AFRICAN ASSOCIATION FOR LEXICOGRAPHY

22nd International Conference

26-29 June 2017

in cooperation with the
CONFERENCE OF THE LANGUAGE ASSOCIATIONS OF SOUTHERN AFRICA (CLASA)

Rhodes University, Grahamstown, South Africa

AFRICAN ASSOCIATION FOR LEXICOGRAPHY

Abstracts

22nd International Conference

in cooperation with the
CONFERENCE OF THE LANGUAGE ASSOCIATIONS OF SOUTHERN AFRICA (CLASA)

Rhodes University, Grahamstown, South Africa
26-29 June 2017

Hosted by: NRF SARChI Chair: Intellectualisation of African Languages, Multilingualism & Education and the School of Languages and Literatures: African Language Studies Section, Rhodes University, Grahamstown, South Africa

Conference coordinator: Prof. Dion Nkomo

Abstract reviewers: Prof. Herman Beyer, Prof. Rufus Gouws, Dr Langa Khumalo, Dr Victor M. Mojela, Dr Paul Achille Mavoungou, Dr Hughes Steve Ndinga-Koumba-Binza, Dr Ketiwe Ndhlovu, Prof. Dion Nkomo, Prof. Thapelo Otlogetswe, Prof. Danie J. Prinsloo, Prof. Elsabe Taljard, Dr Michele van der Merwe, Mr Tim van Niekerk

Abstract booklet editors: Prof. Sonja E Bosch and Prof. Dion Nkomo

ISBN 978-0-620-75209-1

TABLE of CONTENTS

AFRILEX HONORARY MEMBERS

Prof. R.H. Gouws

Prof. A.C. Nkabinde

Dr J.C.M.D. du Plessis

Dr M. Alberts

AFRILEX BOARD

2015 – 2017

President:	Dr M.V. (Victor) Mojela
Vice-President:	Prof. D.J. (Danie) Prinsloo
Secretary:	Prof. H.L. (Herman) Beyer
Treasurer:	Prof. E. (Elsabé) Taljard
Editor Lexikos:	Dr H.S. (Steve) Ndinga-Koumba-Binza
Members:	Prof. S.E. (Sonja) Bosch
	Dr L. (Langa) Khumalo
Conference organiser:	Prof. D. (Dion) Nkomo

MESSAGE FROM THE AFRILEX PRESIDENT

On behalf of the AFRILEX Board, I would like to welcome all of you to the *22nd Annual International Conference of the African Association for Lexicography*, also known as 'AFRILEX 2017'. This year's edition takes place here in the Eastern Cape's historical town of Grahamstown, at Rhodes University. For the first time in the history of AFRILEX, we are having our conference jointly with four other associations working on language issues under the collective banner of the Conference of the Language Associations of Southern Africa (CLASA). This conference follows the successful AFRILEX 2016 which was hosted by the Xitsonga National Lexicography Unit at Karibu Hotel and Leisure Resort in Tzaneen. As an association that aims to bring together all lexicographic activities that take place on the African continent, as well as all friends of AFRILEX from further afield, the AFRILEX Board is pleased to see many scholars from outside and within South Africa attending and participating actively in the AFRILEX conferences, in particular, the scholars from America, Asia, Europe, Namibia, Gabon, Botswana and Zimbabwe, who are always with us in every conference. As in previous years, we are still inviting more lexicographic scholars from the entire African continent to form part of the membership of AFRILEX. Since AFRILEX 2011, which was held at the University of Namibia, all other subsequent AFRILEX International Conferences were held within the Republic of South Africa, which is not appropriate if we regard AFRILEX as a lexicography association for Africa. We are still appealing to all of our AFRILEX members beyond the borders of South Africa to invite this Association to be hosted in their institutions. After this conference we hope that we will get more invitations to host some of our future AFRILEX

conferences outside South Africa. AFRILEX 2017 has been meticulously prepared and coordinated by a local organising team under the leadership of Prof. Dion Nkomo, whom I also congratulate on his promotion to the position of Associate Professor during 2016. At this stage I once again want to thank Prof. Danie Prinsloo, the AFRILEX Deputy President who actively participated in the preparations leading to this Conference. The abstract adjudication process for AFRILEX 2017 was expertly managed and carried out by Prof. Sonja Bosch, assisted by the following abstract reviewers: Prof. Herman Beyer, Prof. Rufus Gouws, Dr. Langa Khumalo, Dr. Victor M. Mojela, Dr. Paul Achille Mavoungou, Dr. Hughes Steve Ndinga-Koumba-Binza, Dr. Ketiwe Ndhlovu, Prof. Dion Nkomo, Prof. Thapelo Otlogetswe, Prof. Danie J. Prinsloo, Prof. Elsabe Taljard, Dr Michele van der Merwe and Mr Tim van Niekerk. As in the past, Prof. Sonja Bosch and Prof. Dion Nkomo once-more did commendable work in the compilation of this Abstract Booklet we are all holding now. A word of appreciation also goes to Dr. Gertrud Faass for designing the template for the abstracts. We want to congratulate and thank them for the job well-done. We also want to say thank you again to Prof. DJ Prinsloo who excellently managed and kept the AFRILEX website up to date and assisted in the compilation of the programme for this conference together with Prof. Dion Nkomo. Not forgetting Prof. Elsabé Taljard, our reliable treasurer who, as always, continuously keeps the AFRILEX moneys safe. Just like in the previous conferences, AFRILEX 2017 promises to be another stellar gathering, with speakers coming from various countries in Africa, North America and Europe, namely Canada, Denmark, Gabon, Georgia, Germany, Namibia, Nigeria, South Africa and Zimbabwe. We also want to thank Prof. Dion Nkomo for effectively coordinating with the organisers of CLASA to make our 2017 unique conference a success. We are also not forgetting the organisers of CLASA, in particular, Prof. Russel Kaschula and the Rhodes University management together with our own local dictionary Unit, the DSAE and its Director, Ms Jill Wolvaardt, for hosting us in this historical university town of Grahamstown. Our international keynote speaker this year is Prof. Tinatin Margalitadze from the Ivane Javakhishvili Tbilisi State University, Tbilisi in Georgia. I want to officially welcome Prof. Margalitadze at AFRILEX 2017 here on the African continent, and in the Eastern Cape town of Grahamstown in particular. The national keynote will be delivered by Ms Jill Wolvaardt, the Executive Director of the Dictionary Unit for South African English, which is also our co-host in this magnificent Institution.

Maropeng Victor Mojela
President: AFRILEX

KEYNOTE PRESENTATION 1

Once again why lexicography is science

'Lexicography is a scientific practice aiming to bring dictionaries into existence'1

Franz Josef Hausmann
Tinatin MARGALITADZE (tinatin.margalitadze@tsu.ge)
Lexicographic Centre, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

XVII International Congress of EURALEX (European Association for Lexicography), held in Tbilisi, Georgia in September 2016 adopted a resolution addressed to UNESCO, national governments throughout the world, research funding agencies, and universities to acknowledge the status of lexicography as an academic discipline and promote the study of words and languages. 'Our multilingual world needs novel types of dictionaries, which requires proper recognition and support'- states the resolution[2].

Prior to the adoption of the resolution, a round table discussion was organized within the framework of the congress which was dedicated to the status of lexicography. 'One of the hot topics today is whether lexicography should be seen merely as a 'craft', or as a scientific academic discipline whose theory should be taught in universities, like mainstream linguistics' – stated the synopsis of the discussion. These statements reveal that we may still come across opinions that working on a dictionary is not a scientific activity. Such views are very damaging to lexicography and hinder its proper development.

Lexicography, which has centuries-old history, has undergone significant evolution. It has always kept abreast of the newest developments in linguistics and allied sciences. The advent of comparative-historical linguistics was manifested in the entries of *Oxford English Dictionary on Historical Principles*; development of electronic corpora and corpus linguistics in the 1980s, was immediately reflected in lexicography; appearance of electronic dictionaries has opened completely new prospects for lexicography turning it into one of the most dynamic and rapidly developing fields of knowledge. Modern lexicography is a complex, multidisciplinary field incorporating multiple components, viz. semantic theories, corpus-based methods, methods and techniques for natural language processing, e-lexicography, research in dictionary use, etymology and so on. Consequently, claims that working on a dictionary does not constitute a scientific activity, or that lexicography has no theory, seem to be an unbelievable misunderstanding.

From my personal observation, the above-mentioned simplistic attitude towards lexicography, stating that it is not a science, stems from the superficial approach to the intricate phenomenon of meaning and related issues. One of the reasons may be descriptive linguistics, which treated lexical level of language as peripheral and non-structural for decades, concentrating on the description of phonological and morphological systems of language. If study of meaning is peripheral, then lexicography, which is primarily involved in the study of words and their meanings, can not be a science. Later, this disregard for the content plane of language has changed, and nowadays different theories of lexical semantics study meaning from many different angles (Geeraerts, 2010), but it has left its mark on the understanding of the essence of lexicography.

Another reason for not regarding lexicography as a science is the view that lexicography has no theory. I fully agree with professor Rufus Gouws (Gouws, 2012) that the authority of

---

[1] Hausmann, F. J. (1985). Lexikographie. In: *Handbuch der Lexikologie* (Hrsg. Christoph Schwarze / Dieter Wunderlich). Königstein/Ts.: Athenäum, 367-411.

[2] Resolution of the XVII Congress of EURALEX (September, 2016). http://euralex.org/resolution2016/

some European scholars, who voice these claims, is responsible for such views. In 1983, at the founding congress of EURALEX, German linguist Herbert Wiegand (Wiegand, 1984) formulated the structure and components of the general theory of lexicography. At the same congress, British scholar, John Sinclair (Sinclair, 1984) raised the issue of setting up a master course in lexicography which would contribute to transforming lexicography from practical activity into an academic discipline and would develop lexicography in close relation with information technologies, computer linguistics, general linguistics and lexicographic practice. Both these papers were excellent points of departure for the further elaboration and development of the unified theory of lexicography, but it has not happened.

Such opinions hinder the proper development of the field and are dangerous. The adverse results of underappreciation of lexicography can be well seen by the observation of the processes taking place in my native language, Georgian.

In my presentation I will give my reasoning why lexicography should be considered to be a science, I will also present my views on the theory of lexicography and its components.

The right approach to lexicography is particularly important in a country where several languages co-exist. True multilingualism does not mean a mere co-existence of a number of languages in any given society and/or state. The true multilingualism sets in only when there is no discrimination between languages and when the same scientific approach serves as a basis for the provision of resources and the creation of dictionaries and the terminology, when everything is done for the full-scale functioning of each particular language in a multilingual environment.

References

Geeraerts, D. (2010). *Theories of Lexical Semantics.* Oxford University Press Inc., New York.

Gouws, R. H. (2012). Theoretical Lexicography and the International Journal of Lexicography. *International Journal of Lexicography*. 25.4, pp. 450 – 463.

Sinclair J. (1984). Lexicography as an Academic Subject. R. R. K. Hartmann (ed.), *LEXeter '83 Proceedings.* Tübingen: Max Niemeyer. pp. 13-30.

Wiegand H. E. (1984). On the Structure and Contents of a General Theory of Lexicography. R. R. K. Hartmann (ed.), *LEXeter '83 Proceedings*. Tübingen: Max Niemeyer, pp. 3-12.

KEYNOTE PRESENTATION 2

South Africa's National Lexicography Units: time for a reboot?

Jill WOLVAARDT (j.wolvaardt@ru.ac.za)
Dictionary Unit for South African English, Rhodes University, Grahamstown, South Africa

How have the flag bearers for South Africa's bold approach to restoring the nation's indigenous languages, become the neglected poor relations of the deeply flawed institution that is the Pan South African Language Board (PanSALB)? How has the national lexicography project, pioneered in the early years of South Africa's democratic transition by some of the country's greatest language activists and academics, been permitted to degenerate into the scattered efforts of a diminishing band of lexicographers? Forced for the last decade into perpetual begging for adequate funding, the National Lexicography Units (NLUs) hover on the verge of extinction. The critical question is, 'does anyone care?'.
Dishearteningly, indications from government seem to imply that the response is, 'Not really.'

It seems, therefore, that the time is ripe for a re-examination of the 'genuine purpose' of the NLUs; whether they are still important in the construction of a South Africa that pays more than lip-service to multilingualism, and – if so – how they need to be adapted to ensure their relevance in the digital age.

In 1996 the National Lexicography Units Bill was presented to parliament, 'To provide for the establishment and management of national lexicography units; to make equitable provision for national general dictionaries for each of the official languages of South Africa and for matters incidental thereto.'

The Bill was not enacted; instead, three years later the National Lexicography Units were bolted on to PanSALB, in an amendment to the PanSALB Act of 1995. This was despite serious misgivings on the part of at least one member of the Language Plan Task Group (LANGTAG), who expressed the prescient view that:

> The cost implications for the PanSALB budget need to be considered … The only way costs would be saved, either directly by government or indirectly via PanSALB, would be if the units were to be under-resourced. If this were to happen it would be better for them not to be established in the first place. (Heugh, K. 1998, personal communication)

And so it came to pass: as PanSALB expanded, so the organisation began to chip into the funds destined for the NLUs, until the share of government funding allocated to lexicography diminished from 33% of PanSALB's income, to the 19% currently divided amongst the eleven NLUs. This diminution in funding was accompanied by increasing ignorance in PanSALB about lexicography and the core business of the NLUs, defined in their founding documents as, 'The continuous and comprehensive collecting, arranging and storing in a lexicographically workable form of the vocabulary of the … language.'

