all be written conjunctively.

3.2.2 Compounding with independent word forms

Compounding in the formation of proper names provides a more familiar indication of the Lusoga compounding system. In (11a-c) the formatives on the left have independent denotative meanings, while they acquire connotative meanings and mutate into proper names conceptualized as single lexical units on the right. Being proper names they are capitalized, unlike in (11d) where the reference is indefinite.

(11) (a) *mwene + iloboozi*   >   *Mweneiloboozi*
           lit. owner + voice   >   name of a specific person
     (b) *gaali + maka*   >   *Gaalimaka*
           lit. it was + home   >   name of a specific person
     (c) *mmele + ya + nkya*   >   *Mmeleyankya*
           lit. food + of + morning   >   name of a specific person
     (d) *ki + mpwitu*   >   *kimpwitu*
           lit. someone + stubbornness >   stubborn person

3.2.3 Compounding and noun gender

In the examples (7) through (11), both singular and plural forms were extracted from the corpus. When person and number come into play, rules for capitalization have to be laid down. If one for example refers to several people called *Gaalimaka*, it is suggested that the plural class prefix be written conjunctively with the compound, and the first letter only be capitalized, as in (12a). Similarly, a form like (12b) takes a word-initial capital letter, while (12c) doesn't.

(12) (a) *Gaalimaka*   + number + capitalization   *Bagaalimaka*
     (b) *Omunauganda*   + number + capitalization   *Abanauganda*
     (c) *kimpwitu*   + number + null capitalization   *bakimpwitu*

3.3   Multiple forms

The mapping between spoken and written forms is unfortunately not always unambiguous in Lusoga. There are two types, each the reverse of the other.

3.3.1 Different words for the same sound

Nabirye (2008) proposed a number of changes to the orthography. Some of these were also carried through to both Namyalo et al. (2008) and to the *Eiwanika*. In most cases only the 'safe' changes were implemented because the writing format for Lusoga was still new and only the cases that were considered to be really important were attended to. As a case in point, /j/ was intro-

duced for use with foreign words with that sound, as in *Janwali* 'January', *jihaadi* 'jihad', the place name *Jinja*, etc. With /j/ added to the alphabet, a frequent rendering like for example *Jinja* could now be accepted as part of the lexicon (in addition to the indigenous *Idhindha*).

While /j/ rather than /dh/ was applied throughout the lexicon, /j/ rather than /gy/ was regrettably not applied consistently, which is especially unfortunate as /j/ and /gy/ are homophones. Existing Lusoga texts as well as early versus more recent transcriptions therefore contain both spellings seen in (13a), as well as both seen in (13b).

(13) (a) (woman who has made   (i) *Hagyati*   /gy/
a pilgrimage to Mecca) (ii) *Hajati*   /j/
(b) (pilgrimage to Mecca)   (i) *higya*   /gy/
(ii) *hija*   /j/

In the *Eiwanika*, *Hagyati* was unfortunately entered rather than *Hajati*, yet at the same time *hija* was (correctly) entered and not *higya*. The existing reference works thus give confusing signals, with multiple spellings for the same words as a result, and this is reflected in the corpus, including the transcribed parts.

Another case of multiple spellings for the same words is the result of a type-writer limitation to representing the velar nasal on the keyboard. The solution was to use /ng/ rather than /ŋ/, a 'solution' which became a tradition with time, even when the limitation had seized to exist. Nabirye (2008) argued for the reintroduction of /ŋ/ in the orthography, and it was introduced in Namyalo et al. (2008). It was also used in the *Eiwanika*. However, in the written sections of the corpus as well as the early transcriptions, the velar nasal is represented with /ng/, rather than /ŋ/, resulting in multiple forms such as those seen in (14).

(14) (a) (a place name)   (i) *Ngando*   / ng /
(ii) *Nŋando*   / ŋ /
(b) relatives   (i) *enganda*   / ng /
(ii) *enŋanda*   / ŋ /

3.3.2 Different sounds for the same word

There are sounds which are dying out and others are taking their place, resulting in spelling variations as shown in (15a) and (15b).

(15) (a) Jinja   (i) *Idinda*
(ii) *Idhindha*   *[dying out]*
(b) nation/country/tribe (i) *eiwanga*
(ii) *eighanga*   *[dying out]*

Although some sounds may be on the wane, as indicated in (15), a single person may still use the various alternatives, even in the same sentence. It is not so that one pronunciation is 'better' Lusoga than the other(s), or that one is 'better' from a linguistic point of view.

3.4   Loanwords

In the absence of any operational language regulatory body for Lusoga, it often seems as if the language spirals out of control, hampering all attempts to streamline and to formalize the writing. When transcribing oral data, foreign adoptions are especially ambiguous to deal with. The linguist's intuition tells her or him to rid the language of these 'intrusions', but then, they are found — often in high numbers — so have clearly become part of the language.

