Online Dictionaries on the Internet: An Overview for the African Languages*

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Ghent, Belgium and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: The main purpose of this research article is rather bold, in that an attempt is made at a comprehensive overview of all currently available African-language Internet dictionaries. Quite surprisingly, a substantial number of such dictionaries is already available, for a large number of languages, with a relatively large number of users. The key characteristics of these dictionaries and various cross-language distributions are expounded on. In a second section the first South African online dictionary interface is introduced. Although compiled by just a small number of scholars, this dictionary contains a world's first in that lexicographic customisation is implemented on various levels in real time on the Internet.

Keywords: LEXICOGRAPHY, TERMINOLOGY, DICTIONARIES, INTERNET, ONLINE, LOOK-UP MODE, BROWSE MODE, AFRICAN LANGUAGES, SESOTHO SA LEBOA, SIMULTANEOUS FEEDBACK, FUZZY SF, CUSTOMISATION

Senaganwa: Dipukuntšu tša online tše di lego mo Inthaneteng: Ponokakaretšo ya maleme a Afrika. Morero wo mogolo wa taodišwana ye ya nyakišišo ke wo o tiilego ka ge teko e dirilwe ka tebelelo ya kakaretšo ye e tletšego go dipukuntšu ka moka tša Inthanete tšeo di šetšego di le gona mo malemeng a Afrika. Sa go makatša ke gore go šetše go na le palo ye ntši ya dipukuntšu tše bjalo mo malemeng a mantši gape di na le badiriši ba bantši. Go hlaloša dipharologantšho tše bohlokwa tša dipukuntšu tše le ka moo diphatlalatšo di dirwago ka gona gare ga maleme a mantši a go fapana. Mo karolong ya bobedi go tsebišwa pukuntšu ya online ye e lego ya pele gape e lego ya makgonthe ya Afrika Borwa. Le ge e le gore pukuntšu ye e hlamilwe ke dirutegi di se kae, e šetše e tšea sefoka lefaseng ka bophara. Se ke ka lebaka la gore pukuntšu ye e dirilwe ka tsela yeo e lego gore dilo di ka beakanywa gore di itšweletše ka botšona gomme tša lokela batho ka moka bao ba e dirišago mo Inthaneteng ka yona nako yeo.

Mantšu a bohlokwa: TLHAMO YA DIPUKUNTŠU, TLHAMO YA MAREO, DIPUKUNTŠU, INTHANETE, ONLINE, MOKGWA WA GO NYAKA, MOKGWA WA GO LEKOLA, MALEME A AFRIKA, SESOTHO SA LEBOA, SIMULTANEOUS FEEDBACK, FUZZY SF, GO BEAKANYA DILO GORE DI BE KA MOKGWA WO O LEGO GORE O TLA GO LOKELA

* An earlier version of this article was presented at the mini-seminar on and launch of *The First South African Online Dictionary Interface*, held at the University of Pretoria, Pretoria, 20 June 2003.

African-language lexicography goes electronic

The face of dictionaries is changing. Rapidly. Whereas paper dictionaries still ruled the market until a decade ago, the large-scale commercial production of electronic dictionaries has boomed since the mid-1990s. Today, dictionaries on CD-ROM typically come in the back pocket of their hardcopy counterparts, while the number of dictionaries on the Internet already runs into tens of thousands. Lexicographers working on the South African languages should not merely watch the unrolling of these events from the sidelines; they *must* and *can* take an active part in writing the future. The wildest futuristic dreams revolve around multimedia Internet dictionaries, for which space restrictions disappear, and for which the output can be tailored to suit each unique user. The aim of this article is therefore twofold. On the one hand, the results are presented of a detailed study of current African-language Internet dictionaries. On the other, the first truly South African online dictionary interface is introduced and contrasted with what is already available on the Internet.

In South Africa, the *Woordeboek van die Afrikaanse Taal* (WAT), a multi-volume overall-descriptive historical dictionary with a paper past of three quarters of a century, recently went electronic. The entire text of the first eleven volumes was computerised and made available on CD-ROM in mid-2003. It is expected that subscription-based intranet and Internet versions will follow suit. The focus in this article will however be on the African languages, and more particularly on those languages that belong to the same language family as the official African languages of South Africa. Outside South Africa, these languages are known as '*Bantu* languages'. Given that this term is stigmatised in South Africa, this language family will be referred to as '*African* languages' below.

When embarking on this research project, colleagues wondered whether just two, a good twenty, or perhaps as many as two hundred African-language Internet dictionaries were already available. Probably the most surprising result of this study is that the actual number is closest to the last estimate. There are indeed nearly two hundred of them, for nearly one hundred and twenty *different* African languages. Considering that there are roughly five to six hundred languages that belong to this family, this means that one fifth are represented. The downside of this extremely positive and, admittedly, surprising outcome of this study is that the sizes of the current African-language Internet dictionaries are generally small, and the contents not often of a high quality.

Guthrie's classification and beyond

Over half a century ago, Malcolm Guthrie 'identified' and 'classified' the languages that are the focus of this article (Guthrie 1948). He used two main and two subsidiary linguistic criteria to identify the languages, and came to the conclusion that this language family covers that part of the African continent

starting from an imaginary line north of the current Democratic Republic of the Congo (DRC) all the way down to the southern tip of the African continent. Roughly speaking, only the languages spoken in the Cape region and north of it (Afrikaans and the Khoesaan languages) do not belong to this family. He then set out to classify all the languages within this region, a classification mainly based on geographical contiguities, and much less on linguistic features. The result consisted of 16 'zones' covering nearly 80 'groups'. The zones start in the northwest (A), go to the northeast (B, C, D and E), then south (F and G), again from west to east (H, K, L, M, N and P), and once more from west to east (R, S and T). These zones are made up of groups (A10, A20, ...; ...; M10, M20, ...; ...), with each group bringing together so-called related languages (A11, A12, ..., A21, A22, ...; ...; M11, M12, ..., M21, M22, ...; ...). Over the years he extensively revised zones A, B and C (Guthrie 1953), and also — all of a sudden, but apparently in response to criticism (Cope 1971: 218) — collapsed the Southern African zones S and T into a single zone S. Guthrie's 'final' classification can be found in the third volume, pages 11 to 15, of his Magnum Opus (Guthrie 1967, 1971, 1970, 1970a).