Which leaves the NLUs where we find them today, desperately trying to justify their existence by producing dictionaries, which, by and large, are based on their feasibility within the constraints of limited funding rather than on any coherent overarching plan. It is doubtful that, with the resources available to them, the majority of NLUs are in a position to 'continuously and comprehensively' collect the vocabulary of their respective languages. Few units have more than three staff and a number of units have only basic information technology to work with. With PanSALB recalcitrant about its funding of the NLUs, clearly another way has to be sought to ensure that South Africa's languages are intensively researched, recorded, preserved and developed in the manner envisaged so optimistically some twenty years ago.

I believe it's critical for a new plan to be drawn up for South African lexicography, with a specific focus on the role of the National Lexicography Units. My presentation intends to raise more questions than it answers. Some are ideological, others pragmatic. For instance:

- Is multilingualism really being served by the preponderance of bilingual dictionaries that use English as the source or target language? Obviously, while English predominates in the spheres of government, education and the media, these bilinguals have a place. But is it the role of the NLUs to develop them?
- Should the NLUs be involved in conventional dictionary publishing at all? Should their efforts rather be directed to lexicographically coding the vocabulary they collect to form part of a much broader digital resource? A resource that would enable the production of monolingual, multi-media online dictionaries, for example, but might also serve a wide variety of other inter-disciplinary language needs.
- Where is the intellectual home of the NLUs? Is the current conformation of one NLU per official language logical, effective or practical? Given the overlap between various languages, is there an argument for re-situating NLUs in hubs that will not only provide a synergy between their own efforts but also with those of institutions with a track record of language development?

In raising these and other questions, I hope that, as a collective of African language lexicographers, we can initiate a process to restore the credibility of a national lexicography project for South Africa, and – by extension – re-imagine, reinvigorate and reboot the National Lexicography Units.

*****

SESSIONS

A usability evaluation of the *Afrikaanse idiome-woordeboek*

Liezl H. BALL (liezl.ball@up.ac.za)
Department of Information Science, University of Pretoria, Pretoria, South Africa.
Theo J.D. BOTHMA (theo.bothma@up.ac.za)
Department of Information Science, University of Pretoria, Pretoria, South Africa.

There are many exciting opportunities that technology brings to the field of lexicography. For example, much more data can be included in an e-dictionary, and as such, words do not need to be abbreviated, e-dictionaries can include or link to more information (De Schryver, 2003: 157) or multimedia can be incorporated (Lew, 2012: 344). Information technology also offers many advantages in terms of access to information. The speed with which information can be retrieved is a considerable advantage (Verlinde & Peeters, 2012: 147) and various search features can be included to enable more efficient search (Lew, 2012: 345, 351; Verlinde & Peeters, 2012: 147). Bothma (2011) also suggests various technologies that could be used to enhance e-dictionaries, such as annotations, decision trees, linked data, recommendations.

When digital tools are developed, it is vital that these tools can be used effectively and efficiently by users, in other words, the usability of a tool is important. Usability becomes more important as products become more complex and can be critical to the success of a product (Tullis & Albert, 2008: 5-7). Usability evaluation is the process where data about how users will use or do use a product is gathered and whether it is suitable and acceptable to users (Preece, Rogers & Sharp, 2011: 433). In order to conduct usability evaluation a set of criteria according to which the evaluation can be done is necessary. Evaluation criteria to evaluate websites exist, but criteria specifically for e-dictionaries were developed by Ball (2016).

The *Afrikaanse idiome-woordeboek* is a prototype e-dictionary of Afrikaans fixed expressions, developed with the intention to test the functionality of the design. The design of this dictionary is based on the function theory of lexicography and presents several dictionaries that are created from one large database (Bergenholtz, Bothma & Gouws, 2011: 36). The different dictionaries are monofunctional and give information relevant to specific situations. The e-dictionary also makes use of various technologies such as advanced search and display options, browsing, multimedia in various articles, links to external sources that provide more information and customisation options. The dictionary was designed in such a way that only information relevant to a specific situation can be given to a user.

Usability evaluation was done on this e-dictionary to determine with what success it can be used. The discount usability methods heuristic evaluation and usability testing were used. In the heuristic evaluation one expert evaluated the e-dictionary according to evaluation criteria. In the usability testing, seven users were asked to complete 16 tasks while being observed.

This paper reports on the findings from the usability tests and are discussed under the categories of content, information architecture, navigation, access (searching and browsing), help, customisation and the use of innovative technologies to manage information in e-dictionaries. The usability evaluation showed that the users did not always use the e-dictionary as the designers intended and various recommendations could be made to the designers of the *Afrikaanse idiome-woordeboek*, as well as to the design of e-dictionaries in general. Recommendations could be made regarding searching in e-dictionaries, the data that can be included in e-dictionaries, further exploration with technologies and theoretical frameworks, training and usability evaluation on e-dictionaries.

References

Ball, LH. 2016. *An evaluative study to determine to what extent technology can be used in e-dictionaries to provide relevant information on demand*. Master's thesis, University of Pretoria.

Bergenholtz, H., Bothma, T. and Gouws, R. 2011. A model for integrated dictionaries of fixed expressions. Kosem, I. and Kosem, K. (eds.) *Electronic lexicography in the 21st century: New applications for new users: Proceedings of eLex 2011, Bled, Slovenia, 10-12 November 2011:* 34-42. Ljubljana: Trojina, Institute for Applied Slovene Studies.

Bothma, T.J.D. 2011. Filtering and adapting data and information in an online environment in response to user needs. In: Bergenholtz, H. and Fuertes-Olivera, P.A. (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London: Continuum International Publishing Group: 71-102.

De Schryver, G. 2003. Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography* (16) 2: 143-199.

Lew, R. 2012. How can we make electronic dictionaries more effective? In: Granger, S. and Paquot, M. (eds.) *Electronic Lexicography*. Oxford: Oxford University Press: 343-361.

Preece, J., Rogers, Y. and Sharp, H. 2011. *Interaction Design. Beyond human-computer interaction*. 3rd ed. Chichester: John Wiley & Sons.

Tullis, T. and Albert, W. 2008. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Burlington: Morgan Kaufmann.

Verlinde, S. and Peeters, G. 2012. Data access revisited: The Interactive Language Toolbox. In: Granger, S. and Paquot, M. (eds.) *Electronic Lexicography*. Oxford: Oxford University Press.

*****

Planning a dialectal dictionary: From user questions to textual structures

Herman L. BEYER (hbeyer@unam.na)
Department of Language and Literature Studies, University of Namibia, Windhoek, Namibia

Based on the premise that lexicography is a scientifically-based discipline and that a theory or theories of lexicography can be formulated (cf., among others, Gouws et al. 2013), this paper sets out to describe elements of the theoretical paradigm being employed in the first stages of the planning of a dialectal dictionary. This paradigm constitutes a communicative metalexicography which is based on the fact that dictionaries are media for indirect communication between the (potential) dictionary user and the lexicographer, albeit a special type of communication in which elements of the full range from interpersonal to mass communication can be shown to apply. This mediated communication is generally manifested in the form of texts.

After the introduction of the planned dictionary's subject matter and the specification of the target user type and dictionary purposes, the exposition of the communicative approach takes the basic notion of question and answer as starting point. One of the central purposes of lexicographic communication is to provide information in answer to specific types of questions posed by users in particular user situations. Allusions to the notion of question and answer abound in the lexicographic literature. The term *search question* (also used in the information sciences) is expressly applied to dictionaries by Wiegand (1987). From a communicative perspective the concept of *raw user question* (RUQ) is introduced. RUQs are indicators of user needs and expectations in user situations which can be empirically established by means of observation protocols and scientific elicitation from (potential) target users in actual or

controlled situations. Collected RUQs are distilled to a set of *user situation questions* (USQs), which can be expressed in different ways.

The paper will list the initial USQs that have been identified for the planned dictionary. It will proceed to show how USQs that are earmarked (according to the data distribution programme) for treatment in the central list can be formalised in terms of predicate calculus, which leads logically to the identification of the primary micro- and addressing structure elements of the dictionary. These microstructural elements and their addressing structures can be presented in the form of a *micro- and addressing structure schema* (MASS). The MASS therefore effectively lists the core microstructural elements that will have the purpose of answering the respective USQs (thereby theoretically fulfilling the immediate purpose of the dictionary). It will also be shown how the expression of USQs by means of predicate calculus relates closely to the processes of textual condensation as described by Wiegand (1996), but with some divergence motivated by communication theory and text linguistics.

In applying a lexicographic interpretation of the theory of conversational implicature (cf. Grice 1991), and particularly that theory's cooperative principle and maxim of Quantity, to the planning of the microstructure, it will be argued that, due to limitations in the nature and media of lexicographic communication, the elements identified in the MASS alone do not in all cases represent the optimal answers to USQs. Consequently, some primary microstructural elements have to be supplemented by secondary microstructural elements to facilitate communicative equivalence and ultimately functional user effects.

Within the framework of a communicative approach that interprets speech act theory (cf. Austin 1962) from a lexicographic perspective, primary and secondary microstructural elements are regarded as signals of lexicographic messages with different illocutionary forces. Primary elements are generally classified as signals of *statements* (in response to USQs), and secondary elements are classified as signals of *advisements* that complement (and are addressed at) statements. The implications of this classification can be manifested in the search area structure in terms of article slot assignment and the employment of differentiated typographical and non-typographical structural markers as well as microarchitectual features.

In conclusion a small number of resulting example dictionary articles will be shown and briefly commented on in the light of the media options for the planned dictionary.

References

Austin, J.L. 1962. *How to Do Things with Words*. Oxford: Clarendon Press.

Gouws, R.H., Heid, U., Schweickard, W. and Wiegand, H.E. (eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin: De Gruyter Mouton.

Grice, P. 1991. *Studies in the Way of Words*. Second edition. Cambridge, Massachusetts and London, England: Harvard University Press.

Wiegand, H.E. 1987. Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. *Lexicographica* 3: 178-227.

Wiegand, H.E. 1996. Textual Condensation in Printed Dictionaries. A Theoretical Draft. *Lexikos* 6: 133-158.

*****

Using corpus query engines for facilitating lexicographical analysis of African languages

Thomas ECKART (teckart@informatik.uni-leipzig.de)
Natural Language Processing Group, University of Leipzig, Leipzig, Germany
Uwe QUASTHOFF (quasthoff@informatik.uni-leipzig.de)
Natural Language Processing Group, University of Leipzig, Leipzig, Germany; and
Department of African Languages, University of South Africa, South Africa
Dirk GOLDHAHN (dgoldhahn@informatik.uni-leipzig.de)
Natural Language Processing Group, University of Leipzig, Leipzig, Germany

Corpus linguistics relies on the availability of, preferably large, text corpora and can be used for a variety of applications, partially dependent on a wide range of linguistic annotations. For the analysis and utilization of these corpora specific query engines have been developed that efficiently support powerful queries which can be designed and adapted even by inexperienced scholars. One of those query engines is the open-source project NoSketchEngine (Rychlý 2007) based on the SketchEngine (Kilgarriff et al. 2004), a powerful corpus management and query system that is especially used in lexicography and corpus linguistics.

The analysis of so called "lesser-resourced languages" often suffers from a lack of large amounts of processable text and of accessible linguistic annotations. However, even for rather standard linguistic annotations - like part-of-speech tags - combined with some background knowledge about a language's general properties much information can be gained by using rather simple queries on acquired text material. In particular, this also contains information relevant for lexicography as it is demonstrated in the following. Necessary background information includes knowledge about typical articles and conjunctions, or about word order as included in the *World Atlas of Language Structures* (Dryer and Haspelmath 2013).

The following example is based on the data of the National Centre for Human Language Technologies (NCHLT) Annotated Corpus of Zulu (2013) using an instance of the NoSketchEngine (available at http://cql.corpora.uni-leipzig.de/?corpusId=zul_rma) for corpus querying. The corpus is a lemmatised, part of speech tagged and morphologically analysed text collection based on documents from the South African government domain crawled from gov.za websites and collected from various language units. Naturally, the used acquisition method cannot guarantee the generation of a balanced and representative text corpus.

However, this exploitation has already proven to be useful especially when dealing with lesser-resourced languages (Goldhahn 2016).

Additionally, the NCHLT tag set was reduced to the core part-of-speech tags of the Universal Dependency project (Nivre et al. 2016), which are easier to handle for users without detailed morphological knowledge about isiZulu. For more advanced users, the NCHLT tag set can also be used in all queries.