3.4.1   Borrowings from neighbouring languages

The current Lusoga orthography strictly specifies that the combination /ny/ is not part of the language; rather, the combination /nh/ is used in Lusoga where neighbouring languages like Luganda have /ny/. However, Luganda forms such as the one shown in (16a) are used so often in Lusoga — more often even than their Lusoga counterparts — that when transcribing oral data one cannot simply ignore or 'correct' these forms. Similarly, in (16b), the prefix is dropped in the spoken language, rendering the Luganda version rather than Lusoga.

(16)		Lusoga	Luganda
(a)	madam	*inhabo*	*nnyabo*
(b)	sir	*isebo*	*ssebo*

More confusing is the use of foreign words to refer to place names, even when referring to places in Busoga. In (17i) the common Lusoga spellings of two place names are shown, as well as the reasons why these spellings are problematic from a Lusoga point of view. Using a Lusoga(ized) spelling, (17ii) is obtained. A brief etymology is also provided in (17ii).

(17)	(i)	<u>*Wanyange*</u>	/ny/ does not exist in Lusoga
		Makerere	/r/ is contested in the Lusoga alphabet
	(ii)	*Wanhange*	named after a bird known as *ennhange*
		<u>*Makeelele*</u>	name derived from a Luganda idiom '*amagezi makeelele*'

When transcribing, it is not simply a matter of choosing either series, say only the commonly-found renderings as in (17i), or the Lusoga(ized) versions as in (17ii). When carefully listening to how these words are actually pronounced, the underlined forms in (17) are used.

3.4.2 Borrowings from the religious sphere

With religious indoctrination not only came new concepts (and a new faith), but also new, previously unseen, sound combinations. Tussling denominations also gave their own twist to word-final vowels. Both these aspects are illustrated in (18a-b).

(18)	(a)	Christ	(i) Catholics	*Kristu*	but not	**Kulisitu*
			(ii) Protestants	*Kristo*	but not	**Kulisito*
	(b)	sacrament	(i) Catholics	*sakramentu*	also as	*saakalamentu*
			(ii) Protestants	*sakramento*	also as	*saakalamento*
	(c)	democracy		*demokrasiya*	also as	*dimokulasiya*

Bantu words normally show a CVCV structure, but forms such as *Kristo/Kristu* are now so well entrenched that they are not only written like that, but also pronounced as they are written (and thus not as **Kulisito/*Kulisitu*, which would have been in accordance with a CVCV structure). Religious publications 'faithfully' stick to using the combination /kr/ throughout, for all loanwords with this foreign sound combination, whether religious (e.g. *sakramento*, as in (18b)), or not (e.g. *demokrasiya*, as in (18c)). When these same words are used in a non-religious context, and/or by non-believers, speakers and writers often resort to variants which do adhere to the CVCV structure however (thus e.g. *saakalamento* and *dimokulasiya* here). An accurate transcription will reflect these differences.

3.4.3 Lexicalization of borrowed abbreviations and acronyms

The wish by speakers of Lusoga to adhere to a CVCV structure is often so strong (i.e. intuitive), that when abbreviations are borrowed wholesale, as in 'FM', some pronounce it as /fa ma/, rather than /ef em/. At that point the transcriber has at least three options, as shown in (19).

(19)	Frequency Mode	(i)	*FM*	/ef em/
		(ii)	*Ef Em*	/ef em/
		(iii)	*Fa Ma*	/fa ma/

Abbreviations and acronyms are not only cited as they are formally known but also made to carry characteristics of person and number as in (20).

(20)	(a)	Resident District Commissioners	*RDC* + number	*baRDC*
	(b)	District Security Officers	*DISO* + number	*baDISO*
	(c)	District Police Commanders	*DPC* + number	*baDPC*
	(d)	District Health Officers	*DHO* + number	*baDHO*

The new abbreviated forms represent a single concept of indefinite office occupants, hence requiring no word-initial capitalization, but they should be written conjunctively. Although the characteristics marked on plurals (i.e. the use of prefixes) are also plausible for singular, the singular form doesn't require additional lexicalization processes to infer the meanings.

3.4.4   Borrowings from English

The hardest borrowings a transcriber has to contend with are not those from neighbouring languages (which at least have a Bantu structure), nor those from the religious sphere (as they are mostly limited to terms from a restricted domain), nor the abbreviations and acronyms (where the spelling issues are minor), but the wholesale borrowings of concepts and words from English that are literally 'dropped' into the language, and on which the full Bantu morphological and morphophonological apparatus is unleashed (including possible phonologization, marking of person and number, prefixation, the addition of verbal extensions, enclitization, etc.).

Example (21) shows a trivial case where some writers will adapt the foreign word to the structure of Lusoga (21i), while others will keep the English spelling intact yet still pluralize it according to the Lusoga morphology rather than to use the English plural suffix *-s* (21ii).

(21)   lawyers   i)   *balooya*   number + phonologized
        ii)   *balawyer*   number + English root

The fully phonologized version (*balooya*) and the version in which the English root is merely prefixed with the cl. 2 plural class prefix (*balawyer*), have the exact same pronunciation. When transcribing Lusoga recordings it is advisable to stick to one spelling only, so either (21i) or (21ii), with our preference going to (21i).