In Tervuren, Belgium, which soon became the mecca of Central-African language studies, a new zone was introduced around the region of the Great Lakes, zone J, consisting of Guthrie's groups E10, E20 and E30, as well as of sections of D40, D50 and D60. The numbering was simply transferred to J10 up to J60 respectively (Bastin 1978). In order to distinguish between neighbouring languages/dialects, extra letters are sometimes added (e.g. L31a for Cilubà spoken by the Balubà, L31b for Cilubà spoken by the Beena Luluwà, etc.). Since Guthrie, some languages have become extinct, while previously undocumented ones have been documented. Languages not originally in Guthrie's list mostly start with the linguistic group to which the extra language seems to be most affiliated, say E40, to which a third digit is added, e.g. E402 for Ikizu. At least, the latter is done by most scholars, such as for instance Lowe and Schadeberg (1996) or Maho (2003).

Nonetheless, in both the Guthrie and Tervuren checklists, the same code sometimes covers different languages. Furthermore, not everyone uses the Tervuren zone J. The result of this state of affairs is that there is considerable confusion as to which language has which code, and *vice versa*. Moreover, many languages often have numerous alternate spellings and/or are simply referred to by means of different names. The existence and status of dialects further complicate the issue. The exact number and location of languages is therefore still not known half a century after Guthrie's pioneering work, yet one generally accepts that there are at least five hundred and less than six hundred. Given all this confusion, it is obviously not truly possible to quantify any claims regarding this family of languages. For one, there is not even a fixed upper limit.

Apart from Guthrie's final classification, and Tervuren's latest checklist (Bastin, Coupez and Mann 1999), there is also a third classification that is often

consulted, viz. the one found in *Ethnologue* (Grimes and Grimes 2000). A highly useful comparison of the three classifications was compiled by Maho (2002). In the discussion below, however, certain decisions had to be made in order to provide for a scientific framework. These decisions were as follows: (1) *Ethnologue* was used as arbiter on language names, (2) the codes for the languages were mainly taken from the Tervuren checklist, (3) wherever Guthrie's data seemed more precise, his 'language name + language code' pair was kept, and (4) where applicable, the current official language names overruled the above.

Internet dictionaries for the South African languages

Now that the term 'African languages' has been delimited for the purposes of this article, one can turn to the concept 'Internet dictionaries'. Such reference works form part of the larger family of human-oriented electronic dictionaries and, within a three-step access dictionary typology, can be characterised as reference works for which 'users worldwide use laptops/desktops to access a dictionary stored on an online server' (De Schryver 2003: 151). Reformulated, these are thus *online* dictionaries for which the data are stored in databases, no matter where these databases are located, and which can be consulted from a *search screen* by anyone from anywhere through the *Internet*. Intranet dictionaries, another type of online electronic dictionary, will thus not be considered. For convenience, however, the terms 'online dictionary' and 'Internet dictionary' are used interchangeably in this article. A comprehensive overview of the features of the various electronic dictionaries, as well as a detailed discussion of their advantages over paper dictionaries, can be found in De Schryver (2003). Suffice it to say here that an electronic dictionary is much more than 'a dictionary in electronic form'. At the very least, the data are stored in a database, to which various (search) indexes are added, with a multitude of links to multimedia, as well as, increasingly, Natural Language Processing (NLP) extensions.

Rather surprisingly, these various aspects *already* exist for some of the African languages spoken in South Africa, albeit not yet all together in one integrated Internet dictionary package. An online dictionary for Tshivenda (S21), for example, is available from *CBOLD*. It contains 8 900 lemma signs, all of them searchable from a search screen, yet only with textual output. Sound files were added to various basic travellers' phrases for Sesotho (S33), among others, at *TravLang*, while full multimedia (i.e. text, audio and computer graphics) can be found at *eLanguage* for isiZulu (S42). Lastly, an example of an online NLP aspect that has been developed for a South African language is the machine translation (MT) system running between isiXhosa (S41) and English at *Xhosa on the Web!* (O'Kennon 1996–2003).

As argued by Varantola (2002: 35) and De Schryver (2003: 167, 169-172), multimedia corpora will increasingly become part and parcel of future electronic dictionaries. This NLP aspect does not yet exist for South African languages, but across the border Internet-searchable *text* corpora are already

available for ChiShona (S11-S12-S14) and SiNdebele (or Zimbabwean Ndebele (S44)). These online corpora of respectively 2.2 million and 0.7 million running words were originally assembled with dictionary compilation in mind, and have now been made available to the wider linguistic community (Ridings 2002).

Although most of the online dictionaries for South African languages have been online for quite some years now, it is somewhat disturbing to note that relatively few people know about their existence. Apart from the fact that the full Internet potential is not used within a single integrated package in any one of them, one of the reasons for their shadowy presence could be that none of these existing online dictionaries was made *in* South Africa, *by* South Africans, *for* South Africans. All these aspects are niches that can be filled by prospective lexicographers, besides the fact that such lexicographers can of course also improve on current size, quality and functionality.

A systematic overview of online African-language dictionaries

In this section, a systematic overview will be presented of currently available Internet dictionaries for the African languages. One immediately notices an uneasy balance between the concepts 'currently available' and 'Internet' here. Indeed, the Internet being an organic medium, its contents literally change every single second. One must therefore put a timestamp on the study, with all claims referring to that time frame. The timestamp is 'April 2003', as this is the period during which the Internet was trawled (with the help of search engines such as *Google*) to trace all available African-language Internet dictionaries. What follows is a summary and a discussion of the main findings, with all claims thus 'valid' for April 2003.

Before the results themselves are presented, it is important to recall that 'Internet dictionaries' in this study are *only* those online dictionaries that can be accessed from a search screen. This thus means that one must be able to type in words or sections of words, potentially including wildcards, followed by a mouse click or 'enter', upon which one or more articles are presented ensuing a page-reload. Based on this premise, the following two types of dictionaries that can be found *en masse* on the Internet have not been included in this study: (1) dictionaries in pdf (Portable Document Format), word processor, or any other downloadable text format — such as for instance Odden's (2002) *Kikerewe–English Dictionary* (J24) in pdf; and (2) dictionaries which are simply plain online HTML (HyperText Markup Language), or HTML-like, files — such as *Ikuska Libros*'s (1997–2003) *Diccionario Lingala–Español–Lingala* (C36d) in HTML, or dictionaries such as those from the *TravLang* series mentioned above, which have no search facilities and can only be 'browsed'.