Based on the isiZulu word *khathi* (time) the following queries show how to extract information about typical word usage from the corpus:

- For sample word usage: Search for the lemma *khathi*: [lemma="khathi"], which gives sample sentences for the inflected forms like *isikhathi*, *sikhathi*, *ngesikhathi*, *nesikhathi*, *izikhathi* etc.
- For verbs used together with *khathi* search for [pos_ud17="VERB"] [lemma="khathi"]. In the sample sentences we see frequent usage of *to take time, to spend time, over time* etc.
- For adjectives used together with *khathi* search for [pos_ud17="ADJ"] [lemma="khathi"]. In the sample sentences we see frequent usage of *for a long time, for how long, how much time, good times* etc.

Apparently it is possible to extract valuable information about word usage and typical word patterns even for rather simple queries and with a minor degree of knowledge about the targeted language. However, the availability of text resources is a crucial precondition for meaningful results as larger corpora will lead to more stable and more complete findings.

Of course the general principle can be extended: more complex queries can be used to identify typical modifiers of nouns and verbs (as part of so called "word sketches") or to identify typical elements involved in or related to standard activities described in the underlying corpus. It therefore can even be used as input for identifying candidates of paradigmatic relations in a semi-automatic approach of synset generation for lexical databases.

References

Dryer, M. S., Haspelmath, M. (eds.) 2013. The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available: http://wals.info. Accessed on 2016-02-09.

Goldhahn, D., Sumalvico, M., Quasthoff, U. 2016. *Corpus collection for under-resourced languages with more than one million speakers*. In: Workshop on Collaboration and Computing for Under-Resourced Languages (CCURL)*, LREC, Portorož*, 2016.

Kilgarriff, A., P. Rychlý, P. Smrz, and D. Tugwell. 2004. The Sketch Engine. In *Proc Eleventh EURALEX International Congress*. Lorient, France.

NCHLT isiZulu Annotated Text Corpora 2013. Available: http:// rma.nwu.ac.za/index.php/resource-catalogue/isizulu-nchlt-annotated-text-corpora.html. Accessed on 12 February 2016.

Nivre, J., Marneffe, M., Ginter, F., Goldberg, Y., Hajič, J., Manning, C., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., Zeman, D. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 2016.

Rychlý, P. 2007. Manatee/Bonito - A Modular Corpus Manager. In *1st Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University, 2007. p. 65-70. ISBN 978-80-210-4471-5.

*****

A portal for -corpus collection for under-resourced languages

Dirk GOLDHAHN (dgoldhahn@informatik.uni-leipzig.de)
Natural Language Processing Group, University of Leipzig, Leipzig, Germany
Thomas ECKART (teckart@informatik.uni-leipzig.de)
Natural Language Processing Group, University of Leipzig, Leipzig, Germany
Uwe QUASTHOFF (quasthoff@informatik.uni-leipzig.de)
Natural Language Processing Group, University of Leipzig, Leipzig, Germany; and
Department of African Languages, University of South Africa, South Africa

There are about 350 languages with more than one million speakers. For about 40 of them, the situation concerning text resources is comfortable: there are corpora of reasonable size and adapted tools like POS taggers or parsers. For the remaining languages, the number of speakers indicates a need for both corpora and tools.

Random Web crawling for smaller languages has several limitations. Among others, they are related to aspects like technical issues, the relatively small amount of Web pages, the inadequate link structure and the ranking on search engines.

Therefore, native speakers with knowledge of Web pages in their language are of invaluable help. In order to facilitate the gathering of URLs for a large number of languages a specific Web portal has been developed, which is available at http://curl.informatik.uni-leipzig.de. It enables scholars and language enthusiasts with knowledge of Web sites in their respective language to contribute to corpus creation or extension by entering a URL into a simple Web Interface. Using this Web portal URLs of interest are collected. They are then downloaded using Heritrix, the crawler of the Internet Archive project, and processed by a standardized corpus processing chain for daily newspaper corpora creation. The processing pipeline was adapted to append newly added Web pages to an increasing corpus for each language. This enables us to collect larger corpora for under-resourced languages with community help.

We apply a standardized, language-independent pipeline for building corpora from raw data used also for the corpus creation at the Leipzig Corpora Collection (Goldhahn et al. 2012). We use self-developed tools for extracting raw text from WARC files (i.e. the Heritrix output) and HTML pages. Then we apply statistical language identification on document level. As a data basis for comparison web corpora or documents from sources such as the Universal Declaration of Human Rights or Watchtower for several hundred languages are utilized. Further processing steps are sentence segmentation, removal of ill-formed sentences based on handwritten regular expressions (Eckart et al. 2012), language identification on sentence basis (Biemann & Teresniak 2005), duplicate sentence removal, tokenization and word co-occurrence calculation. Finally, the corpora are stored as MySQL databases with a standardized schema. In addition to the basic workflow, additional (possibly language-specific) tools can be applied to some corpora, like POS-tagging, which results in additional database tables. Tests on various input data have shown that our processing chain handles data volumes of up to 200 million sentences. For corpora of 100K - 1M sentences, the running times are typically less than an hour.

For smaller languages, individual Web sites in the corresponding language are often not linked and hence difficult to collect. If larger Web sites are available, they are often driven by governmental organizations or newspapers. The language Kirundi (ISO 639-3: run) is spoken by about 6 million people in Burundi. A random crawl of its top-level domain .bi in 2015 lead to only about 2,000 sentences in Kirundi. Language identification was based on Kirundi Bible texts. There were no newspapers in Kirundi mentioned in ABYZ News Links (http://abyznewslinks.com), one of the largest international newspaper directories. Having found the IGIHE news site with Kirundi texts by manual effort, the CURL crawling tool was started with the corresponding URL and three additional Web links (http://www.igihe.bi, http://burundiimage.info/public_html/kirundi, http://indundi.com/news/page/category/kirundi, http://ikirundi.com.au). The crawler collected 178 MB of HTML pages. After preprocessing and language identification, the resulting corpus contained more than 16,000 sentences with about 340,000 running words.

In summary, this paper describes a corpus collection initiative for lesser resourced languages, enabling scholars or language enthusiasts to create and extend corpora by simply entering URLs into a Web interface. Using this Web portal URLs of interest are collected with the help of the respective communities. As a result, we are able to collect larger corpora for under-resourced languages by a community effort. These corpora are made publicly available.

References

Biemann, C., Teresniak, S. 2005. Disentangling from Babylonian Confusion – Unsupervised Language Identification. In: Gelbukh A. (eds) *Computational Linguistics and Intelligent Text Processing. CICLing 2005. Lecture Notes in Computer Science, vol 3406.* Berlin, Heidelberg: Springer.

Eckart, T., Quasthoff, U., Goldhahn, D. 2012. Language Statistics-Based Quality Assurance for Large Corpora. In: *Proceedings of Asia Pacific Corpus Linguistics Conference*. 2012, Auckland, New Zealand.

Goldhahn, D., Eckart, D., Quasthoff, U. 2012. Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. 2012, Istanbul, Turkey.

*****

Small-scale evaluation of the perceived impact of the *Oxford Bilingual School Dictionary: isiXhosa and English* on learners and teachers in Port Elizabeth

Megan HALL (megan.hall@oup.com)
Oxford University Press South Africa, Cape Town, South Africa
Nontsikelelo NTUSIKAZI (ntsikie.berlin@gmail.com)
Teacher mentor at Edupeg; freelance researcher and publisher, Cape Town, South Africa

Research was conducted in the Eastern Cape to find out about teachers' perceptions of the impact of using the *Oxford Bilingual School Dictionary: isiXhosa and English* (2014) on teachers and learners at school.

The study was an exemplar that formed part of the development of the Oxford Impact Framework. Oxford Impact is a unique way of evaluating the impact that educational products and services from Oxford University Press (OUP) have on teaching and learning. At its heart is the Oxford Impact Framework; a structured process developed with the National Foundation for Educational Research (NFER), and supported by Oxford University Department of Education (OUDE). By carefully assessing the impact of products, OUP is able to develop the resources that make the most positive difference to teaching and learning.

In South Africa, children at school must study two languages – one at Home Language (HL) level, and one at First Additional Language (FAL) level. While a child's mother tongue may be used as the Language of Learning and Teaching (LOLT) in grades 1 to 3, most schools will choose to use English as the LOLT from grade 4 onwards. Proficiency in English is therefore crucial to a child's success in almost all subjects, a particular challenge when more than 90% of the population speaks a language other than English as their mother tongue (extrapolation from Table 2.6, Statistics South Africa:25).

In 2004, Oxford University Press South Africa (OUPSA) started developing a range of bilingual dictionaries to support the learning of English as an additional language, especially by learners with an African language as a mother tongue. By 2016, OUPSA were keen to find out how, if at all, the latest dictionary in the range (*Oxford Bilingual School Dictionary: isiXhosa and English*) was making a difference in the classroom. We chose to carry out a Perceptions of Impact study to explore this question.

Telephone interviewing in the mother tongue of the teachers was chosen as a suitable method. In terms of the geographical scope, we decided to focus on primary schools offering grade 7, on the basis of previous unpublished research conducted in the province prior to publication of the dictionary. We selected a single educational district (Port Elizabeth) in the province of the Eastern Cape, since that province has the largest number of mother-tongue isiXhosa-speakers (5.09m according to Table 2.5, Statistics South Africa:23).

The Provincial Education Department was asked to identify schools in the district that had ordered more than 10 copies of the dictionary in 2014 (after ordering, quantities were reduced by the Department due to budget constraints, hence the use of quantity as a criterion). A total of ten schools met the criteria; this subsequently reduced to 6, after permission from

principals was sought. In all, six interviews with grade 7 teachers across 5 schools were completed in March 2016.

Key findings cover three main impact themes:

- perceived impact on learners,
- perceived impact on teachers,
- and perceptions of the counterfactual (i.e. what would happen if dictionaries were not used).

In addition, we collected process information about how teachers use the dictionary, how dictionaries are bought, what subjects the teachers teach, and how many dictionaries they had access to.

All research participants noted a positive impact on learners. Most participants commented that learners' understanding and comprehension had improved, that the dictionary supported them and that they had become more independent, no longer needing "spoon-feeding" by the teachers. Several teachers noted that learners could use the dictionary on their own to find the meaning of words they didn't understand.

All research participants said that using the dictionary had had a positive impact on them as teachers. Most said that it helped them teach content (or non-language) subjects, such as Maths, Natural Sciences, and Life Orientation.

We asked research participants what they felt would happen if their learners did not have access to dictionaries in class. Most said that teaching and learning would become "difficult"' or "very difficult", or that teaching and learning would be a struggle.

This evaluation is based on a small number of teachers and schools in a single educational district. Further research with a larger number of schools would be necessary to confirm whether the perceptions of these teachers are common. Further research is planned for 2017.

References

Statistics South Africa. 2012. *Census 2011: Census in brief*. Pretoria: Statistics South Africa.

*****

The design and implementation of a corpus management system for the isiZulu National Corpus

Langa KHUMALO (khumalol@ukzn.ac.za)
Linguistics Program, School of Arts, University of KwaZulu-Natal, South Africa

The imperative exists to develop computational tools to support the 11 official languages in South Africa. Hitherto text-based human language technologies in South Africa have been developed by CTexT through the Autshumato project, whereas speech technologies have been developed by the Meraka Institute, which include Automatic Speech Recognition (ASR), pronunciation dictionaries and text-to-speech (TTS) technologies under the auspices of the Lwazi project. This paper discusses an open-source technology for exploring the isiZulu National Corpus (INC) designed to meet the exigence to develop computational tools. We discuss the design and functionalities of a corpus management system (CMS) for the INC. The INC has an impressive 20.5 million tokens, which is a significant milestone towards the intellectualization of isiZulu. The INC's CMS is part of a series of technologies designed as enablers in the development and intellectualization of isiZulu (Khumalo, 2017). IsiZulu is currently an under-resourced language in terms of computational information and knowledge processing (Keet and Khumalo, 2014, 2017; Spiegler et al., 2010). The CMS provides

researchers and end-users with an interface for performing searches and analyzing statistics for metadata. The CMS thus has three critical suites that allows for *wordlist* and frequency searches, *concordance* function and *keyword* extraction. Thus this paper describes the corpus query engine and the user management system, implemented as an open-source web-based application. The application consists of two main parts, first is the server-side corpus query engine, which handles storage of the corpus and processes queries. Second is the client-side user interface, with which users interact (*see appendix 1. showing the schematic representation of the CMS system architecture*). The application was developed based on Python version 2.7 with Django framework version 1.9.8. Other application utilities employed include MySQL for python, Natural Language Toolkit (NLTK) (Bird, 2016) and Pygeoip geographical IP address locator (*see appendix 2. showing the application data model for the software*). The component entities fall in three categories. First the authentication entities, which handle user authentication and authorization, denoted by prefix 'auth'. Second is application-related information, which handle application-related information like corpus files and extra user information, denoted by prefix 'app'. Third is the settings entities, which control application behavior and track changes within the data model, denoted by prefix 'Django'. Functionally the system is composed of a user management module which manages three role levels – users, sub-administrators and administrators - and the corpus query engine composed of the main functions that the software offers to users. The user management module is only accessible to (sub-) administrators, including backend facilities like dashboard monitoring, site activity, email handling, and file and user management. The software interface can be accessed via most web browsers, either on a PC or mobile device, by typing the url (https://iznc.ukzn.ac.za/) in the address bar of the browser. To create a word list, there is a Wordlist button on the left of the panel or the Wordlist hyperlink in the main panel of the page. There is a facility provided for the user to choose whether to sort the words by frequency or alphabetically. The default is to sort by frequency. The word list is sorted by descending order of frequency of occurrence within the corpus. The keyword function allows users to extract keyword from a corpus. It works similar to keyword extraction in WordSmith (Scott, 1996). The user supplies a generic file and a corpus file in order to extract keywords. Just like Wordlist, a function is provided for exporting the output of keyword extraction as a .csv file. Concordance is one of the most popular uses of a corpus software. Some of these software were designed mainly for that purpose (Wiechmann and Fuhs, 2006; Reppen, 2001). The CMS offers this function, accessible from different interfaces, namely concordance for a word from the Wordlist output, concordance of a word from the keyword extraction output, and concordance of a word supplied into the concordance search engine. The concordance output can also be exportable as .csv. The paper will be able to show that the CMS is a functioning solution for querying the INC.