Depending on the noun class a particular borrowed word ends up in, both singular and plural forms may of course also look the same; compare (22a-c) with (22d-e) in this respect.

(22)   (a)   hello(s)   *alo* (sg.)/*baalo* (pl.)   gender 1a/2a
        (b)   district(s)   *district* (sg.)/*madistrict* (pl.)   gender 9/6
        (c)   government(s)   *gavumenti* (sg.)/*dhigavumenti* (pl.)   gender 9/10
        (d)   bye(s)   *bbaayi* (both sg. and pl.)   gender 9/10
        (e)   septrine(s)   *septrine* (both sg. and pl.)   gender 9/10

Furthermore, any type of prefixes (thus also other than noun class prefixes) may precede the borrowed material, as in (23).

(23)   pertaining to the post of a Minister   *obwa- + Minister*   *obwaminister*

Usages of such constructions have a considerable occurrence in the oral part of the corpus. A specification of how to address these forms in writing is thus warranted. The argument of full word conceptualization as a single lexical form lends itself easily here, while the non-capitalization seen is in harmony with similar cases in the lexicon.

Once a word is borrowed into Lusoga, it may also be moved around the noun classes, and for example be made to take on degree assessments, such as the diminutives seen in (24).

(24) (a) a bit of airtime   ***ka****airtime*   gender **12**/14
     (b) short programs   ***bu****programs*   gender 12/**14**

In the process, speakers of Lusoga may combine two grammatical systems, as in (24b), where one notices a double plural marking: one from Lusoga, i.e. the cl. 14 plural noun prefix *bu-*, and one from English, i.e. the plural suffix *-s*. A truthful transcription will reflect such idiosyncrasies.

Words from word classes other than nouns in English may also end up being nominalised in Lusoga, as seen in (25).

(25) (a) small plus   ***ka****plus*   gender **12**/14
     (b) *extensives   *ebyo **bi**extensive ebyo*   gender 7/**8**

(25a) shows an example of the nominalization of a conjunction/adjective, while (25b) shows an example of the nominalization of an adjective/adverb. The task of the transcriber is again to record all such instances, leaving their analysis for a later phase.

Lastly, English verbs too may be borrowed wholesale. They are typically adopted in their infinitive form, and then 'Lusogaized' by means of the standard verbal morphology of Lusoga. As such, Lusogaized verbs may take on verbal extensions, as the applicative in (26a), or the perfective forms reflected on the verb ending vowels (including sound changes) in (26b) and (26c), they may include reflexive markers as in (26d), they may mark aspect as in (26e) and (26f), they may accommodate a subject concord and an enclitic as in (26g), etc. Cases of elision, which are quite dominant in the Lusoga corpus, also occur with the Lusogaized word forms, as seen in (26h).

(26) (a) and planning for her/him   *n'**okumu**<u>planing</u>**ila***
     (b) they pronounced   *ba<u>pronouncin</u>**ze***
     (c) I did not monitor   *t**ya**<u>monitorin</u>**ze***
     (d) it comes   *nee<u>coming</u>a*
     (e) they still lack   *ba**kaa**<u>lacking</u>a*
     (f) it/which will always subsidize *e**naa**<u>subsidizing</u>a*
     (g) she/he categorized them   *yaa**dhi**<u>categorizing</u>a**mu***
     (h) I am petitioning   *nd**ii**<u>petitioning</u>a*

Language purists may prefer to ignore such borrowings, but given their increasingly frequent use, any description of the current language cannot ignore them.

3.5   Formality levels

Spoken language is by and large less formal than written language. This difference is also apparent in the transcriptions.

3.5.1   The use of the apostrophe

As it turns out, the grammatical function of the use of the apostrophe to indicate shortening in Lusoga has remained unspecified in the existing reference works on Lusoga. For example, (27i) and (27ii) are cases that should not be mistaken for (2a) and (2b), as they do not result in any semantic change but should be considered as a distinction between formal and informal usages.

(27)   as a person (i)   *nga omuntu*   (formal representation)
                   (ii)   *ng'omuntu*   (informal representation)

3.5.2   The reduplication of noun and verb stems

Stem reduplication, too, has remained unspecified in the existing reference works on Lusoga. Various examples have been extracted from the corpus and are shown in (28), where the parts between brackets are the reduplicated parts. From these examples we see that stem reduplication typically occurs twice or three times, and with both noun and verb stems.

(28) (a)   numerous small wounds   *obubwa(bwa)(bwa)*
     (b)   somewhere in the middle   *wagati(gati)*
     (c)   somehow of black colour   *obwilugavu(ilugavu)*
     (d)   I have redistributed them   *mbakutwile(kutwile)mu*
     (e)   they just drop in   *basuula(suula)mu*
     (f)   small amounts (of liquids)   *matono(tono)*
     (g)   they do not repeatedly fall sick *tibalwala(lwala)*
     (h)   numerous small envelopes   *obubbaasa(bbaasa)*
     (i)   *lit*. small eye eye   *kaliiso(liiso)*

Carefully listening to the recordings moreover reveals that there are environments where the vowel between the reduplicated parts is lengthened. See for instance (29c-f) vs. (29a-b).