The following three types, on the other hand, were considered for this study: (1) *online* dictionaries, i.e. dictionaries stored in databases over the Internet; (2) *pop-up* dictionaries, i.e. dictionaries with which, once one has down-

loaded a small piece of software, one can move the mouse over words online, upon which the relevant articles pop up in dedicated screens; and (3) *PC* dictionaries, i.e. dictionaries for which a piece of software *cum* one or more lexica are downloaded from the Internet, to be used as offline PC dictionaries. Note that the lexica in (2) can also be downloaded to the hard drive of a PC, at which point they become, in addition, functional as *offline* pop-up PC dictionaries.

Following the investigation, an impressive number of 182 African-language Internet dictionaries were found, 165 of the 'online' type, 8 of the 'pop-up' type, and 9 of the 'PC' type. All major characteristics of these 182 dictionaries have been tabulated in the Appendix, and as such this appendix — which is sorted by the *names* of the languages — should be considered the basis of the analysis that is to follow. These 182 dictionaries cover 117 *different* languages, as well as Common Bantu (CB) and Proto Bantu (PB). PB is the hypothetical language to which all current languages within this family can be traced back, while CB are the c. 2 800 series of comparative forms that were used by Guthrie to reconstruct PB. The distribution of the number of Internet dictionaries per language is as follows: Swahili (G42): 20 x, Chagga (E62): 14 x, Lingala: 5 x, Ganda (J15) and isiZulu: 4 x each, Meru (E61): 3 x, 18 other languages + PB: 2 x each, and 93 other languages + CB: 1 x each. As for many other real-world phenomena, one notices a Zipfean distribution, i.e. the number of Internet dictionaries is extremely high for just a small number of languages, while the frequency for the great majority is very low. That there are relatively many dictionaries for languages such as Swahili, Lingala and isiZulu is understandable; these are the languages that also receive much academic (and other) attention. That a language such as Chagga scores high, however, is out of proportion.

Indeed, there is some serious skewing in the geographical dispersion as a result of one single source that contains over a hundred African-language Internet dictionaries. In the early 1970s Derek Nurse and Gérard Philippson surveyed the languages of Tanzania and neighbouring countries — their study is known as the *Tanzania Language Survey* (TLS, Nurse and Philippson 1975) — and this resulted in 124 parallel c. 1 000-word wordlists. For some of the languages, however, different dialects were recorded — in the case of Chagga, 14, in the case of Meru, 3, etc. In all, there are lexica for 97 *different* languages, as well as one for PB and one for English. Given this, it is thus clear that there is a significant bias towards the languages of Tanzania and East Africa. The fact moreover that Swahili is mainly spoken in Tanzania, pushes the distribution even more into that region of the African continent.

Despite the bias, and despite the small size of the TLS lexica, they are as a whole an interesting application of the *hub-and-spoke model* (Martin 1996: 209, 214). Indeed, with English/Swahili as hub, all the other 122 lexica are linked to it as spokes, and as a result an online dictionary for each and every language pair, triple, quadruple, etc. can now also be 'created', passing through the hub. The number of permutations, and thus the potential number of different *multilingual* dictionaries one can generate in this way, is virtually unlimited. The

basic hub-and-spoke framework is actually becoming ever more popular online for dictionaries involving the languages used in the European Community (EC). In one set of applications, viz. *Ergane* and *Majstro*, Esperanto was chosen as hub with, besides mostly EC languages, Swahili, isiZulu and Setswana (S31) as spokes. From a sound metalexicographic point of view, there are many good reasons to have reservations when it comes to the hub-and-spoke model. Yet choosing an artificial language as hub, thus one where the level of polysemy is virtually non-existent, definitely goes some way to avoid a number of the theoretical problems.

While learners might find it most useful that English was included as one of the parallel lexica of TLS, comparative linguists surely appreciate the fact that Guthrie's PB reconstructions were also added, so that reflexes across the various languages can be directly compared. From the time when Guthrie worked on PB, reconstructions have mainly been drawn up in Tervuren, with Meeussen's (1980, based on a manuscript from 1969) *BLR* and Coupez, Bastin and Mumba's (1998) *BLR 2* the two major releases so far. BLR 2, with 9 800 reconstructed forms, is the backbone of the ambitious *CBOLD* project, originally located in Berkeley, now transferred to Lyon. This research team collected a manifold of dictionaries, mostly as downloadable text files only however, and containing many errors resulting from the use of optical character recognition (OCR) on poor-quality scans. As pointed out above, such dictionaries have not been considered in this study. A total of 22 other dictionaries, as well as BLR 2, *can* be queried online though. Reconstructions to PB for these dictionaries, with BLR index numbers and Guthrie codes, is still ongoing. Note that, at the time of writing, a web site dedicated to BLR 3 is in the making (Bastin et al. 2003).

The CBOLD web site also houses the TLS data, which effectively makes this single site the 'major collection', at least quantity-wise, with 146 online dictionaries for 111 different languages and 2 for PB. In April 2003, the largest Internet dictionary for this language family, however, was located at Yale University, where *The Kamusi Project* contained 58 038 Swahili and 58 041 English 'articles' (Kamusi 1994–2001). These values were arrived at by simply counting the number of entries, and do not reflect the true sizes since a new entry is used for each new synonym, for each new sense, etc. If the number of truly unique lemma signs is summed, regardless of part of speech (POS), then the Swahili to English side turns out to contain 18 411 items, and the English to Swahili side 26 970 items. This dictionary is a prototypical example of bottom-up lexicography (Carr 1997: 214), which means that it is being compiled by Netizens. The contents should thus be consulted with caution.

The second-largest online African-language dictionary, for Lozi (S34), contains 24 000 items. Then follow dictionaries for ChiShona with 15 000 items, for Nyankore (J13) with 12 500 items, etc. At the other end of the spectrum, some of the online dictionaries contain as few as 100 items (for Ganda), 186 items (for Setswana), 300 items (for Lingala), etc. The average number of items in the 182 online African-language dictionaries is 1 978.

It has already been pointed out that the 182 dictionaries cover 117 different languages. Many of these languages are spoken across country borders, such as Chewa (N31b) which is spoken in both Malawi and Botswana, or Fipa (M13) in Tanzania and Malawi, Luyia (J32) in Kenya and Uganda, Yaka (H31) in the DRC and Angola, etc. If one studies the distribution of the number of languages that have online dictionaries *per country*, the data shown in Table 1 are arrived at.