References

Khumalo, L. 2017. Disrupting language hegemony: intellectualizing african languages, in *Disrupting Higher Education Curriculum: Undoing Cognitive Damage*, pp. 247-264.

Keet, C. M. and Khumalo L. 2014. Basics for a grammar engine to verbalize logical theories in isiZulu. *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pp. 216-225.

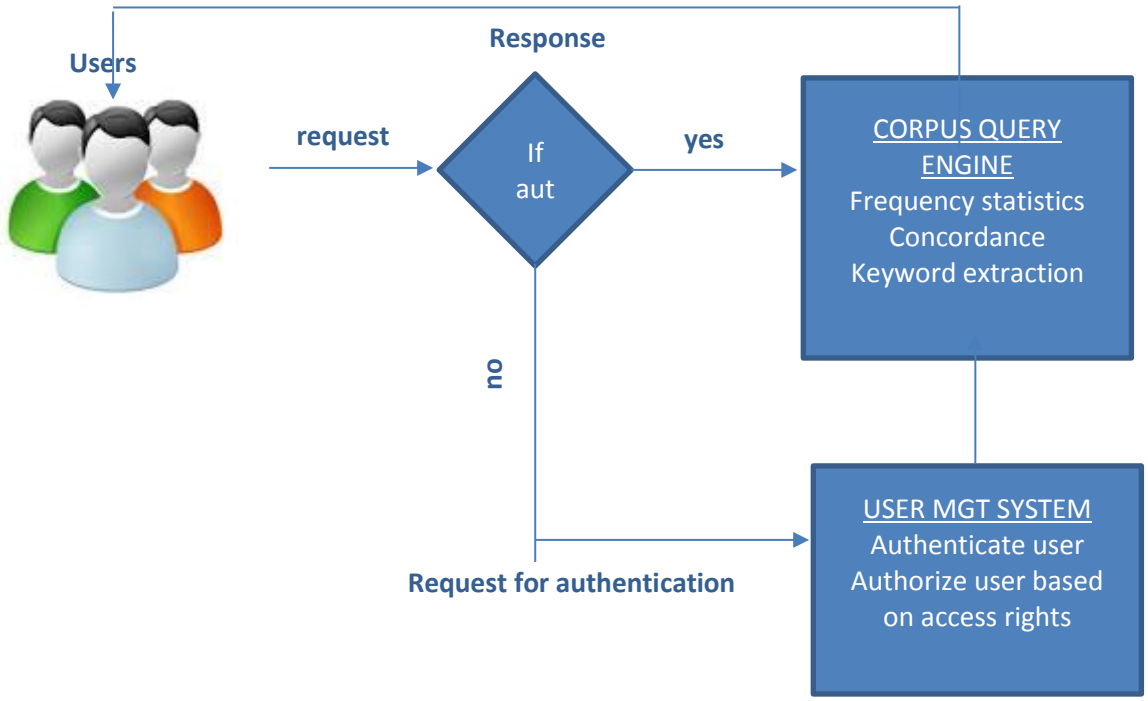
Spiegler, S., Van Der Spuy A. and Flach, P.A. 2010. Ukwabelana: An open-source morphological Zulu corpus, in *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 1020-1028.

Keet, C.M. and Khumalo, L. 2017. Toward a knowledge-to-text controlled natural language of isiZulu, *Language Resources and Evaluation*, pp. 1-27, pp. 131-157.

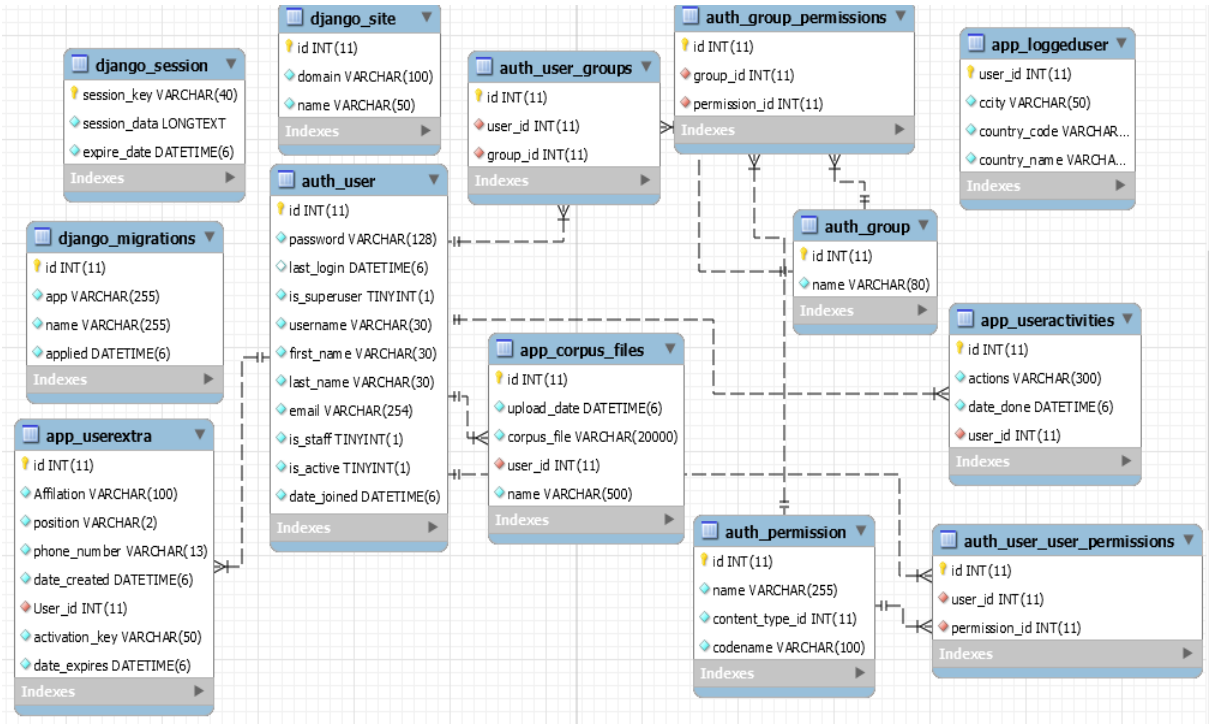
Bird, S. 2006. NLTK: the natural language toolkit, in *Proceedings of the COLING/ACL on*

Interactive presentation sessions, pp. 69-72.
Scott, M. 1996. *WordSmith Tools.* Oxford: Oxford University Press.
Wiechmann, D. and Fuhs, S. 2006. Concordancing software. Corpus Linguistics and Linguistic Theory, 2(1), pp. 107-127.
Reppen, R. 2001. Review of MonoConc Pro and WordSmith Tools. *Language Learning and Technology*, vol. 5, pp. 32-36.

Appendix 1. Schematic representation of the CMS system architecture



Appendix 2. Application data model



*****

Data visualisation in the online Dictionary of South African English

Bridgitte LE DU (b.ledu@ru.ac.za)
Dictionary Unit for South African English, Rhodes University, Grahamstown, South Africa
Tim VAN NIEKERK (t.vanniekerk@ru.ac.za)
Dictionary Unit for South African English, Rhodes University, Grahamstown, South Africa

The online *Dictionary of South African English* (DSAE, http://dsae.co.za) is an electronic version of *A Dictionary of South African English on Historical Principles* (Silva et al., Oxford University Press, 1996). The print edition, produced by the Dictionary Unit for South African English (DUSAE) at Rhodes University in Grahamstown, South Africa was the culmination of 25 years' research resulting in a 1.7 million-word text with 4 600 main entries documenting the development of South African English from its origins in the late 17th Century to 1995. Entries emphasise word history and show etymologies, variant spellings, compounds, derivatives and phrases. In total 14 700 word forms are represented, reflecting diverse borrowings from other South African languages; notably, the dictionary is rooted in quotation evidence, reproducing 44 000 bibliographically-documented citations. Since the publication of an initial pilot online edition, the dictionary has gained collaborators from the University of Hildesheim, Germany (HU) and Stellenbosch, South Africa (SU) working towards a thoroughly adapted digital version which makes full use of data presentation possibilities offered by the electronic medium (see Du Plessis & van Niekerk, 2016; van Niekerk et al., 2016).

This paper will give a brief description of the project, highlighting key areas of the print-to-digital adaptation of the DSAE, with a focus on data visualisation. In this dictionary, which supports *cognitive* uses to an unusual degree, visual devices provide alternate views of lexical data and expanded possibilities for navigation via multiple browsing pathways (see

Bergenholtz & Tarp, 2002, 2003; Tarp, 2008). As a historical variety dictionary, the DSAE represents a unique source of information for knowledge-based enquiries by ethnologists, historians, literary scholars and those interested in South African history and culture. However, beyond the level of specific word searches, users' potential cognitive enquiries prompt the mapping of relationships between entries, realizing complex threads of continuity along semantic, historical, cultural and pragmatic relations of meaning (Tasovac & Petrović, 2015). In response, we have introduced visual strategies for content display that not only simplify presentation but also allow the visual representation of lexicographical content at different levels of abstraction. These visual displays act on two levels: *macro-visualisation* techniques offering overviews of and alternate access routes to the overall lexicographical structure and content; and *micro-visualisation* techniques that provide simplified representations of complex entry microstructure. Examples of macro-visualisations include treemap structures, which allow browsing of the dictionary content by feature such as date of first use or subject category, while examples of micro-visualisations are a time bar reflecting the historical range of citations and a histogram showing the distribution of citations over time. Both macro- and micro-visualisation strategies introduce advanced browsing functionality and enhance usability in contexts of cognitive enquiry, the former by offering (interactive) overviews of content, and the latter by providing rapid comprehensibility of entry-specific information.

While the introduction of visual devices opens new possibilities for viewing, browsing and navigating the lexicographical content of the DSAE, it also begins to demonstrate how a historical variety dictionary can be transformed from an 'extended wordlist' into an accessible linguistic, cultural and encyclopaedic inventory. We present these strategies as examples of alternative presentations of text-heavy lexicographical information in feature-rich lexical datasets, as well as to invite feedback.

References

Bergenholtz, H. and Tarp, S. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 18: 253-263.

Bergenholtz, H., Tarp, S. 2003. Two opposing theories: On HE Wiegand's recent discovery of lexicographic functions. *Hermes, Journal of Linguistics* 31: 171-196.

Du Plessis, A., van Niekerk, T., 2016. Adapting a Historical Dictionary for the Modern Online User: The Case of the *Dictionary of South African English on Historical Principles's* Presentation and Navigation Features. *Lexikos* 26: 82-102.

Silva, P., Dore, W., Mantzel, D., Muller, C., Wright, M. 1996. *A Dictionary of South African English on Historical Principles*. Cape Town: Oxford University Press.

Tarp, S. 2008. Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner's lexicography. In *Lexicographica*. Series Maior 134. Tübingen: Max Niemeyer.

Tasovac, T., Petrović, S. 2015. Multiple Access Paths for Digital Collections of Lexicographic Paper Slips. In *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference*, 11-13 August, 2015. Herstmonceux Castle, United Kingdom: 384-396. Available: https://elex.link/elex2015/conference-proceedings. Accessed on 06/02/2017.

Van Niekerk, T., Stadler, H., Heid, U. 2016. Enabling Selective Queries and Adapting Data Display in the Electronic Version of a Historical Dictionary. In: *XVII EURALEX International Congress*. 6-10 September, 2016. Tbilisi, Georgia: 635-646. Available: http://euralex2016.tsu.ge/publication.html. Accessed on 21/02/2017.

*****

The design of a text-reception-oriented dictionary app based on data from the DSAE

Elisabeth LEMKE (lemkee@uni-hildesheim.de)
Department of Information Science and Natural Language Processing, University of Hildesheim, Germany
Ulrich HEID (heid@uni-hildsheim.de)
Department of Information Science and Natural Language Processing, University of Hildesheim, Germany
Tim VAN NIEKERK (t.vanniekerk@ru.ac.za)
Dictionary Unit for South African English, Grahamstown, South Africa

The 'Dictionary of South African English on historical principles' (DSAE) is a monovolume monolingual historical dictionary of South African English. It contains about 14.000 items which are specific to the South African regional variety and its subvarieties, and it indicates their word class and meaning, etymology and word history, as well as diasystematic and domain marking. Most notably, it contains around 40.000 fully documented historical and contemporary quotations

DSAE appeared as a book in 1996 and was turned into an XML-based online version in 2014 (http://www.dsae.co.za). By means of responsive design, a corresponding (lexicographically almost unchanged) mobile website was made available in 2015. In parallel, however, the XML data of the online version underwent substantial computational lexicographic revisions (cf. Van Niekerk, Stadler & Heid, 2016) that lead to an enrichment at two major levels: (i) the introduction of ontological classes for many entries, and (ii) the separation of form- and word-history-related indications which can now each be used as separate search criteria.