(29) (a)   *lit*. we close close   *tuwumba(wumba)*
     (b)   *lit*. to frighten frighten them *okubatiisa(tiisa)*

(c)	*lit*. to try try	*okuwad<u>aa</u>(wada)*
(d)	*lit*. someone trick tricks	*abuz<u>aa</u>(buza)*
(e)	*lit*. to look look after	*okubud<u>aa</u>(buda)*
(f)	*lit*. to suffer suffer	*okubon<u>aa</u>(bona)*

3.5.3   The use of enclitics

Enclitics slightly modify the meanings of expressions, as in (30). They can be repeated and/or combined two or three times, as in (31), respectively (32). The recordings also reveal that any additional enclitic on a word brings about the doubling of the vowel before the onset of the next enclitic. In some cases (underlined in (31) and (32)) the verb ending vowel is equally doubled before the onset of the enclitic(s), a case similar to that noted in (29c-f). In (33a) the enclitic *-ku* is further preceded by the conjunctive *na*, and in (33b) the enclitic *-yo* is preceded by the aspect feature *nga*.

(30)	(a)	we also have there	*twalina**yo***
	(b)	a bit up	*wangulu**ku***
	(c)	a bit down	*wansi**ku***
	(d)	a bit near/nearby	*kumpi**ku***
(31)	(a)	*lit*. to bathe a bit a bit	*okunaaba**kuu(ku)***
	(b)	*lit*. we have some bit bit	*tubaay<u>oo</u>**kuu(ku)***
	(c)	*lit*. to converse a bit bit	*okuway<u>aa</u>**muu(ku)***
	(d)	*lit*. you are of a bit bit number	*muwelaawela**muu(ku)***
	(e)	*lit*. s/he still has a remaining bit bit	*afiisiiz<u>aa</u>**muu(ku)***
	(f)	*lit*. let us make it a bit bit smoother	*tutandaaz<u>ee</u>**woo(ku)***
(32)	(a)	*lit*. we respond to some bit bit bit	*twilanga**muu(kuu)(ku)***
	(b)	*lit*. you prevent some bit bit bit	*waaziyiz<u>aa</u>**kuu(kuu)(ku)***
(33)	(a)	I enable them to get at least a bit	*mbafunhise**(na)(ku)***
	(b)	let's proceed for the moment	*tweyongele**(nga)(yo)***

Forms with multiple enclitics and imbedded formatives only occur in the oral part of the corpus, which leads to the assumption that the reduplication of enclitics is more regulated in the written language. In other words, it seems as if the use of multiple enclitics is exploited in the spoken language to bring about meanings that best serve informal contexts.

3.6   Concords

A core feature of all Bantu languages is the existence of a noun class system with linked concordial agreement. This basically means that nouns, verbs, ad-

jectives, and many other parts of speech, are all 'in harmony' with one another on phrase level. Thus, if a noun belongs to cl. 4, then the adjective modifying it will use an adjective concord (AC) of cl. 4, the verb referring to it will use a subject concord (SC) of cl. 4, etc. Concordial agreement is thought to be inviolable. In the oral part of the corpus, however, this system is occasionally defied, as may be seen from the selection shown in (34).

(34)

#	Phrase	Gender	Concord mismatch
(a)	*omwana gw'endiga* child of a lamb	**1**/2	for cl. **1,** PC cl. 3 *gwe-* is used not cl. 1 *wa-*
(b)	*Mbeebaziiza **emi**limo **dhe** mukoze* I thank you for the work you have done	3/**4**	for cl. **4**, RC cl. 10 *dhe-* is used not cl. 4 *gye-*
(c)	*emilimo **edhi**ndi* some other jobs/work	3/**4**	for cl. **4**, SC cl. 10 *dhi-* is used not cl. 4 *gi-*
(d)	*n'**emi**ti Katonda* *n'eya**dhi**tonda* *lit.* even trees God created them	3/**4**	for cl. **4**, SC cl. 10 *dhi-* is used not cl. 4 *gi-*
(e)	*emikolo **dhaa**tebwa* celebrations were held	3/**4**	for cl. **4**, SC cl. 10 *dhi-* is used not cl. 4 *gi-*
(f)	*emyoyo **dha**ife* our souls	3/**4**	for cl. **4**, PC cl. 10 *dhi-* is used not cl. 4 *gya-*
(g)	*mwino akuwa **gy'o**wa* *lit.* you give to the one who gives you	**1**/2	for cl. **1,** RC cl. 4 *gye-* is used not cl. 1 *gwe-*

In (34a), for example, *omwana* 'child' is a noun in cl. 1, with its plural *abaana* 'children' in cl. 2, thus together gender 1/2. In the phrase, the possessive concord (PC) that follows should thus also be the one from cl. 1, yet here the PC from cl. 3 is used. In (34b), *emilimo* 'jobs' is a noun in cl. 4, yet the relative concord (RC) that follows and which should be in accordance with it is the one from cl. 10 rather than cl. 4. And so on for the other examples.