Table 1: Distribution of the number of African languages with Internet dictionaries per country

#C	Country	#Ls with Ds	%
1	Angola	1	0.66
2	Botswana	3	1.99
3	Burundi	2	1.32
4	Congo	2	1.32
5	DRC	6	3.97
6	Gabon	2	1.32
7	Kenya	14	9.27
8	Malawi	7	4.64
9	Mozambique	5	3.31
10	Namibia	2	1.32
11	Rwanda	2	1.32
12	South Africa	4	2.65
13	Tanzania	81	53.64
14	Uganda	10	6.62
15	Zambia	5	3.31
16	Zimbabwe	5	3.31
		151	100.00

From Table 1 it is clear that the greatest allocation is once more to be found in Tanzania, with as many as 81 languages covered. Neighbouring countries such as Kenya with online dictionaries for 14 languages, and Uganda for 10 languages, also score high. In Southern Africa, countries like Zambia, Zimbabwe, Malawi and Mozambique, each cover more languages than South Africa, where there are but 4 languages with Internet dictionaries.

Based on the data found in *Ethnologue*, the 117 languages are spoken by over 100 million people. The dispersion once more moves between extremes. At one extreme, some languages covered are nearly extinct (Geviya (B30)), or are spoken by only a few (Zalamo (G33)), up to a few thousand people (Mpongwe (B11a), Kahe (E64), etc.). At the other extreme, some languages are spoken as primary language by over 5 million (Swahili, Sukuma (F21) and Gikuyu (E51)), over 6 million (Rundi (J62) and isiXhosa), over 7 million (ChiShona and Rwanda (J61)), up to over 9 million (isiZulu) people. Very roughly speaking, the average number of primary speakers per language for which there is at least one Internet dictionary is 1 million.

If one looks at dictionary typology, one notices that all but one of the 182 dictionaries is bilingual or multilingual. The only monolingual dictionary is the *Duramazwi ReChiShona* 'General Shona Dictionary' (Chimhundu 1999). Ironically, however, the interface of this monolingual dictionary is entirely in English. A full breakdown of the gloss and/or hub languages is shown in Table 2.

Table 2: Breakdown of the gloss and/or hub languages for all African-language Internet dictionaries

Gloss/Hub language	**#**	**%**
English/Swahili	123	67.58
English	25	13.74
French	17	9.34
Esperanto	7	3.85
English/French	4	2.20
German	3	1.65
English/etc. (9 in all)	1	0.55
Italian	1	0.55
— (*monolingual*)	1	0.55
	182	100.00

As one could have expected, roughly nine out of ten dictionaries use English, and only one out of ten use French as the gloss/hub language. Unexpectedly, however, is the relatively large number of dictionaries that involve Esperanto.

None of the 182 dictionaries is stored on a computer in Africa. Even the electronic version of the *Duramazwi ReChiShona* was developed by The Norwegian Documentation Project, and is stored on a server in Oslo. Moreover, very few Africans were involved in the computerisation and creation of these online dictionaries. If one studies the various providers, one notices a clear bias towards academic institutions, which are responsible for eight out of every ten dictionaries. Dotcoms provide one out of seven dictionaries, and less than five percent are personal efforts. The exact distribution has been calculated in Table 3.

Table 3: Providers of African-language Internet dictionaries

Provider	**#**	**%**
Academic	149	81.87
Dotcom	24	13.19
Personal	9	4.95
	182	100.00

In general, the soundest contents can be found for the Internet dictionaries compiled by academics, while the most versatile and appealing interfaces are those brought together by dotcoms. The average compilation year is 1981, with the distribution per decade as listed in Table 4.

Table 4: Number of African-language Internet dictionaries compiled per decade

Decade	#Ds	%
1930s	1	0.55
1940s	—	—
1950s	3	1.65
1960s	2	1.10
1970s	127	69.78
1980s	—	—
1990s	18	9.89
2000s	30	16.48
s.d.	1	0.55
	182	100.00

The number of users of the current online dictionaries is much higher than anticipated. For Swahili, for example, *The Kamusi Project* has received over 1.1 million visitors since mid-1995, the *Freedict* dictionary handles 700 visitors per day, while the *Kamusi Kiswahili–Kiesperanto* (Vessella 2001) is accessed at least 1 000 times per month. The online pop-up dictionaries for African languages available from *Babylon* have an average number of 1 400 users each. Lastly, *Xhosa on the Web!* (O'Kennon 1996–2003) has welcomed nearly 30 000 visitors so far.

The first South African online dictionary interface

From the overview presented above, at least two conclusions can be drawn. On the one hand, African-language lexicographers will have to admit that quite a substantial body of Internet dictionaries is *already* available. On the other hand, and this primarily from a South African perspective, one cannot deny the fact that the South African languages should and could be better represented as far as Internet dictionaries are concerned. *TshwaneDJe*, a Human Language Technology (HLT) development team, based in Pretoria and consisting of David Joffe, Gilles-Maurice de Schryver, D.J. Prinsloo and Salmina Nong, therefore decided to bring together all the material for the first South African Internet dictionary.

The choice fell on Sesotho sa Leboa (S32) as the first language for which to compile a dictionary, given that no online dictionaries were found for this language during the course of the Internet study summarised above. The gained expertise would then be applied to the compilation of other African-language Internet dictionaries. The starting point was Prinsloo and De Schryver's (2000) *SeDiPro 1.0*, a Sesotho sa Leboa to English dictionary available to the team in Microsoft Word format. Joffe wrote a parser to transfer the data to *TshwaneLex*, a novel and professional South African software application for dictionary compilation (Joffe, De Schryver and Prinsloo 2003, 2003a). TshwaneLex was

designed in such a way that it can be used to produce hardcopy, CD-ROM as well as online dictionaries. On 22 April 2003, the first version of an *Online Sesotho sa Leboa–English Dictionary* was uploaded (De Schryver and Joffe 2003). Two months later, on 20 June 2003, the online dictionary was officially launched at the University of Pretoria.

Between the first upload and the launch, several adaptations were made and numerous extra features were added to the online dictionary. As such this dictionary is a direct implementation of the concept known as *Simultaneous Feedback* (De Schryver and Prinsloo 2000, 2000a), a methodology whereby especially *indirect feedback* is near-instantly 'fed back' into the compilation process of a dictionary. The lexicographic contents are currently being updated by Nong.