On this basis, a prototype of a native app has been derived from the current XML data of the DSAE. Our design study was complemented by website usage data from Google analytics and by the results of a small questionnaire survey carried out in early 2015 (ca. 75 respondents).

While, according to this survey, most users of the online DSAE are language or culture experts (teachers, historians, linguists) who, in terms of the Lexicographic Function Theory (Tarp, 2008), have cognitive needs, the app is designed to address lay persons who want to look up specific South-African words in a text reception situation, mainly to understand their meaning.

To serve this public and this dictionary function, short versions of the existing meaning explanations are displayed by default when a word is searched. Variant orthographic forms are allowed as search criteria and linked to the preferred or current form. From DSAE's quotations we only select the newest one for display in the default setting, and for more quotations, we experiment with several information-on-demand devices. The design adapts to the different screen sizes offered by current smartphone models that use android.

As a basis for the design decisions and for the evaluation of the prototypes, a persona-based requirements definition was prepared. The information architecture is designed to meet the mental model of the target user in terms of intuitive use and information overload. In particular, Donald Norman's (1988) Design Principles influenced the anatomy of the interaction patterns of the app. Here, we experiment with visibility on demand for cognitive dictionary functions on a word's origin and history as well as for the quotations. This way, the user is presented with a simple information structure while having the option to expand visible data. Two types of navigation concepts are being tested, one where cognitively oriented data are displayed in a horizontal access structure via tabs and another one where they are accessible vertically via expanders. A consistent use of interaction patterns throughout the app ensures easy learnability and an ergonomic use.

A formative evaluation of the prototypes is carried out in two consecutive steps. The first one is to verify the information architecture's coherence. Therefore, the app is reviewed in sequences of action by an expert in a cognitive walkthrough. At a later time in the design process, the navigation concepts and their conformity with the user's mental model are tested in a usability test with six participants, who correspond to the defined personas. The test is scenario-based and comprises three main tasks and subtasks with increasing complexity. One scenario is targeted at testing the browsing functions by encouraging an explorative approach, while the others request the user to look up specific words in a text reception situation.

In the paper, we will describe and motivate our design decisions, discuss the results of both usability tests and present typical workflows where the app is being used.

References

Tarp, S. 2008. Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner's lexicography. In: *Lexicographica*. Series Maier 134. Thübingen, Max Niemeyer.

Norman, D. A. 1988. *The psychology of everyday things*. New York, Basic Books.

Van Niekerk, T., Stadler, H. and Heid, U. 2016. Enabling Selective Queries and Adapting Data Display in the Electronic Version of a Historical Dictionary. In: *Proceedings of the XVII EURALEX international congress. Lexicography and Linguistic Diversity*. September 6-10th, 2016. Tbilisi, Georgia: 635-646.

*****

Lemmatization of shortenings in indigenous South African languages, especially Xitsonga

Ximbani Eric MABASO (mabasxe@unisa.ac.za)
Department of African languages, University of South Africa, Pretoria, South Africa

It is often stated that the indigenous South African languages (IndiSAL) lack specialized terminology. One of the many areas where such an assertion is true is in the case of shortenings such as abbreviations and acronyms which constitute an important aspect of linguistic wealth. The aim of this study is to ascertain the extent to which abbreviations and acronyms have been accommodated or have not been included as lemmata in reference works. These could be dictionaries, word lists and encyclopedias in the official indigenous South African languages. The official languages concerned are Xitsonga, isiZulu, Sesotho, Sesotho sa Leboa, Tshivenda, Setswana, isiNdebele, Siswati and isiXhosa. The research approach that will be adopted will be firstly to take stock of abbreviations and acronyms from the different languages and draw up a corpus from various written sources. This corpus will serve as benchmark in the second step to assess whether the available stock of abbreviations and acronyms has been incorporated or not in the reference works of these languages. Thirdly, the study will try to find out what motivates the presence or absence of these shortenings in reference works. In conclusion, and fourthly, how such state of affairs impacts on the users of the IndiSAL.

The lack of dictionaries in the IndiSAL is well documented. The situation is worse for the more 'previously marginalized languages' like Xitsonga which prior to the new democratic South Africa did not have dictionary units (Mashele, 2016:17). This state of affairs relates to monolingual, bilingual, trilingual and multilingual dictionaries and word lists.

Preliminary observation is that whereas abbreviations and acronyms are included as lexical items/entries in English dictionaries, e.g. "**adj**. *abbrev* for adjective" (Collins, 2006:18). This is not the case with African languages. For example, National Lexicography Unit (2005: vii-viii) lists 64 Xitsonga abbreviations but none appears as a lemma. The same applies to Marhanele and Bila (2016: 18-19) under "*minkomiso ya swin'wana na swin'wana*". English

also boasts several specialized dictionaries of abbreviations and acronyms (e.g. Dale and Puttick, 1999). There are no such dictionaries for African languages. This state of affairs is evidence of the diminished use and status of the IndiSAL. This study will highlight the lack of sources that focus and deal specifically with this linguistic genre which language users can consult for e.g. Xitsonga-English equivalents or for correctness and authenticity verification. By using corpus from published reference works, the paper will bring to light how consequently very often translators experience serious challenges relating to how they should handle the translation of shortenings such as abbreviations and acronyms. Lehohla (2013) will be used as a multilingual terminology list to highlight this point. Some translators create their own shortenings which result in a multiplicity of non-standard forms, some appropriate but others completely out of line and ungrammatical. For instance, whereas it is perfectly in order to translate human titles such as 'Mr' (Mister) to '*Tat*' (*Tatana*) or '*Nk*' (*Nkulukumba*), it is unacceptable to translate abbreviations for words like 'ABSA' (Amalgamated Banks of South Africa) and render it as *'*THAD*' (*Tibangi leti Hlanganisiweke ta Afrika-Dzonga*). Another related problem is that of a lack of standardization guidelines. These abbreviations need to be standardized because in some texts there are different abbreviations for the plural form of '*Tat*' namely '*Vatatana*' (Messrs) which in one case appears as 'Vat' and in other places as '*Vatat*' (Mabaso, 2016). The study will recommend that standardizing procedures need to be applied in order to determine the linguistically correct form for authentication. Such problems need to be solved. Gouws and Prinsloo (2005) assert: "One of the salient features of dictionaries throughout many centuries is their function to assist users to resolve linguistic problems." This paper will apply the problem-based and content analysis theory (Leedy and Ormrod, 2005) to stress the need for the inclusion of shortenings in dictionaries and for availing specialized dictionaries in indigenous South African languages.

References

Collins. 2006. *Collins English Dictionary*. Concise Edition (6th ed). Glasgow: HarperCollins.

Dale, R. and Puttick, S. 1999. *The Wordsworth Dictionary of Abbreviations and Acronyms*. Great Britain: Wordsworth.

Gouws, R.H. and Prinsloo, D.J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Leedy, P.D. and Ormrod, J.E. 2005. *Practical research: planning and design*. Upper Saddle River, N.J.: Prentice Hall.

Lehohla, P. 2013. *Multilingual Statistical Terminology*. Pretoria: Statistics South Africa.

Mabaso, X.E. 2016. Nkomiso eka Xitsonga: Nxopaxopo wa Ntivoririmi (The Shortened Form in Xitsonga: A linguistic analysis). Unpublished Doctoral Thesis. Pretoria: University of South Africa.

Marhanele, M.M. and Bila, V. 2016. *Tihlùngù ta Rixaka*. Polokwane: Timbila Poetry Project.

Mashele, H.T. 2016. Towards Corpus-based Dictionaries for Xitsonga. Unpublished Masters Dissertation. Pretoria: University of Pretoria.

Xitsonga National Lexicography Unit. 2005. *Xitsonga-English, English-Xitsonga Dictionary*. Cape Town: Phumelela Books.

*****

Motivating the development of a parallel corpus: towards automated machine translation

Njabulo MANYONI (Manyonin@ukzn.ac.za)
Language Planning and Development Office, University of KwaZulu-Natal Durban, South Africa

The University of KwaZulu-Natal (henceforth UKZN) is committed to the intellectualization of isiZulu so that it can function at par with English in all areas of administration, research, teaching and learning (cf. Khumalo, 2016). This commitment is expressed in the University's language policy and plan of 2006, which was revised in 2014. As a result UKZN is advancing the development of human language technologies for isiZulu in order for the language to be used in higher functions and to contribute effectively in knowledge production, knowledge dissemination, and the knowledge economy. To this end, the University has been involved in the production of literature in both languages as required by policy through various translations that include annual reports (or sections thereto), media communiques, speeches, lectures, research proposals, student rules and handbooks, bio-notes and campus signage. It is argued in this paper that this massive production of texts in isiZulu forms a compelling basis for the development of an English-isiZulu parallel corpus. Parallel corpora are collections of identical texts in two languages (Aijmer and Altenberg: 1996), processed and stored in machine readable formats. A parallel corpus is useful as a basis for the development of automated machine translations. In order to "train" the machine to recognize the various texts, there needs to be an exact translation of each file in languages. It is also an important resource that can be made public and thus verifiable by others. A parallel corpus must be comprised of a variety of text types (in the case of UKZN, reports, study material, news articles, forms, abstracts, papers, speeches etc). The focus for text in UKZN is more on texts that is utilizing standardized terminology, allowing for the terminology used to be of an official and linguistically correct nature. The Pan South African Language Board (PanSALB) is the legislated body that is responsible for ensuring that languages enjoy equal status and that language rights as enshrined in the South African Constitution Section 6(4) of 1996, are protected. It is PanSALB's responsibility through it National Language Bodies to ensure that proper terminology is developed for all official languages. UKZN has generated a wide range of terminology for specific academic fields which enables the texts from those fields to be translated using officially recognized terminology. In corpus linguistics, the general trend is that the bigger the size of the corpus the better it is in terms of quality. The size of the parallel corpus will thus influence its effectiveness as a tool that forms the basis of an automated translation machine. Currently, the (massive) production of the translations between English and isiZulu languages happens manually, which is a slow, error prone, arduous and time consuming process. Whilst the text production processes are necessary towards the intellectualization of a language and are in keeping with the University Language Policy, they are manual and therefore inhibiting. It is our argument in this paper that the current text production can be used to motivate a process towards building a parallel corpus, which can be used as the basis for the development of Automated Machine. According to (Grover, 2011), the Human Language Technologies Audit of 2009 conducted by the National Human Language technology Network, the results reflected a disparity in terms of HLT's that exist compared to the size of the population that uses those particular languages. Therefore the development of parallel corpora not only bridges the digital divide between underdeveloped languages and developed ones but also assists in the development of effective HLT's aimed at bringing about parity of esteem and usage of South African official languages. It can also be argued that the effort to develop parallel corpora and specialized terminologies in both English and isiZulu can be used to develop specialized bi-lingual glossaries that are crucial in teaching and learning at UKZN. Specialized bi-lingual

dictionaries (cf. Khumalo 2015) based on the parallel corpus will be developed as enablers in improving epistemic access in specialized disciplines for which they are crafted.

References

Aijmer, K., Altenberg, B., and Johansson, M. (eds) (1996). *Languages in contrast: Papers from a Symposium on text-based cross-linguistic studies*, Lund: Lung University Press.
Grover, A.,Huyssteen, B., and Pretorius M., (2011). *The South African Human Language Technology Audit*, New York: Springer-Verslag.
Khumalo, L. "*Semi-automatic Term Extraction for an isiZulu Linguistics Terms Dictionary Using a Corpus Linguistic Method*", Lexikos 25: 495-506, 2015.
Khumalo, L. "Disrupting language hegemony: intellectualizing African languages," in Samuel et.al. *Disrupting Higher Education Curriculum: Undoing Cognitive Damage*, Boston: Sense Publishers. 247-264, 2017.

*****

The role of translation in lexicography with special reference to Tshivenḓa-English dictionaries in the promotion of multilingualism

Mashudu MAṬHABI (Mashudu.nthambeleni@univen.ac.za)
MER Mathivha Centre of African Languages Arts and Culture, University of Venḓa

Given South Africa's multilingualism, dictionaries have become one of the ways of promoting communication among and between speakers of different languages. In recognition of the historical imbalances of the past, South Africa has committed itself to taking "practical and positive measures to elevate the status and advance the use of indigenous African languages" (*Constitution of South Africa,* 1996:4). Furthermore, the *Constitution* (1996:15) points out that "everyone has the right to cultural life of their choice and that a person belonging to a cultural, religious or linguistic community may not be denied the right to enjoy their culture, to practise their religion and use their language". This shows South Africa's commitment to multilingualism. Multilingualism and translation go hand in hand as their existence is mutually inclusive.