It is of course not so that the entire concordial agreement system is broken when mother-tongue speakers speak, but the instances shown in (34) are not isolated examples: similar cases are found in the speech of different speakers, in recordings made across different regions. Actually, there were likely more instances than those captured which were 'corrected' in error.

Thus is the burden of the transcriber: While reducing the spoken to the written form of a language, all existing orthography rules need to be applied, mapping undocumented writing problems onto it, so as to be as consistent as possible. However, consistency does not imply that one can intervene to 'correct' what is *thought* to be wrong. Imposing too rigid a structure may even create more structure than there actually is in a language.

4.   Implications for Lexicography

AFRILEX members were introduced to the notion of 'proscriptive lexicography' during the keynote address of Henning Bergenholtz at the *7th International AFRILEX Conference*, in July 2002. Although Gregory James questioned the choice of the term itself during question time (and suggested 'praeterscriptive lexicography', cf. Bergenholtz 2003: 80), and although similar doubts have periodically been raised over the years, amongst others during a workshop on "Proscription, Prescription and Description" at the *14th International AFRILEX Conference*, in July 2009, the concept and the need for it are not in dispute.

The most powerful single example of what proscriptive lexicography entails has been provided by Bergenholtz himself. Consider the following entry in a hypothetical Danish LSP dictionary (Bergenholtz 2003: 78):

> **kraftvarmeværk** noun <et; -et, -er, -erne>
> Other spellings with hyphens are possible: *kraft-varmeværk* or *kraft-varme-værk*. They are not recommended. They are quite rare in language use, e.g. by special field experts. The Danish Language Council allows only the spelling with two hyphens: *kraft-varme-værk*.

Note, here, that even though the Danish Language Council allows the use of one and only one spelling for this word, namely *kraft-varme-værk* (i.e. with two hyphens), the lexicographer explicitly states that this form is "not recommended" (in lines 2 and 3 of the entry). Rather, the lexicographer recommends the spelling without any hyphens, thus *kraftvarmeværk*, as seen in the lemma sign, the argument being:

> Contrary to prescriptive dictionary articles, [the dictionary user] is advised about language use similar to the normal language use in society.
> Bergenholtz (2003: 78)

Indeed, both a linguistic survey and a Google search revealed an overwhelming preference (respectively 85.0% and 98.3%) for the form without any hyphens (Bergenholtz 2003: 69). In proscriptive lexicography, a *recommendation* is thus given, and that recommendation is based on true language usage, regardless of norms and regulations.

Contrast this with prescriptive and descriptive lexicography. Using a prescriptive approach, *kraft-varme-værk* would have been the lemma sign, with the user being told that all other spellings are disallowed. Using a descriptive approach, the user would have been provided with all the facts, and all the spelling options would also have been entered as lemma signs and cross-referred in the dictionary.

Proscriptive lexicography thus takes both data and actual language usage extremely seriously. With regard to the data available to the lexicographer who wants to pursue proscription, Bergenholtz distinguishes the following possibilities:

(a)   introspection,

(b)   analysis of a linguistic survey,

(c)   involvement of descriptions in existing dictionaries, grammars, mono-
graphs, articles, etc.,

(d)   analysis of a number of examples which have been randomly chosen from
random texts (corresponding with the practice of dictionary making before
the age of computers),

(e)   analysis of a specifically constructed text corpus, and

(f)   analysis of usage found in texts in the examined language in all available
web-sites on the Internet.
Bergenholtz (2003: 77)

During the compilation of the *Eiwanika*, possibilities (a) through (d) were used
(cf. Nabirye 2008, 2009a, and Nabirye and De Schryver 2010), and since then
possibilities (e) and (f) have been added to the mix (cf. De Schryver and
Nabirye 2010, as well as Table 1 in the present article). Lusoga lexicography is
thus in the unique position where so-called 'total proscription' is not only a
possibility, but also where it is actually put in practice.

This unique position could, in theory, lead to the perfect dictionary, or
rather, "the perfect *set* of dictionaries". Indeed, given a certain type of problem,
a certain type of user in a certain type of user situation (cf. Bergenholtz 2003:
68), will fair best when consulting a *monofunctional* dictionary:

[A] monofunctional dictionary contain[s] as much data as necessary but as little
as possible to guarantee a rapid access that is not impeded by unnecessary data
or that leads to information stress or even information death. [… However:] The
default approach of many lexicographers is that they are producing a polyfunc-
tional dictionary which should assist the user in satisfying at least a cognitive
function, a text reception function and a text production function.
Bergenholtz and Gouws (2010: 41, 43)

With regard to the latter two dictionary functions, Bergenholtz and Gouws
(2010) have made a case that (text) reception dictionaries need to be *descriptive*,
but that for (text) production, *description* is not viable if more than one variant
prevails, that *prescription* could be viable, and that *proscription* is likely the best
option.

