Corpus-driven Bantu Lexicography Part 1: Organic Corpus Building for Lusoga

Gilles-Maurice de Schryver, *BantUGent, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Department of African Languages, University of Pretoria, Pretoria, South Africa (gillesmaurice.deschryver@UGent.be)*
and
Minah Nabirye, *BantUGent, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Department of Teacher Education and Development Studies, Kyambogo University, Kampala, Uganda (minah.nabirye@UGent.be)*

Abstract: This article is the first in a trilogy that deals with corpus-driven Bantu lexicography, which is illustrated for Lusoga. The focus here is on the building of a so-called 'organic corpus' from scratch, while the next two instalments will deal with the use of that corpus on the macro-structural and microstructural levels, respectively. Not many detailed descriptions of corpus-building efforts exist for Bantu languages, so each and every step is discussed in detail, paying particular attention to the parameters that have to be taken into account, while not losing sight of the need to log the metadata either.

Keywords: BANTU, LUSOGA, CORPUS BUILDING, ORGANIC CORPUS, ORAL, WRITTEN, SOURCE, PERIOD, GENRE, TOPIC, METADATA

Obufunze: Omutengeso gw'eitu ogukozesebwa mu namawanika w'ennimi dha Bantu. Ekitundu 1: Okuzimba namukyukilo w'eitu ly'Olusoga. Olupapula luno n'olusooka ku isatu edhinaayogela ku musomo gw'omutengeso gw'eitu ogukozesebwa mu namawanika w'ennimi dha Bantu nga gulaga omulimu ogw'akolebwa ku Lusoga. Mu lupapula luno, eisila lili ku nzimba ya itu namukyukilo okuva ku ntandiiko. Ebitundu ebinaaba mu lupapula olw'okubili n'olw'okusatu biidha kugema ku nkozesa ya itu lino ku isa ly'omutindiigo ogw'ebizimbibwa mu mutegeko n'eisa elilaga eitu lino mu mwoleko ogw'azimbibwa mu mutindiigo n'engeli omusingi ogulimu bwe gulagibwa mu iwanika. Mu nnimi dha Bantu, emilimu egilaga omusingi guno tigitela kuwandiikibwaku mu butongole okusobola okumanhisa abo abayinza okuba nga bagasibwa. Kale buli kitundu ekiteesebwaku mu nnambika eli mu mpapula eisatu dhino kitoolayo buli kanhomelo ka bukodyo n'emitendela egy'agobelebwa ela gy'akozesebwa mu kusenvula omulimu gw'okuzimba omutimbo gw'ekyebungo ky'olulimi Olusoga gwonagwona.

Ebigambo ebikulu: BANTU, LUSOGA, OKUZIMBA EITU, EITU NAMUKYUKILO, ENDHOGELA, EMPANDIIKA, OBUVO, EKISEELA, ENNAMBIKA, EKINHUMYO, OMUTIMBO GW'EKYEBUNGO

1.   Goal of the present study

In this article we wish to show how an electronic corpus for a Bantu language, especially an under-resourced Bantu language, may be assembled from scratch. We have lexicographic applications in mind, but such corpora may also be used (and *have* successfully been used) for Bantu corpus linguistics studies more generally. While Bantu corpora have been built for about two decades now, explicit descriptions of their composition are rare in the literature. For instance, in his MA dissertation de Schryver (1999: 103-117) devotes about 14 pages to the design, structure, contents and text collection of a 300 000-word Cilubà corpus, but to this date that study remains unpublished. When it comes to the descriptions of the corpora that have been assembled for the South African Bantu languages, these are typically less than a page long (de Schryver and Prinsloo 2000). On the other hand, corpus stability tests have been carried out for the South African Bantu languages (Prinsloo and de Schryver 2001, Prinsloo 2015), as well as attempts at multilingual corpus building and multi-lingual data extraction (de Schryver 2002, Prinsloo and de Schryver 2005). Scientific articles on the Zimbabwean corpora built under the umbrella of ALLEX/ALRI tend to focus on specific topics, such as tagging issues for a Shona corpus (Chabata 2000) or the sociolinguistic, political and economic con-siderations that influence the contents of a corpus of Zimbabwean Ndebele (Hadebe 2002). Even the latest version of the widely-used *Helsinki Corpus of Swahili* is not accompanied by a proper description (Hurskainen 2016).

The only exceptions to this pattern seem to be the corpora built to carry out corpus linguistics studies at BantUGent (i.e., the UGent Centre for Bantu Studies) where, for instance, the PhDs of Mberamihigo (2014), Nshemezimana (2016) and Misago (2018) describe the various Kirundi corpora built, or where the PhD of Kawalya (2017) describes the Luganda corpus that he used for his study. The building of a Lingála corpus may be found in the PhD of Sene-Mongaba (2013), reworked and expanded as Sene-Mongaba (2015). Our effort (Nabirye 2016), on which the Lusoga case study presented below is based, is also the result of PhD research undertaken at BantUGent.

With regard to corpus-building efforts for Lusoga, only one exploratory study has appeared so far (Nabirye and de Schryver 2011). In that study, the main focus was on the writing problems that the corpus builder encounters during the transcription of oral material and the implications for the corpus lexicographer when data is extracted from such a corpus. In contrast, of par-ticular interest in the present study will be the parameters/axes that can be used to characterise the composition of a Bantu-language corpus, these being, in addition to oral vs. written, also the distribution of the sources, the periods, the genres and the topics. Orthographic issues will only briefly be recapped here. Furthermore, the value of detailed corpus documentation will be exempli-fied; this will be done by means of the inclusion of and reference to a compre-hensive addendum. Corpus-query software will be mentioned in passing.

2.   The Lusoga language and publications in Lusoga

Lusoga is a largely undocumented Great Lakes Bantu language classified as JE16 (Guthrie 1948, Maho 2009). According to the Uganda Bureau of Statistics, 2 062 920 people identified themselves as Basoga in 2002 (UBOS 2006: 12), a figure that grew by nearly half to a respectable 2 960 890 by 2014 (UBOS 2016: 71). While immediately acknowledging that not all people who claim to be Basoga also necessarily speak 'Lusoga', however defined,[1] one should still realise that several million people currently speak Lusoga, of which about two million are monolingual. While it might surprise that a language with up to three million speakers may be largely undocumented, it is fitting to recall that there are even endangered languages with millions of speakers (Adelaar 2014).

Lusoga was first reduced to writing near the end of the 19th century, as pointed out by Condon a century ago:

> The Basoga Batamba had no written characters. Nor do any writings on rocks or pictorial characters exist. According to native report — and I mean natives of a ripe old age — there never was, as far as they remember, any means whatever of placing down their verbal utterances. All messages from one chief to another were committed to a trustworthy man, who learned the communication by heart, and so delivered the message by word of mouth. It is only within the last 15 years that the language of this people has been put in book form.
> (Condon 1911: 368)

The very first language data for Lusoga may be found in the 'vocabularies' included in Johnston (1902: 980-991) as well as in Condon (1911). However, we have found no evidence to suggest that Lusoga was documented in earnest prior to the 1960s. The earliest reference uncovered so far with an exclusive focus on Lusoga is the orthography of Byandala (1963). That booklet was followed by the documentation of Lusoga proverbs and riddles in Lyavala-Lwanga (1967, 1969). There is no record of Lusoga materials produced during the 1970s or the 1980s. Writing on and in Lusoga was again picked up in the 1990s. The first Lusoga publication in this period was the second version of the Lusoga orthography: Kajolya (1990). It was followed by two attempts at publishing a newspaper, which faltered shortly after: *Kodh'eyo* (1997–98) and *Ndimugezi* (1998–99). From the late 1990s and early 2000s onwards, the main output in Lusoga has come from the *Cultural Research Centre* (CRC), a religious body based in Jinja (e.g., CRC 1998a, 1999a, b, c, d, e, f, g, h, 2000a, b, 2002, 2005a, Kaluuba et al. 2010, CRC 2011).[2] Also, one very prolific writer is Gulere who, amongst others, self-published ten children's story books, which he placed online in various locations at various times and in various formats (Gulere 2011a, b, c, d, e, f, g, h, i, j). Gulere moreover self-published two translations, one of *Antigone*, a tragedy by the ancient Greek playwright Sophocles from 441 BC (Gulere 2007a), another of *The Bride*, a play in English by the Ugandan Austin L. Bukenya from 1987 (Gulere 2007b).[3] Lastly, a first novel has now been published in Lusoga, written by Kuunya (2011a).

3.   Building a corpus for Lusoga

3.1   Towards an organic (but structured), general-language, synchronic Lusoga corpus

The basics of corpus building for the Bantu languages have been described by de Schryver and Prinsloo (2000). The two important concepts that also applied to the building of our Lusoga corpus are that of an 'organic corpus' and that of a 'structured corpus'. An 'organic corpus' has been defined by Atkins, Clear and Ostler as follows:

> [...] a corpus may be thought of as organic, and must be allowed to grow and live if it is to reflect a growing, living language. [...] In order to approach a 'balanced' corpus, it is practical to adopt a method of successive approximations. First, the corpus builder attempts to create a representative corpus. Then this corpus is used and analysed and its strengths and weaknesses identified and reported. In the light of this experience and feedback the corpus is enhanced by the addition or deletion of material and the cycle is repeated continually. [...] In our ten years' experience of analysing corpus material for lexicographical purposes, we have found any corpus — however 'unbalanced' — to be a source of information and indeed inspiration. Knowing that your corpus is unbalanced is what counts.
> (Atkins et al. 1992: 1, 4, 6)

De Schryver and Prinsloo link this to what they call a 'structured corpus' as follows:

> Formulated differently, it is any corpus compiler's task to attempt to assemble a representative corpus for his/her specific need(s). Subsequent additions and deletions of sections should be seen as a balancing activity to rectify initial weaknesses, but more importantly, also to take account of and track a growing, living language. As such, there is no such thing as 'the' corpus of a certain language (variety). Rather, at any point in time one selects a certain number of texts from the range of available electronic texts (which might or might not be grouped together into sub-corpora), and uses 'a' corpus for the specific research one wishes to pursue. The minimum requirement for any organic corpus is thus that the corpus compiler(s) will have attempted to put some structure in assembling the range of electronic texts. Within this framework, any first attempt at compiling an organic corpus will at least result in a structured corpus.
> (de Schryver and Prinsloo 2000: 92)

Our Lusoga corpus is both structured and organic. On the whole, the organicity means that the overall size has increased and decreased over the years.

Corpus building for the Bantu languages is always slightly opportunistic, in that one adds the little existing written material one can get hold of, except when a serious imbalance results. In other words, to get going, one often makes do with an 'imperfect corpus', which is then modified later on, when 'better' data becomes available. Over and above this balancing act, the corpus used

should always attempt to be representative of the population that is the subject of the planned description or research. For a general-language corpus, the goal is consequently to acquire as many different genres as possible, that deal with as wide a topic range as possible. Existing written material for all but a few Bantu languages is unfortunately biased in this respect. Most are the result of (modern) missionary activities, so the genre *Biblical documents* tends to be over-represented in many Bantu corpora. Conversely, for Bantu languages with a varied, vibrant and ongoing online media presence, the genre *Journalism* may be overrepresented, and within that, topics such as *Sports* and *Politics*. Of course, when the aim is to describe features of biblical works or journalistic texts, then such types of corpora may indeed be 'representative', and when multiple sources have been equally sampled, these corpora may also be 'balanced'. But if the goal is to describe the general language, then an effort needs to be made to achieve both representativeness and balance in another way. It is here that the material found in the oral component of a corpus may bring a solution, as it did for our Lusoga corpus (cf. infra, §3.5.1).

Another important point concerns the time period covered by a Bantu corpus. In all but a few cases, this will be 'the present', with that present optionally stretching back to a number of decades, maximum half a century. Although attempts are being made to build Bantu corpora with time-depths of at least half a century down to a century — such as for Zulu (de Schryver and Gauton 2002), Kirundi (Mberamihigo et al. 2016) and Luganda (Kawalya et al. 2018) — the only Bantu corpus containing substantial amounts of diachronic data that has been built (and used)[4] is the set of corpora for the *Kikongo Language Cluster*, where some parts are up to four centuries old, while others go back to around 250 years ago (Bostoen and de Schryver 2015). For Lusoga, the aim has always been to build a synchronic corpus covering the general language. Material older than a few decades is in any case extremely rare for Lusoga (cf. supra, §2). When available, it was nonetheless included in an attempt to widen the genre/topic range.

3.2   The 0.5m Lusoga corpus

A first Lusoga corpus, of about half a million words, was built as part of the research leading to an MA dissertation. Its composition is as shown in Table 1 (adapted from Nabirye (2008: 70)).

Table 1:   Genre distribution in the 0.5m Lusoga corpus

Genre	Tokens	%
Journalism (*Kodh'eyo*, *Ndimugezi*)	187 393	34.84%
Biblical documents (*New Testament* and others)	199 853	37.16%
Short stories and idioms (*Kintu*, *Ababita Ababiri*, etc.)	150 560	28.00%
SUM	**537 806**	**100.00%**

3.3   The 0.9m Lusoga corpus

For a corpus-based study of the Lusoga noun (de Schryver and Nabirye 2010) the Lusoga 'MA corpus' was supplemented with the full text of the *Eiwanika ly'Olusoga* (Nabirye 2009), being a monolingual Lusoga dictionary compiled without the use of a corpus. The reasoning at the time was that because the example sentences from that dictionary were the result of original fieldwork, they could as well form part of a Lusoga corpus. A number of reports written in Lusoga (from the Busoga clan leaders, the private sector, academia, etc.) were also added, as was the initial impetus for a true oral part of the Lusoga corpus (i.e., the first few transcriptions of conversations, interviews and songs). The make-up of this Lusoga 'noun corpus' is as shown in Table 2 (taken from de Schryver and Nabirye (2010: 100)).

Table 2:   Genre distribution in the 0.9m Lusoga corpus

Genre	Tokens	%
Reference work (*Eiwanika ly'Olusoga*)	305 660	35.00%
Journalism (*Kodh'eyo*, *Ndimugezi*)	187 393	21.46%
Biblical documents (*New Testament* and others)	199 853	22.88%
Reports (from the Busoga clan leaders, private sector, academia, etc.)	24 166	2.77%
Short stories and idioms (*Kintu*, *Ababita Ababiri*, etc.)	150 560	17.24%
Transcriptions of conversations, interviews and songs	5 716	0.65%
SUM	**873 348**	**100.00%**

This version of the Lusoga corpus contained about 870 000 running words (tokens), and about 150 000 orthographically different words (types). Not only the transcriptions of conversations, interviews and songs but also the dictionary examples (together close to a third of the total) could be considered reductions of spoken data to text; the other genres being written texts from the start. From Table 3 (also taken from de Schryver and Nabirye (2010: 100)) one may further deduce that most sources are recent to very recent, with over 98% produced during the past two decades.

Table 3:   Period distribution in the 0.9m Lusoga corpus

Period	Tokens	%
1960s	16 822	1.93%
1970s	—	—
1980s	—	—
1990s	457 978	52.44%
2000s	398 548	45.63%
SUM	**873 348**	**100.00%**

3.4   The 1.1m Lusoga corpus

Following the Lusoga noun study, and with the acquisition of more data to compensate for it, the dictionary data was again dropped from the Lusoga corpus. Although based on natural language production, the dictionary examples lacked the original context, and had in any case been 'selected' for their pedagogical value. As such, they did not have their place in a proper text corpus, that is, one that consists of large sections of free-flowing, running text. Instead, the symbolic oral section of about 6 000 tokens in the Lusoga 'noun corpus' was enlarged to well over 400 000 tokens. Furthermore, various texts translated from English, as well as digital-born Lusoga material, were also added, to obtain the corpus that was used for the study of the writing problems in a Lusoga corpus (Nabirye and de Schryver 2011). The composition of that new corpus is as shown in Table 4 (adapted from Nabirye and de Schryver (2011: 123)).

Table 4:   Genre distribution in the 1.1m Lusoga corpus

Genre	Tokens	%
Journalism (*Kodh'eyo, Ndimugezi*)	187 393	17.07%
Biblical documents (*New Testament* and others)	199 853	18.20%
Reports (from the Busoga clan leaders, private sector, academia, etc.)	24 166	2.20%
Short stories and idioms (*Kintu, Ababita Ababiri*, etc.)	150 560	13.71%
Transcriptions of conversations, interviews and songs, as well as traditional ceremonies, speeches, sermons, radio broadcasts, etc.	413 827	37.69%
Translations from English (*PEAP (Poverty Eradication Action Plan)*, *ICEE (International Centre for Eye Education)*, *FIDA/PLAN (inheritance laws)*, etc.)	19 814	1.80%
Electronic texts (e-mails, mailing lists, *Facebook*, etc.)	102 365	9.32%
SUM	**1 097 978**	**100.00%**

This 1.1m Lusoga 'writing-problems corpus' — just as the earlier 0.9m Lusoga 'noun corpus' and the even earlier 0.5m Lusoga 'MA corpus' — was not annotated for any linguistic features. As such, these corpora were not tagged for parts of speech, nor lemmatised. They are known as 'raw corpora'.

3.5   The 1.7m Lusoga corpus

The latest iteration of the Lusoga corpus stands at over 1 700 000 tokens and about 200 000 types. The various text files of the 1.1m Lusoga 'writing-problems corpus' were cleaned up, re-assembled and renamed. New material was added for each genre except *Journalism*. For the latter, however, all the newspa-

per clippings were reprocessed with better software (cf. infra, §3.5.2). It is this version of the Lusoga corpus that we will now study in more detail.

3.5.1   Oral vs. written distribution

In contrast to the 0.5m Lusoga corpus, which had no transcribed text, and the 0.9m one with just 5 716 such tokens, a major effort in building the 1.7m Lusoga corpus went to expanding the oral component even further compared to the 1.1m Lusoga corpus. While the model of all modern corpora, the 100m *British National Corpus* (BNC 1994–2018), has set the standard for general-language corpora to contain 10% spoken material vs. 90% written material (Rundell and Stock 1992: 46), we managed to triple this conventional allocation of the spoken part in the total. In all, 216 audio files were transcribed, amounting to well over half a million tokens, as may be seen from Table 5, which corresponds to 31% of the total corpus, illustrated graphically in Figure 1.