A Lack of dictionaries with properly translated lexical items in African languages such as Tshivenḓa is a matter of great concern to the users of the language as communication in the present age of information technology is crucial. Nkomo (2010:371) states that lack of well-prepared dictionaries "results in users consulting any available but inappropriate dictionaries''. Mafela (2005:276) also notes that "Dictionary users find it difficult to use the bilingual Venḓa dictionaries because they are confronted with equivalents which they cannot distinguish''. A good dictionary plays an important role in the achievement of good communication. Dictionaries, therefore, have become a *sine qua non* for bringing about effective communication among the different ethnic groups in South Africa.

It has come to the attention of various scholars that many bilingual dictionaries in South Africa are of poor quality as far as translation of lexical items is concerned (Mabasa, 2009; Rapotu, 2011). For instance, many Tshivenḓa-English/English-Tshivenḓa bilingual dictionaries reflect unsatisfactory translation of lexical items, for example:

	Tshivenḓa	**English**
(1)	Mbeu	seed (Van Warmelo, 1989:191; Tshikota, 2006:44).

The translation in (1) above, although correct, is not sufficient as it has excluded many other usages. It has only provided a literal translation, without considering the other communicative aspects associated with the lemma. For instance, *mbeu* in Tshivenḓa may also refer to gender, semen or female egg. For a person who is learning Tshivenḓa, the translation in (1) above is highly inadequate because the person would not have an idea of the other meanings expressed by the lemma *mbeu*. What this entails is that poor translation of lemmas in dictionaries leads to miscommunication and misunderstanding.

Although the above-mentioned dictionaries were written by different scholars, they are never the less similar because all of them used the literal method to achieve equivalence between the source language (SL) and the target language (TL). Therefore, there is still a need to compile dictionaries that take into account the communicative context. The aim of this paper is to examine the role of translation in lexicography with special reference to selected Tshivenḓa-English dictionaries (bilingual dictionaries). In order to achieve this aim, the paper will need to answer the following questions:

- What strategies can be used to have effective translation of lemmas in dictionaries?
- How has translation been applied in selected Tshivenḓa bilingual dictionaries?

It is against this backdrop that the researcher attempt to conduct a study on whether translation in the compilation of selected Tshivenḓa English dictionaries has been applied properly or not. Attention will be given to treatment of translation of nouns according to class prefixes as well as the treatment of nouns according to translation equivalents. The selected dictionaries are: Tshikota (2006) Ṱhalusamaipfi Tshivenḓa/English Dictionary and Van Warmelo (1989) Venḓa Dictionary.

This study will utilised a qualitative method to collect the data and interviews will be conducted with lexicographers, university lecturers, language practitioners and students which will be randomly selected. This study will be a benefit to lexicographers and they will be able to conduct user-friendly dictionaries.

References

Mabasa, P.T. 2009. *An Evaluation of Translation Procedures with Special Reference to Xitsonga and English: The Case of Natural Science and Technology Dictionary*. Unpublished Master's Dissertation. Polokwane: University of Limpopo.

Mafela, M.J. 2005. *Making Discrimination in Bilingual Venḓa Dictionaries.* Lexikos, Vol 15: pp 276-285.

Nkomo, D. 2010. Affirming the Role of Specialised Dictionaries in Indigenous African Languages. *Lexikos, Vol 20*: pp. 371-389.

Rapotu, M.E. 2011. *Retranslation of Lexical Items as Translation Equivalents: A Lexicographic Analysis*. Unpublished Master's Dissertation. Polokwane: University of Limpopo.

South African Government, 1996. *Constitution of the Republic of South Africa.* www.info.gov.za/documents/constitution/index.htm. Accessed on the 16 February 2017.

Tshikota, S.L. 2006. *Tshivenḓa/English Ṱhalusamaipfi Dictionary*. Cape Town: Phumelela

Van Warmelo, N.J. 1989. *Venḓa Dictionary.* Pretoria: J.L.Van Schaik.

*****

Perspectives for Lexicography Units in multilingual Gabon

Hugues Steve Ndinga-Koumba-Binza (nkbinza@uwc.ac.za)
Department of Language Education, University of the Western Cape, South Africa
Blanche Nyangone Assam (bassam@uwc.ac.za)
Department of Foreign Languages, University of the Western Cape, South Africa
Virginie Ompoussa (vompoussa@yahoo.com)
Département des Sciences du Langage, Université Omar Bongo, Libreville, Gabon

In the inception of the strategic planning for Gabonese lexicography (Emejulu, 2001 & 2003; Ndinga-Koumba-Binza, 2005), it was suggested that for its development Gabonese lexicography should adopt certain procedural steps that led to the development of South African lexicography. One of these processes is the establishment of National Lexicography Units (NLUs). In fact, the importance of NLUs has been recognized in the promotion and development of lexicographic activities for a number of African languages in post-apartheid South Africa (Alberts, 2011; Mongwe, 2006).

However, the suggestion to establish lexicography units for Gabonese languages comes against various points of concern. Two of these issues[3] should be considered in the focus of this study. First, no comprehensive reflection has ever been conducted so as to implement such a proposal. Suggestions by Emejulu (2001 & 2003) and Ndinga-Koumba-Binza (2005) limit themselves to what should be the point of departure for setting up lexicography or dictionary units for Gabonese native languages, i.e. the *language-units*[4] as identified by Kwenzi Mikala (1988, 1990 & 1998). Further suggestions by other Gabonese scholars such as Afane Otsaga (2004) and Tomba Moussavou (2007) restrict themselves to making the case that Kwenzi Mikala's *language-units* appear to be a stepping stone for language standardization, which in turn should be continued and promoted through a lexicography unit for each *language-unit*. Consequently, since Kwenzi Mikala (1988, 1990 & 1998) enumerates 10 *language-units* which comprise 62 speech forms (referring to both languages and dialects), the common proposal is that Gabon should establish 11 lexicography units (including Gabonese French).

The second issue which comes on the way of establishing lexicography units for Gabonese language in the manner of South Africa is the fact that this proposal fails to consider practical differences between Gabon and South Africa. The dissimilarities between the two countries would have been noticed if strategic procedures for instituting Gabonese lexicography units had been thoroughly thought of. For instance, although both Gabon and South Africa are primarily language diversity countries, one of the differences between the two is that the latter has a constitutionally-recognized multilingualism with 11 official languages, while the former has a former colonial language as sole official language despite the abundance and the actual use of native languages (Ndinga-Koumba-Binza, 2007 & 2011). Furthermore, while South Africa has an official language policy and a language planning put in place, Gabon has none of these, except for some government initiatives that have "*had no real effect on the status of native languages*" (Ndinga-Koumba-Binza, 2007: 107).

The present paper intends to provide strategic elements for setting up Gabonese lexicography units. First, it refutes the principle of one lexicography unit for each Kwenzi Mikala's *language-unit*. It rather suggests two analysis frameworks for Kwenzi Mikala's *language-units*. These analysis frameworks may lead to establishing two or more lexicography

[3] These are mainly linguistic issues. Other practical issues such as financial support, management and computerization of proposed lexicography units will be the concern of a further study.

[4] Direct translation from Kwenzi Mikala's concept of "*unités-langues*" in French. Kwenzi Mikala (1990: 122) labels as "*unités-langues*" (language-units) a set of various languages and dialects mutually comprehensible.

units for a number of the *language-units*. The first framework is concerned with the issue of homogeneity in Gabonese *language-units*. The second framework deals with the conceptualization of language and dialect, seeing that no clear distinction is made between language and dialect within Kwenzi Mikala's *language-units*. Thus, this issue of language versus dialect is dealt with within the historico-dialectological framework.

Finally, it is the view of this study that the establishment of Gabon's lexicography units is intimately related to language inventory in the country. However, as it has been variously mentioned in many studies on the language situation of Gabon, an exact number of Gabonese native languages is unknown. The present paper should therefore also contribute towards solving the issue of Gabonese language inventory by providing a necessary strategy for setting up lexicography units on the basis of Kwenzi Mikala's *language-units*.

References

Afane Otsaga, T. 2004. *The standard translation dictionary as an instrument in the standardization of Fang*. PhD thesis. Stellenbosch: University of Stellenbosch.

Alberts, M. 2011. National Lexicography Units: Past, Present, Future. *Lexikos* 21: 23-52.

Emejulu, J.D. 2001. Lexicographie multilingue et multisectorielle au Gabon: planification, stratégie et enjeux. Emejulu, J.D. (ed.). *Eléments de lexicographie gabonaise*. New York: Jimacs-Hillman Publishers. Tome 1: 38-57.

Emejulu, J.D. 2003. Challenges and promises of a comprehensive lexicography in the developing world: The case of Gabon. Botha, W.F. (ed.). *'n Man wat beur. Huldinggsbudenl vir Dirk van Schalkwyk*. Stellenbosch: Bureau of the WAT. 195-212.

Kwenzi Mikala, J.T. 1988. L'identification des unités-langues bantu gabonaises et leur classification interne. *Muntu* 8: 54-64.

Kwenzi Mikala, J.T. 1990. Quel avenir pour les langues gabonaises ? *Revue Gabonaise des Sciences de l'Homme* 2: 121-124.

Kwenzi Mikala, J.T. 1998. Parlers du Gabon: classification du 11.12.97. *Les Langues du Gabon*, edited by A. Raponda-Walker. Libreville: Editions Raponda-Walker. 271-220.

Mongwe, M.J. 2006. *The role of the South African National Lexicography Units in the planning and compilation of multifunctional bilingual dictionaries*. Unpublished MPhil thesis. Stellenbosch: University of Stellenbosch.

Ndinga-Koumba-Binza, H.S. 2005. Considering a lexicographic plan for Gabon within the Gabonese language landscape. *Lexikos* 15: 132-150.

Ndinga-Koumba-Binza, H.S. 2007. Gabonese language landscape: Survey and perspectives. *South African Journal of African Languages*, 27(3): 97-116.

Ndinga-Koumba-Binza, H.S. 2011. From foreign to national: a review of the status of French in Gabon. *Literator* 32(2): 135-150.

Tomba Moussavou, F. 2007. *Metalexicographic criteria for a monolingual descriptive dictionary presenting the standard variety of Yipunu*. PhD thesis. Stellenbosch: University of Stellenbosch.

*****

Lemmatisation of Zulu and Zimbabwean Ndebele nouns using the stem method: A proposal for criteria for ensuring consistency in its use

Eventhough NDLOVU (evennthough@yahoo.co.uk)

Unit for Language Facilitation and Empowerment, University of the Free State, Bloemfontein, South Africa

An examination of Zulu and Zimbabwean Ndebele dictionaries shows that nouns have been lemmatised using the initial letter of the stem, the initial letter of the prefix proper,

the singular and plural forms, the singular form, the plural form and the initial vowel of the noun prefix. (See: Doke and Vilakazi (1948), Doke, Malcom and Sikakana (1958), Dent and Nyembezi (1969), Pelling (1965), Nkabinde (1982, 1985), Nyembezi (1992) Hadebe (2001) and Nkomo and Moyo (2006). This paper examines the validity of the criticism that the stem method cannot be applied with consistency (Van Wyk, 1995; Maphosa, 1997; De Schryver, 2008; 2010). A close analysis of the dictionaries which lemmatise using the stem method show that the source of the inconsistencies in lemmatising using the stem method largely derive from the challenge of identifying and ascertaining the noun stem. This paper therefore proposes criteria for identifying and ascertaining the Zulu and Zimbabwean Ndebele noun stem to show that the stem method can be applied with consistency. It also argues that the choice of the method for lemmatising Zulu and Zimbabwean Ndebele nouns should be dictated largely by the target users and the dictionary type being compiled. In light of this, despite the said problems of the stem method, this method can be an ideal alternative method for lemmatising nouns in specialised dictionaries of linguistics, which target students and scholars of Zulu and Zimbabwean Ndebele linguistic structure. In this paper we therefore argue that there is need to avoid completely discarding the stem method, but to appreciate that it can be consistently used and is a suitable method for certain dictionary types and target users. The key participants in this study were Ordinary and Advanced Level Ndebele majors in two selected secondary schools in Gwanda urban, 2015 and 2016 BAA, BA Honours, BA Dual Honours Ndebele and Linguistics students as well as Master of Arts degree in African Languages and Literature students at the University of Zimbabwe and Ndebele lecturers in the Department of African Languages and Literature at the University of Zimbabwe. Questionnaires completed by these participants showed that users have difficulties in morphologically segmenting the Zulu/Zimbabwean Ndebele noun. Users with some linguistic training were quick to morphologically segment the nouns and highlighting the criteria they use to ascertain the noun stem. However, the majority of high school learners and a sizeable number of both university students and lecturers, failed to morphologically segment the nouns; a clear indication of the dire need for criteria for identifying and ascertaining the noun stem. Consequently, the study proposes at least three criteria that can be employed to identify and ascertain the Zulu and Zimbabwean Ndebele noun stem, namely the singularity and plurality criterion, the subject concord criterion and the morphophonological criterion. It was noted that these criteria provide a reliable way of identifying and ascertaining the Zulu and Zimbabwean Ndebele noun stem and this will make it easy to lemmatise using the stem method in a very consistent manner.