While all of this makes perfect sense *in abstractum*, it is instructive to re-
analyze the various transcription problems enumerated in section 3 above, and
to see how they could be handled in a database from which a variety of mono-
functional dictionaries could subsequently be extracted.[4]

The first type is the easy one: A monofunctional reception dictionary is
ideally descriptive throughout. Each of the grouped examples in (1) through
(34) can indeed be described in utmost detail in a dictionary database: this form
is correctly spelled and means X; this form has a more/lesser frequently used
variant X; this form is used in domain X, for domain Y use …; don't confuse

this form with X; note the capital letter in this form; note the sound change in this form; this form can be repeated, with the meaning/function X; this form is constructed as follows …; this form is an obsolete spelling for X; this form is derived from language X; this form is actually language X, but used; this form is spelled wrongly (according to the official terminology), the correct one is X; the plural of this form is X; the diminutive of this form is X; although this is not according to the grammar, it is used; etc. At face value, the lexicographer did not introduce any value judgements here; in truth, however, both implicit prescription and implicit proscription are at work, amongst others because the most recent and now standard orthography for Lusoga was proposed by the same lexicographer.

For the second type, namely the monofunctional production dictionary, a hybrid approach has been put forward. While proscription is seen as the best option, this needs to be understood as a situation in which the lexicographer recommends throughout. At times, these recommendations may overlap with what is actually prescribed, and where there are no variants, one is in effect describing. Or in the words of Bergenholtz and Gouws:

> In trying to satisfy a text production function, the lexicographer should pay careful attention to the application of an approach characterised by either description, prescription or proscription or a hybrid application in which more than one of these approaches can be combined. This decision should not be made in a haphazard way.
> Bergenholtz and Gouws (2010: 47)

While our database contained 'mere' descriptions in the slots meant for the creation of a reception dictionary, the situation is truly hybrid in the database slots meant for the creation of a production dictionary. The latter is illustrated visually in the Addendum to this article.

By way of conclusion, then, the metalexicographic explorations by Bergenholtz and his colleagues have found an application — and indeed confirmation — in the emerging field of dictionary making for Lusoga. Reducing a language to writing before it has even been standardized is a daunting undertaking, but knowing that after moving from oral data to transcription, there is a theoretical framework consisting of a hybrid use of lexicographic prescription, description and proscription, to subsequently move to a dictionary database and finally to a set of monofunctional dictionaries, is sufficient consolation to carry on with the work.

Endnotes

1. For the definition of 'organic corpus' (and the related issues of 'representativeness' and 'balance'), see Atkins et al. (1992: 1, 4, 6).
2. Namely the first author of this article, who is also a father-tongue speaker of the language.

3.   Similarly, *Omusoga* 'Musoga' is also found as **omusoga* and **omuSoga* in the e-mail and Facebook sections of the corpus.

4.   For the concept of "one database, many dictionaries", see De Schryver and Joffe (2005).

References

Atkins, B.T.S., J. Clear and N. Ostler. 1992. Corpus Design Criteria. *Journal of Literary and Linguistic Computing* 7(1): 1-16.

Bergenholtz, H. 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.

Bergenholtz, H. and R.H. Gouws. 2010. A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography. *Lexikos* 20: 26-51.

Byandala, G.I. 1963. *The Lusoga Orthography*. Iganga.

De Schryver, G.-M. and D. Joffe. 2005. One Database, Many Dictionaries — Varying Co(n)text with the Dictionary Application TshwaneLex. Ooi, V.B.Y., A. Pakir, I. Talib, L. Tan, P.K.W. Tan and Y.Y. Tan (Eds.). 2005. *Words in Asian Cultural Contexts. Proceedings of the 4th Asialex Conference, 1-3 June 2005, M Hotel, Singapore*: 54-59. Singapore: Department of English Language and Literature & Asia Research Institute, National University of Singapore.

De Schryver, G.-M. and M. Nabirye. 2010. A Quantitative Analysis of the Morphology, Morphophonology and Semantic Import of the Lusoga noun. *Africana Linguistica* 16: 97-153.

Gonza, R.K. 2007. *Lusoga–English Dictionary and English–Lusoga Dictionary*. (Revised edition of P. Korse's (1999a) dictionary.) Kampala: MK Publishers.

Kajolya, J.B.N. 1990. *The Lusoga Orthography*. (Revised edition of G.I. Byandala's (1963) Lusoga orthography.) Jinja: Lusoga Ecumenical Committee.

Korse, P. 1999a. *Dictionary Lusoga–English/English–Lusoga*. Jinja: Cultural Research Centre.

LULANDA and CRC. 2004. *Empandiika y'Olulimi Olusoga Enkalamu/Standard Lusoga Orthography*. Jinja: Lusoga Language Authority.

Nabirye, M. 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished MA dissertation. Kampala: Makerere University.

Nabirye, M. 2009. *Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwega Olusoga* [A Dictionary of Lusoga. For speakers of Lusoga, and for those who would like to learn Lusoga]. Kampala: Menha Publishers.