Table 5:   Statistics for the oral vs. written distribution in the 1.7m Lusoga corpus

Medium	No. of files	%	Tokens	%
Oral	216	55.24%	541 129	31.39%
Written	175	44.76%	1 182 562	68.61%
SUM	**391**	**100.00%**	**1 723 691**	**100.00%**

Oral vs. Written in the 1.7m Lusoga corpus
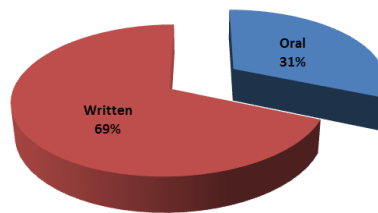
Oral 31%

Written 69%

Figure 1:   Pie chart showing the oral vs. written distribution in the 1.7m Lusoga corpus

There is nothing magic about attaining over half a million words of spoken data,[5] nor about reaching a division of a third for oral vs. two-thirds for written data, but for a language which to this date is chiefly an oral language, it simply looked like a necessity in order to ensure that any explanations drawn from this corpus would also reflect real language usage. The oral component is size-able enough so as to feature in every screenful of concordance lines, where oral and written material is instantly juxtaposed and may be cross-compared to

make sure there are no differences between oral vs. written language use that would need to be reported.

What is true is that there is an addictive aspect to corpus building, so a goal was set to reach about '100 hours of audio'. Indeed, the 541 129 tokens of transcribed material correspond to exactly 98 hours, 42 minutes, and 38 seconds of audio files. Transcribing half an hour of audio took on average two hours, which means that 400 hours were required for all the transcriptions (not counting the fieldwork and hours spent recording in the first place, nor the many hours to collect and log all the metadata and consent forms). The types of audio recorded and transcribed are varied, and include modern and traditional songs, radio talk shows, traditional ceremonies (as currently being performed), business meetings, interviews and dialogues.

3.5.2 Source distribution

The bulk of the written part of the 1.7m Lusoga corpus was assembled through the digitization of more or less every work, down to every snippet, ever written and published in Lusoga, whether commercially or produced as grey literature. A total of 85 sources were scanned in high resolution, after which the optical character recognition (OCR) tool of OmniPage (1995–2018) was utilised to turn the images into machine-readable texts.[6] These 85 sources were good for about 670 000 tokens. OCR was also used to re-digitise large parts of the two short-lived Lusoga newspapers: *Kodh'eyo: Busoga etebenkere* (Kodh'eyo 1997–98) and *Ndimugezi n'omukobere: The factfinder* (Ndimugezi 1998–99). Due to the poor quality of the printing of these newspapers, the OCR output required substantial clean-up. The result was about 200 000 tokens of newspaper articles. A further 62 files were obtained electronically. These included self-published works found on the Internet, unpublished material from friends, private e-mail and mailing list communications, translations into Lusoga taken from government, NGO and commercial websites, as well as some religious material found online. All these texts together came to about 260 000 tokens. The translations we ourselves had made over the years, 15 of them, were also added, which contributed a further 25 000 tokens, as well as some of our own writings, six texts with just 2 500 tokens. The remainder consisted of low-resolution images of texts found online, as well as a single hand-written document, which were all retyped, adding another 25 000 tokens.

An overview of these various sources may be seen in Table 6. For a mostly undocumented and oral language like Lusoga, we must admit that we never expected to be able to reach nearly 1.2m tokens of material that had been written in one way or another. Extending the corpus building effort beyond the more obvious transcriptions and OCR, as seen in the last five bullets of Table 6, clearly helped in this regard (and in effect resulted in about a quarter of the written data).

Table 6:   Statistics for the source distribution in the 1.7m Lusoga corpus

Source	No. of files	%	Tokens	%
ORAL				
• Transcriptions of audio	216	55.24%	541 129	31.39%
WRITTEN				
• OCR (optical charac-ter recognition)	85	21.74%	669 320	38.83%
• OCR + Retyping	2	0.51%	201 664	11.70%
• Electronic transfers	62	15.86%	258 990	15.03%
• Translations	15	3.84%	25 365	1.47%
• Own writings	6	1.53%	2 568	0.15%
• Retyping of images	4	1.02%	24 436	1.42%
• Retyping of hand-written document	1	0.26%	219	0.01%
SUM	**391**	**100.00%**	**1 723 691**	**100.00%**

3.5.3 Period distribution

As may be seen from the data presented in Table 7 and the bar chart shown in Figure 2, the 1.7m Lusoga corpus is essentially a synchronic corpus with a time-depth of just over 20 years.

Table 7:   Statistics for the period distribution in the 1.7m Lusoga corpus

Period	No. of files	%	Tokens	%
1940s	1	0.26%	1 325	0.08%
1950s	—	—	—	—
1960s	2	0.51%	36 065	2.09%
1970s	—	—	—	—
1980s	1	0.26%	16 657	0.97%
1990s	44	11.25%	417 837	24.24%
2000s	139	35.55%	398 153	23.10%
2010s (to 2013)	204	52.17%	853 654	49.52%
SUM	**391**	**100.00%**	**1 723 691**	**100.00%**

Only four files represent the 1940s, 1960s and 1980s.[7] The 1990s and 2000s are equally represented, with about 400 000 tokens each, while the 2010s (and only up to August 2013 at that) is represented by as many as 850 000 tokens. While each of the past two periods and the present one cover both oral and written material, up to 70% of the transcriptions concern spoken data from the 2010s, which is the main reason why the 2010s contain more material than any other period. Another is the flurry of primers that were produced in the 2010s, in the

wake of the recognition of Lusoga as a medium of instruction in 2005 (NCDC 2006: 5).
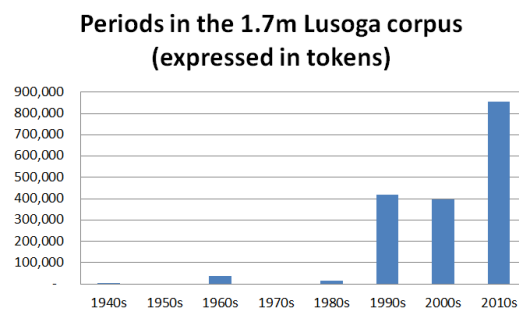

Figure 2:   Bar chart showing the period distribution in the 1.7m Lusoga corpus

3.5.4   Genre distribution

The 391 files of the 1.7m Lusoga corpus were also grouped into 12 broadly-defined genres, as summarised in Table 8 and shown graphically in Figure 3. Three genres dominate, making up more than half the corpus: *Biblical documents* (23% of the tokens),[8] *Literature* (16%) and *Radio talk shows* (15%). Also sizable are *Journalism* (12%) and *E-mails* (9%). Each of the next five genres contains about a twentieth (5%) of the total corpus: *Policy documents*, *Interviews*, *Songs*, *Celebrations*, and *Academic documents*. *Newsletters* and *Advertisements* each represent less than 1% of the total.

Table 8:   Statistics for the genre distribution in the 1.7m Lusoga corpus

| Genre	No. of files	%	Tokens	%
Biblical documents	44	11.25%	388 026	22.51%
Literature	64	16.37%	271 701	15.76%
Radio talk shows	41	10.49%	265 726	15.42%
Journalism	2	0.51%	201 664	11.70%
E-mails	18	4.60%	153 563	8.91%
Policy documents	19	4.86%	101 029	5.86%
Interviews	11	2.81%	94 693	5.49%
Songs	155	39.64%	86 028	4.99%
Celebrations	10	2.56%	82 138	4.77%
Academic documents	19	4.86%	68 662	3.98%
Newsletters	6	1.53%	10 027	0.58%
Advertisements	2	0.51%	434	0.03%
SUM	**391**	**100.00%**	**1 723 691**	**100.00%**

Genres in the 1.7m Lusoga corpus
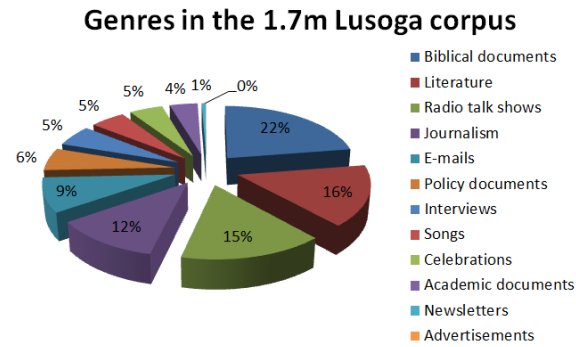
Figure 3:   Pie chart showing the genre distribution in the 1.7m Lusoga corpus

3.5.5   Topic distribution

The different files in the Lusoga corpus were also grouped into 18 broadly-defined topics. To do so, related subjects were brought together, such as:

Health:
— Health planning
— Ill-health & death
— Rural health management
— Traditional healing
— AIDS scourge
— Eye-care education ...

Inspirational:
— Self-appreciation
— Jubilation
— Honouring activity
— Hope message
— Graduation ceremony ...

Even though a strict division between genre and topic is not always possible, and even though some files actually deal with various topics, the data shown in Table 9 may be considered to be a good approximation of the actual topics covered in the corpus.

While a quarter of the Lusoga corpus deals with *Religion*, the inverse also means that three-quarters does not, which is fine given the usual bias in Bantu-language corpora. The topic *Networking* actually covers such varied items as newspaper texts, mailing-list messages, songs about networking, and even advertisements. The other topic labels are self-explanatory. The data is shown graphically in Figure 4.

Table 9:   Statistics for the topic distribution in the 1.7m Lusoga corpus

Topic	No. of files	%	Tokens	%
Religion	55	14.07%	439 915	25.52%
Networking	22	5.63%	355 761	20.64%
Health	41	10.49%	153 588	8.91%
Language[9]	26	6.65%	126 449	7.34%
Sensitization	35	8.95%	102 061	5.92%
Politics	16	4.09%	97 785	5.67%
Fables	41	10.49%	92 470	5.36%
Marriage	36	9.21%	79 839	4.63%
Life	19	4.86%	75 237	4.36%
History	12	3.07%	59 739	3.47%
Proverbs	5	1.28%	45 556	2.64%
Inspirational	13	3.32%	31 443	1.82%
Science	5	1.28%	19 784	1.15%
Riddles	3	0.77%	15 694	0.91%
Relationships	31	7.93%	12 203	0.71%
Rehabilitation	13	3.32%	7 400	0.43%
Money	13	3.32%	6 469	0.38%
Gratitude	5	1.28%	2 298	0.13%
SUM	**391**	**100.00%**	**1 723 691**	**100.00%**

While the percentages for each of the broadly-defined topics as seen in Figure 4 may or may not reflect the actual allocation to each of these topics in the way Lusoga is used by millions of speakers on a daily basis in Busoga, what is relatively certain is that the coverage of the range and variation is rather wide in the 1.7m Lusoga corpus.
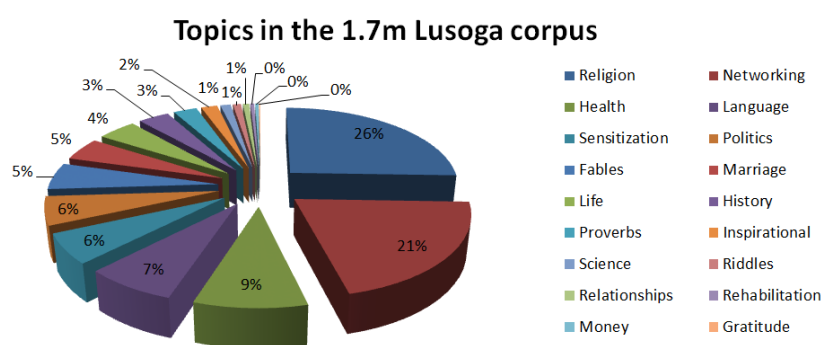
Figure 4:   Pie chart showing the topic distribution in the 1.7m Lusoga corpus

3.5.6 The orthography in the corpus

Important to observe at this point is that the various orthographies as seen in the original written sources were left intact. Bar a few exceptions, there are no tone markings in the corpus.

This implies that the stated number of types (i.e. the orthographically unique words) is always slightly inflated compared to a corpus in which the spelling would have been homogenised. Working with a corpus that contains various spellings for some of the same words is not an insurmountable hurdle; it only means that one is dealing with some (evenly spread) noise as far as the type counts are concerned; the token counts, however, are (mostly) correct.

Although a number of Lusoga orthography guides exist, one must conclude that they did not have much impact on helping the different authors streamline their writing in Lusoga. But then, the majority of the texts which are now in the corpus were not necessarily meant for formal usage, so their authors did not adhere to a strict application of any orthographic rules. For example, biblical prayer books are in-house documents that are only employed for the purposes of religious teaching. The different short stories and the novel in Lusoga have all been produced informally and are written in a style that the authors feel is most appropriate at the time of writing. E-mails and website texts in Lusoga display a severely unregulated use of written Lusoga. Also, the type of written Lusoga found in this category of sources is often mixed with English. In addition, Lusoga is borrowing sounds from neighbouring languages, such as the palatal nasal [ɲ] which is not an indigenous Lusoga sound. One also notices a switch between the voiced labio-velar approximant [w] and the velar fricative [ɣ]; and the fact that the Lusoga dental sounds are being relegated to neighbouring alveolar sounds (which are easier to pronounce for non-native speakers). Most prominent is an ongoing discussion on whether Lusoga really has a trill [r], only a flap [ɾ], or neither of the two — which results in inconsistent uses of /r/ and /l/ in the orthography.[10]

Instances of orthography-based problems in writing Lusoga are shown in examples (1) & (2). (For the abbreviations in the glosses, see the explanations at the end.)

(1) *enhyandhula*   instead of *ennhandhula*   'introduction'
     okuhwunga   instead of *okuwunga*   'to catch an object mid-air'
     cyatulirwa   instead of *kyatulilwa*   'it is pronounced/spoken'
     [File ID: KiyinKbi | W • Literature • Language • 1969]

(2) a. *Me Enterprise development oyinza okufuna bakakensa [...]*
     me   **...**   **o-yinza**   **oku-fun-a**   **ba-kakensa**
     CON   ...   SM2SG-can   15-get-FV   2-expert
     But for Enterprise development you can get experts [...]
     [File ID: Mail1306 | W • E-mails • Networking • 2013]

b. *What did I ng'omuntu do?*
    ...   nga   o-mu-ntu   ...
    ...   adv   AUG-1-person   ...
    What did I as a person do?
    [File ID: Mail1306 | W • E-mails • Networking • 2013]

In the examples in (1) the author decided to write the dental nasal as /nhy/, the voiced labio-velar approximant as /hw/, and the voiceless palatal plosive as /c/, as well as making distinctions in writing the trill after /i, e/ and the lateral flap after /u/ and /a/. The orthographic problems seen in examples of this nature seem to arise out of a need to use a phonetic-inspired orthography. Such orthographic interpretations may simply be idiosyncratic improvisations made in the absence of a proper (and popular) phonetic description of the sounds of Lusoga.

On the other hand, the examples in (2) reflect a user who is continuously code switching, and missing out on a few basic grammatical forms in the writing system. This is probably due to ignorance or the lack of a proper grounding in writing Lusoga.

The type of issues seen in the two examples can be generalised as occurring rather often in the informal written texts included in the corpus. While the spelling of the original texts was left intact, recognition errors might have been introduced during the OCR process, with some of the letters being machine unreadable and interpreted differently, even though we did our utmost to read through the OCRed material.

It is also probable that some 'errors' were introduced during the transcription process: while we tried to steer away from it, there was a tendency to 'over-correct' misspoken sections and hesitations, as the goal of our corpus-building efforts is not to use the material for, say, sociolinguistic studies of detailed turn-taking, but to use the material to uncover language as it was meant to be (Hanks 2012: 416).[11]

We do trust that these 'inconsistencies' and 'errors' have not obscured the proper usages of Lusoga.

3.5.7 Querying the corpus

The 391 files of the 1.7m Lusoga corpus are stored as *plain text files*, and as such this 1.7m Lusoga corpus is also a 'raw corpus'. Raw corpora may successfully be searched using off-the-shelf corpus-query software like WordSmith Tools (Scott 1996–2018). WST was indeed used in this way to present the various corpus counts above, and will also be used for the macrostructural and microstructural illustrations in the next two parts of this set of three articles.

However, and as we will explain in Part 2, the 1.7m Lusoga corpus was also part-of-speech tagged and lemmatised for lexicographic purposes. Either or even both of these levels (i.e., the part-of-speech labels and/or the lemmas of

each orthographic word) may also be added as tags to all (or part of the) 1.7m tokens of the Lusoga corpus. Software such as WST is able to handle such *marked-up text files* as well.

3.5.8   Corpus file IDs, corpus filename bibliography and corpus metadata database

As could be seen in examples (1) and (2), for material excerpted from the corpus, it is good practice to mention the source from which it was taken. In (1) this information was presented following all the examples, and in (2) this was done on the line following the interlinear glossing and translation. In all cases, the corpus details are presented between square brackets.

In actual fact, for all material that is quoted from a corpus, whether for lexicographic purposes or more generally in corpus linguistics, three distinct levels of supplementary information may be provided for each source. At the quoted material itself a *File ID* may be provided, together with 'minimal information', here on whether the treated example is either taken from the written or the oral section of the corpus, and further information on the genre and topic, as well as the year or period, in the following format:

at examples

[File ID: Filename	W(ritten) *or* O(ral) • Genre • Topic • Year *or* Period]

The *Filename* also serves as the entry point to Addendum 1, where further details on each source may be found. The author (or for audio, performer) as well as the title of the work (either as published or as given by us), the number of types and tokens for the work, the source of the work, the place of publication and publisher, as well as the number of pages of the work (or for audio, length of the recording) are all provided in that addendum. The format used for the twelve slots of information in Addendum 1 is always as follows:

in Addendum 1

Filename Author *or* Performer Year *or* Period (Title) • Genre / Topic • Tokens / Types • W(ritten) *or* O(ral) ~ Source • Place: Publisher • Pages *or* Length of recording

For instance, the *Filename* for (1) above reveals the following in Addendum 1:

KiyinKbi   Lyavala-Lwanga, E.J. 1969 (Kiyini Kibi) • Literature / Language • 19,256 / 7,737 • W ~ Retyping of image • Kampala: Milton Obote Foundation • 123 pp.

This type of information includes what one would find in a traditional bibliography (before the first bullet, after the penultimate and last bullets), but adds corpus-specific information to that (all the rest in-between).

in the corpus metadata database

Addendum 1 is an extract from a larger database, which, for the written sources and when relevant, also includes the translator and date of translation, as well as the edition number and year of original publication. For the oral material, that database additionally includes the date of the recording, and the names of the recorders and transcribers. Lastly, for each source the standardised type-token ratio (with a base of 1 000) and the standard deviation thereof are also given.[12]

A notes field is used for any additional information that needs to be mentioned. This *corpus metadata database*, which brings together all the metadata of the corpus in a structured format, is available electronically and may be consulted at BantUGent together with the corpus itself.

3.5.9   Original data database

While it is feasible to store all of the 391 files in one single folder, much more intelligent is to arrange the files into various folders and sub-folders, for instance reflecting the different genres (12 sub-folders) or topics (18 sub-folders). How this is organised for a particular corpus depends on the use that will be made of that corpus. Another division could be oral vs. written, or the use of sub-folders that reflect the different time periods in the corpus, or even combinations of all of the above using tiered sub-folders. What has furthermore proven to be very useful is to keep several copies of the corpus at hand: in each, one finds the same data, but structured differently.

What is of paramount importance, however, is to keep a parallel version of one of these corpus structures in a different (off-site) location, where all the *original files* are kept. There the original audio (.wav, .mp3, ...) and at times even videos (.mp4, .webm, ...) are stored, as well as the original web pages (.htm, .html, ...), documents (.doc, .pdf, ...) and images (.jpeg, .png, ...). Temporary files such as those used to turn scanned material ('image pdfs') with OCR software (e.g., .opd) into machine-readable images ('searchable pdfs') should also be kept there. This parallel version of the corpus, or *original data database*, not only functions as a backup from which the corpus files could be regenerated whenever this would prove to be necessary, but it is also the first place to go to whenever in doubt about a certain transcription (audio) or the orthography in an automatically-recognised (written) work. Published texts, with their formatting, and multimedia files furthermore contain more information than the text (.txt) versions in the corpus, which may at times and for certain purposes be useful to consult.

4. Discussion

In this article we have given a detailed description of the building of a general-language corpus for Lusoga, an under-resourced Bantu language. We showed that it is indeed possible to reach a substantial size, in this case 1.7 million tokens, a third of which consists of oral data, even though the building of this corpus has basically been a one- to two-person effort. This stands in sharp contrast to for instance the ALLEX/ALRI corpora, for which scores of students were sent into the field and as many were enlisted to transcribe the recordings.

Our corpus is an 'organic corpus', as material has not only been added over the years, but some of it has also been taken away, while still other parts were replaced after being reworked. Merely having more data does not necessarily mean one has better data, as one should keep an eye on balance as well. In the overview presented in the present article, the 1.7m Lusoga corpus is a 'raw corpus', in that it has not been annotated; but it was pointed out that with the results from Part 2, part-of-speech tags and/or lemma tags could enrich this corpus linguistically.

We also illustrated the importance of *knowing* one's corpus, not only in terms of the oral vs. written distribution, but similarly with regard to the distribution of the sources, periods, genres, and topics. Variations on our presentation are of course possible, and indeed in the PhDs of Mberamihigo (2014), Nshemezimana (2016) and Misago (2018) for Kirundi, as well as the PhD of Kawalya (2017) for Luganda, three-dimensional graphs are shown in addition, the third dimension representing the diachronic aspects of their corpora. The point, however, is that a detailed description of a corpus is needed if one is to make intelligent use of it.

As the details in the addendum indicate, we further place particular importance on the metadata of a corpus. Metadata may evidently be put to good use when actually *using* a corpus: for lexicographic ends, but also far beyond in the wider discipline of linguistics. There are no doubt differences between the spoken and the written forms of a language, and certain phenomena may be realised slightly differently depending on the genre or topic, just as word use differs with register. Likewise, for differences in word use depending on the author or performer, or even the publishing house of a certain work (each with their own style guide and own approach to copy-editing), and so on. Sub-corpora may indeed be assembled along such lines.

Reformulated, depending on how one intends to use a corpus, all the categorisations given so far may play an important role. But they do not inform each study in the same way. Within the field of lexicography, the first two and main uses of a corpus have to do with the creation of the macrostructure of a dictionary on the one hand, and the compilation of the articles in the microstructure on the other. These two topics will now be looked into, and illustrated for Lusoga, in two follow-up studies.

Abbreviations

#	noun class number	FV	final vowel
ADV	adverb	SG	singular
AUG	augment	SM_x	subject marker (of cl. or
cl.	class		person x)
CON	connective		

Acknowledgements

Endnotes

1. In our work Lusoga, as in all subsequent mentions of 'Lusoga corpus', narrowly refers to the Lutenga variety only (Nabirye et al. 2016).
2. At the CRC library in Jinja, a substantial amount of grey literature may also be found, either written by the CRC staff itself, or facilitated by them. These works are mostly for internal use, of a religious nature and typically do not have a stated publisher, but may be 'assigned' to the CRC (e.g., CRC 1998b, Kasozi 2000, CRC 2003a, b, c, 2005b, 2008, Wabugoyera et al. 2008, CRC 2010, 2012a, b, c, d, e, f, g). Other religious works often do not have publication years, such as Mwesigwa (s.d.), except for those published by The Bible Society of Uganda, for which, see Endnote 8. Lately, the CRC has begun rejacketing earlier works, including CRC (2009) and Kaluuba and Korse (2010). The CRC also played a pioneering role in producing the first grammars for Lusoga (Korse 1999, CRC 2004, Wambi et al. 2005, Kuunya 2011b), the first bilingual Lusoga–English dictionaries (Korse 2000, Gonza 2007), new orthographies (LULANDA and CRC 2001, 2004), as well as readers (e.g., Gulere and Wambi 2011).
3. Gulere also compiled a bilingual Lusoga–English dictionary (Gulere 2009).
4. At BantUGent a diachronic corpus for Swahili with a time-depth of up to two centuries is under construction. Research articles have not yet been published, however, although preliminary results have been presented at conferences (Devos and de Schryver 2013, 2016).
5. While not magic, Rundell and Stock (1992: 46) refer to this part of a corpus as the 'Holy Grail': 'Truly spontaneous speech, however — the everyday conversation of ordinary members of the public — has so far been available only in very small quantities and for lexicographers this remains the "Holy Grail".'
6. In earlier descriptions of corpus building for the Bantu languages, some attention was paid to the type of OCR errors one needs to attend to (de Schryver 1999: 116). Today's OCR software is however so performant that all one needs to remember is that the letter combination read as 'rn' should often be corrected to the single letter 'm'.
7. Observe that material for the 1980s was found after all, in an academic publication (Cohen 1986), following a memorable search (Nabirye 2016: 25-27). Although eventually published in 1986, this edited material is based on recordings made two decades earlier, in 1966–1967.

8. A late entrant — in the sense that it came too late to be added to the 1.7m Lusoga corpus (apart from the fact that it may not have been desirable for reasons of representativeness and balance) — is the full Bible in Lusoga, which became available in 2014 (BSU 2014). As is normally the case with biblical works, the full Bible (BSU 2014) incorporates the New Testament (BSU 1998) — published earlier and included in the 1.7m Lusoga corpus. The New Testament itself incorporated the even earlier Gospel of Mark (BSU 1996), which in turn incorporated the still earlier Chapters 4 and 5 of the same gospel (BSU 1994). After the New Testament was released, at least one other edition appeared, with the addition of the Psalms from the Old Testament (BSU 2011).

9. The topic *Language* mainly includes material about teaching the language of Lusoga and instructional material for Lusoga (written in Lusoga), as well as website texts and journal abstracts on Lusoga (written in Lusoga).

10. See Nabirye et al. (2016) for more on these phonetic issues.

11. Or, as Kennedy (1998: 82) writes: 'A transcription is an imperfect written approximation of a speech event which exists initially as a dance of air molecules. The level of delicacy or amount of detail in a transcription is [...] related to the use to which the transcription will be put'.

12. As defined by Scott (1996–2018) 'the *standardised type/token ratio* (STTR) is computed every **n** words as Wordlist goes through each text file. By default, n = 1,000. In other words the ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of your text or corpus. A running average is computed, which means that you get an average type/token ratio based on consecutive 1,000-word chunks of text. (Texts with less than 1,000 words (or whatever n is set to) will get a standardised type/token ratio of 0.)'.

References

Adelaar, W.F.H. 2014. Endangered Languages with Millions of Speakers: Focus on Quechua in Peru. *JournaLIPP* 3: 1-12.

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

BNC. 1994–2018. British National Corpus. Available online at: http://www.natcorp.ox.ac.uk/.

Bostoen, K. and G.-M. de Schryver. 2015. Linguistic Innovation, Political Centralization and Economic Integration in the Kongo Kingdom: Reconstructing the Spread of Prefix Reduction. *Diachronica* 32(2): 139-185 + 13 pages of supplementary material online.

BSU. 1994. *Mariko Omutwe Ogwokuna n'Ogwokutaanu mu Lusoga [A Selection from St. Mark's Gospel Chapters 4 and 5 in Lusoga].* Kampala: The Bible Society of Uganda.

BSU. 1996. *Mariko. Amawulire Amalungi mu Lusoga [The Gospel of Mark in Lusoga].* Kampala: The Bible Society of Uganda.

BSU. 1998. *Endagaano Empyaka [New Testament].* Kampala: The Bible Society of Uganda.

BSU. 2011. *Endagaano Empyaka ni Zabbuli [New Testament and Psalms].* Kampala: The Bible Society of Uganda.

BSU. 2014. *Baibuli. Ekibono kya Katonda. Omuli n'ebitabo ebyetebwa deuterokanoniko/apokurifa [Bible. The Word of God, Which Also has the Books Known as Deuteronomy/Apocrypha].* Kampala: The Bible Society of Uganda.

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

Chabata, E. 2000. The Shona Corpus and the Problem of Tagging. *Lexikos* 10: 75-85.

Cohen, D.W. 1986. *Towards a Reconstructed Past: Historical Texts from Busoga, Uganda* (Union Académique Internationale, Fontes Historiae Africanae Series Varia III). Oxford: Oxford University Press (for the British Academy).

Condon, M.A. 1911. Contribution to the Ethnography of the Basoga-Batamba Uganda Protectorate, Br. E. Africa. Part 2. *Anthropos: International Review of Ethnology and Linguistics* 6(2): 366-384.

CRC. 1998a. *Kintu*. Jinja: Cultural Research Center.

CRC. 1998b. *Priestly Ordination of Rev. Richard Kayaga Gonza, Rev. Silvester Makwali*. Bugembe: Diocese of Jinja.

CRC. 1999a. *Ababita Ababiri*. Jinja: Cultural Research Centre.

CRC. 1999b. *Akatabo Akasooka ak'Enfumo edh'Abasoga*. Jinja: Cultural Research Center.

CRC. 1999c. *Amagezi Tigamalwayo*. Jinja: Cultural Research Center.

CRC. 1999d. *Ensambo edh'Abasoga*. Jinja: Cultural Research Center.

CRC. 1999e. *Mwidhe Tufume*. Jinja: Cultural Research Center.

CRC. 1999f. *Obufunvu Magezi*. Jinja: Cultural Research Centre.

CRC. 1999g. *Omuvangano mu Busoga*. Jinja: Cultural Research Centre.

CRC. 1999h. *Twire ku Butaka*. Jinja: Cultural Research Centre.

CRC. 2000a. *Enkabi Ekifiini mu Busoga*. Jinja: Cultural Research Centre.

CRC. 2000b. *Lwaki Abakazi Tibabeeda Mulambo*. Jinja: Cultural Research Centre.

CRC. 2002. *Ebikoiko eby'Abasoga*. Jinja: Cultural Research Centre.

CRC. 2003a. *Diocesan Family Day*. Iganga: Diocesan Printery.

CRC. 2003b. *The Priestly Ordination for Rev. Deacon Mbaziira Henry Jude, Rev. Deacon Musana Paul*. Bugembe: Diocesan Printery.

CRC. 2003c. *Priestly Ordination of Rev. Serapio Kasuura Wamara Araali*. Bugembe: Diocese of Jinja.

CRC. 2004. *A Lusoga Grammar* (Revision of Korse 1999). Jinja: Cultural Research Centre.

CRC. 2005a. *Ebindi kw'Idembe ery'Obw'omuntu mu Nsi Yoonayoona*. Kisubi: Marianum Publishing Company.

CRC. 2005b. *Priestly Ordination of Deacon Mwangi Simon Gitua, Deacon Mugabe Paschal Atwooki, Deacon Jenga Fred*. Jinja: Little Sisters of St. Francis.

CRC. 2008. *Enhembo mu Mikolo Emitukuvu Egyo Busaserdooti, Obudyakoni ne Miruka e Kyebando Parish nga 02-08-2008*. Kyebando.

CRC. 2009. *Ensambo dh'Abasoga (Kisoga Proverbs)* (Revision of CRC 1999). Kisubi: Marianum Publishing Company.

CRC. 2010. *Installation of Rt. Rev. Bishop Charles Martin Wamika as Bishop of Diocese of Jinja*. Kisubi: Marianum Publishing Company.

CRC. 2011. *Ensengeka y'Omusomo gw'Ekikulu (Curriculum for Functional Adult Learners in Lusoga)*. Kisubi: Marianum Publishing Company.

CRC. 2012a. *Basoga Catholics in and Around Kampala*. Nsambya: Diocese of Jinja.

CRC. 2012b. *Missa mu Lusoga Ebiseera eby'Amatuuka n'Amazaalibwa*. Jinja: Diocese of Jinja.

CRC. 2012c. *Missa mu Lusoga Ebiseera eby'Amazuukira*. Jinja: Diocese of Jinja.

CRC. 2012d. *Missa mu Lusoga Ebiseera eby'Ekisiibo*. Jinja: Diocese of Jinja.

CRC. 2012e. *Missa mu Lusoga Ebiseera eby'Omwaka n'Enaku edh'Abatuukirivuu*. Jinja: Diocese of Jinja.

CRC. 2012f. *Missa mu Lusoga Ensengeka y'Emikolo gya Wiiki Entukuvu*. Jinja: Diocese of Jinja.

CRC. 2012g. *Thanks Giving Mass. for Rev. Sr. Restetuta Wangoye*. Jinja: Diocese of Jinja.

de Schryver, G.-M. 1999. *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: Ghent University.

de Schryver, G.-M. 2002. Web for/as Corpus: A Perspective for the African Languages. *Nordic Journal of African Studies* 11(2): 266-282.

de Schryver, G.-M. and R. Gauton. 2002. The Zulu Locative Prefix ku- Revisited: A Corpus-based Approach. *Southern African Linguistics and Applied Language Studies* 20(4): 201-220.

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.

de Schryver, G.-M. and D.J. Prinsloo. 2000. The Compilation of Electronic Corpora, With Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.

Devos, M. and G.-M. de Schryver. 2013. From 'habitually going' to 'maybe': Grammaticalization and Lexicalization of an Epistemic Sentence Adverb in Swahili. Anon. (Ed.). 2013. *Abstracts of The 21st International Conference on Historical Linguistics*: 29. Oslo: University of Oslo.

Devos, M. and G.-M. de Schryver. 2016. From Usually Going to Epistemic Possibility. Origin and Development of an Epistemic Sentence Adverb in Swahili. Anon. (Ed.). 2016. *6th International Conference on Bantu Languages, Workshop on the Expression of Mood and Modality in Bantu Languages*: 11. Helsinki: University of Helsinki.

Facebook. 2004–18. Facebook Online Social Media and Social Networking Service. Available online at: https://www.facebook.com.

Gonza, R.K. 2007. *Lusoga–English Dictionary and English–Lusoga Dictionary* (revision of Korse 2000). Kampala: MK Publishers.

Gulere, C.W. 2007a. *Nantamegwa* (translation into Lusoga of *Antigone*, a play by Sophocles from 441 BC). Busembatya: Lusoga Language Academic Board.

Gulere, C.W. 2007b. *Omugole* (translation into Lusoga of *The Bride*, a play by Bukenya from 1987). Busembatya: Lusoga Language Academic Board.

Gulere, C.W. 2009. *Lusoga–English Dictionary / Eibwanio*. Kampala: Fountain Publishers.

Gulere, C.W. 2011a. *Abasikawutu*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011b. *Amagelo mu Nsiko*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011c. *Ebikete bya Busoga*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011d. *Ekidhuubo kya Giligoori*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011e. *Engabo ya Busoga*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011f. *Lusoga Nguli Namanha*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011g. *Ndi ni Mukazi Wange*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011h. *Nsobola Nsobola*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011i. *Ogusolo ni Ekikaadho*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. 2011j. *Okusanhusa Tikwesanhusa*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Gulere, C.W. and M. Wambi. 2011. *Lusoga Olusonhe*. Busembatia: Association of Lusoga Language Educationists, Researchers and Translators.

Guthrie, M. 1948. *The Classification of the Bantu Languages* (Handbook of African Languages). London: Oxford University Press (for the International African Institute).

Hadebe, S. 2002. The Ndebele Language Corpus: A Review of Some Factors Influencing the Content of the Corpus. *Lexikos* 12: 159-170.

Hanks, P. 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.