References

De Schryver GM. 2010. Revolutionizing Bantu lexicography – A Zulu case study. *Lexikos* 20:161–201.

De Schryver GM. 2008 A new way to lemmatize adjectives in a user-friendly Zulu – English dictionary. *Lexikos* 18: 63 – 93.

Dent. G.R., Nyembezi, C.L.S. 1969. *Scholar's Zulu dictionary: English – Zulu, Zulu – English*. Pietermaritzburg: Shuter and Shooter.

Doke, C.M., Vilakazi, B.W. 1948. *Zulu – English dictionary*. Johannesburg: Witwatersrand University Press.

Doke, C.M., Malcom, McK. and Sikakana, J.M.A. 1958. *English – Zulu dictionary*. Johannesburg: Witwatersrand University Press.

Hadebe, S.*et.al*. 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.

Maphosa M. 1997. The morphological structure of the noun in Ndebele and its implications

on the ordering of entries in Ndebele dictionaries. BA Honours Dissertation. University of Zimbabwe, Zimbabwe.
Nkabinde, A.C. 1982. *Isichazamazwi 1*. Pietermaritzburg: Shuter & Shooter.
Nkabinde, A.C. 1985. *Isichazamazwi 2*. Cape Town: Oxford University Press.
Nkomo, D. and Moyo, N. 2006. *Isichazamazwi SezoMculo*. Gweru: Mambo Press.
Nyembezi, S.L. 1992. *A-Z Isichazamazwi Sanamuhla Nangomuso*. Pietermaritzburg: Reach Out.
Pelling, J.N. 1965. *A Practical Ndebele dictionary*. Harare: Longman Publishers.
Van Wyk, E.B. 1995. Linguistic assumptions and lexicographical traditions in the African languages. *Lexikos* 5: 82–96.

*****

Cross-referencing in *Isichazamazwi SesiNdebele*

Eventhough NDLOVU (evennthough@yahoo.co.uk)
Unit for Language Facilitation and Empowerment, University of the Free State,
Bloemfontein, South Africa
Thompson NDLOVU (thompsonndlovu@gmail.com)
Department of African Languages and Literature, University of Zimbabwe,
Harare, Zimbabwe

This paper examines cross-referencing in *Isichazamazwi SesiNdebele* (henceforth: *ISN*) 2001. *ISN* is the first Zimbabwean Ndebele monolingual general-purpose dictionary, showing that dictionary-making in Zimbabwean Ndebele is still in its infancy and a lot of work lies ahead in this area. Reviews of this pioneering work will go a long way in improving the compilation of future Zimbabwean Ndebele dictionaries, especially from a user-oriented perspective. This paper focuses on how cross-referencing was employed to save space and enhance the microstructure, mediostructure and macrostructure of *ISN*. In this study, cross-referencing is examined with respect to headword selection, definitions and cross-referencing of synonyms, variants and illustrations. Maphosa and Nkomo's (2009) study shows that synonyms and variants top the list of the information categories required by Ndebele dictionary users. As such, this paper closely examines the use of cross-referencing in the treatment of these information categories. According to Gouws and Prinsloo (2005:177), cross-referencing is a lexicographic device that is used to establish relations between different components of a dictionary and save space therein. It interconnects the knowledge elements represented in different sectors of the dictionary's levels of lexicographic description to form a network and to enhance acceptability of the dictionary by including entries from other varieties of the language as variants and/ or synonyms. Gouws and Prinsloo (2005:177 – 192) identify the following pitfalls of cross-referencing: circular cross-referencing, dead cross-referencing, failure to utilise cross-referencing where needed, cross-referencing to the wrong address, cross-referencing that misguides the user during information retrieval and the use of cross-referencing to avoid a full treatment of the lemma. In this paper, cross-referencing in *ISN* is assessed in terms of its user-friendliness, accessibility, and its ability to meet user needs and user perspectives. Through the findings of this paper, the researchers hope to suggest ways of improving the use of cross-referencing in dictionary-making in Zimbabwean Ndebele. The key participants of the study were Ordinary and Advanced Level Ndebele majors in two selected secondary schools in Gwanda district, 2015 and 2016; BAA, BA Honours, BA Dual Honours Ndebele and Linguistics students; students of the Master of Arts degree in African Languages and Literature at the University of Zimbabwe; and Ndebele lecturers in the Department of African Languages and Literature at the University of Zimbabwe, Lupane State University,

Midlands State University, and Great Zimbabwe University. An in-depth analysis of *ISN* and interviews with *ISN* compilers and users were conducted to gather the perspectives of compilers and users on *ISN*'s use of cross-referencing in terms of how it enhanced dictionary user-friendliness, accessibility acceptability and how it satisfied user needs and met user perspectives. A critical analysis of *ISN* and data gathered through interaction with study participants reveals that in some cases, cross-referencing was not effectively and properly employed. It was observed that in *ISN*, there are cases of circular cross-referencing, dead cross-referencing and unnecessary repetition of definitions of cross-referenced lemmata. These pitfalls also waste space, since no dictionary is spared the necessity of saving space. They also led to the failure to satisfy user needs and meet user perspectives. These pitfalls of cross-referencing can be attributed to lack of thorough editing of the dictionary and the heavy reliance on the traditional or intuitive approach to dictionary-making, among other things. This paper stresses the need for compilers to establish from the onset how cross-referencing will be employed so as to ensure consistency in its use, save space and enhance the dictionary's user-friendliness and accessibility. Emphasizing the value of consistency in lexicography, Zgusta (1971) and Maphosa and Nkomo (2009) note that for the user to master the microstructure once and for all and continue using the dictionary efficiently, there is need to ensure consistency in the presentation of information categories, which would go a long way in educating users on how to use the dictionary. We underscore the need for thorough editing to ensure consistency in the use of cross-referencing and avoid the afore-mentioned pitfalls associated with cross-referencing. We also recommend the adoption of the mixed approach to dictionary-making, where the intuitive approach is complemented by the corpus based approach or vice versa.

References

Gouws, R.H. and Prinsloo, D.J. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN MeDIA.

Hadebe, S.*et.al*. 2001. *IsichazamazwiSesiNdebele*. Harare: College Press.

Maphosa, M. and Nkomo, D. 2009. The Microstructure of IsichazamazwiSesiNdebele. *Lexikos*, 19: 38 – 50.

Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

*****

A perspective on online dictionaries for African languages

Danie PRINSLOO (danie.prinsloo@up.ac.za)
Department of African Languages, University of Pretoria, South Africa
Jacobus PRINSLOO (jacobus.rezist@gmail.com)
Aerosud, National Laser Centre, CSIR, Pretoria, South Africa
Daniel PRINSLOO (dprinsloo@gmail.com)
Entelec, Johannesburg, South Africa

African language lexicography does not stand in isolation – dictionaries for these languages are influenced by trends, changes and developments in international lexicography, especially in the major languages of the world such as English, French and German. African language lexicography however has certain unique challenges that cannot be met by a simple one-size-fits-all lexicographic approach. To the root of the problems lies what could in basic terms be described as complex grammatical systems, the classification of nouns into different classes, a complex concordial and pronominal system, problematic lemmatisation traditions and orthographic systems. African language lexicographers have to fulfil the role of mediators between such complex systems on the one hand and their dictionary users on the other. It will

be shown that online dictionaries offer more ways to the lexicographer to deal with such complex grammatical systems and also have the potential to solve the difficulties caused by e.g. stem lemmatisation in paper dictionaries.

One of the major developments in lexicography is the compilation of electronic dictionaries. Currently available online dictionaries (accessible on the internet) for African languages such as Wolof, Yoruba, Kinyarwanda, Hausa, isiZulu, Sesotho, Sepedi and Tshivenda were studied and the aim of this paper is to give a critical evaluation of African language dictionaries in the era of the internet. Prominent expectations for online dictionaries include new designs of electronic dictionaries for these languages, the utilization of electronic features enabled by the computer era and the maximum utilization of speed and space. Prinsloo (2011) even expects electronic dictionaries to solve lexicographic problems such as stem identification in conjunctively written languages. The aim of this paper is not to attempt a comprehensive overview of African language dictionaries or to calculate the number of such dictionaries, e.g. as done by De Schryver (2003), but rather to reflect on the quality of current online dictionaries for the African languages.

Different stages in the development of electronic dictionaries can be distinguished. First, for many languages of the world paper dictionaries were compiled which could be described as of low lexicographic achievement. This was followed by a stage in which paper dictionaries reached a level of real lexicographic achievement, cf. for example the so-called "Big five" dictionaries CALD, COBUILD, LDOCE, MED and OALD.

At the dawn of the electronic era, dictionary compilers tried to get the best of both worlds by putting a CD ROM dictionary in the back pocket of the dictionary. This was followed by a strategy of merely presenting the content of the paper dictionary online, sugar-coated by additional search functions and clickable icons for e.g. pronunciation guidance. Until recently publishers were hesitant to make the quantum leap from paper dictionaries to fully-fledged online dictionaries. Macmillan, however, took such a bold step in 2012.

It will be shown in this paper that for online African language dictionaries a wide spectrum of achievement exists, ranging from mere word lists with or without translations, very small dictionaries with limited treatment, technically fully functional dictionaries but with empty alphabetical stretches, to dictionaries of high lexicographic achievement employing innovative electronic technology.

For the African languages the development from paper dictionaries to online dictionaries was perhaps more traumatic than for the major languages of the world because the internet era dawned on the African languages at a time when the compilation of paper dictionaries of high lexicographic standards such as the "Big five" had not yet been fully achieved. The pressure to produce electronic dictionaries came at a time when most dictionaries for African languages had not yet reached a high level of sophistication.

It will be argued and illustrated by means of examples that currently available dictionaries for African languages can be categorised into online dictionaries that are:

- merely scanned images of paper dictionary pages,
- word lists with or without basic translation equivalents,
- identical to the paper dictionary but in electronic format with search functions added, and
- dictionaries of high lexicographic achievement

The following online dictionaries will be discussed: *Bilingo Multilingual South African Dictionary*, *cBold*, *Dicts.info*, *Freedict.com*, *Freelang.net*, *Hausa Dictionary*, *isiZulu.net*, *Kinyarwanda Dictionary*, *Macmillan online dictionary* and *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete*.

Specific attention will be given to the presentation of the data in online dictionaries for these languages.

This paper is presented against the background of African languages being lesser resourced languages and often lacking a strong dictionary culture. The under-development of human languages technologies also contributes to the challenges faced by African language lexicographers.

References

(CALD) McIntosh, Colin (Ed.). 2013. Cambridge Advanced Learner's Dictionary. Cambridge: Cambridge University Press.
(COBUILD) Carroll, Katherine (Ed.). 2012. Collins COBUILD Advanced Dictionary of English. Glasgow: HarperCollins.
(LDOCE) Mayor, Michael (Ed.). 2009. Longman Dictionary of Contemporary English. Harlow, Essex: Pearson Education.
(MED) Rundell, Michael (Ed.). 2007. Macmillan English Dictionary for Advanced Learners. Oxford: Macmillan Education.
(OALD) Deuter, Margaret, Jennifer Bradbery and Joanna Turnbull (Eds.). 2015. Oxford Advanced Learner's Dictionary of Current English. Oxford: Oxford University Press.
De Schryver, G-M. 2003. Online Dictionaries on the Internet: An Overview for the African Languages. Lexikos 13: 1-20.
Prinsloo, D.J. 2011. A critical analysis of the lemmatisation of nouns and verbs in isiZulu. Lexikos 21. 169-193.

Online dictionaries
Bilingo Multilingual South African Dictionary: http://www.bilingo.co.za
Bukantswe: http://bukantswe.sesotho.org/
cBold: Venda.Murphy1997.txt: http://www.cbold.ish-lyon.cnrs.fr/
Dicts.info: http://dicts.info/dictlist1.php
Freedict.com: (http://freedict.com/onldict/afr.html).
Freelang.net: http://www.freelang.net/online/afrikaans.php?lg=gb
Hausa Dictionary: http://maguzawa.dyndns.ws/frame.html
isiZulu.net: https://isizulu.net/
Kinyarwanda Dictionary: http://kinyarwanda.net/
Macmillan online dictionary: http://www.macmillandictionary.com/
Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete: http://africanlanguages.com/psl/.
Webster's Online Dictionary: http://www.websters-online-dictionary.org/. Consulted 25/8/2012.
Wolof Dictionary, Sierra Dem, Peace Corps. 1995
http://www.africanculture.dk/gambia/ftp/wollof.pdf
Yoruba – English dictionary:
http://www.yorubadictionary.com/Yoruba_English/yoruba_p.htm

*****

Dictionaries in the knowledge age: What must lexicographers do in Zimbabwe?