Nabirye, M. 2009a. Compiling the First Monolingual Lusoga Dictionary. *Lexikos* 19: 177-196.

Nabirye, M. and G.-M. de Schryver. 2010. The Monolingual Lusoga Dictionary Faced with Demands from a New User Category. *Lexikos* 20: 326-350.

Namyalo, S., L. Walusimbi, G. Bukenya, M.W. Masakala, M. Nabirye and F. Kiingi. 2008. *A Unified Standard Orthography of Eastern Interlacustrine Bantu Languages*. Monograph Series 68. Cape Town: The Centre for Advanced Studies of African Society.

NCDC. 2006. *THEMA. The newsletter of the Thematic Primary Curriculum. Issue 1. August 2006*. Kampala: National Curriculum Development Centre.

NCDC. 2006a. *THEMA News Letter. Issue 2. December 2006*. Kampala: National Curriculum Development Centre.

Addendum: Prescription (Pre), Description (De) and Proscription (Pro) in a Monofunctional Production Dictionary for Lusoga
[*with ✓ = the inclusion of the lemma sign; (✓) = the inclusion of the lemma sign, but with a treatment limited to a cross-reference to the main/ correct variant only*]

N #	English	Lusoga	Pre	De	Pro
1 a	today	*(o)lwaleelo*	✓		
b	the day of today	*(o)lwa leelo*	✓		
c	the day of Sunday	*(o)lwa Saabbiiti*	✓		
2 a	March	*(O)gwokusatu*	✓		
b	a third	*ogw'okusatu*	✓		
3 a	*lit.* month **second time** 2001	*mwezi **Gwakubili** <u>nkumi ibili ikumi na mulala</u>*	✓		
b	*lit.* day **fifth**	*olunaku **lwakutaanu***	✓		
c	*lit.* before we go in **first**	*nga tukaali gya mu **Lubelyebelye***	✓		
4 a	*lit.* (something in cl. 7) of **first** > primary 1	*ekyo**lubelyebelye***	✓		
b	*lit.* (something in cl. 7) of **daylight** > lunch	*ekyo**musana***	✓		
5 a	also	*boona*	✓		
b	all	*boonaboona*	✓		
c	all (+ emphasis)	*boona boona …*	✓		
6 a	grandparents	*badhaadha*	✓		
b	grandparents of the grandparent	*badhaadha ba dhaadha*	✓		
c	grandparents of the grandparent's grandparent	*badhaadha ba dhaadha ba dhaadha*	✓		
7 a	someone from Bilumba	*Munabilumba*	✓		
b	people from Mayuge	*Banamayuge*	✓		
c	someone from Masese	*Munamasese*	✓		
8 i	Ugandan	*(O)munauganda*	✓		
ii	also found as:	**(o)muna Uganda*	(✓)		
iii	also found as:	**(o)munaUganda*	(✓)		
9 a	of the Menha clan	*Omwisemenha*	✓		
b	of the Igaga clan	*Abaiseigaga*	✓		
c	of the Mususwa clan	*Abaisemususwa*	✓		
10	supporter of the NRM	*mwiseNRM*	✓		
11 a	*lit.* owner + voice	*Mweneiloboozi*	✓		
b	*lit.* it was + home	*Gaalimaka*	✓		
c	*lit.* food + of + morning	*Mmeleyankya*	✓		
d	*lit.* someone + stubbornness	*kimpwitu*	✓		
12 a	*Gaalimaka* + number + capitalization	*Bagaalimaka*	✓		
b	*Omunauganda* + number + capitalization	*Abanauganda*	✓		
c	*kimpwitu* + number + null capitalization	*bakimpwitu*	✓		
13 a_i	(woman who has made a pilgrimage to Mecca)	*Ha**gy**ati*			(✓)
a_ii		*Ha**j**ati*			✓
b_i	(pilgrimage to Mecca)	*hi**gy**a*			(✓)
b_ii		*hi**j**a*			✓

142   Minah Nabirye and Gilles-Maurice de Schryver

				C1	C2	C3
14	a_i	(a place name)	*Ngando*	(✓)		
	a_ii		*Nŋando*	✓		
	b_i	relatives	*enganda*	(✓)		
	b_ii		*eŋŋanda*	✓		
15	a_i	Jinja	*Idinda*		✓	
	a_ii	[dying out]	*Idhindha*		(✓)	
	b_i	nation/country/tribe	*eiwanga*		✓	
	b_ii	[dying out]	*eighanga*		(✓)	
16	a	madam	*inhabo > nnyabo*			✓
	b	sir	*isebo > ssebo*			✓
17		*Wanyange*	*Wanhange*			✓
		Makerere	*Makeelele*			✓
18	a_i	Christ (Catholics)	*Kristu* / ***Kulisitu***		✓/(✓)	
	a_ii	Christ (Protestants)	*Kristo* / ***Kulisito***		✓/(✓)	
	b_i	sacrament (Catholics)	*sakramentu* / **saakalamentu**		✓/(✓)	
	b_ii	sacrament (Protestants)	*sakramento* / **saakalamento**		✓/(✓)	
	c	democracy	*demokrasiya* / **dimokulasiya**		✓/(✓)	
19	i	Frequency Mode	*FM*		(✓)	
	ii		*Ef Em*		(✓)	
	iii		*Fa Ma*		✓	
20	a	*RDC* + number	*baRDC*	✓		
	b	*DISO* + number	*baDISO*	✓		
	c	*DPC* + number	*baDPC*	✓		
	d	*DHO* + number	*baDHO*	✓		
21	i	lawyers (number + phonologized)	*balooya*	✓		
	ii	lawyers (number + English root)	*balawyer*	✓		
22	a	hello(s)	*alo* (sg.) / *baalo* (pl.)		✓	
	b	district(s)	*district* (sg.) / *madistrict* (pl.)		✓	
	c	government(s)	*gavumenti* (sg.) / *dhigavumenti* (pl.)		✓	
	d	bye(s)	*bbaayi* (both sg. and pl.)		✓	
	e	septrine(s)	*septrine* (both sg. and pl.)		✓	
23		pertaining to the post of a Minister	*obwaminister*		✓	
24	a	a bit of airtime	***ka****airtime*		✓	
	b	short programs	***bu****programs*		✓	
25	a	small plus	***ka****plus*		✓	
	b	*extensives	*ebyo **bi**extensive ebyo*		✓	
26	a	and planning for her/him	*n'okumuplaningila*			✓
	b	they pronounced	*bapronouncinze*			✓
	c	I did not monitor	*tyamonitorinze*			✓
	d	it comes	*neecominga*			✓
	e	they still lack	*bakaalackinga*			✓
	f	it/which will always subsidize	*enaasubsidizinga*			✓
	g	she/he categorized them	*yaadhicategorizingamu*			✓
	h	I am petitioning	*ndiipetitioninga*			✓
27	i	as a person (formal representation)	*nga omuntu*	✓		
	ii	as a person (informal representation)	*ng'omuntu*	✓		
28	a	numerous small wounds	*obubwa(bwa)(bwa)*		✓	
	b	somewhere in the middle	*wagati(gati)*		✓	
	c	somehow of black colour	*obwilugavu(ilugavu)*		✓	
	d	I have redistributed them	*mbakutwile(kutwile)mu*		✓	
	e	they just drop in	*basuula(suula)mu*		✓	
	f	small amounts (of liquids)	*matono(tono)*		✓	

				✓
g	they do not repeatedly fall sick	*tibalwala(lwala)*	✓	
h	numerous small envelopes	*obubbaasa(bbaasa)*	✓	
i	*lit*. small eye eye	*kaliiso(liiso)*	✓	
29 a	*lit*. we close close	*tuwumba(wumba)*	✓	
b	*lit*. to frighten frighten them	*okubatiisa(tiisa)*	✓	
c	*lit*. to try try	*okuwad<u>aa</u>(wada)*	✓	
d	*lit*. someone trick tricks	*abuz<u>aa</u>(buza)*	✓	
e	*lit*. to look look after	*okubud<u>aa</u>(buda)*	✓	
f	*lit*. to suffer suffer	*okubon<u>aa</u>(bona)*	✓	
30 a	we also have there	*twalina**yo***	✓	
b	a bit up	*wangulu**ku***	✓	
c	a bit down	*wansi**ku***	✓	
d	a bit near / nearby	*kumpi**ku***	✓	
31 a	*lit*. to bathe a bit a bit	*okunaaba**kuu(ku)***	✓	
b	*lit*. we have some bit bit	*tubaay<u>oo</u>**kuu(ku)***	✓	
c	*lit*. to converse a bit bit	*okuway<u>aa</u>**muu(ku)***	✓	
d	*lit*. you are of a bit bit number	*muwelaawela**muu(ku)***	✓	
e	*lit*. s/he still has a remaining bit bit	*afiisiiz<u>aa</u>**muu(ku)***	✓	
f	*lit*. let us make it a bit bit smoother	*tutandaaz<u>ee</u>**woo(ku)***	✓	
32 a	*lit*. we respond to some bit bit bit	*twilanga**muu(kuu)(ku)***	✓	
a	*lit*. you prevent some bit bit bit	*waaziyiz<u>aa</u>**kuu(kuu)(ku)***	✓	
33 a	I enable them to get at least a bit	*mbafunhise**(na)(ku)***	✓	
b	let's proceed for the moment	*tweyongele**(nga)(yo)***	✓	
34 a	child of a lamb	***omwana gw'endiga***		✓
b	I thank you for the work you have done	*Mbeebaziiza **emi**limo dhe mukoze*		✓
c	some other jobs / work	***emi**limo **edh**indi*		✓
d	*lit*. even trees God created them	*n'**emi**ti Katonda n'eyadhitonda*		✓
e	celebrations were held	***emi**kolo **dhaa**tebwa*		✓
f	our souls	***emy**oyo **dha**ife*		✓
g	*lit*. you give to the one who gives you	***mwi**no akuwa **gy'o**wa*		✓