During the first two months, users primarily learned about the new online dictionary through word of mouth. On the eve of the launch, 366 different users had searched for 3 341 items, or on average 9.12 searches per person. This was equivalent to more than 50 searches by more than 7 different users per day. The first media release appeared two weeks later, on 4 July 2003 (cf. e.g. *Mail and Guardian Online* 2003). At the end of that day, the number of searches had already reached 5 779 by 802 different users, or an average of 78.09 searches by 12.15 persons per day. The great majority of these searches had been made from hosts *in* South Africa. This clearly exceeded even the wildest expectations at TshwaneDJe.

From a metalexicographic perspective, this online dictionary deserves some extra discussion. Firstly, it is the first African-language Internet dictionary that can be accessed in all languages covered by the dictionary. In this case, this means that all interface pages are available in both Sesotho sa Leboa and English. Primary speakers of Sesotho sa Leboa can thus for the first time consult a dictionary in their own language.

Secondly, although actually only the direction Sesotho sa Leboa to English exists, an English search index (which also includes support for multi-word units) has been added which makes it possible to search the dictionary as if the reverse side were also available. The layout of the output is also a first, as it shows how the senses in one language are spread all over the lexicon in another, and how these then again spread out, etc. With 24 921 items on the Sesotho sa Leboa side and 28 198 in the English index, this online dictionary becomes the largest African-language Internet dictionary.

Thirdly, besides a general-language dictionary, this is also the first online dictionary that includes a dedicated terminology list for an African language. The terminology list that has currently been added is one for linguistics, containing over 300 terms, and more terminology lists are planned.

Fourthly, when consulting the terminology list, users can choose between look-up and browse mode. This is thus an original implementation of Atkins's (1996) innovative view of future electronic dictionaries. According to her, "the user is in search of a specific piece of information" in look-up mode, while "a more relaxed reading takes place" in browse mode (1996: 529). In look-up mode

users are furthermore re-routed from (potentially) incorrectly to correctly spelled items for words involving the letters s/š, e/ê and o/ô.

Lastly, and also most importantly, the terminology list contains a world's first for an online dictionary, namely the customisation of the output of part-of-speech (POS) tags, usage labels and cross-references depending on the language chosen. As such, this is the first step towards one concept of the dictionary of the future, viz. *Fuzzy SF* (De Schryver and Prinsloo 2001). In Fuzzy SF, or Fuzzy Simultaneous Feedback, "log-file based Artificial Intelligence components enable the implicit retrieval of personalised user feedback with which the package customises each user's own and unique dictionary" (De Schryver 2003: 189).

Conclusion

In this article a near-exhaustive overview was presented of the current state-of-the-art of African-language Internet dictionaries. The concepts 'African languages' and 'Internet dictionaries' were first defined for the purposes of this article. All currently available African-language Internet dictionaries were then reviewed, listed and compared to one another. Various statistics were calculated and distributions shown, from which one may conclude that there is a geographic bias towards the languages of East Africa, especially Tanzania. Among the most successful implementations one must count the hub-and-spoke model as used for the presentation of the data from the *Tanzania Language Survey*, now part of the *CBOLD* web site.

A surprising number of 182 dictionaries were uncovered, for 117 different languages. The South African share was shown to be small. Although an estimated 100 million people speak the languages covered, just one of the dictionaries is a monolingual one. None of the dictionaries is stored in Africa, and few Africans contributed to the computational creation of these dictionaries. Most dictionaries are the output of academic institutions, are relatively recent, and have a higher-than-expected number of users. The most popular dictionaries are those for Swahili, for which there are as many as 20.

In order to turn the relatively inactive online lexicographic tide for the languages spoken in South Africa, it was indicated how the HLT development team *TshwaneDJe* decided to produce the first truly South African online dictionary interface. The language embarked upon is Sesotho sa Leboa. Compilation is undertaken within the frameworks of *Simultaneous Feedback* (SF) and *Fuzzy SF*, and it was shown how, in less than three months, the number of searches and users had already reached unexpected heights. The dictionary is currently the largest online African-language Internet dictionary. Among the novelties of the online Sesotho sa Leboa dictionary, the dual dictionary interface language (including the first in an African language), a layout inherently departing from an African language, the first searchable African-language Internet terminology list, the optional look-up and browse modes, as well as

the first steps towards user customisation, were highlighted. As such, South African lexicography is already writing the future.

References

Atkins, B.T. Sue. 1996. Bilingual Dictionaries: Past, Present and Future. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Röjder Papmehl (Eds.). 1996. *Euralex '96 Proceedings I–II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.

Babylon. 1997–2003. Online Dictionary Software by Babylon [online]. Available: <http://www.babylon.com/>.

Bastin, Yvonne. 1978. Les langues bantoues. Barreteau, Daniel (Ed.). 1978. *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur Madagascar*: 123-185. Paris: Conseil international de la langue française.

Bastin, Yvonne, André Coupez and Michael Mann. 1999. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study*. Human Sciences Annals of the Royal Museum for Central Africa, Series in -8°, 162. Tervuren: RMCA.

Bastin, Yvonne, André Coupez, Evariste Mumba and Thilo C. Schadeberg. 2003. *Reconstructions lexicales bantoues 3 / Bantu Lexical Reconstructions 3* [online]. Available: <http://linguistics.africamuseum.be/BLR3.html>.

Carr, Michael. 1997. Internet Dictionaries and Lexicography. *International Journal of Lexicography* 10(3): 209-230.

CBOLD. 1997–2003. Comparative Bantu OnLine Dictionary [online]. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/>.

Chimhundu, Herbert (Ed.). 1999. *Duramazwi ReChiShona* (General Shona Dictionary) [online]. Available: <http://www.dokpro.uio.no/allex/gsd.html>.

Cope, Anthony T. 1971. A Consolidated Classification of the Bantu Languages. *African Studies* 30(3-4): 213-236.

Coupez, André, Yvonne Bastin and Evariste Mumba. 1998. *Reconstructions lexicales bantoues 2 / Bantu Lexical Reconstructions 2* [online]. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/>.

De Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

De Schryver, Gilles-Maurice and David Joffe. 2003. *Online Dictionary Interface Sesotho sa Leboa (Northern Sotho)–English* [online]. Available: <http://africanlanguages.com/sdp/>.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000. Dictionary-Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth Euralex International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 197-209. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000a. The Concept of 'Simultaneous Feedback': Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2001. Fuzzy SF: Towards the Ultimate Customised

Dictionary. *Studies in Lexicography* 11(1): 97-111. [Also published in Lee, Sangsup (Ed.). 2001. *ASIALEX 2001 PROCEEDINGS, Asian Bilingualism and the Dictionary*: 141-146. Seoul: Center for Linguistic Informatics Development, Yonsei University.]

eLanguage. 2002–2003. Riverdeep Interactive Learning Limited [online]. Available: <http://www.elanguage.com/eng/default.asp>.

Ergane. 2003. Ergane Free Multi-lingual Dictionary [online]. Available: <http://www.travlang.com/Ergane/>.

Freedict. 2003. Free Online Swahili–English Dictionary [online]. Available: <http://www.freedict.com/onldict/swa.html>.

Freelang. 1997–2003. Free Dictionaries to Download and Free Human Translation [online]. Available: <http://www.freelang.net/>.

Glossword. 2002–2003. Glossword Development [online]. Available: <http://dev.glossword.ru/?il=en>.

Google. 2003. Google: Internet Search Engine [online]. Available: <http://www.google.com>.

Grimes, Barbara F. and Joseph E. Grimes (Eds.). 2000. *Ethnologue. Volume 1: Languages of the World. Volume 2: Maps and Indexes* (14th Edition) [online]. Dallas: SIL International. Available: <http://www.ethnologue.com/>.

Grouselle, Pascal. 2002–2003. *Initiation au lingala, langue bantoue* [online]. Available: <http://perso.wanadoo.fr/pascal.grouselle>.

Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. London: Oxford University Press.

Guthrie, Malcolm. 1953. *The Bantu Languages of Western Equatorial Africa*. London: Oxford University Press.

Guthrie, Malcolm. 1967. *Comparative Bantu 1: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg Press.

Guthrie, Malcolm. 1970. *Comparative Bantu 3: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International Publishers.

Guthrie, Malcolm. 1970a. *Comparative Bantu 4: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International Publishers.

Guthrie, Malcolm. 1971. *Comparative Bantu 2: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International Publishers.

Ikuska Libros. 1997–2003. Diccionario Lingala–Español–Lingala [online]. Available: <http://www.ikuska.com/Africa/Lenguas/lingala/diccionario.htm>.

Joffe, David, Gilles-Maurice de Schryver and D.J. Prinsloo. 2003. Introducing TshwaneLex — A New Computer Program for the Compilation of Dictionaries. De Schryver, Gilles-Maurice (Ed.). 2003. *TAMA 2003 South Africa: CONFERENCE PROCEEDINGS*: 97-104. Pretoria: $(SF)^2$ Press.

Joffe, David, Gilles-Maurice de Schryver and D.J. Prinsloo. 2003a. Tree Structure, I/O Architecture, Database Loading and Cross-referencing in the Dictionary Application "TshwaneLex". *Southern African Linguistics and Applied Language Studies* 21(3).

Kamusi. 1994–2001. The Kamusi Project. Internet Living Swahili Dictionary [online]. Available: <http://www.yale.edu/swahili/>.

Lowe, John B. and Thilo C. Schadeberg. 1996. *Bantu MapMaker version 3.1* [online]. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/Maps/BMM1.html>.

Maho, Jouni F. 2002. *Bantu Line-up: Comparative Overview of Three Bantu Classifications* [online].

Gothenburg: Department of Oriental and African Languages, Göteborg University. Available: <http://www.african.gu.se/maho/jfmpub.html>.

Maho, Jouni F. 2003. *The Languages of Tanzania: A Web Links Collection* [online]. Available: <http://www.african.gu.se/tanzania/weblinks.html>.

Mail and Guardian Online. 2003. Online Sotho–English Dictionary Launched [online]. Friday, July 4, 2003; South African Press Association. Available: <http://www.mg.co.za/Content/l3.asp?ao=16666>.

Majstro. 2003. A Multilingual Translation Dictionary [online]. Available: <http://www.majstro.com/>.

Maniacky, Jacky. 2003. *Bantu Web Site, Site Web Bantu* [online]. Available: <http://www.bantu.ovh.org>.

Martin, Willy. 1996. Lexicographical Resources in a Multilingual Environment: An Orientation. *Lexikos* 6: 199-214.

Meeussen, Achille E. 1980 [1969]. *Bantu Lexical Reconstructions.* Archives d'Anthropologie 27. Tervuren: RMCA.

Nurse, Derek and Gérard Philippson. 1975. The North-East Bantu Languages of Tanzania and Kenya: A Classification. *KiSwahili* 45(2): 1-28.

Odden, David. 2002. *Kikerewe–English Dictionary* [online]. Available: <http://www.ling.ohio-state.edu/~odden/>.

O'Kennon, Martha. 1996–2003. *Xhosa on the Web! Xhosa to English and English to Xhosa Translator* [online]. Available: <http://mokennon.albion.edu/>.

Prinsloo, D.J. and Gilles-Maurice de Schryver (Eds.). 2000. *SeDiPro 1.0, First Parallel Dictionary Sepêdi–English*. Pretoria: University of Pretoria.

Ridings, Daniel. 2002. *Shona and Ndebele Corpus Resources from the African Languages Lexicon Project (ALLEX)* [online]. Available: <http://www.ling.gu.se/~ridings/africanlang.html>.

TLS. 1999 [1970s]. Tanzania Language Survey [online]. WWW version 1.0. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/Docs/TLS.html>.

TravLang. 1995–2003. Foreign Languages for Travellers [online]. Available: <http://www.travlang.com/languages/>.

Varantola, Krista. 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility? Corréard, Marie-Hélène (Ed.). 2002. *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins*: 30-44. EURALEX. Information: <http://www.ims.uni-stuttgart.de/euralex/>.

Vessella, Nino. 2001. *Kamusi Kiswahili–Kiesperanto / Vortaro Esperanto–Suahila* [online]. Available: <http://kamusi.vessella.net/>.

WAT. 1926– . Woordeboek van die Afrikaanse Taal. Information: <http://www.sun.ac.za/wat/>.