Emmanuel SITHOLE (emmanesu@gmail.com)
School of Languages and Literature, Rhodes University, Grahamstown, South Africa

Lexicography has always played a central role towards the development of human societies since its inception as a practice some four centuries ago. Scholars express unanimity about the importance of dictionaries in performing linguistic, communicative and cognitive functions in society (Zgusta, 1971; Gouws, 2007; Hadebe, 2006; Nkomo, 2017). Scholars such as Tarp

(2008, p. 22) offered remarkable views that the compilation of dictionaries has been "a problem solving activity" thereby resonating with Landau's (2001, p. 21) postulation that dictionaries are both "containers of knowledge" and "storehouses of information". Executing such functions and more, it is evident that dictionaries are utility products meant to fulfill a myriad of societal needs (Tarp, 2008) across different epochs. However, in light of the ever-changing technological conditions, it is imperative that lexicographic products retain their relevance by transforming to match prevailing environmental changes in society. In view of the contemporary knowledge age, it is essential to critically re-examine the roles that lexicographers execute towards creating information and knowledge in the world.

Conceived against a background where there are few quality dictionaries in majority languages, low dictionary culture, poor dictionary using skills, less comprehensive dictionaries partnered by a dwindling motivation to compile dictionaries in Zimbabwe, it is critical to probe how lexicography can play a more instrumental function in developing societies especially in the contemporary technology era. The information and knowledge gaps identified raise pertinent questions such as the following: Is lexicography still relevant in Zimbabwe? What measures can expedite the optimal use of lexicographic products in the country? What new roles must Zimbabwean lexicographers assume to retain their relevance in today's information and/or knowledge age? What lexicographical devices can consolidate lexicographers' roles in the country? While these general questions may broadly apply to many African countries in similar predicaments, they are particularly significant to Zimbabwe which faces a cocktail of lexicographic problems described above. Specifically, they seek to re-examine and re-configure lexicographers' functions to bridge the identified gaps across high status domains. In that sense, the paper intends to urge lexicographers to extend their focus to adopt and adapt environmental and/or technological changes to not only retain their relevance but also to benefit the Zimbabwean society and its respective languages.

It is interesting to notice that questions posed above highlight the importance of understanding lexicography from a functional perspective as expressed in Bergenholtz and Tarp's (1995) and Tarp's (2008) lexicographical function theory. The lexicographical function theory places agency on dictionary users and the specific types of information and knowledge needed in society. In this paper, the emphasis on the user does not only determine the appropriateness of lexicographical interventions but also places an obligation on lexicographers to devote time, expertise and resources to create information and knowledge in domains and environments where it is mostly needed. In the case of Zimbabwe, focus needs to be channelled towards: utilizing new and emerging human language technologies to expedite dictionary compilation; producing dictionaries to develop majority and minority languages; improving dictionary compilation and dictionary user skills; and enhancing dictionary culture in society.

The notion that lexicographic materials can be developed and utilized to achieve specific goals is not peculiar to Zimbabwe. Throughout Africa, missionaries compiled early dictionaries to pursue evangelical and pedagogical functions (Doke, 1931). Apart from serving as missionaries' handbooks for disseminating Christianity among local people, early dictionaries simultaneously functioned as indigenous language instruction manuals to fellow white settlers who needed to communicate with their local servants. Therefore, it can be speculated that the success of Christianity among other aspects of foreign (western) culture in Africa cannot be objectively divorced from the production of evangelical, dictionaries and other works that purveyed them. Equally, early dictionaries among other works immensely contributed towards the development and standardization of Shona and Ndebele in Zimbabwe (Hadebe, 2006). Based on that, this paper argues that Zimbabwean lexicographers can create an enabling platform for the production of new information to benefit marginalized societies and their respective languages through dictionaries. What is needed is a holistic approach in

Zimbabwean lexicography, for example, publishing general and specialized lexicography; monolingual and bi/multilingual lexicography; and embracing majority and minority languages to provide information and knowledge across all sectors and ethnolinguistic groups in Zimbabwe.

The researcher will gather information through purposive sampling. Semi-structured interviews will be used to gather data from five lexicographers representing Shona, Ndebele, Ndau, Shangani and Tonga. Shona and Ndebele lexicographers have practical experience in compiling general and specialized dictionaries in their mother tongues, Ndau and Shangani lexicographers boast of recently published general dictionaries whereas Tonga lexicographers in Zimbabwe are "still hard at work to produce one" (Mumpande 2015, personal communication). Taken together, the researcher will blend insights from Shona and Ndebele lexicographers with new minority language lexicographers' experience to argue for the need to widen the scope and activities to usher in a new lexicographic practice in Zimbabwe.

References

Doke, C. (1931). *The report on the unification of Shona dialects*. Hereford: Austin and Sons.

Bergenholtz, H., and Tarp, S. (1995). *Manual of Specialised Lexicography*. Amsterdam: John Benjamins.

Gouws, R. H. (2007). A transtextual approach to lexicographic functions. In *Lexikos* 17: 77 - 87.

Hadebe, S. (2006). *The standardisation of the Ndebele language through dictionary making.* Oslo: The ALLEX Project.

Landau, S. I. (1989). *Dictionaries. The art and craft of lexicography*. New York/Cambridge: Cambridge University Press.

Mumpande, I. (2015, July 18). Personal communication.

Nkomo, D. (2017). Dictionaries and language policy. In P. Fuertes-Olivera (Ed.). *Routledge Handbook of Lexicography* (In press). London: Routledge.

Tarp, S. (2008). *Lexicography in the Borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography*. Tübingen: Max Niemeyer.

Zgusta, L. (1971). *Manual of lexicography*. The Hague: Mouton and Company.

*****

Dictionary criticism and lexicographical function theory

Sven TARP (St@cc.au.dk)

Centre for Lexicography, Aarhus University, Aarhus, Denmark

Since Paolo Beni's famous *L'anticrusca*, published by the Academy of Crusca in 1612 as a response to the *Vocabulario* compiled by the academicians from the very same Academy, dictionary criticism has played an important role in the development of dictionaries as well as lexicographical theory. The paper will discuss dictionary criticism in the light of the function theory of lexicography (cf. Tarp 2017) and will argue that this criticism is an essential lexicographical activity which constantly needs to be enhanced if the discipline should continue to develop. In this respect, the paper will above all treat dictionary criticism as a *research activity* that encompasses a number of specific problems, of which the most important are:

- the purpose of dictionary criticism;
- the main types of dictionary criticism;
- the concept of dictionary criticism;
- the required knowledge;

- the main elements to be criticized;
- the method to be applied;
- the presentation of the results and conclusions;
- the theoretical and practical implications.

Like any other lexicographical activity, dictionary criticism should always be performed with a specific and well-considered purpose in mind. The paper will therefore discuss the most important purposes with which criticism has been and can be made:

1. To raise a debate among lexicographers about concrete dictionaries.
2. To raise a debate among lexicographers about specific topics.
3. To make recommendations to the author(s) of the criticized dictionary.
4. To inspire other lexicographers when preparing similar dictionary projects.
5. To prepare or improve own dictionaries.
6. To transmit interesting experiences.
7. To initiate students into the world of lexicography.
8. To recommend or not recommend a dictionary to its potential users.

The paper will provide some examples of dictionary criticism carried out with the above purposes as well as their theoretical and practical outcome. In this connection it will also discuss the two main types of dictionary criticism, namely criticism of other authors' dictionaries and self-criticism of one's own dictionaries, where the second one is frequently ignored in the academic literature in spite of the fact that a self-critical and fearless approach to one's own dictionaries is of fundamental importance when preparing new projects or new editions (updating) of already published works.

Based on this discussion, the paper will proceed to a definition of the concept of dictionary criticism. It will argue that it cannot be reduced to neither a specific genre as it is defined by text linguistics nor an independent research area as defined by Wiegand (1989). Instead it will define dictionary criticism as a *theory-based activity* conducted within various lexicographical research areas, and the outcome of which may be expressed in texts belonging to *many different genres* or even kept indoors depending on the specific purpose of the criticism.

Moreover, the contribution will briefly discuss the various types of knowledge and skills required to make a comprehensive criticism, especially in terms of modern online dictionaries; the issues which may be criticized (in this respect it will distinguish between lexicographical and non-lexicographical issues); the overall method applied by the supporters of the function theory, and the way dictionary criticism could be presented in order to create debate. Finally, the paper will indicate the important role dictionary criticism has had, and still has, in the development of the lexicographical theory and practice, and will therefore endorse an open and critical discussion culture within the discipline.

References

Beni, P. 1612: *L'anticrusca*. Florence: Accademia della Crusca, 1612.

Tarp, S. 2017: Dictionary criticism and lexicographical function theory. In M. Bielińska and S.J. Schierholz (eds.): *Wörterbuchkritik*. *Dictionary Criticism*. Berlin, Boston: De Gruyter.

Wiegand, H.E. 1989: Der gegenwärtige Status des Lexikographie und ihr Verhältnis zu anderen Disziplinen. In F.J. Hausmann, O. Reichmann, H.E. Wiegand and L. Zgusta, (eds.): *Wörterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography*. Berlin/New York: Walter de Gruyter, 246-280.

*****

An African word list proposal using NSM as a lexicographic starting point

Bruce WIEBE (bruce.wiebe@sil.org)
SIL Nigeria, Jos, Nigeria

The central claim of this paper is that the semantic theory of Natural Semantic Metalanguage (NSM) (Goddard, 2011; Goddard & Wierzbicka, 2014a) offers some lexicographic insights that aid (1) the choice of foundational / core lexemes for a beginning dictionary (or defining vocabulary) (Atkins & Rundell, 2008:448-450), and (2) the writing of definitions that are more (a) clear, (b) easily and directly translatable, and (c) free of Anglocentrism / ethnocentrism. Support for these four sub-claims come from the results of the NSM research program. It has specifically aimed at uncovering semantic dependencies (expressing concepts in terms of simpler concepts) and ultimately the semantic core, the foundational concepts, of a language. By principle it restricts definitions to use of either universal concepts, or semantic bundles from the language being described, and disallows definitions that are circular or increase semantic complexity. This promotes (1) inclusion and layering of increasingly core lexemes according to semantic dependency, and (2) definitions that (a) avoid semantic complexity and circularity, (b) use easily translated universal concepts, or concepts that can be broken down into such, and (c) avoid concepts not found in the language being described, rather than importing difficult, foreign, technical terms (Goddard & Wierzbicka, 2014b:80-82). A word list proposal is presented, together with the beginnings of work on a dictionary for it, which apply the theory and demonstrates how these four goals are accomplished.

NSM is a theory that proposes, among other things, (1) that semantic explication should be done by reductive paraphase, expressing concepts in simpler concepts (Goddard, 2011:64-65), and (2) that each language is sufficient to describe its own semantics, because there is a core of 65 semantic primes, semantically irreducible concepts, that exist as lexical items (words, set phrases / idioms, or morphemes) in every language, and all other concepts can be built up from these (Goddard, 2011:65-67). The claim to the existence of these primes has been empirically verified in about 30 languages, from a wide variety of geographical areas and language families (cf. Goddard, 2011:68). Much work has been done over the 45 year life span (cf. Wierzbicka, 1972) of this theory, on semantic explications in various languages using these semantic primes. We follow the basic theoretical claims evidenced by these and other NSM researchers, and look at their implications for lexicographic practice. Goddard and Wierzbicka (2014b:89) have said, "In our view, the inventory of semantic primes should be an item in the toolkit (and backpack) of every field linguist."

I am in a working group at CanIL that is generating a revision (reduced subset) of the SIL Comparative African Word List (SILCAWL, Snider & Roberts, 2004). The working group includes an original author of SILCAWL, and members with fieldwork experience in a variety of languages from around the African continent (Cameroon, Nigeria, Ghana, DR Congo, Kenya, and Mozambique). SILCAWL already largely avoids culture-specific or domain-specific words (cf. Atkins & Rundell, 2008:24,33) to get at a common core of language.

My proposed English / French wordlist, branching from this, includes all NSM semantic primes, many of which were missing from SILCAWL, as well as universal or near-universal NSM semantic molecules (explained next), which were in SILCAWL and needed to be retained. Semantic primes are like atoms, and semantic molecules are bundles of semantic primes that form concepts that are important in a language, and useful in explicating other concepts. A number of these have been investigated and found to be universal or near-universal (Goddard, 2011:375-383). Semantic primes, semantic molecules, and the semantic grammar for combining one into the other (Goddard, 2011:69-71) will be surveyed, and the resulting core lexeme inclusion and layering shown.

The accompanying dictionary (in process) follows principles and methodology outlined in Newell (1995), Atkins and Rundell (2008), and Goddard (2011), and serves as an example of the application of NSM theory to the creation of a beginning dictionary, with clear, easily translatable, and culturally appropriate definitions, that will transfer well to African languages.

References

Atkins, B T Sue and Rundell, Michael. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Goddard, C. 2011. *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.

Goddard, C and Wierzbicka, A. 2014a. *Words and Meanings: Lexical Semantics across Domains, Languages, and Cultures*. Oxford: Oxford University Press.

Goddard, C and Wierzbicka, A. 2014b. Semantic fieldwork and lexical universals. *Studies in Language* 2014 (38) 1:80-127.

Newell, L. E. 1995. *Handbook on Lexicography for Philippine and Other Languages*. Manila: Linguistic Society of the Philippines.

Snider, K and Roberts, J. 2004. SIL Comparative African Word List (SILCAWL). *The Journal of West African Languages* (31) 2:73-122.

Wierzbicka, A. 1972. *Semantic Primitives*. Frankfurt: Athenaum.