Hurskainen, A. 2016. Helsinki Corpus of Swahili 2.0 (HCS 2.0) Annotated Version. Available online at: http://urn.fi/urn:nbn:fi:lb-2016011301.

Johnston, H.H. 1902. *The Uganda Protectorate. An Attempt to Give Some Description of the Physical Geography, Botany, Zoology, Anthropology, Languages and History of the Territories under British Protection in East Central Africa, between the Congo Free State and the Rift Valley and between the First Degree of South Latitude and the Fifth Degree of North Latitude*. London: Hutchinson & Co.

Kajolya, J.B.N. 1990. *The Lusoga Orthography* (revision of Byandala 1963). Jinja: Lusoga Ecumenical Committee.

Kaluuba, J.P., J. Kivuunike, J.S. Dhizaala and C. Nabirye. 2010. *Gw'Olekera Abato: Okusoma kuleeta obusobozi*. Jinja: Cultural Research Centre, Marianum Publishing Company and Literacy and Adult Basic Education.

Kaluuba, J.P. and P. Korse. 2010. *Kintu* (revision of CRC 1998). Jinja: Cultural Research Centre.

Kasozi, J. 2000. *Mwidhe Tugye Tusenge*. Jinja: Diocese of Jinja.

Kawalya, D. 2017. *A Corpus-driven Study of the Expression of Modality in Luganda (Bantu, JE15)*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Kawalya, D., G.-M. de Schryver and K. Bostoen. 2018. From Conditionality to Modality in Luganda (Bantu, JE15): A Synchronic and Diachronic Corpus Analysis of the Verbal Prefix *-andi-*. *Journal of Pragmatics* 127: 84-106.

Kennedy, G. 1998. *An Introduction to Corpus Linguistics* (Studies in Language and Linguistics). Harlow, Essex: Addison Wesley Longman.

Kodh'eyo. 1997–98. *Kodh'eyo: Busoga etebenkere* (a short-lived newspaper in Lusoga). Kampala: Kodh'eyo Publications.

Korse, P. 1999. *A Lusoga Grammar*. Jinja: Cultural Research Centre.

Korse, P. 2000. *Dictionary Lusoga–English / English–Lusoga*. Jinja: Cultural Research Centre.

Kuunya, C. 2011a. *Agakuba Omughafu*. Kisubi: Marianum Publishing Company.

Kuunya, C. 2011b. *Gulaama ey'Olusoga [Grammar of Lusoga]*. Kisubi: Marianum Publishing Company.

LULANDA and CRC. 2001. *Empandiika ey'Olulimi Olusoga Enkalamu [Standard Lusoga Orthography]*. Jinja: Lusoga Language Development Academy & Cultural Research Centre.

LULANDA and CRC. 2004. *Empandiika ey'Olulimi Olusoga Enkalamu [Standard Lusoga Orthography]* (2nd edition). Jinja: Lusoga Language Development Academy & Cultural Research Centre.

Lyavala-Lwanga, E.J. 1967. *Endheso dh'Abasoga [Proverbs of the Basoga]*. Kampala: Milton Obote Foundation.

Lyavala-Lwanga, E.J. 1969. *Kiyini Kibi [on language and Lusoga riddles]*. Kampala: Milton Obote Foundation.

Maho, J.F. 2009. NUGL Online: The Online version of the New Updated Guthrie List, a Referential Classification of the Bantu Languages. Available online at: http://goto.glocalnet.net/ mahopapers/nuglonline.pdf.

Mberamihigo, F. 2014. *L'expression de la modalité en kirundi : Exploitation d'un corpus électronique*. Unpublished Ph.D. dissertation. Brussels; Ghent: Université libre de Bruxelles; Ghent University.

Mberamihigo, F., G.-M. de Schryver and K. Bostoen. 2016. Entre verbe et adverbe : Grammaticalisation et dégrammaticalisation du marqueur épistémique *umeengo/umeenga* en kirundi (bantou, JD62). *Journal of African Languages and Linguistics* 37(2): 247-286.

Misago, M.-J. 2018. *Les verbes de mouvement et l'expression du lieu en kirundi (bantou, JD62) : Une étude linguistique basée sur un corpus*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Mwesigwa, R. s.d. *Mu Bigere Bye. Olugero lw'Abasoga ku kugoberera Yesu*. Jinja: Church of Christ.

Nabirye, M. 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished M.A. 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. 2016. *A Corpus-based Grammar of Lusoga*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Nabirye, M. and G.-M. de Schryver. 2011. From Corpus to Dictionary: A Hybrid Prescriptive, Descriptive and Proscriptive Undertaking. *Lexikos* 21: 120-143.

Nabirye, M., G.-M. de Schryver and J. Verhoeven. 2016. Illustrations of the IPA: Lusoga (Lutenga). *Journal of the International Phonetic Association* 46(2): 219-228 (+ supplementary audio online).

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

Ndimugezi. 1998–99. *Ndimugezi n'omukobere: The Factfinder* (a short-lived newspaper, with sections in Lusoga). Jinja: Ndimugezi Publications.

Nshemezimana, E. 2016. *Morphosyntaxe et structure informationnelle en kirundi : Focus et stratégies de focalisation*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

OmniPage. 1995–2018. Optical character recognition (OCR) software now available from Nuance Communications. Available online at: http://www.nuance.com/for-individuals/by-product/ omnipage/.

Prinsloo, D.J. 2015. Corpus-based Lexicography for Lesser-resourced Languages — Maximizing the Limited Corpus. *Lexikos* 25: 285-300.

Prinsloo, D.J. and G.-M. de Schryver. 2001. Monitoring the Stability of a Growing Organic Corpus, with Special Reference to Sepedi and Xitsonga. *Dictionaries: Journal of The Dictionary Society of North America* 22: 85-129.

Prinsloo, D.J. and G.-M. de Schryver. 2005. Managing Eleven Parallel Corpora and the Extraction of Data in All Official South African Languages. Daelemans, W., T. du Plessis, C. Snyman and L. Teck (Eds). 2005. *Multilingualism and Electronic Language Management. Proceedings of the 4th International MIDP Colloquium, 22–23 September 2003, Bloemfontein, South Africa* (Studies in Language Policy in South Africa 4): 100-122. Pretoria: Van Schaik Publishers.

Rundell, M. and P. Stock. 1992. The Corpus Revolution 3. A Consideration of the Prospects and Potential of Corpus-and-concordance Lexicography (third article of three). *English Today, The International Review of the English Language* 8(4): 45-51.

Scott, M. 1996–2018. WordSmith Tools. Available online at: http://www.lexically.net/wordsmith/.

Sene-Mongaba, B. 2013. *Le lingála dans l'enseignement des sciences dans les écoles de Kinshasa : Une approche socioterminologique*. Unpublished Ph.D. dissertation. Ghent: Ghent University.

Sene-Mongaba, B. 2015. The Making of Lingala Corpus: An Under-resourced Language and the Internet. *Procedia — Social and Behavioral Sciences* 198: 442-450.

UBOS. 2006. *The 2002 Uganda Population and Housing Census, Analytical Report, Population Composition*. Kampala: Uganda Bureau of Statistics.

UBOS. 2016. *The National Population and Housing Census 2014 — Main Report*. Kampala: Uganda Bureau of Statistics.

Wabugoyera, J.B. Kasubi, J.P. Kaluuba, Mukama, M. Maganda and M. Maganda. 2008. *Ekimuliikirira, August–October 2008*. Jinja: Diocese of Jinja.

Wambi, M., R. Naigaga and CRC. 2005. *Idha Tusome [Come and We Read]*. Jinja: Lusoga Language Authority.

YouTube. 2005–18. YouTube video-sharing website. Available online at: https://www.youtube. com.

Addendum 1:   Corpus filename bibliography for the 391 sources in the 1.7m Lusoga corpus

`Filename`	**Author** *or* **Performer Year** *or* **Period (Title)** ● **Genre / Topic** ● **Tokens / Types** ● **W(ritten)** *or* **O(ral) ~ Source** ● **Place: Publisher** ● **Pages** *or* **Length of recording**
`1Bakyaba`	Ambassador Institute 2012 (Eiterekero: Eriya ku lusozi kalameri) ● Biblical documents / Religion ● 741 / 380 ● W ~ e-Transfer ● Internet: Ambassador Institute ● 2 pp.
`Ababala`	Malagala, Stephen 2000s (Ababala Emilimu) ● Songs - Traditional / Sensitization ● 660 / 434 ● O ~ Transcription ● Anon.: - ● 0:08:08
`AbabitAb`	Ssajabi, Sophronius 1999 (Ababita Ababiri) ● Literature / Fables ● 5,063 / 1,825 ● W ~ OCR ● Jinja: CRC ● 38 pp.
`Abadhel`	Malagala, Stephen 2000s (Abadhelega Emilimu) ● Songs - Traditional / Sensitization ● 551 / 364 ● O ~ Transcription ● Anon.: - ● 0:06:34
`Abalamu`	Baisi 2010s (Abalamu Tusaanila Tukole) ● Songs - Traditional / Sensitization ● 633 / 308 ● O ~ Transcription ● Anon.: - ● 0:08:49
`ABamBamu`	Mata, Nassani & Isiko 1990s (Bamusabire) ● Songs - Traditional / Money ● 441 / 153 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:13:00
`ABamEita`	Mata, Nassani & Isiko 1990s (Eitaka) ● Songs - Traditional / Rehabilitation ● 252 / 140 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:07:15
`ABamKate`	Mata, Nassani & Isiko 1990s (Katengeke) ● Songs - Traditional / Relationships ● 331 / 161 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:10:14
`Abantub`	Mugwisa 2000s (Abantu Beebisa Bulala) ● Songs - Traditional / Rehabilitation ● 609 / 359 ● O ~ Transcription ● Anon.: - ● 0:08:57
`Abasikaw`	Gulere, Cornelius 2011 (Abasikawutu) ● Literature / Fables ● 1,885 / 851 ● W ~ OCR ● Internet: Google books ● 19 pp.
`Abasoga`	Salimu 2010 (Abasoga) ● Songs - Modern / Sensitization ● 793 / 238 ● O ~ Transcription ● Jinja: CRC ● 0:05:51
`Abatool`	Malagala, Stephen 2000s (Abatoolamu Embuto) ● Songs - Traditional / Health ● 697 / 391 ● O ~ Transcription ● Anon.: - ● 0:07:46
`Abatwes`	Malagala, Stephen 2000s (Abatwesimbamu) ● Songs - Traditional / Relationships ● 377 / 262 ● O ~ Transcription ● Anon.: - ● 0:04:35
`Abeelad`	Malagala, Stephen 2000s (Abeeladha ku Nsolo) ● Songs - Traditional / Rehabilitation ● 724 / 375 ● O ~ Transcription ● Anon.: - ● 0:07:11
`AEGY1`	Various 2010 (Aids Education Group for Youths 1) ● Radio talk shows / Health ● 3,668 / 1,380 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:45:16
`AEGY2`	Various 2010 (Aids Education Group for Youths 2) ● Radio talk shows / Health ● 4,979 / 1,679 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:59:26
`AEGY3`	Various 2010 (Aids Education Group for Youths 3) ● Radio talk shows / Health ● 6,126 / 2,152 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:31:15
`AEGY4`	Various 2010 (Aids Education Group for Youths 4) ● Radio talk shows / Health ● 1,679 / 801 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:13:52

`Filename`	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
`AgakbOmu`	Kuunya, Christopher 2012 (Agakuba Omughafu) ● Literature - Novels / Life ● 47,964 / 12,828 ● W ~ OCR ● Jinja: Marianum Publishing Company ● 310 pp.
`Akabend`	Mugwisa 2000s (Akabendhe) ● Songs - Traditional / Money ● 590 / 320 ● O ~ Transcription ● Anon.: - ● 0:09:07
`Akalango`	Orange 2009 (Akalango ka Orange) ● Advertisements / Networking ● 347 / 193 ● W ~ Translation ● Kampala: Orange mobile phone network in Uganda ● 1 p.
`Akaleed`	Malagala, Stephen 2000s (Akaleediyo) ● Songs - Traditional / Marriage ● 746 / 469 ● O ~ Transcription ● Anon.: - ● 0:07:07
`Akaleky`	Anon. 2010s (Akalelelo ka Kiyingi) ● Songs - Modern / Politics ● 501 / 96 ● O ~ Transcription ● Anon.: - ● 0:05:13
`akalelel`	Anon. 2010s (Leeta Akalelelo) ● Songs - Traditional / Money ● 677 / 116 ● O ~ Transcription ● Iganga: - ● 0:07:40
`AkatAkas`	Ssajabi, Sophronius 1999 (Akatabo Akasooka ak'Enfumo edh'Abasoga) ● Literature / Fables ● 5,831 / 2,107 ● W ~ OCR ● Jinja: CRC ● 38 pp.
`akatiko`	Geo Bless 2010s (Akatiko) ● Songs - Modern / Relationships ● 322 / 196 ● O ~ Transcription ● Iganga: - ● 0:03:59
`Akatook`	Anon. 2010s (Akatooke k'Endala) ● Songs - Traditional / Politics ● 1,075 / 143 ● O ~ Transcription ● Anon.: - ● 0:08:25
`Akeeyo`	Bamulumba, Yasiini 2012 (Akeeyo) ● Songs - Modern / Rehabilitation ● 642 / 317 ● O ~ Transcription ● internet: Intangible Culture Heritage Conservation Project ● 0:08:46
`ALwaLwak`	Kabugu, Jessica & Kabugu, Milton Peter 1990s (Lwaki Tosulanga y'Ogunhwa) ● Songs - Traditional / Rehabilitation ● 1,076 / 565 ● O ~ Transcription ● Jinja: Ali Mukembo and Sons ● 0:11:21
`ALwaOmwe`	Kabugu, Milton Peter 1990s (Omwenge Taabbu) ● Songs - Traditional / Rehabilitation ● 717 / 258 ● O ~ Transcription ● Jinja: Ali Mukembo and Sons ● 0:09:05
`ALwaSili`	Kabugu, Milton Peter 1990s (Siliimu) ● Songs - Traditional / Health ● 549 / 309 ● O ~ Transcription ● Jinja: Ali Mukembo and Sons ● 0:08:21
`Amaadhi`	Anon. 2010s (Amaadhi) ● Songs - Traditional / Relationships ● 515 / 274 ● O ~ Transcription ● Anon.: - ● 0:05:14
`AmagelM`	Gulere, Cornelius 2007 (Amagelo mu Nsiko) ● Literature / Fables ● 1,185 / 653 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 20 pp.
`Amagelom`	Gulere, Cornelius 2011 (Amagelo mu Nsiko) ● Literature / Fables ● 1,084 / 607 ● W ~ OCR ● Internet: Google books ● 18 pp.
`AMagEnfu`	Kabugu, Milton Peter 1990s (Enfuna y'Esente) ● Songs - Traditional / Money ● 503 / 265 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:08:26
`AMagInha`	Kabugu, Milton Peter 1990s (Inhazaala Ghange) ● Songs - Traditional / Marriage ● 426 / 248 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:10:56
`AMagObuf`	Kabugu, Milton Peter 1990s (Obufumbo) ● Songs - Traditional / Marriage ● 1,196 / 551 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:11:46

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
`AmagTigm`	Kyakulaga, Zion 1999 (Amagezi Tigamalwayo) ● Literature / Fables ● 4,198 / 1,735 ● W ~ OCR ● Jinja: CRC ● 38 pp.
`Amateeka`	Justice Law and Order Sector (JLOS) 2008 (Amateeka Agagema ku Kusiba Abantu mu Byalo) ● Policy documents - Government / Sensitization ● 89 / 61 ● W ~ Translation ● Kampala: JLOS ● 0.5 pp.
`Archbis`	Orombi, Henry Luke 2009 (Speech of the Archbishop of Uganda during his Visit to the Diocese of Jinja) ● Celebrations / Religion ● 4,574 / 1,786 ● O ~ Transcription ● Jinja: Church of Uganda ● 1:25:33
`Artbase`	Artbase 2010s (Artbase Anthem) ● Songs - Modern / Inspirational ● 322 / 132 ● O ~ Transcription ● Anon.: - ● 0:05:46
`Asaanak`	Ntende, Monika 2010 (Asaana Kwebaza) ● Songs - Gospel / Religion ● 404 / 169 ● O ~ Transcription ● Internet: YouTube ● 0:05:47
`Ateoba`	Kigenyi, Amos 2010 (Ate oba Wankyawa) ● Songs - Modern / Relationships ● 57 / 29 ● O ~ Transcription ● Internet: YouTube ● 0:04:15
`ATirEkya`	Mata, Nassani 1990s (Ekyanguza Empale) ● Songs - Traditional / Inspirational ● 192 / 128 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:05:57
`ATirKawo`	Mata, Nassani 1990s (Kawoiwolo) ● Songs - Traditional / Relationships ● 368 / 222 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:05:28
`ATirMump`	Mata, Nassani 1990s (Munpe Omwana) ● Songs - Traditional / Marriage ● 343 / 180 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:07:08
`ATirOmwo`	Mata, Nassani 1990s (Omwoyo Fiitina) ● Songs - Traditional / Relationships ● 171 / 102 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:02:15
`ATirSula`	Mata, Nassani 1990s (Sulaayi) ● Songs - Traditional / Relationships ● 436 / 307 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:06:09
`ATirTiil`	Mata, Nassani 1990s (Tiilime) ● Songs - Traditional / Money ● 249 / 144 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:04:50
`AVATVAT`	Kirimungu, Siragi 2000s (VAT) ● Songs - Traditional / Sensitization ● 553 / 251 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:08:30
`AVATVAT2`	Kirimungu, Siragi 2000s (VAT Vol. 2) ● Songs - Traditional / Sensitization ● 278 / 184 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:05:45
`Babalan`	Baisi 2010s (Babalanda) ● Songs - Traditional / Inspirational ● 870 / 296 ● O ~ Transcription ● Anon.: - ● 0:08:39
`Bakalakt`	Gulere, Cornelius 2006 (Bakalakatana) ● Literature / Fables ● 3,902 / 1,770 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 32 pp.
`Bakulim`	Magoola, Racheal 2010 (Bakulimba) ● Songs - Modern / Relationships ● 173 / 93 ● O ~ Transcription ● Internet: YouTube ● 0:05:00
`Bakyali`	Musooko 2010s (Bakyali) ● Songs - Modern / Relationships ● 701 / 396 ● O ~ Transcription ● Iganga: - ● 0:06:43
`Balocle`	Bujagaali's daughters 2006 (Ennhemba dh'Abalongo) ● Interviews / Health ● 310 / 185 ● O ~ Transcription ● Jinja: - ● 0:09:20

Filename	**Author** *or* **Performer Year** *or* **Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten)** *or* **O(ral) ~ Source ● Place: Publisher ● Pages** *or* **Length of recording**
`Balodis`	Bujagaali's client 2006 (Balongo Discussion) ● Interviews / Marriage ● 1,757 / 719 ● O ~ Transcription ● Jinja: - ● 0:33:24
`Bascath`	Cultural Research Centre (CRC) 2012 (Basoga Catholics in and Around Kampala) ● Biblical documents / Networking ● 426 / 203 ● W ~ OCR ● Nsambya: Diocese of Jinja ● 16 pp.
`Bbaabba`	Kigenyi, Amos 2010 (Bbaabba Toyombesa Maama) ● Songs - Modern / Marriage ● 311 / 114 ● O ~ Transcription ● Internet: YouTube ● 0:04:28
`BBamAkat`	Mata, Nassani & Isiko 1990s (Akatooke) ● Songs - Traditional / Rehabilitation ● 427 / 208 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:12:06
`BBamBali`	Mata, Nassani & Isiko 1990s (Balizanila) ● Songs - Traditional / Relationships ● 306 / 141 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:10:15
`BBamEndo`	Mata, Nassani & Isiko 1990s (Endoola) ● Songs - Traditional / Life ● 346 / 195 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:08:45
`BEL09-Q2`	Bujagali Hydropower Project (BHPP) 2009 (Bujagali Project Newsletter Q2 - July, 2009) ● Newsletters / Sensitization ● 1,645 / 784 ● W ~ e-Transfer ● Internet: Bujagali Energy Ltd ● 5 pp.
`BEL09-Q3`	Bujagali Hydropower Project (BHPP) 2009 (Bujagali Project Newsletter Q3 - 30th September, 2009) ● Newsletters / Sensitization ● 1,532 / 717 ● W ~ e-Transfer ● Internet: Bujagali Energy Ltd ● 4 pp.
`BEL09-Q4`	Bujagali Hydropower Project (BHPP) 2009 (Bujagali Project Newsletter Q4 - December, 2009) ● Newsletters / Sensitization ● 1,796 / 804 ● W ~ e-Transfer ● Internet: Bujagali Energy Ltd ● 5 pp.
`BEL10-Q4`	Bujagali Hydropower Project (BHPP) 2010 (Bujagali Project Newsletter Q4 - 31st December, 2010) ● Newsletters / Sensitization ● 1,991 / 749 ● W ~ e-Transfer ● Internet: Bujagali Energy Ltd ● 5 pp.
`BEL11-Q1`	Bujagali Hydropower Project (BHPP) 2011 (Bujagali Project Newsletter Q1 - 31st March, 2011) ● Newsletters / Sensitization ● 1,692 / 737 ● W ~ e-Transfer ● Internet: Bujagali Energy Ltd ● 5 pp.
`BEL11-Q3`	Bujagali Hydropower Project (BHPP) 2011 (Bujagali Project Newsletter Q3 - 30th September, 2011) ● Newsletters / Sensitization ● 1,371 / 585 ● W ~ e-Transfer ● Internet: Bujagali Energy Ltd ● 4 pp.
`Betty`	Malagala, Stephen 2000s (Betty) ● Songs - Traditional / Marriage ● 563 / 373 ● O ~ Transcription ● Anon.: - ● 0:05:56
`BibChEas`	Hughes, Edward 2013 (Embaga ey'Amazuukira Eyasooka) ● Biblical documents / Religion ● 1,030 / 561 ● W ~ e-Transfer ● Internet: Bible for Children ● 25 pp.
`BibChGod`	Hughes, Edward 2013 (Katonda nga Bweyatonda Buli Kintu) ● Biblical documents / Religion ● 921 / 446 ● W ~ e-Transfer ● Internet: Bible for Children ● 26 pp.
`BibChHea`	Hughes, Edward 2013 (Eigulu, Amaka ga Katonda Agaboneka Obulungi Einho) ● Biblical documents / Religion ● 928 / 496 ● W ~ e-Transfer ● Internet: Bible for Children ● 22 pp.

`Filename`	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
`BibChJes`	Hughes, Edward 2013 (Okuzaalibwa kwa Yesu) ● Biblical documents / Religion ● 815 / 449 ● W ~ e-Transfer ● Internet: Bible for Children ● 29 pp.
`BibChNoa`	Hughes, Edward 2013 (Ebigema ku Noah n'Omwidhuzo ogw'Amaadhi Omukologho) ● Biblical documents / Religion ● 730 / 413 ● W ~ e-Transfer ● Internet: Bible for Children ● 25 pp.
`BibChSad`	Hughes, Edward 2013 (Amainhama ag'Okunakughala okw'Omuntu) ● Biblical documents / Religion ● 744 / 427 ● W ~ e-Transfer ● Internet: Bible for Children ● 25 pp.
`Biblest2`	Various 2010 (Bible Story 2) ● Biblical documents / Religion ● 6,344 / 1,917 ● O ~ Transcription ● Internet: YouTube ● 0:44:48
`Biblest3`	Various 2010 (Bible Story 3) ● Biblical documents / Religion ● 2,785 / 1,034 ● O ~ Transcription ● Internet: YouTube ● 0:28:09
`Birugrd`	Various 2011 (Graduation Ceremony in Buwaabe) ● Celebrations / Inspirational ● 5,858 / 2,114 ● O ~ Transcription ● Iganga: - ● 1:04:38
`BLwaBana`	Kabugu, Milton Peter 1990s (Banamwandu ni Bamulekwa) ● Songs - Traditional / Marriage ● 774 / 434 ● O ~ Transcription ● Jinja: Ali Mukembo and Sons ● 0:09:25
`BLwaNgol`	Kabugu, Jessica 1990s (Ngoli Namala Naidha Luvanhuma) ● Songs - Traditional / Marriage ● 854 / 421 ● O ~ Transcription ● Jinja: Ali Mukembo and Sons ● 0:10:27
`BLwaOmun`	Kabugu, Jessica 1990s (Omuntu gh'Ensi Muzibu) ● Songs - Traditional / Life ● 691 / 354 ● O ~ Transcription ● Jinja: Ali Mukembo and Sons ● 0:08:04
`BMagEbiz`	Kabugu, Milton Peter 1990s (Ebizibu eby'Ensi) ● Songs - Traditional / Life ● 488 / 259 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:10:09
`BMagOmul`	Kabugu, Milton Peter 1990s (Omulamu Asaalilwa) ● Songs - Traditional / Relationships ● 445 / 234 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:09:30
`BMagRose`	Kabugu, Milton Peter 1990s (Rose Mary) ● Songs - Traditional / Relationships ● 480 / 250 ● O ~ Transcription ● Jinja: Sanyu Music Studios ● 0:20:13
`BTirBand`	Mata, Nassani 1990s (Bando Asiliile) ● Songs - Traditional / Marriage ● 196 / 131 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:02:12
`BTirIdha`	Mata, Nassani 1990s (Idha Ompelekeleku) ● Songs - Traditional / Relationships ● 264 / 116 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:04:35
`BTirNang`	Mata, Nassani 1990s (Nangobi) ● Songs - Traditional / Relationships ● 448 / 329 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:05:37
`BTirNinz`	Mata, Nassani 1990s (Ni Nze Mbeese) ● Songs - Traditional / Money ● 388 / 236 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:05:23
`BTirObug`	Mata, Nassani 1990s (Obugumba) ● Songs - Traditional / Marriage ● 234 / 124 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:05:15
`BTirOmuk`	Mata, Nassani 1990s (Omukazi Omwenzi) ● Songs - Traditional / Money ● 251 / 179 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:03:06
`BTirWabu`	Mata, Nassani 1990s (Wabukala Bando) ● Songs - Traditional / Marriage ● 244 / 131 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:06:11

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
BucakaML	Various 2013 (Bucaka Mails) ● E-mails / Networking ● 5,535 / 2,201 ● W ~ e-Transfer ● Internet: Bucaka ● 57 pp.
Bujagal1	Bujagaali & First wife 2006 (Bujagaali Interview 1) ● Interviews / Health ● 917 / 431 ● O ~ Transcription ● Jinja: - ● 0:32:30
Bujagal2	Bujagaali & First wife 2006 (Bujagaali Interview 2) ● Celebrations / Health ● 2,280 / 1,056 ● O ~ Transcription ● Jinja: - ● 0:43:40
Bultg10	Various 2010 (Busoga-Bulleting 1) ● E-mails / Networking ● 4,567 / 2,045 ● W ~ e-Transfer ● Internet: Yahoo! ● 27 pp.
Bultg11.6	Various 2011 (Busoga-Bulleting 2) ● E-mails / Networking ● 6,842 / 2,716 ● W ~ e-Transfer ● Internet: Yahoo! ● 32 pp.
Bultg11.7	Various 2011 (Busoga-Bulleting 3) ● E-mails / Networking ● 537 / 261 ● W ~ e-Transfer ● Internet: Yahoo! ● 4 pp.
Bultg12	Various 2012 (Busoga-Bulleting 4) ● E-mails / Networking ● 3,854 / 1,691 ● W ~ e-Transfer ● Internet: Yahoo! ● 17 pp.
BusogaCh	The Cross-Cultural Foundation of Uganda (CCFU) 2012 (The Uganda Clan Leaders' Charters) ● Policy documents - Busoga Kingdom / Sensitization ● 1,339 / 787 ● W ~ e-Transfer ● Kampala: The Cross-Cultural Foundation of Uganda ● 86 pp.
Buwaabe	Nabirye, Minah 2009 (Buwaabe Sunday Church Service) ● Biblical documents / Religion ● 8,788 / 2,936 ● O ~ Transcription ● Iganga: - ● 1:44:26
BuwaabGr	Nabirye, Minah 2010 (Buwaabe Graduation Ceremony) ● Celebrations / Inspirational ● 16,370 / 4,605 ● O ~ Transcription ● Iganga: - ● 3:51:17
BVATObuf	Kirimungu, Siragi 2000s (Obufumbo Buzibu) ● Songs - Traditional / Marriage ● 349 / 182 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:14:26
BVATOmwo	Kirimungu, Siragi 2000s (Omwoyo Fiitina) ● Songs - Traditional / Relationships ● 1,153 / 513 ● O ~ Transcription ● Jinja: Ali Mukembo Studio ● 0:14:05
Bwoteef	Mugwisa 2000s (Bw'oteefaaku) ● Songs - Traditional / Sensitization ● 428 / 219 ● O ~ Transcription ● Anon.: - ● 0:08:48
Bwozaal	Mugwisa 2000s (Bw'ozaala n'Abaawo) ● Songs - Traditional / Inspirational ● 562 / 308 ● O ~ Transcription ● Anon.: - ● 0:09:06
Byaif09	Various 2009 (Busogayaife 1) ● E-mails / Networking ● 2,020 / 947 ● W ~ e-Transfer ● Internet: Yahoo! ● 21 pp.
Byaif10	Various 2010 (Busogayaife 2) ● E-mails / Networking ● 29,743 / 9,845 ● W ~ e-Transfer ● Internet: Yahoo! ● 142 pp.
Byaif11.6	Various 2011 (Busogayaife 3) ● E-mails / Networking ● 24,805 / 8,320 ● W ~ e-Transfer ● Internet: Yahoo! ● 98 pp.
Byaif11.7	Various 2011 (Busogayaife 4) ● E-mails / Networking ● 5,289 / 2,440 ● W ~ e-Transfer ● Internet: Yahoo! ● 25 pp.
Byaif12	Various 2012 (Busogayaife 5) ● E-mails / Networking ● 41,983 / 13,218 ● W ~ e-Transfer ● Internet: Yahoo! ● 169 pp.
ByaKfaKl	Gulere, Cornelius 2010 (Bya Kufa Kuleka 5) ● Literature / Fables ● 16,109 / 5,449 ● W ~ OCR ● Busembatya: Mpolyabigere RC – RICED Center ● 72 pp.

Filename	**Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording**
`ChildAct`	Uganda Legal Information Institute (ULII) 2012 (Children's Act in Lusoga) ● Policy documents - Government / Sensitization ● 16,211 / 2,354 ● W ~ OCR ● Internet: Government of Uganda Parliamentary Act on Human Rights ● 68 pp.
`Cohen86`	Cohen, David 1986 (Towards a Reconstructed Past: Historical texts from Busoga, Uganda) ● Academic documents / History ● 16,657 / 5,742 ● W ~ OCR ● London: Oxford University Press ● 54 pp.
`Communit`	Various 2009 (Community Development) ● Radio talk shows / Inspirational ● 5,164 / 1,955 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:32:17
`Diocesan`	Cultural Research Centre (CRC) 2003 (Diocesan Family Day) ● Biblical documents / Religion ● 4,201 / 1,486 ● W ~ OCR ● Iganga: Diocesan Printery ● 48 pp.
`Ebibiin`	Kirimungu, Siragi 2000s (Ebibiina Biyamba) ● Songs - Traditional / Networking ● 484 / 208 ● O ~ Transcription ● Anon.: - ● 0:06:48
`Ebigkbal`	Bujagaali's client 2006 (Ebigema ku Balongo) ● Interviews / Marriage ● 1,954 / 815 ● O ~ Transcription ● Jinja: - ● 0:14:24
`Ebikemo`	Soyinka, Wole 2010 (Ebikemo by'Owoluganda Yero) ● Literature - Plays / Life ● 7,092 / 2,725 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 39 pp.
`Ebikete`	Gulere, Cornelius 2011 (Ebikete bya Busoga) ● Literature / Fables ● 2,572 / 1,430 ● W ~ OCR ● Internet: Google books ● 40 pp.
`EbikoikE`	Cultural Research Centre (CRC) 2002 (Ebikoiko eby'Abasoga) ● Literature / Riddles ● 6,963 / 2,818 ● W ~ OCR ● Jinja: CRC ● 126 pp.
`Ebikoiko`	Gulere, Cornelius 2008 (Ebikoiko mu Lusoga) ● Literature / Riddles ● 7,797 / 2,796 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 35 pp.
`Ebikolwa`	Wabugoyera; Kasubi, J.B.; Kaluuba, John Patrick; Mukama; Maganda, Matia & Maganda, Matayo 2010 (Ebikolwa bya Sapuli) ● Biblical documents / Religion ● 17,887 / 6,001 ● W ~ OCR ● Jinja: Our Lady of Fatima Parish Church ● 49 pp.
`EbikolwE`	Uganda Gazette 2008 (Ebikolwa Eby'ongelwaaku) ● Policy documents - Government / Politics ● 19,064 / 2,728 ● W ~ OCR ● Internet: Ministry of Education ● 159 pp.
`EbikolWK`	Gulere, Cornelius 2007 (Ebikolwa bya Wankembo) ● Literature / Fables ● 1,087 / 588 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 19 pp.
`Ebilung`	Mugwisa 2000s (Ebilungi Tibikoma) ● Songs - Traditional / Sensitization ● 535 / 322 ● O ~ Transcription ● Anon.: - ● 0:09:11
`EbindKuI`	Cultural Research Centre (CRC) 2005 (Ebindi kw'Idembe ery'Obw'omuntu mu Nsi Yoonayoona) ● Policy documents - Human rights / Sensitization ● 26,520 / 4,891 ● W ~ OCR ● Kisubi: Marianum Publishing Company ● 119 pp.
`Ebintub`	Malagala, Stephen 2010s (Ebintu Bisingagana) ● Songs - Traditional / Life ● 815 / 447 ● O ~ Transcription ● Kamuli: - ● 0:09:23
`Ebizibu`	Mugwisa 2000s (Ebizibu mu Duniya) ● Songs - Traditional / Health ● 576 / 349 ● O ~ Transcription ● Anon.: - ● 0:09:07
`Ebyensi`	Crado 2010s (Eby'ensi) ● Songs - Modern / Rehabilitation ● 474 / 251 ● O ~ Transcription ● Anon.: - ● 0:06:00

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Egyamw	Crado 2010s (Egya Mwete) ● Songs - Modern / Health ● 314 / 181 ● O ~ Transcription ● Anon.: - ● 0:05:03
Eidemban	Cultural Research Centre (CRC) 2010 (Eidembe ly'Abantu) ● Songs - Modern / Sensitization ● 3,360 / 1,224 ● O ~ Transcription ● Jinja: CRC ● 0:31:17
Eidembbw	Cultural Research Centre (CRC) 2010 (Eidembe ly'Obw'obuntu) ● Songs - Modern / Sensitization ● 2,744 / 1,034 ● O ~ Transcription ● Jinja: CRC ● 0:29:04
Eidembun	Anon. 2010 (Eidembe ly'Obuntu) ● Songs - Modern / Sensitization ● 946 / 240 ● O ~ Transcription ● Jinja: CRC ● 0:07:23
Eifumbi	Malagala, Stephen 2000s (Eifumbilo) ● Songs - Traditional / Marriage ● 537 / 374 ● O ~ Transcription ● Anon.: - ● 0:06:32
Eighali	Gulere, Cornelius 1998 (Eighali Lirikwisa) ● Literature / Fables ● 988 / 630 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 7 pp.
Eisomoly	International Centre for Eye Education (ICEE) 2008 (Eisomo ly'Okugezesa Obwangu bw'Enkyukakyuka mu Kubona) ● Policy documents - NGOs / Health ● 576 / 289 ● W ~ Translation ● Kampala: International Centre for Eye Education ● 0.5 pp.
Eisuubi	Gulere, Cornelius 2013 (Eisuubi Okusaaka Obusomi) ● Celebrations / Politics ● 955 / 541 ● W ~ OCR ● Internet: Tarehe sita ● 9 pp.
Ekibila	Kabugu, Milton Peter 2010 (Ekibila) ● Songs - Modern / Sensitization ● 784 / 130 ● O ~ Transcription ● Jinja: CRC ● 0:06:13
Ekidhuub	Gulere, Cornelius 2011 (Ekidhuubo) ● Literature / Fables ● 1,829 / 1,027 ● W ~ OCR ● Internet: Google books ● 22 pp.
Ekikwek	Malagala, Stephen 2000s (Ekikwekabya) ● Songs - Traditional / Marriage ● 748 / 470 ● O ~ Transcription ● Anon.: - ● 0:08:36
Ekimlik	Wabugoyera; Kasubi, J.B.; Kaluuba, John Patrick; Mukama; Maganda, Matia & Maganda, Matayo 2008 (Ekimuliikirira, August-October 2008) ● Biblical documents / Religion ● 18,452 / 4,839 ● W ~ OCR ● Jinja: Diocese of Jinja ● 54 pp.
Ekinait	Malagala, Stephen 2000s (Ekinaita Embwa) ● Songs - Traditional / Health ● 657 / 378 ● O ~ Transcription ● Anon.: - ● 0:07:52
Ekirangi	Abakulu b'ebika bya Busoga 2009 (Ekirangiriro eri Obusoga n'Ensi Yoonayoona) ● Policy documents - Busoga Kingdom / History ● 2,710 / 980 ● W ~ Translation ● Jinja: Katukiro w'olukiiko lw'abakulu b'ebika bya Busoga ● 14 pp.
Ekiwandi	International Centre for Eye Education (ICEE) 2010 (Ekiwandiiko Ekilaga Ennambuula y'Amaka) ● Policy documents - NGOs / Health ● 1,612 / 525 ● W ~ Translation ● Kampala: International Centre for Eye Education ● 5 pp.
EliinaEl	International Centre for Eye Education (ICEE) 2010 (Eliina Elisooka) ● Policy documents - NGOs / Health ● 195 / 109 ● W ~ Translation ● Kampala: International Centre for Eye Education ● 1 p.
Embeeke	Malagala, Stephen 2000s (Embeekela) ● Songs - Traditional / Relationships ● 634 / 381 ● O ~ Transcription ● Anon.: - ● 0:07:08

Filename	**Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording**
`Empambo`	Gulere, Cornelius 2007 (Empambo) ● Literature / Fables ● 2,335 / 1,391 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 43 pp.
`EmpisaB`	Gulere, Cornelius 2007 (Empisa n'Obuntubulamu) ● Literature / Fables ● 1,271 / 776 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 16 pp.
`Endagaan`	Bible Society Uganda (BSU) 1998 (Endagaano Empyaka) ● Biblical documents / Religion ● 150,223 / 19,829 ● W ~ OCR ● Jinja: The Bible Society of Uganda ● 518 pp.
`EndhesE2`	Gulere, Cornelius 2007 (Endheso Ennhimpi) ● Literature / Proverbs ● 2,659 / 1,877 ● W ~ OCR ● Busembatya: Lusoga Language Academic Board (LLAB) ● 34 pp.
`EndhesoD`	Lyavala-Lwanga, E.J. 1967 (Endheso dh'Abasoga) ● Literature / Proverbs ● 16,809 / 7,289 ● W ~ OCR ● Kampala: Milton Obote Foundation ● 97 pp.
`Endhesoe`	Gulere, Cornelius 2011 (Endheso Ennhimpi) ● Literature / Fables ● 2,723 / 1,900 ● W ~ OCR ● Internet: Google books ● 40 pp.
`Endhesul`	Gulere, Cornelius 2012 (Endheso Ensuusulemu) ● Literature / Proverbs ● 2,125 / 1,066 ● W ~ OCR ● Internet: Anon. ● 31 pp.
`Endhiiy1`	Private Sector Uganda (PSU) 2009 (Endhiiya y'Obukodyo bw'Enkulankulana 1) ● Policy documents - NGOs / Money ● 602 / 299 ● W ~ Translation ● Kampala: Private Sector Uganda ● 5 pp.
`Endhiiy2`	Private Sector Uganda (PSU) 2009 (Endhiiya y'Obukodyo bw'Enkulankulana 2) ● Policy documents - NGOs / Money ● 602 / 300 ● W ~ Translation ● Kampala: Private Sector Uganda ● 4 pp.
`EndyaBul`	Gulere, Cornelius 2007 (Endya Bulamu) ● Literature / Fables ● 1,101 / 644 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 13 pp.
`EnfumKay`	Kisubi, Alfred James Igaga 2009 (Enfumitiriza Kayingo n'Entegeka luv'Okwaaya y'Olukiiko lw'Abakulu b'Ebika) ● Policy documents - Busoga Kingdom / History ● 2,550 / 1,049 ● W ~ OCR ● Jinja: The Re-Unification of the Clans of Busoga ● 14 pp.
`Engabo`	Gulere, Cornelius 2011 (Engabo ya Busoga) ● Literature / Fables ● 765 / 431 ● W ~ OCR ● Internet: Google books ● 19 pp.
`Engedh`	Anon. 2010s (Engeli Dhaimwe) ● Songs - Modern / Relationships ● 463 / 237 ● O ~ Transcription ● Anon.: - ● 0:05:13
`EnhemboM`	Cultural Research Centre (CRC) 2008 (Enhembo mu Mikolo Emitukuvu) ● Biblical documents / Religion ● 3,771 / 1,486 ● W ~ OCR ● Jinja: CRC ● 35 pp.
`EnkEkifn`	Cultural Research Centre (CRC) 2000 (Enkabi Ekifiini mu Busoga) ● Literature / Proverbs ● 7,842 / 2,768 ● W ~ OCR ● Jinja: CRC ● 30 pp.
`Ennakun`	Mugwisa 2000s (Ennaku Namugalula) ● Songs - Traditional / Life ● 486 / 300 ● O ~ Transcription ● Anon.: - ● 0:09:51
`Ennhemba`	Nabirye, Minah 2000 (Ennhemba dh'Olusoga) ● Songs - Traditional / Relationships ● 653 / 284 ● O ~ e-Transfer ● Jinja: - ● (own writing from memory recollections)
`Ennhonh`	Mugwisa, Andy Cooke 2010 (Ennhonhi ku Lugyo) ● Songs - Traditional / Science ● 262 / 161 ● O ~ Transcription ● Internet: YouTube ● 0:04:00

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
EnsambDh	Cultural Research Centre (CRC) 1999 (Ensambo edh'Abasoga) ● Literature / Proverbs ● 16,121 / 7,810 ● W ~ OCR ● Jinja: CRC ● 90 pp.
Ensiek	Malagala, Stephen 2000s (Ensi Ekyuse) ● Songs - Traditional / Life ● 472 / 315 ● O ~ Transcription ● Anon.: - ● 0:06:08
Ensieno	Baisi 2010s (Ensi eno Weetuuse) ● Songs - Traditional / Rehabilitation ● 676 / 347 ● O ~ Transcription ● Anon.: - ● 0:08:55
Ensinzb	Salimu 2010 (Ensi Nzibu) ● Songs - Modern / Life ● 1,206 / 234 ● O ~ Transcription ● Jinja: CRC ● 0:06:31
Erikwain	Various 2009 (Erikwaine) ● Radio talk shows / Politics ● 5,885 / 2,077 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:56:23
Esaalmk1	Mbutu, Rose & Musooko, Paulo 2010 (Esaala ey'Amaka 1) ● Radio talk shows / Religion ● 6,281 / 1,951 ● O ~ Transcription ● Bugembe: Holy Christ Family Ministry ● 0:48:06
Esaalmk2	Mbutu, Rose & Musooko, Paulo 2010 (Esaala ey'Amaka 2) ● Radio talk shows / Religion ● 6,624 / 2,260 ● O ~ Transcription ● Bugembe: Holy Christ Family Ministry ● 0:54:08
Eyalyaom	Gulere, Cornelius 2007 (Eyalya Omuunhu) ● Literature / Fables ● 512 / 340 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 10 pp.
Eyeesig	Geo Bless 2010s (Eyeeesigibwa) ● Songs - Modern / Relationships ● 245 / 133 ● O ~ Transcription ● Anon.: - ● 0:04:16
Ezilamul	Kabugu, Milton Peter 2010 (Ezila Mulungi ku Nsi) ● Songs - Modern / Life ● 986 / 150 ● O ~ Transcription ● Jinja: CRC ● 0:05:39
Ezirakkw	Gulere, Cornelius 2012 (Ezira Kyetaagisibwa Kwongeraku) ● Literature / Fables ● 352 / 211 ● W ~ OCR ● Internet: Tarehe sita ● 9 pp.
Facebook	Various 2009 (Posting) ● E-mails / Language ● 54 / 51 ● W ~ e-Transfer ● Internet: Facebook ● 1 p.
Fiida	Malagala, Fiida 2010s (Fiida) ● Songs - Traditional / Marriage ● 268 / 190 ● O ~ Transcription ● Anon.: - ● 0:05:02
GulwOlAs	Dhizaala, John Stephen 2011 (Gulaama w'Olulimi Olusoga Asookerwaku) ● Academic documents / Language ● 11,311 / 3,464 ● W ~ OCR ● Jinja: CRC ● 57 pp.
Gw'olile	Anon. 2010s (Gw'Olilekela Omwana) ● Songs - Traditional / Marriage ● 244 / 133 ● O ~ Transcription ● Iganga: - ● 0:04:34
Gw'olkba	Kaluuba, John Patrick; Kivuunike, James; Dhizaala, John Stephen & Nabirye, Christine 2010 (Gw'Olekera Abato: Okusoma kuleeta obusobozi) ● Literature / Language ● 5,766 / 2,170 ● W ~ OCR ● Jinja: CRC, Marianum Publishing Company and LABE ● 55 pp.
HonKiyg	Various 2003 (Hon Kiyingi) ● Radio talk shows / Politics ● 1,068 / 550 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:11:45
HonKizge	Various 2009 (Hon Kizige) ● Radio talk shows / Politics ● 7,932 / 2,715 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:09:59

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Idembeb	Kabugu, Milton Peter 2010 (Eidembe ly'Abaana) ● Songs - Modern / Sensitization ● 599 / 142 ● O ~ Transcription ● Jinja: CRC ● 0:06:06
IdhaTusm	Wambi, M.; Naigaga, R. & CRC 2005 (Idha Tusome) ● Academic documents / Language ● 3,059 / 1,458 ● W ~ OCR ● Jinja: Lusoga Language Authority (LULA) ● 80 pp.
Immunisa	Various 2009 (Immunization) ● Radio talk shows / Health ● 6,544 / 2,353 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:07:21
ImweMwOk	Bible Society Uganda (BSU) 1994 (Imwe Mwagheebwa Okumanha) ● Biblical documents / Religion ● 1,681 / 828 ● W ~ OCR ● Kampala: The Bible Society of Uganda ● 8 pp.
InstallC	Cultural Research Centre (CRC) 2010 (Installation of Rt. Rev. Bishop Charles Martin Wamika as Bishop of Diocese of Jinja) ● Biblical documents / Religion ● 1,897 / 611 ● W ~ OCR ● Jinja: Marianum Publishing Company ● 48 pp.
IntBilaa	Various 2008 (Introduction Ceremony 1) ● Celebrations / Marriage ● 11,716 / 3,555 ● O ~ Transcription ● Iganga: - ● 4:35:34
IntHadij	Nabirye, Minah 2008 (Introduction Ceremony 2) ● Celebrations / Marriage ● 12,098 / 3,558 ● O ~ Transcription ● Iganga: - ● 3:33:41
Isabiry	Malagala, Stephen 2000s (Isabirye ni Bbaabba We) ● Songs - Traditional / Marriage ● 441 / 301 ● O ~ Transcription ● Anon.: - ● 0:06:15
Isatifik	Foundation for Endangered Languages (FEL) 2012 (Isatifikeeti lya Bughanzi) ● Celebrations / Inspirational ● 41 / 36 ● W ~ OCR ● Internet: Foundation for Endangered Languages ● 1 p.
Isebantu	Kabugu, Milton Peter 2010 (Isebantu) ● Songs - Modern / History ● 920 / 127 ● O ~ Transcription ● Jinja: CRC ● 0:06:18
Judicial	Various 2010 (Judicial Service Commission) ● Radio talk shows / Sensitization ● 7,721 / 2,461 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:00:41
Kabili	Malagala, Stephen 2000s (Kabili Ndeese) ● Songs - Traditional / Sensitization ● 320 / 228 ● O ~ Transcription ● Anon.: - ● 0:04:03
Kabindi1	Kabindi, Erukaana 2008 (Kabindi Interview 1) ● Interviews / Health ● 4,313 / 1,504 ● O ~ Transcription ● Bugiri: - ● 0:40:05
Kabindi2	Kabindi, Erukaana 2009 (Kabindi Interview 2) ● Interviews / Health ● 2,549 / 1,026 ● O ~ Transcription ● Bugiri: - ● 0:44:49
KadokInt	Kadooko, John 2012 (Interview on the Kyabazingaship and the Status of Lusoga) ● Interviews / History ● 9,414 / 2,999 ● O ~ Transcription ● Jinja: - ● 1:35:07
Kaibutag	Kai Butagaya women 2010 (Katonda n'Agaba) ● Songs - Traditional / Gratitude ● 366 / 110 ● O ~ Transcription ● Internet: YouTube ● 0:04:28
Kaleebi	Kaleebi, George 2010 (Kaleebi) ● E-mails / Networking ● 1,437 / 809 ● W ~ e-Transfer ● Local: Self ● 7 pp.
Kalikub	Anon. 2010s (Kali Kubayiiga) ● Songs - Modern / Health ● 510 / 111 ● O ~ Transcription ● Anon.: - ● 0:06:00

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
KamuliTC	Various 2009 (Kamuli Town Council) ● Radio talk shows / Sensitization ● 5,369 / 1,774 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:55:05
KasreeOl	Various 2009 (Posting) ● E-mails / Language ● 71 / 63 ● W ~ e-Transfer ● Internet: Facebook ● 1 p.
Katondk	Baisi 2010s (Katonda ky'Akuwa Osiima) ● Songs - Traditional / Gratitude ● 539 / 255 ● O ~ Transcription ● Anon.: - ● 0:08:25
Kawoiwol	Gulere, Cornelius 2007 (Kawoiwolo Ayenda Kubayiza) ● Literature / Fables ● 1,123 / 670 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 18 pp.
KayondOm	Gulere, Cornelius 2011 (Kayondo Omuyondho) ● Literature - Booklets / Sensitization ● 709 / 315 ● W ~ e-Transfer ● Busembatya: Mpolyabigere RC – RICED Center ● 11 pp.
Kibbaaly	Various 2009 (Kibbaalya) ● Radio talk shows / Politics ● 9,707 / 3,059 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:21:06
Kibbumb	Anon. 2010 (Kibbumba) ● Songs - Traditional / Religion ● 675 / 142 ● O ~ Transcription ● Internet: YouTube ● 0:08:30
Kilikum	Anon. 2010s (Kili ku Mwino) ● Songs - Traditional / Life ● 306 / 141 ● O ~ Transcription ● Iganga: - ● 0:05:12
Kintu	Cultural Research Centre (CRC) 1998 (Kintu) ● Literature - Plays / History ● 3,785 / 1,503 ● W ~ OCR ● Jinja: CRC ● 34 pp.
Kiriggwa	Kiriggwajjo 2007 (Akalango ka Kiriggwajjo) ● Advertisements / Networking ● 87 / 61 ● W ~ Translation ● Kampala: - ● 0.5 pp.
Kisaati	Various 2010 (Kisaati Kawooya Mugainho) ● Radio talk shows / Politics ● 5,155 / 2,041 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:44:45
KisambiA	Kisambira, Amurafeeri 2004 (Kisambira Amurafeeli) ● Academic documents / Language ● 219 / 161 ● W ~ Retyping of hand-written document ● Kampala: - ● 2 pp.
Kisiki	Gulere, Cornelius 2006 (Kisiki) ● Literature / Fables ● 954 / 554 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 7 pp.
KiyinKbi	Lyavala-Lwanga, E.J. 1969 (Kiyini Kibi) ● Literature / Language ● 19,256 / 7,737 ● W ~ Retyping of image ● Kampala: Milton Obote Foundation ● 123 pp.
KKhst	Kirunda, Kivejinja 2012 (Interview on the History of Busoga and Lusoga) ● Interviews / History ● 8,296 / 2,659 ● O ~ Transcription ● Kampala: - ● 1:25:07
KKspb	Kirunda, Kivejinja 2012 (Interview on the Sapoba Legacy) ● Interviews / History ● 5,801 / 2,239 ● O ~ Transcription ● Kampala: - ● 0:56:10
Kodh'eyo	Various 1997–1998 (1997) (Kodh'eyo: Busoga etebenkere) ● Journalism / Networking ● 185,843 / 45,945 ● W ~ OCR & Retyping ● Kampala: Kodh'eyo Publications ● 341 pp.
Kolatug	Anon. 2010s (Kola Tugyeyo) ● Songs - Modern / Relationships ● 158 / 65 ● O ~ Transcription ● Anon.: - ● 0:04:15
KufaLeka	Gulere, Cornelius 2006 (Kufa, Leka Kweghaana) ● Literature / Fables ● 1,091 / 653 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 11 pp.

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
`Kuwagil`	Kigenyi, Amos 2000s (Kuwagila Kaguta) ● Songs - Modern / Politics ● 161 / 100 ● O ~ Transcription ● Anon.: - ● 0:04:50
`Kyandib`	Salimu 2000s (Kyandibaile Kilungi) ● Songs - Traditional / Life ● 683 / 360 ● O ~ Transcription ● Anon.: - ● 0:12:35
`Ky'oyend`	Anon. 2010 (Ky'oyenda) ● Songs - Modern / Inspirational ● 357 / 173 ● O ~ Transcription ● Internet: YouTube ● 0:03:31
`LandPoli`	The Uganda National Land Policy 2010 (Ekighandiiko Ekigema ku Itaka) ● Policy documents - NGOs / Sensitization ● 7,602 / 2,680 ● W ~ e-Transfer ● Kampala: The Uganda Land Alliance, supported by Concern Worldwide ● 40 pp.
`Lexiko09`	Nabirye, Minah 2009 (Eiwanika ly'Olusoga Elyasookela Ilala) ● Academic documents / Language ● 323 / 187 ● W ~ Own writing ● Stellenbosch: WAT ● 0.5 pp.
`Lexiko10`	Nabirye, Minah 2010 (Eiwanika ly'Olusoga Lizuuseeku Omukozesa Ataali Muluubilile) ● Academic documents / Language ● 332 / 203 ● W ~ Own writing ● Stellenbosch: WAT ● 0.5 pp.
`Lexiko11`	Nabirye, Minah 2011 (Okulondoola Engeli Eitu ly'Olusoga bwe Linaatuusibwa mu Iwanika: Omutindo ogulaga olulimi bwe luli, bwe luteekwa okuba oba bwe lube lutwalibwe) ● Academic documents / Language ● 386 / 223 ● W ~ Own writing ● Stellenbosch: WAT ● 0.5 pp.
`Lexiko13`	Nabirye, Minah 2013 (Okuta Eiwanika ly'Olusoga mu Mbeela y'Omutegekowaziso Ogusomwa ku Kompyuta: Ebizibu n'ebiluubililwa) ● Academic documents / Language ● 174 / 121 ● W ~ Own writing ● Stellenbosch: WAT ● 0.5 pp.
`Lukabyo`	Various 2012 (Lukabyo Eulogy) ● Biblical documents / Religion ● 358 / 218 ● W ~ OCR ● Ibulanku: - ● 12 pp.
`LusgOls2`	Gulere, Cornelius 2012 (Olusoga Olusookelwaaku (Level 2)) ● Academic documents / Language ● 4,570 / 2,022 ● W ~ OCR ● Busembatya: Lusoga Language Academic Board (LLAB) ● 37 pp.
`LusHyms`	Cultural Research Centre (CRC) 2012 (Lusoga Hymns) ● Biblical documents / Religion ● 477 / 299 ● W ~ OCR ● Jinja: Diocese of Jinja ● 4 pp.
`LusLdPr`	Bible Society Uganda (BSU) 2012 (The Lord's Prayer in Lusoga) ● Biblical documents / Religion ● 114 / 91 ● W ~ OCR ● Internet: Bible Society Uganda ● 1 p.
`Lusmades`	Kagoya, Michelle Johnson 2011 (Lusoga Made Simple) ● Academic documents / Language ● 2,181 / 800 ● W ~ OCR ● Jinja: CRC ● 134 pp.
`LusMath1`	Gulere, Cornelius 2006 (Lusoga Mathematics Primer 1) ● Academic documents / Science ● 5,081 / 259 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 105 pp.
`Lusterm1`	Jore, Nathan D. 2011 (Obutonde Okutuuka ku Kuva Ekiketezo 1) ● Biblical documents / Religion ● 7,046 / 2,258 ● W ~ e-Transfer ● Plymouth, MN: Ambassador Institute ● 84 pp.

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
`Lusterm2`	Jore, Nathan D. 2011 (Ensi Ensubize Okutuuka ku Banabbi Ekiketezo 2) ● Biblical documents / Religion ● 9,937 / 3,094 ● W ~ e-Transfer ● Plymouth, MN: Ambassador Institute ● 84 pp.
`Lusterm3`	Jore, Nathan D. 2011 (Obulamu bwa Yesu Kiketezo 3) ● Biblical documents / Religion ● 6,032 / 2,328 ● W ~ e-Transfer ● Plymouth, MN: Ambassador Institute ● 75 pp.
`Lusterm4`	Jore, Nathan D. 2011 (Amagezi Ekiketezo 4) ● Biblical documents / Religion ● 4,702 / 1,913 ● W ~ e-Transfer ● Plymouth, MN: Ambassador Institute ● 69 pp.
`Lusterm5`	Jore, Nathan D. 2011 (Enono Ekiketezo 5) ● Biblical documents / Religion ● 6,481 / 2,356 ● W ~ e-Transfer ● Plymouth, MN: Ambassador Institute ● 89 pp.
`Lusterm6`	Jore, Nathan D. 2011 (Obuwereza Ekiketezo 6) ● Biblical documents / Religion ● 6,599 / 2,675 ● W ~ e-Transfer ● Plymouth, MN: Ambassador Institute ● 96 pp.
`Luthour`	Lutheran Church Ministry 2010 (Lutheran Hour) ● Radio talk shows / Religion ● 7,604 / 2,441 ● O ~ Transcription ● Jinja: NBS FM ● 0:59:14
`LwakAbTb`	Cultural Research Centre (CRC) 2000 (Lwaki Abakazi Tibabeeda Mulambo) ● Literature / Fables ● 8,086 / 2,615 ● W ~ OCR ● Jinja: CRC ● 77 pp.
`M&G-Mkwa`	Various 2009 (Introduction Ceremony 3) ● Celebrations / Marriage ● 11,194 / 3,520 ● O ~ Transcription ● Kampala: - ● 2:00:00
`MAAppdx3`	Nabirye, Minah 2008 (Test 1A Questionnaire) ● Academic documents / Language ● 657 / 343 ● W ~ Own writing ● Kampala: Makerere Institute of Languages ● 4 pp.
`MAAppdx6`	Nabirye, Minah 2008 (Okugezesa Eiwanika ly'Olusoga) ● Academic documents / Language ● 696 / 363 ● W ~ Own writing ● Kampala: Makerere Institute of Languages ● 4 pp.
`MagezinK`	Gulere, Cornelius 2007 (Magezi ni Kasilu) ● Literature / Fables ● 2,094 / 907 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 17 pp.
`Mail13-5`	Various 2013 (Lusoga Mails 1) ● E-mails / Networking ● 13,538 / 5,367 ● W ~ e-Transfer ● Internet: Yahoo! ● 69 pp.
`Mail13-6`	Various 2013 (Lusoga Mails 2) ● E-mails / Networking ● 9,446 / 4,079 ● W ~ e-Transfer ● Internet: Yahoo! ● 31 pp.
`MaisoTig`	Gulere, Cornelius 2006 (Maiso Tigalya Guba Mwoyo) ● Literature / Fables ● 1,377 / 855 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 10 pp.
`MarikoA`	Bible Society Uganda (BSU) 1996 (Mariko. Amawulire Amalungi mu Lusoga) ● Biblical documents / Religion ● 12,416 / 3,743 ● W ~ OCR ● Kampala: The Bible Society of Uganda ● 54 pp.
`Mazima`	Various 2010 (Mazima) ● Radio talk shows / Politics ● 7,423 / 2,375 ● O ~ Transcription ● Jinja: NBS FM ● 1:11:16
`MenhaW1`	Nabirye, Minah 2008 (Website Information) ● Academic documents / Language ● 1,674 / 782 ● W ~ e-Transfer ● Internet: Menha Publishers ● 2 pp.
`MenhaW2`	Nabirye, Minah 2010 (Engeli Kampuni bwe Yatandiika) ● Academic documents / Language ● 1,140 / 575 ● W ~ e-Transfer ● Internet: Menha Publishers ● 3 pp.

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Missa1	Cultural Research Centre (CRC) 2012 (Missa mu Lusoga Ebiseera eby'Omwaka n'Enaku edh'Abatuukirivuu) ● Biblical documents / Religion ● 31,122 / 4,019 ● W ~ OCR ● Jinja: Diocese of Jinja ● 195 pp.
Missa2	Cultural Research Centre (CRC) 2012 (Missa mu Lusoga Ebiseera eby'Amatuuka n'Amazaalibwa) ● Biblical documents / Religion ● 12,344 / 2,249 ● W ~ OCR ● Jinja: Diocese of Jinja ● 65 pp.
Missa3	Cultural Research Centre (CRC) 2012 (Missa mu Lusoga Ensengeka y'Emikolo gya Wiiki Entukuvu) ● Biblical documents / Religion ● 7,165 / 1,887 ● W ~ OCR ● Jinja: Diocese of Jinja ● 38 pp.
Missa4	Cultural Research Centre (CRC) 2012 (Missa mu Lusoga Ebiseera eby'Amazuukira) ● Biblical documents / Religion ● 10,593 / 1,489 ● W ~ OCR ● Jinja: Diocese of Jinja ● 73 pp.
Missa5	Cultural Research Centre (CRC) 2012 (Missa mu Lusoga Ebiseera eby'Ekisiibo) ● Biblical documents / Religion ● 7,979 / 1,639 ● W ~ OCR ● Jinja: Diocese of Jinja ● 53 pp.
MpeeBulm	Ministry of Health 2010 (Mpeereza ya Bulamu) ● Policy documents - Government / Health ● 9,773 / 2,305 ● W ~ OCR ● Internet: Ministry of Health ● 29 pp.
Mpuuta	Malagala, Stephen 2010s (Mpuuta na Mwogo) ● Songs - Traditional / Marriage ● 951 / 497 ● O ~ Transcription ● Kamuli: - ● 0:09:23
MuBigrBy	Mwesigwa, Roy 2000 (Mu Bigere Bye. Olugero lw'Abasoga ku kugoberera Yesu) ● Biblical documents / Religion ● 25,870 / 5,519 ● W ~ OCR ● Jinja: Church of Christ ● 111 pp.
Mukamug	Mukaabya, Willy 2010 (Muka Mugandawo Twala Butwale) ● Songs - Modern / Marriage ● 521 / 252 ● O ~ Transcription ● Internet: YouTube ● 0:07:37
Mukazimk	Anon. 2010s (Omukazi Muka Beene) ● Songs - Modern / Rehabilitation ● 365 / 169 ● O ~ Transcription ● Iganga: - ● 0:04:22
Muko	Mukaabya, Willy 2010 (Muko) ● Songs - Modern / Marriage ● 586 / 350 ● O ~ Transcription ● Internet: YouTube ● 0:08:23
Mulinaa	Mugwisa 2000s (Mulinaanwa) ● Songs - Traditional / Networking ● 476 / 179 ● O ~ Transcription ● Anon.: - ● 0:09:17
Musoke	Various 2010 (Eby'omu Ndhu) ● Radio talk shows / Marriage ● 7,407 / 2,514 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:04:40
MutMalym	Bible Society Uganda (BSU) 2012 (Mutendwa Malyaamu) ● Biblical documents / Religion ● 55 / 47 ● W ~ OCR ● Internet: Bible Society Uganda ● 1 p.
Muwuliil	Kigenyi, Amos 2010 (Muwuliile Bulungi) ● Songs - Modern / Gratitude ● 297 / 79 ● O ~ Transcription ● Internet: YouTube ● 0:04:45
Mwebale	Mugwisa 2000s (Mwebale Ssaba) ● Songs - Traditional / Gratitude ● 560 / 341 ● O ~ Transcription ● Anon.: - ● 0:07:43
Mwenewo	Geo Bless 2010s (Mwenewo) ● Songs - Modern / Marriage ● 288 / 202 ● O ~ Transcription ● Anon.: - ● 0:04:39

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
MwidhTgT	Kasozi, John 2000 (Mwidhe Tugye Tusenge) ● Literature / Religion ● 1,568 / 604 ● W ~ Retyping of image ● Jinja: Diocese of Jinja ● 12 pp.
MwidTufm	Ssajabi, Sophronius 1999 (Mwidhe Tufume) ● Literature / Fables ● 5,867 / 2,087 ● W ~ OCR ● Jinja: CRC ● 62 pp.
Mwino	Various 2010 (Mwino Akuwa y'Owa) ● Radio talk shows / Politics ● 9,211 / 2,666 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:19:37
Mwinoak	Anon. 2000s (Mwino Akuwa y'Owa) ● Songs - Traditional / Networking ● 151 / 76 ● O ~ Transcription ● Anon.: - ● 0:08:20
NAAD113	Various 2009 (NAADS 1) ● Radio talk shows / Science ● 5,797 / 2,092 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:51:05
NAAD1130	Various 2009 (NAADS 2) ● Radio talk shows / Science ● 6,357 / 2,087 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:55:56
NabM10	Nabirye, Minah 2010 (Nabirye 1) ● E-mails / Networking ● 2,530 / 1,172 ● W ~ e-Transfer ● Local: Self ● 8 pp.
NabM11.6	Nabirye, Minah 2011 (Nabirye 2) ● E-mails / Language ● 840 / 533 ● W ~ e-Transfer ● Local: Self ● 3 pp.
NabM11.7	Nabirye, Minah 2011 (Nabirye 3) ● E-mails / Language ● 472 / 287 ● W ~ e-Transfer ● Local: Self ● 2 pp.
Nakoowa	Crado 2010s (Nakoowa) ● Songs - Modern / Marriage ● 242 / 184 ● O ~ Transcription ● Anon.: - ● 0:04:06
Nantaga	Mugwisa 2000s (Nantagalagilwa) ● Songs - Traditional / Marriage ● 434 / 223 ● O ~ Transcription ● Anon.: - ● 0:07:48
Nantameg	Gulere, Cornelius 2007 (Nantamegwa) ● Literature - Plays / Life ● 9,702 / 3,522 ● W ~ OCR ● Busembatya: Lusoga Language Academic Board (LLAB) ● 48 pp.
NantamuD	Nantamu, Dyogo Peter 2011 (Factors Associated With Male Involvement in Maternal Health Care Services in Jinja District, Uganda) ● Academic documents / Health ● 1,328 / 634 ● W ~ e-Transfer ● Kampala: Makerere University School of Public Health ● 108 pp.
Ndimbonk	Gulere, Cornelius 2006 (Ndimubonakuuli) ● Literature / Fables ● 1,192 / 714 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 11 pp.
Ndimugez	Various 1998–1999 (1998) (Ndimugezi n'Omukobere: The factfinder) ● Journalism / Networking ● 15,821 / 7,159 ● W ~ OCR & Retyping ● Jinja: Ndimugezi Publications ● 42 pp.
Ndinimuk	Gulere, Cornelius 2011 (Ndi ni Mukazi Wange) ● Literature / Riddles ● 934 / 509 ● W ~ OCR ● Internet: Google books ● 12 pp.
Ngulina	Gulere, Cornelius 2011 (Lusoga: Nguli namanha) ● Literature / Fables ● 1,445 / 531 ● W ~ OCR ● Internet: Google books ● 21 pp.
Nibwonv	Kigenyi, Amos 2010 (Ni bw'Onvuma Agaiso) ● Songs - Modern / Relationships ● 393 / 107 ● O ~ Transcription ● Internet: YouTube ● 0:04:01
NkontaM	Gulere, Cornelius 2007 (Nkontamuti) ● Literature / Fables ● 1,082 / 585 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 20 pp.

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Nsangail	Magoola, Racheal 2010s (Nsangaile) ● Songs - Traditional / History ● 311 / 163 ● O ~ Transcription ● Iganga: - ● 0:06:09
Nsobola	Gulere, Cornelius 2011 (Nsobola Nsobola) ● Literature / Fables ● 979 / 596 ● W ~ OCR ● Internet: Google books ● 16 pp.
Obbangai	Afrigo Band 2010 (Obbangaina) ● Songs - Modern / Relationships ● 232 / 109 ● O ~ Transcription ● Internet: YouTube ● 0:04:07
Obufumbo	Gulere, Cornelius 2012 (Obufumbo) ● Literature - Booklets / Sensitization ● 1,306 / 669 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 8 pp.
Obughan	Baisi 2010s (Obughangwa Bwaife) ● Songs - Traditional / Sensitization ● 226 / 148 ● O ~ Transcription ● Anon.: - ● 0:06:42
Obugumb	Anon. 2000s (Obugumba Buluma) ● Songs - Traditional / Marriage ● 297 / 154 ● O ~ Transcription ● Anon.: - ● 0:07:43
Obukyay	Geo Bless 2010s (Obukyayi) ● Songs - Modern / Relationships ● 340 / 185 ● O ~ Transcription ● Anon.: - ● 0:04:37
Obululu	Malagala, Stephen 2000s (Obululu) ● Songs - Traditional / Politics ● 678 / 426 ● O ~ Transcription ● Anon.: - ● 0:06:51
Obwende	Malagala, Fiida 2010s (Obwende Mpisa) ● Songs - Traditional / Marriage ● 327 / 249 ● O ~ Transcription ● Anon.: - ● 0:05:51
Ogunguma	Babirye, Judith 2010 (Ogungumale Kibbumba) ● Songs - Gospel / Religion ● 357 / 77 ● O ~ Transcription ● Internet: YouTube ● 0:03:16
Ogusolo	Gulere, Cornelius 2011 (Ogusolo n'Ekikaadho) ● Literature / Fables ● 1,030 / 594 ● W ~ OCR ● Internet: Google books ● 15 pp.
Okozeewo	Crado 2010s (Okozeewo Ki) ● Songs - Modern / Gratitude ● 536 / 306 ● O ~ Transcription ● Anon.: - ● 0:06:08
Okukonk	Mugwisa 2000s (Okukonkona Embaile) ● Songs - Traditional / Inspirational ● 464 / 280 ● O ~ Transcription ● Anon.: - ● 0:08:59
Okukyal	Various 2011 (Okukyala kw'Abasiki e Buwaabe) ● Celebrations / Politics ● 17,052 / 4,973 ● O ~ Transcription ● Iganga: - ● 3:28:26
OkusanT1	Gulere, Cornelius 2007 (Okusanhusa Tikwesanhusa 1) ● Literature / Language ● 614 / 357 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 15 pp.
OkusanT2	Gulere, Cornelius 2007 (Okusanhusa Tikwesanhusa 2) ● Literature / Language ● 219 / 144 ● W ~ OCR ● Internet: Google books ● 15 pp.
OkusBkUg	Mbowa, Rose 2013 (Okusaaka kwa Bakazi ba Uganda) ● Policy documents - Human rights / Sensitization ● 605 / 363 ● W ~ OCR ● Internet: Human Rights Advocacy ● 5 pp.
Okuwasa	Geo Bless 2010s (Okuwasa) ● Songs - Modern / Money ● 408 / 276 ● O ~ Transcription ● Anon.: - ● 0:04:53
OkwEniBz	Uganda Human Rights Commission (UHRC) 2008 (Okwanganga Ennimi dh'Obuzaale) ● Policy documents - Endangered languages / Language ● 1,857 / 801 ● W ~ OCR ● Internet: Anon. ● 11 pp.

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Olimund	Salimu 2010 (Oli mu Ndoolo) ● Songs - Modern / Sensitization ● 934 / 188 ● O ~ Transcription ● Jinja: CRC ● 0:06:08
Olugelsk	Various 2010 (Bible Story 1) ● Biblical documents / Religion ● 691 / 358 ● O ~ Transcription ● Internet: YouTube ● 0:08:48
Olukiiko	Wangoola, Paulo 2009 (Olukiiko Oluluubililia Okugaita Eitwale lya Busoga) ● Policy documents - Busoga Kingdom / History ● 1,710 / 764 ● W ~ Translation ● Kampala: Task Force on the Principled Cultural Unity - Busoga Kingdom ● 6 pp.
Olumbe	Malagala, Stephen 2000s (Olumbe) ● Songs - Traditional / Health ● 467 / 311 ● O ~ Transcription ● Anon.: - ● 0:06:01
Olumbel	Anon. 2000s (Olumbe Lulaile) ● Songs - Traditional / Health ● 507 / 215 ● O ~ Transcription ● Anon.: - ● 0:08:50
Olumbes	Anon. 2000s (Olumbe Siliimu) ● Songs - Traditional / Health ● 423 / 242 ● O ~ Transcription ● Anon.: - ● 0:09:34
Olusoga1	Gulere, Cornelius 2007 (Olusoga Olusookelwaku (Level 1)) ● Literature / Language ● 2,800 / 1,380 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 33 pp.
Olusoga3	Gulere, Cornelius 2007 (Olusoga Olusookelwaku (Level 3)) ● Literature / Language ● 3,015 / 1,418 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 36 pp.
Omudaal	Salimu 2000s (Omudaala) ● Songs - Traditional / Life ● 702 / 358 ● O ~ Transcription ● Anon.: - ● 0:11:37
Omugole	Bukenya, Austin 2007 (Omugole) ● Literature - Plays / Marriage ● 11,898 / 3,753 ● W ~ OCR ● Busembatya: Lusoga Language Academic Board (LLAB) ● 45 pp.
Omukazi	Gulere, Cornelius 2010 (Omukazi) ● Literature / Fables ● 227 / 154 ● W ~ OCR ● Internet: Anon. ● 1 p.
Omukonk	Geo Bless 2010s (Omukonkoonhia) ● Songs - Modern / Relationships ● 272 / 162 ● O ~ Transcription ● Anon.: - ● 0:04:38
Omulamu	Mugwisa 2000s (Omulamu Tiyeesigika) ● Songs - Traditional / Relationships ● 425 / 254 ● O ~ Transcription ● Anon.: - ● 0:07:58
Omulilo	Gulere, Cornelius 2007 (Omulilo) ● Literature / Fables ● 629 / 398 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 20 pp.
Omulyam	Kirimungu, Siragi 2000s (Omulya Mmele) ● Songs - Traditional / Life ● 493 / 284 ● O ~ Transcription ● Anon.: - ● 0:07:59
Omumbeed	Gulere, Cornelius 2007 (Omumbeedha Omutuufu) ● Literature / Fables ● 728 / 458 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 15 pp.
Omusaad	Anon. 2000s (Omusaadha Awalamula Egaali) ● Songs - Traditional / Money ● 135 / 55 ● O ~ Transcription ● Anon.: - ● 0:06:47
OmusAkb	Luboga, Sam 2012 (Omusoga Akoba) ● Literature / Fables ● 1,839 / 1,216 ● W ~ OCR ● Busembatya: Lusoga Language Academic Board (LLAB) ● 15 pp.
Omutmuz	Salimu 2010 (Omuntu Muzibu) ● Songs - Modern / Life ● 683 / 264 ● O ~ Transcription ● Jinja: CRC ● 0:05:07
OmuvangL	Kirimungu, Siragi 2000s (Omuvangano) ● Songs - Traditional / Rehabilitation ● 353 / 175 ● O ~ Transcription ● Anon.: - ● 0:07:01

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
OmuvngMB	Cultural Research Centre (CRC) 1999 (Omuvangano mu Busoga) ● Biblical documents / Religion ● 6,927 / 2,304 ● W ~ OCR ● Jinja: CRC ● 28 pp.
Omuzail	Mugwisa 2000s (Omuzaile Muwe Ekitiibwa) ● Songs - Traditional / Marriage ● 344 / 168 ● O ~ Transcription ● Anon.: - ● 0:07:28
OmwanaK	Gulere, Cornelius 2007 (Omwana Kwania) ● Literature / Fables ● 253 / 190 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 26 pp.
Omwenge	Salimu 2000s (Omwenge Seneta) ● Songs - Traditional / Rehabilitation ● 531 / 299 ● O ~ Transcription ● Anon.: - ● 0:11:20
Omwenka	Kabugu, Milton Peter 2010 (Omwenkanonkano) ● Songs - Modern / Sensitization ● 884 / 167 ● O ~ Transcription ● Jinja: CRC ● 0:07:14
OrderoM1	Cultural Research Centre (CRC) 2012 (Order of Mass 1) ● Biblical documents / Religion ● 1,846 / 903 ● W ~ OCR ● Namwendwa: Diocese of Jinja ● 15 pp.
OrderoM2	Cultural Research Centre (CRC) 2012 (Order of Mass 2) ● Biblical documents / Religion ● 872 / 434 ● W ~ OCR ● Jinja: Diocese of Jinja ● 16 pp.
Otabona	Mata, Nassani 2010 (Otabona Bukaile na Nvu) ● Songs - Traditional / Inspirational ● 574 / 179 ● O ~ Transcription ● Internet: YouTube ● 0:07:08
Otawuli	Kirimungu, Siragi 2000s (Otawulila Boogezi) ● Songs - Traditional / Life ● 813 / 409 ● O ~ Transcription ● Anon.: - ● 0:12:42
Otelaok	Gulere, Cornelius 2011 (Otela Okwila) ● Literature / Fables ● 315 / 229 ● W ~ OCR ● Internet: Google books ● 14 pp.
Owayang	Baisi 2010s (Owayanga) ● Songs - Traditional / Health ● 358 / 219 ● O ~ Transcription ● Anon.: - ● 0:08:14
PEAP	Ministry of Finance, Planning and Economic Development 2005 (Poverty Eradication Action Plan (PEAP)) ● Policy documents - Government / Sensitization ● 2,476 / 1,022 ● W ~ Translation ● Kampala: Makerere Institute of Languages ● 12 pp.
Petsg091	Various 2009 (Lusoga Songs Performed in Twin Ceremonies 1) ● Songs - Traditional / Marriage ● 3,214 / 864 ● W ~ Translation ● Jinja: - ● 13 pp.
Petsg092	Various 2010 (Lusoga Songs Performed in Twin Ceremonies 2) ● Songs - Traditional / Marriage ● 1,201 / 444 ● W ~ e-Transfer ● Jinja: - ● 14 pp.
PFExtaud	Nabirye, Minah 2012 (Phonetics Fieldwork - Extra Audio files) ● Interviews / Language ● 52,098 / 10,051 ● O ~ Transcription ● Busoga: - ● 8:55:38
PIbaale1	Pastor Ibaale 2010 (Pastor Ibaale 1) ● Radio talk shows / Religion ● 6,352 / 1,935 ● O ~ Transcription ● Jinja: NBS FM ● 0:59:24
PIbaale2	Pastor Ibaale 2010 (Pastor Ibaale 2) ● Radio talk shows / Religion ● 5,728 / 1,955 ● O ~ Transcription ● Jinja: NBS FM ● 1:01:23
PIbaale3	Pastor Ibaale 2010 (Pastor Ibaale 3) ● Radio talk shows / Religion ● 5,676 / 1,941 ● O ~ Transcription ● Jinja: NBS FM ● 0:59:42
PIbaale4	Pastor Ibaale 2010 (Pastor Ibaale 4) ● Radio talk shows / Religion ● 6,472 / 2,051 ● O ~ Transcription ● Jinja: NBS FM ● 1:00:29

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Pililya	Geo Bless 2010s (Pililiya) ● Songs - Modern / Inspirational ● 314 / 183 ● O ~ Transcription ● Anon.: - ● 0:05:13
Pl101021	Various 2010 (Plan Water and Sanitation 1) ● Radio talk shows / Health ● 9,534 / 2,843 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:29:37
Pl101215	Various 2010 (Plan Water and Sanitation 2) ● Radio talk shows / Health ● 5,500 / 1,961 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:49:26
Pl101216	Various 2010 (Plan Water and Sanitation 3) ● Radio talk shows / Health ● 5,900 / 1,990 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:50:48
PriestO1	Cultural Research Centre (CRC) 1998 (Priestly Ordination of Rev. Richard Kayaga Gonza, Rev. Silvester Makwali) ● Biblical documents / Religion ● 517 / 252 ● W ~ OCR ● Bugembe: Diocese of Jinja ● 26 pp.
PriestO2	Cultural Research Centre (CRC) 2003 (The Priestly Ordination for Rev. Deacon Mbaziira Henry Jude, Rev. Deacon Musana Paul) ● Biblical documents / Religion ● 2,071 / 794 ● W ~ OCR ● Bugembe: Diocesan Printery ● 34 pp.
PriestO3	Cultural Research Centre (CRC) 2003 (Priestly Ordination of Rev. Serapio Kasuura Wamara Araali) ● Biblical documents / Religion ● 1,343 / 640 ● W ~ OCR ● Bugembe: Diocese of Jinja ● 31 pp.
PriestO4	Cultural Research Centre (CRC) 2005 (Priestly Ordination of Deacon Mwangi Simon Gitua, Deacon Mugabe Paschal Atwooki, Deacon Jenga Fred) ● Biblical documents / Religion ● 1,446 / 575 ● W ~ OCR ● Jinja: Little Sisters of St. Francis ● 20 pp.
Publiche	School of Public Health (SPH) 2009 (Baseline Survey on Institutional Deliveries 1) ● Academic documents / Health ● 2,870 / 985 ● W ~ Translation ● Kampala: Makerere University School of Public Health ● 28 pp.
Queenw	Crado 2010s (Queen Wange) ● Songs - Modern / Relationships ● 210 / 101 ● O ~ Transcription ● Anon.: - ● 0:04:26
SafedelQ	School of Public Health (SPH) 2010 (Baseline Survey on Institutional Deliveries 2) ● Academic documents / Health ● 3,339 / 1,182 ● W ~ Translation ● Kampala: Makerere University School of Public Health ● 28 pp.
Safedelv	Various 2010 (Safe Deliveries) ● Radio talk shows / Health ● 8,453 / 2,661 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:11:05
Sente	Salimu 2000s (Sente) ● Songs - Traditional / Money ● 868 / 448 ● O ~ Transcription ● Anon.: - ● 0:13:32
Sentene	Kirimungu, Siragi 2000s (Sente n'Ekola) ● Songs - Traditional / Money ● 755 / 394 ● O ~ Transcription ● Anon.: - ● 0:08:33
Soyabean	Anon. 2010 (Soya Bean in Lusoga) ● Literature - Booklets / Science ● 2,287 / 968 ● W ~ Retyping of image ● Internet: - ● 32 pp.
StarEC1	Various 2010 (Star EC 1) ● Radio talk shows / Health ● 6,997 / 2,224 ● O ~ Transcription ● Jinja: NBS FM ● 0:52:46
StarEC2	Various 2010 (Star EC 2) ● Radio talk shows / Health ● 7,364 / 2,208 ● O ~ Transcription ● Jinja: NBS FM ● 0:58:01

`Filename`	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
`StarEC3`	Various 2010 (Star EC 3) ● Radio talk shows / Health ● 6,603 / 2,296 ● O ~ Transcription ● Jinja: NBS FM ● 0:53:05
`StarEC4`	Various 2010 (Star EC 4) ● Radio talk shows / Health ● 6,430 / 2,312 ● O ~ Transcription ● Jinja: NBS FM ● 0:58:35
`StarEC5`	Various 2010 (Star EC 5) ● Radio talk shows / Health ● 7,149 / 2,208 ● O ~ Transcription ● Jinja: NBS FM ● 0:58:28
`StarEC6`	Various 2010 (Star EC 6) ● Radio talk shows / Health ● 6,939 / 2,117 ● O ~ Transcription ● Jinja: NBS FM ● 1:00:36
`StarEC7`	Various 2011 (Star EC 7) ● Radio talk shows / Health ● 7,103 / 2,147 ● O ~ Transcription ● Jinja: NBS FM ● 0:58:09
`Strides1`	Various 2010 (Strides 1) ● Radio talk shows / Sensitization ● 6,859 / 2,257 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:59:31
`Strides2`	Various 2010 (Strides 2) ● Radio talk shows / Health ● 7,285 / 2,276 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 1:04:51
`Sunpanel`	Various 2011 (Sunday Panel) ● Radio talk shows / Politics ● 11,466 / 3,460 ● O ~ Transcription ● Jinja: NBS FM ● 1:28:50
`ThanksGv`	Cultural Research Centre (CRC) 2012 (Thanksgiving) ● Biblical documents / Religion ● 659 / 353 ● W ~ OCR ● Jinja: Diocese of Jinja ● 32 pp.
`Tubeepn`	Geo Bless 2010s (Tubeepene) ● Songs - Modern / Politics ● 452 / 265 ● O ~ Transcription ● Anon.: - ● 0:04:40
`Tuboin1`	Salimu 2000s (Tuboineboine 1) ● Songs - Traditional / Life ● 564 / 343 ● O ~ Transcription ● Anon.: - ● 0:09:55
`Tuboin2`	Salimu 2000s (Tuboineboine 2) ● Songs - Traditional / Life ● 745 / 414 ● O ~ Transcription ● Anon.: - ● 0:11:09
`Tusanga`	Anon. 2000s (Tusangaile) ● Songs - Traditional / Relationships ● 730 / 301 ● O ~ Transcription ● Anon.: - ● 0:10:42
`Twaghaya`	Baisi 2010s (Twaghanga) ● Songs - Traditional / Rehabilitation ● 554 / 293 ● O ~ Transcription ● Anon.: - ● 0:10:09
`Twebaze`	Salimu 2000s (Twebaze Katonda) ● Songs - Traditional / History ● 301 / 171 ● O ~ Transcription ● Anon.: - ● 0:08:20
`Twekubi`	Geo Bless 2010s (Twekubile Dance) ● Songs - Modern / Inspirational ● 355 / 179 ● O ~ Transcription ● Anon.: - ● 0:03:59
`TwireKBu`	Ssajabi, Sophronius 1999 (Twire ku Butaka) ● Literature / Fables ● 5,108 / 1,838 ● W ~ OCR ● Jinja: CRC ● 58 pp.
`TwoLusFb`	Gumbo, E. & Kafuho, E. 1946 (Two Lusoga Fables) ● Literature / Fables ● 1,325 / 791 ● W ~ Retyping of image ● Kampala: The Uganda Journal ● 16+24 pp.
`Vooto`	Anon. 2010s (Vooto) ● Songs - Modern / Relationships ● 338 / 70 ● O ~ Transcription ● Iganga: - ● 0:05:06
`Waalink`	Magoola, Racheal 2010 (Waalinkobye) ● Songs - Modern / Relationships ● 289 / 129 ● O ~ Transcription ● Internet: YouTube ● 0:04:58

Filename	Author *or* Performer Year *or* Period (Title) ● Genre / Topic ● Tokens / Types ● W(ritten) *or* O(ral) ~ Source ● Place: Publisher ● Pages *or* Length of recording
Walgund	Gulere, Cornelius 2007 (Walugundhu) ● Literature / Fables ● 380 / 280 ● W ~ OCR ● Internet: Mpolyabigere RC – RICED Center ● 9 pp.
WangoInt	Wangoola, Paulo 2012 (Interview on the Evolution of the Dialects of Busoga) ● Interviews / History ● 7,284 / 2,312 ● O ~ Transcription ● Kampala: - ● 1:18:05
Wankoko	Gulere, Cornelius 2007 (Wankoko ni Wamusota) ● Literature / Fables ● 554 / 288 ● W ~ e-Transfer ● Cape Town: CASAS (pre-publication) ● 15 pp.
Water1	Various 2009 (District Water 1) ● Radio talk shows / Health ● 5,077 / 1,743 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:44:49
Water2	Various 2009 (District Water 2) ● Radio talk shows / Health ● 5,138 / 1,705 ● O ~ Transcription ● Kamuli: Kamuli Broadcasting Service ● 0:45: 31
Weebale	Kiirya, Maurice 2010s (Weebale Okundoga) ● Songs - Modern / Relationships ● 274 / 131 ● O ~ Transcription ● Iganga: - ● 0:04:18
WorCulLe	Federation of Female Lawyers (FIDA) and Plan Uganda 2010 (Enkolagana y'Okutumbula Eidembe) ● Policy documents - NGOs / Marriage ● 4,936 / 1,562 ● W ~ Translation ● Kampala: FIDA and PLAN Uganda ● 16 pp.
WSGextr	Nabirye, Minah 2009 (Ebikookelo mu Eiwanika ly'Olusoga) ● Academic documents / Language ● 12,665 / 4,039 ● W ~ e-Transfer ● Kampala: Menha Publishers ● 79 pp.