Appendix: African-language Internet dictionaries – An overview (April 2003)

#D	#L	Language/Dialect	Guthrie code	Country(ies)	Speakers	Type	Lemma signs	Gloss/Hub language	Compiler(s)	Year	Provider
1	1	Bena	G63	Tanzania	568 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
2	2	Bende	F12	Tanzania	20 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
3	3	Bondei	G24	Tanzania	80 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
4						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
5	4	Bukusu	J31c	Kenya	565 000	Online	5900	English	KWL	1998	CBOLD
6						Online	900	English	Deblois	*s.d.*	CBOLD
7	5	Bungu (Wungu)	F25	Tanzania	36 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
8						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
9	6	Chagga (Keni, Kibosho, Kimochi, Kiseri, Lema, Machame, Mamba, Mkuu, Siha, Vunjo)	E62	Tanzania	400 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
10						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
11						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
12						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
13						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
14						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
15						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
16						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
17						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
18						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
19						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
20						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
21						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
22						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
23	7	Chewa	N31b	Malawi, Botswana	3 958 000	Online	6200	English	Scott & Hetherwick	1957	CBOLD
24	8	Chiga (Rukiga)	J14	Uganda	1 391 442	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
25	9	ChiShona	S11-S12-S14	Zimbabwe, Zambia	7 000 000	Online	15000	— (*monolingual*)	Chimhundu	1999	Norwegian Doc. Project
26	10	Chuka	E53	Kenya	70 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
27	11	Doe	G301	Tanzania	24 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
28	12	Embu	E52	Kenya	429 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
29	13	Fipa	M13	Tanzania, Malawi	200 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
30	14	Ganda (Luganda)	J15	Uganda, Tanzania	3 025 000	Online	6300	English	Snoxall	1967	CBOLD
31						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
32						PC	100	French	Ramel	2001	Freelang
33						PC	100	(French →)	Ramel	2001	Freelang
34	15	Geviya	B30	Gabon	*near extinct*	Online	1450	French	Van der Veen	1994	CBOLD

35	16	Gikuyu (Kikuyu)	E51	Kenya	5 347 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
36	17	Gusii	E42	Kenya	1 582 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
37	18	Gweno	E65	Tanzania	*unknown*	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
38	19	Gwere	J17	Uganda	275 608	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
39	20	Ha (Kiha)	J66	Tanzania	800 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
40	21	Hangaza	J65	Tanzania	150 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
41	22	Haya	J22	Tanzania	1 200 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
42	23	Hehe	G62	Tanzania	750 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
43	24	Ikizu	E402	Tanzania	28 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
44	25	Ikoma (Nata)	E45	Tanzania	15 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
45	26	isiXhosa	S41	South Africa	6 876 000	Online	3000	English	O'Kennon	1996	Personal
46						Online	3000	(English →)	O'Kennon	1996	Personal
47	27	isiZulu	S42	South Africa	9 142 000	Online	2000	English	Anon.	2002	eLanguage
48						Online	1001	Esperanto	Van Wilgen	2003	Majstro
49						PC	835	Esperanto	Van Wilgen	2002	Ergane
50						Pop-up	805	Italian	Toscano	2001	Babylon
51	28	Jita	J25	Tanzania	217 000	Online	2200	English	Downing	1999	CBOLD
52						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
53	29	Kahe	E64	Tanzania	2 700	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
54	30	Kalanga	S16	Botswana, Zimbabwe	321 000	Online	3000	English	Mathangwane	1994	CBOLD
55	31	Kamba (Kamba Kitu/Mach)	E55	Kenya	2 448 302	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
56						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
57	32	Kami	G36	Tanzania	10-20 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
58	33	Kara (Kilegi)	J252	Tanzania	86 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
59	34	Kerebe (Kerewe)	J24	Tanzania	100 000	Online	1500	English	Odden	1994	CBOLD
60						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
61	35	Kimbu	F24	Tanzania	78 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
62	36	Kinga	G65	Tanzania	65 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
63	37	Kinyaturu (Nyaturucha, Nyaturuwil)	F32	Tanzania	556 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
64						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
65	38	Kisi	G67	Tanzania	13 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
66	39	Koyo	C24	Congo	*unknown*	Online	1700	French	Gazania & Hyman	1996	CBOLD
67	40	Kuria (Kuria Mago/Tari)	E43	Tanzania, Kenya	348 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
68						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
69	41	Kutu	G37	Tanzania	45 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
70	42	Kwaya (Mkwaya)	J251	Tanzania	102 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
71	43	Lambya	N26	Tanzania, Malawi	81 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
72	44	Langi	F33	Tanzania	310 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
73	45	Lingala	C36d	DRC	309 100	Online	539	French	Grouselle	2002	Personal

74						Online	693	(French →)	Grouselle	2002	Personal
75						PC	508	French	Grouselle	2002	Freelang
76						PC	656	(French →)	Grouselle	2002	Freelang
77						Online	300	English/French	Maniacky	2002	Personal
78	46	Logooli (Maragoli)	E41	Kenya	197 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
79	47	Lozi	S34	Zambia, Namibia, Zimbabwe	557 000	Online	24000	English	Jalla	1937	CBOLD
80	48	Luyia	J32	Kenya, Uganda	3 593 461	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
81	49	Makhua (Emakhua)	P31	Mozambique, Tanzania	2 500 000	Online	7200	English	Kisseberth	1996	CBOLD
82	50	Makonde	P23	Tanzania, Mozambique	1 260 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
83	51	Malila	M24	Tanzania	52 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
84	52	Mambwe-Lungu (Mambwe, Rungu)	M15-M14	Zambia, Tanzania	359 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
85						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
86	53	Manda	N11	Tanzania	18 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
87	54	Masaba (Lumasaaba)	J31	Uganda	751 253	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
88	55	Matengo	N13	Tanzania	150 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
89	56	Matumbi	P13	Tanzania	72 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
90	57	Maviha (Mabia)	P25	Tanzania, Mozambique	*cf. Makonde*	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
91	58	Mbunga	P15	Tanzania	29 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
92	59	Meru (Meruimenti, Merutig)	E61	Kenya	1 305 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
93						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
94						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
95	60	Mpongwe	B11a	Gabon	1-4 000	Online	2600	French	Mouduiama	1994	CBOLD
96	61	Mpoto	N14	Tanzania, Malawi	80 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
97	62	Munukutuba	Kongo (H16) based Creole	Congo	1 156 800	Online	300	English/French	Maniacky	2002	Personal
98	63	Mwanga (Namwanga)	M22	Zambia, Tanzania	256 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
99	64	Mwera	P22	Tanzania	400 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
100						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
101	65	Nandi (Nande)	J42	DRC	903 000	Online	2100	French	Kavutirwaki	1978	CBOLD
102	66	Ndali (Ndari)	M301	Tanzania	150 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
103						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
104	67	Ndamba	G52	Tanzania	55 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
105	68	Ndengereko (Ndengeleko)	P11	Tanzania	110 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
106	69	Ngh'wele (Kwere)	G32	Tanzania	98 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
107	70	Ngindo	P14	Tanzania	220 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
108	71	Ngoni	N12	Tanzania, Malawi, Mozambique	205 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
109	72	Ngulu (Nguungulu)	G34	Tanzania	132 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS

110	73	Ngurimi (Ngoreme)	E401	Tanzania	32 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
111	74	Nyakyusa	M31	Tanzania, Malawi	1 050 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
112	75	Nyambo (Runyambo)	J21	Tanzania	7 000	Online	1500	English	Rugemalira	1993	CBOLD
113						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
114	76	Nyamwezi	F22	Tanzania	926 000	Online	2000	English	Maganga & Schadeberg	1992	CBOLD
115						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
116	77	Nyankore (Runya-nkore, Kiga-Nkore)	J13	Uganda	1 643 193	Online	12500	English	Taylor	1959	CBOLD
117						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
118	78	Nyiha	M23	Tanzania, Zambia	626 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
119	79	Nyiramba (Nilamba)	F31	Tanzania	440 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
120	80	Nyoro (Runyoro)	J11	Uganda	495 443	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
121	81	Pangwa	G64	Tanzania	177 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
122	82	Phende (Pende)	K52	DRC	420 000	Online	8200	French	Gusimana	1972	CBOLD
123	83	Pimbwe	M11	Tanzania	29 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
124	84	Pogolo (Pogoro)	G51	Tanzania	185 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
125	85	Rufiji	P12	Tanzania	200 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
126	86	Ruguru (Luguru)	G35	Tanzania	520 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
127	87	Rundi	J62	Burundi, Rwanda, Tanzania, Uganda	6 000 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
128	88	Rungwa (Lungwa)	M12	Tanzania	18 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
129						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
130	89	Rwanda (Munyarwanda)	J61	Rwanda, Burundi, DRC, Tanzania, Uganda	7 362 800	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
131	90	Saamia (Samialugwe)	J34	Uganda	50 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
132	91	Safwa	M25	Tanzania	158 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
133	92	Sangu	G61	Tanzania	75 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
134	93	Setswana	S31	South Africa, Botswana, Namibia	4 000 000	PC	186	Esperanto	Van Wilgen	1998	Ergane
135	94	Shambala (Sambaa)	G23	Tanzania	550 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
136	95	Shi	J53	DRC	654 000	Online	2500	French	Polak-Bynon	1978	CBOLD
137	96	Shubi	J64	Tanzania	153 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
138	97	SiNdebele (Zim-babwean Ndebele)	S44	Zimbabwe	1 502 000	Online	5000	English	Pelling	1971	CBOLD
139	98	Sizaki (Shashi Siz)	E404	Tanzania	82 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
140	99	Soga (Lusoga)	J16	Uganda	1 370 845	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
141	100	Suba	E403	Kenya, Tanzania	159 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
142	101	Sukuma (Ntuzu)	F21	Tanzania	5 000 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
143						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
144	102	Sumbwa	F23	Tanzania	191 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS

145			G42	Tanzania, Kenya	5 000 000	Online	18411	English	Netizens	2001	Kamusi
146						Online	26970	(English →)	Netizens	2001	Kamusi
147						Online	8545	Esperanto	Vessella	2001	Personal
148						Online	6547	(Esperanto →)	Vessella	2001	Personal
149						Online	5000	English/etc. (9)	Anon.	2002	eLanguage
150						Pop-up	2503	German	Alim	2001	Babylon
151						Pop-up	684	French	Omankoy	2001	Babylon
152						Pop-up	1213	(French →)	Omankoy	2001	Babylon
153						Pop-up	1092	German	Madete	2002	Babylon
154	103	Swahili				Pop-up	1045	(German →)	Madete	2002	Babylon
155						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
156						Pop-up	644	English	Van der Meijden	2000	Babylon
157						Pop-up	949	(English →)	Van der Meijden	2000	Babylon
158						Online	750	English	Anon.	2003	Freedict
159						Online	750	(English →)	Anon.	2003	Freedict
160						PC	570	French	Anon.	1999	Freelang
161						PC	749	(French →)	Anon.	1999	Freelang
162						PC	736	Esperanto	Van Wilgen	2002	Ergane
163						Online	665	Esperanto	Van Wilgen	2003	Majstro
164						Online	645	English	Shilnikov	2002	Glossword
165	104	Taita (Dawida)	E74a	Kenya	203 389	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
166	105	Temi (Sonjo)	E46	Tanzania	20 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
167	106	Tharaka	E54	Kenya	112 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
168	107	Tooro (Rutooro)	J12	Uganda	488 024	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
169	108	Tshivenda	S21	South Africa, Zimbabwe	750 000	Online	8900	English	Murphy	1997	CBOLD
170	109	Vinza (Kivinza)	J67	Tanzania	10 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
171	110	Wanda	M21	Tanzania	24 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
172	111	Wanji	G66	Tanzania	60 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
173	112	Yaka	H31	DRC, Angola	150-200 000	Online	3900	French	Ruttenberg	1969	CBOLD
174	113	Yao	P21	Malawi, Mozambique, Tanzania	1 597 000	Online	7400	English	Sanderson	1954	CBOLD
175						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
176	114	Zalamo (Zaramo)	G33	Tanzania	*a few*	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
177	115	Zanaki	E44	Tanzania	62 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
178	116	Zigula (Zigua)	G31	Tanzania	355 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
179	117	Zinza	J23	Tanzania	138 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
180	118	* Common Bantu	* CB	*N.A.*	*N.A.*	Online	2783	English/French	Maniacky	2002	Personal
181	119	* Proto Bantu	* PB	*N.A.*	*N.A.*	Online	9800	English/French	Tervuren Bantu Group	1998	CBOLD
182						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS