AFRICAN ASSOCIATION FOR LEXICOGRAPHY
21st International Conference

4 - 7 July 2016

XITSONGA NATIONAL LEXICOGRAPHY UNIT
in cooperation with
SESOTHO SA LEBOA NATIONAL LEXICOGRAPHY UNIT

Tzaneen, South Africa
AFRICAN ASSOCIATION FOR LEXICOGRAPHY

Programme and Abstracts

21st International Conference

XITSONGA NATIONAL LEXICOGRAPHY UNIT in cooperation with SESOTHO SA LEBOA NATIONAL LEXICOGRAPHY UNIT,

Tzaneen, South Africa

4 - 7 July 2016

Hosted by: XITSONGA NATIONAL LEXICOGRAPHY UNIT in cooperation with SESOTHO SA LEBOA NATIONAL LEXICOGRAPHY UNIT, Karibu Hotel and Leisure Resort, Tzaneen, South Africa.

Conference organiser: Mr H.T. Mashele

Abstract reviewers: Prof. Herman L. Beyer, Mr André du Plessis, Dr Hughes Steve Ndinga-Koumba-Binza, Dr Langa Khumalo, Dr Dion Nkomo, Prof. Thapelo J. Otlegetswe, Prof. Annel Otto, Dr Michelle van der Merwe.

Abstract booklet editors: Dr Dion Nkomo & Prof. Sonja E Bosch

© 2016 AFRILEX, African Association for Lexicography
ISBN 978-0-620-71514-0
**TABLE of CONTENTS**

AFRILEX HONORARY MEMBERS ........................................................................................................... 5

AFRILEX BOARD ..................................................................................................................................... 6

MESSAGE FROM THE AFRILEX PRESIDENT .......................................................................................... 6

PROGRAMME ........................................................................................................................................ 8

Monday 4 July 2016 ................................................................................................................................ 8

Pre-conference workshop and tutorial on corpus collection and corpus manipulation 8

Parallel Sessions ...................................................................................................................................... 9

Tuesday 5 July 2016 .............................................................................................................................. 9

Parallel Sessions ...................................................................................................................................... 9

Wednesday 6 July 2016 .......................................................................................................................... 10

Parallel Sessions ...................................................................................................................................... 11

Thursday 7 July 2016 ............................................................................................................................. 11

KEYNOTE PRESENTATION 1 .................................................................................................................. 12

KEYNOTE PRESENTATION 2 .................................................................................................................. 14

PARALLEL SESSIONS .......................................................................................................................... 16

The lexicographic treatment of isiZulu negation

Sonja BOSCH & Gertrud FAASS ............................................................................................................. 16

Demonstration of the pilot project for a new online *Woordeboek van die Afrikaanse Taal*

Willem F. BOTHA & Hanelle FOURIE BLAIR ......................................................................................... 18

Are we there yet? An investigation of the VivA-app as Afrikaans’ answer to the mobile dictionary market

André H. DU PLESSIS ............................................................................................................................. 19

Redesigning the DSAE’s presentation and navigation features: adapting a historical dictionary for the modern online user

André H. DU PLESSIS & Tim VAN NIEKERK .......................................................................................... 21

Integrating online reference sources into a comprehensive dictionary culture

Rufus H. GOUWS ................................................................................................................................... 23

The lemmatisation of Ideophones and Interjections in Cuenod’s *Tsonga–English Dictionary*

Madala C. HLUNGWANI & Memory MPAPA ......................................................................................... 24

Intellectualization through terminology development

Langa KHUMALO ..................................................................................................................................... 25

Term creation through translation: An evaluation of Shona phonetics and phonology
termology at GZU
Penniah MABASO-SHAMANO
Reflections towards the revision of *Dikixinari ya Xitsonga ya ka Lingua Franca* 28

Ximbani E. MABASO & Mangalani J. HLONGWANE
Xitsonga dictionaries a boom or doom for the standardisation of Xitsonga: A linguistic introspection 29

Ximbani E. MABASO
A discussion of isiNdebele terminology 31

Katjie S. MAHLANGU
The use of a corpus in the development of an electronic monolingual Sesotho school learners’ dictionary 32

Malefu J. MAHLOANE
Powers and problems of language modernisation in lexicography: A review of terminological approaches in Northern Sotho dictionaries 33

Kegaugetjwe M. MAKGOPA
Efficacy of bilingual dictionaries as translator’s aid: A case of Sesotho sa Leboa and isiXhosa 35

Kwena MASHAMAITE & Feziwe SHOBA
The role of translation in lexicography with special reference to Tshivenda-English dictionaries 36

Mashudu MAŢHABI
Problems and perspectives in Civilii lexicography 38

Paul A. MAVOUNGOU & Hugues S. NDINGA-KOUMBA-BINZA
Electronic dictionaries and lexicographical theory: A case study of the *Dictionnaire Intercatif pour Jeunes Apprenants* 39

Ludwine MBINDI ANINGA
Developing an LSP bilingual glossary to support international students at the University of Johannesburg 41

Guy-Roger MIHINDOU & Mapelo TLOWANE
The treatment of synonym paradigm members and plural nouns in the macrostructure of Thanodi ya Setswana and strategies to enhance its reference quality 42

Nickey S. MOKONE
Pronunciation in African language dictionaries 44

Hugues S. NDINGA-KOUMBA-BINZA
Lexicography as a discipline and profession in Zimbabwe: A critical appraisal 46

Eventhough NDLOVU
S’ncamtho lexico-semantics and terminology development in Ndebele: Implications on lexicography and standardisation 47

Sambulo NDLOVU
Lexicography and language revitalisation for the Tjwao language in Zimbabwe 49
Shumirai NYOTA & Zvinashe MAMVURA
How much specialised data can and should be incorporated into meaning explanations in general language dictionaries? A discussion based on sports terms

Heidi A.P. PEDERSEN
The Sepedihelper as a support tool for text production in a dictionary use situation

Daniel J. PRINSLOO, J.V. PRINSLOO & Danie J. PRINSLOO
Lexicographic approach to spoken versus written corpora for Setswana dictionaries

Fannie SEBOLELA & Danie J. PRINSLOO
Lost and found: Assessing the value of an unknown bilingual dictionary towards the development of Ndau

Emmanuel SITHOLE
Advantages and disadvantages in the use of the internet as a corpus. The case of the online dictionaries of Spanish “University of Valladolid”

Sven TARP
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:** Prof. D.J. (Danie) Prinsloo  
**Members:** Prof. S.E. (Sonja) Bosch  
Dr L. (Langa) Khumalo  
Dr H.S. (Steve) Ndinga-Koumba-Binza  
Dr D. (Dion) Nkomo  
**Conference organiser:** Mr H.T. (Hlengiwe) Mashele

### MESSAGE FROM THE AFRILEX PRESIDENT

On behalf of the AFRILEX Board, I would like to welcome all of you to the 21\textsuperscript{st} Annual International Conference of the African Association for Lexicography, also known as ‘AFRILEX 2016’. This year’s edition takes place in the magnificent Karibu Hotel & Leisure Resort (Karibu, a Kiswahili word for ‘Welcome’), in the warm and tropical Mopani District of the Limpopo Province, next to cleanest town of Tzaneen. Our conference this year is hosted by the Xitsonga National Lexicography Unit which is hosted in the Tivhumbeni Multipurpose Centre, in collaboration with the Sesotho sa Leboa National Lexicography Unit, which is hosted in the University of Limpopo. This conference follows the successful AFRILEX 2015 which was hosted by the Howard College Campus of the University of Kwa-Zulu Natal in Durban. 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 Europe, Namibia, Gabon and Botswana, Zimbabwe, who are always with us in every conference. We want to invite many lexicographic scholars from the entire African continent to form part of the membership of AFRILEX. Since AFRILEX 2011, which was held in the University of Namibia, all other subsequent Annual 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. Presently, the Board is still busy considering an invitation to a back-to-back conference with
other language associations at Rhodes University for AFRILEX 2017. If the Rhodes invitation can be positively accepted we hope that AFRILEX 2018 will possibly be held outside this country, on invitations from our colleagues.

AFRILEX 2016 has been meticulously prepared by a local organising team under the leadership of Mr Hlengiwe Mashele, as the local organiser. At this moment I also want to thank Prof. Danie Prinsloo, the AFRILEX Deputy President who, as always, actively participated in the preparations leading to this Conference. The abstract adjudication process for AFRILEX 2016 was expertly managed and carried out by Dr Dion Nkomo, Prof. Sonja Bosch and all the AFRILEX Board members, including the following: Prof. Herman L. Beyer, Mr André du Plessis, Dr Hughes Steve Ndinga-Koumba-Binza, Dr Langa Khumalo, Dr Dion Nkomo, Prof. Thapelo J. Otlegetswe, Prof. Annél Otto, and Dr Michelle van der Merwe for serving as peer reviewers of the abstracts. Prof. Sonja Bosch and Dr Dion Nkomo also did commendable work in the compilation of this Abstract Booklet we are holding now. We want to congratulate and thank them for the job well-done. We also want to say thank you once-more to Prof. DJ Prinsloo who excellently managed and kept the AFRILEX website up to date and managed the compilation of the programme for this conference. Not forgetting Prof. Elsabé Taljard, our reliable treasurer who, as always, continuously keeps the AFRILEX moneys safe.

Just like the previous conferences, AFRILEX 2016 promises to be another stellar gathering, with speakers coming from a dozen different countries in Africa, the USA and Europe, namely Belgium, Botswana, Denmark, DR Congo, Gabon, Germany, Namibia, Poland, South Africa, Spain, Zimbabwe and the Czech Republic. We also want to thank Mr Hlengiwe Mashele and the Board of the Ngula ya Xitsonga National Lexicography Unit for hosting our conference in this magnificent forest paradise. In conclusion, AFRILEX wants to thank Ms Nyeleti Mushwana, the owner of the Karibu Hotel & Leisure Resort for hosting our conference and accommodating almost all of our members in this magnificent hotel.

We also continue the tradition of giving the floor to dictionary publishers during a Publishers’ Session. Our international keynote speaker this year is Prof. Serge Verlinde from the Leuven Language Institute, whose presentation is entitled ‘(E-)lexicography and beyond’. I want to welcome Prof. Verlinde on the African continent and at AFRILEX 2016, and in the Limpopo Province in particular. The national keynote will be delivered by Dr Victor Mojela, from the Sesotho sa Leboa National Lexicography Unit, with a presentation entitled ‘Lexicography versus standardization: a huge barrier in the pace of lexicography for the developing SA indigenous languages’.

Maropeng Victor Mojela
President: AFRILEX
## PROGRAMME

### Monday 4 July 2016

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30 – 09:55</td>
<td><strong>Registration and tea</strong></td>
</tr>
</tbody>
</table>
| 10:00 – 12:55  | **Pre-conference workshop and tutorial on corpus collection and corpus manipulation**  
**Session 1: Presenter: Uwe Quasthoff, University of Leipzig, Germany**  
Introduction to corpora collection:  
Corpus collection: Crawling and pre-processing (HTML stripping, language identification, sentence identification, pattern-based corpus cleaning)  
- Calculation and usage of word co-occurrences  
- Corpus presentation in the Web  
- Part of speech tagging  
- Using the corpus for lexicography  
**Session 2: Presenter: Dirk Goldhahn, University of Leipzig, Germany**  
Tutorial with hands-on exercises:  
- The connection between relevance and frequency for lexicography  
- Use the NoSketchEngine for advanced corpus search  
- Explain the Corpus Query Language CQL  
- Identify queries interesting for different applications  |
| 13:00 – 13:55  | **Lunch**                                                                 |
| 14:00 – 14:25  | **Official Opening**                                                      |
|                | **Victor Mojela – President of AFRILEX**                                  |
|                | A word of welcome                                                        |
|                | **Local hosts and organisers**                                           |
| 14:30 – 15:25  | **Keynote Address 1:**                                                   |
|                | **Serge Verlinde, Leuven Language Institute**                            |
|                | (E-)lexicography and beyond                                               |
| 15:30 – 15:55  | **Tea**                                                                   |
## Parallel Sessions

### 16:00 – 16:25
- **Are we there yet? An investigation of the VivA-app as Afrikaans’ answer to the mobile dictionary market**
  - Du Plessis, A.H.
- **Xitsonga dictionaries a boom or doom for the standardisation of Xitsonga: a linguistic introspection**
  - Mabaso, X.E.

### 16:30 – 16:55
- **Intellectualization through terminology development**
  - Khumalo, L.
- **A discussion of isiNdebele terminology**
  - Mahlangu, K.S.

### 18:00 – Cocktail Party

---

## Tuesday 5 July 2016

### Parallel Sessions

#### 09:00 – 09:25
- **Advantages and disadvantages in the use of the internet as a corpus. The case of the online dictionaries of Spanish “University of Valladolid”**
  - Tarp, S.
- **Lexicography as a discipline and profession in Zimbabwe: a critical appraisal**
  - Ndlovu, E.

#### 09:30 – 09:55
- **Demonstration of the pilot project for a new online Woordeboek van die Afrikaanse Taal**
  - Botha, W.F. and Fourie Blair, H.
- **Lexicography and language revitalisation for the Tjwao language in Zimbabwe**
  - Nyota, S. and Mamvura, Z.

#### 10:00 – 10:25
- **Powers and problems of language modernisation in lexicography: A review of terminological approaches in Northern Sotho dictionaries**
  - Makgopa, K.M.
- **Term creation through translation: an evaluation of Shona phonetics and phonology terminology at GZU**
  - Mabaso-Shamano, P.

#### 10:30 – 10:55
- **Tea**

#### 11:00 – 11:25
- **Lemmatizing syntactico-intonational units in Gabonese native language dictionaries**
  - Ella, E.M.
- **Lexicographic approach to spoken versus written corpora for Setswana dictionaries**
  - Sebolela, F. and Prinsloo, D.J.
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 11:30 – 11:55 | Developing an LSP bilingual glossary to support international students at the University of Johannesburg  
Mihindou, G.R. and Tlowane, M.                                                                 |
|              | Electronic dictionaries and lexicographical theory: a case study of the Dictionnaire Intercatif pour Jeunes Apprenants  
Mbindi Aninga, L.                                                                 |
| 12:00 – 12:25 | Pronunciation in African language dictionaries  
Ndinga-Koumba-Binza, H.S.                                                                 |
|              | Lost and Found: assessing the value of an unknown bilingual dictionary towards the development of Ndua  
Sithole, E.                                                                 |
| 12:30 – 13:25 | Lunch                                                                 |
| 13:30 – 13:55 | Integrating online reference sources into a comprehensive dictionary culture  
Gouws, R.H.                                                                 |
|              | The lemmatization of ideophones and interjections in Cuenod’s  
Tsonga – English Dictionary  
Hlungwani, M.C. and Mpapa, M.                                                                 |
| 14:00 – 14:25 | Reflections towards the revision of Dikixinari ya Xitsonga ya ka  
Lingua Franca  
Mabaso, X.E.                                                                 |
|              | The lexicographic treatment of isiZulu negation  
Bosch, S.E. and Faaß, G.                                                                 |
| 14:30 – 14:55 | Efficacy of bilingual dictionaries as translator’s aid: a case of  
Sesotho sa Leboa and isiXhosa  
Mashamaite, K.J. and Shoba, F.                                                                 |
|              | The role of translation in lexicography with special reference to Tshivenḓa-English dictionaries  
Maṱhabi, M.                                                                 |
| 15:00 – 15:25 | Tea                                                                 |
| 15:30 – 17:00 | Annual General Meeting                                                                 |
| 18:00        | Conference Dinner                                                                 |

**Wednesday 6 July 2016**

**Keynote Address 2:**

09:00 – 09:55 Victor Mojela, Sesotho sa Leboa National Lexicography Unit  
Lexicography versus standardization: a huge barrier in the pace of lexicography for the developing SA indigenous languages.
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 – 10:25</td>
<td>Tea</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Parallel Sessions</strong></td>
<td></td>
</tr>
<tr>
<td>10:30 – 10:55</td>
<td><strong>S’ncamtho lexico-semantics and terminology development in Ndebele:</strong></td>
<td>How much specialized data can and should be incorporated into meaning explanations in general language dictionaries? A discussion based on sports terms</td>
</tr>
<tr>
<td></td>
<td>implications for lexicography and standardization</td>
<td>Pedersen, H.A.P.</td>
</tr>
<tr>
<td></td>
<td>Ndlovu, S.</td>
<td></td>
</tr>
<tr>
<td>11:00 – 11:25</td>
<td>The treatment of synonym paradigm members and plural nouns in the macrostructure of Thanodi ya Setswana and strategies to enhance its reference quality</td>
<td>Redesigning the DSAE’s presentation and navigation features: adapting a historical dictionary for the modern online user</td>
</tr>
<tr>
<td></td>
<td>Mokone, N.S.</td>
<td>Du Plessis, A.H. and Van Niekerk, T.</td>
</tr>
<tr>
<td></td>
<td>Mavoungou, P.A. and Ndinga-Koumba-Binza, H.S.</td>
<td>Mahloane, M.J.</td>
</tr>
<tr>
<td>12:00 – 12:25</td>
<td>The Sepedihelper as a support tool for text production in a dictionary use situation.</td>
<td>TBA</td>
</tr>
<tr>
<td></td>
<td>Prinsloo, Daniel, Prinsloo J.V. and Prinsloo, D.J.</td>
<td></td>
</tr>
<tr>
<td>12:30 – 12:40</td>
<td>Closure</td>
<td>D.J. Prinsloo – Vice-President of AFRILEX</td>
</tr>
<tr>
<td>13:00 – 14:00</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Thursday 7 July 2016</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Post-conference excursion</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Kruger Park</strong></td>
<td></td>
</tr>
</tbody>
</table>
In his keynote at the Afrilex conference last year, Michael Rundell outlined how dictionaries should be redefined to include thesauri, corpora, data from the Web, and language resources, and how research should focus more on the use being made of online dictionaries in order to optimize dictionaries for the digital era. However, it seems lexicographers (and editors) still do not really know where to start or how to rethink the dictionary, and have so far not addressed the questions raised by Sue Atkins (1996:516) exactly twenty years ago:

It is up to us to take up the real challenge of the computer age, by asking not how the computer can help us to produce old-style dictionaries better, but how it can help us to create something new: to look at the needs of dictionary users of every language, and every walk of life, users as diverse as people themselves, and give them the kind of information they need for whatever they are using the dictionary for [...].

Perhaps lexicographers (and editors) should draw inspiration from what has happened with the camera: very sophisticated ones are available for the (amateur) photographer, as dictionaries often are for linguists or highly educated users, while everybody (else) uses a mid tech version on their smartphone.

This raises the question of what kind of tools our lexicographic data could be integrated in so that this information could be made available to a much wider audience than today and precisely when it is needed.

We believe this could well be in (smart) writing assistants, which add a layer of (lexicographic) information on top of the text while it is written by the user. To fully achieve this goal, we still need more lexicographic data. Additional analysis of corpus evidence is necessary, mainly in the field of word combinations, locally and at the sentence level, as we will illustrate with examples. However, for a large number of languages we already have enough data to start building such tools.

At our Institute, writing assistants have been developed or are ‘under construction’ for various languages (Academic Dutch, Academic English, Dutch as a foreign and native language, Afrikaans and French). They are designed for both language learners and native speakers, as well as for writing professionals (e.g. journalists). Their main goal is to support the writing process rather than to evaluate the final text.

Content wise, a writing assistant does not only rely on a huge amount of available lexicographic data. Rather, like a smartphone, it is an integrated tool also including corpus analysis data and learner corpora evidence, enhanced with the use of natural language processing techniques.

The design and use of our writing assistants will be illustrated.
References


Lexicography is the process or profession of writing or compiling dictionaries. Dictionary.com (2016) defines lexicography as ‘the principles and procedures involved in writing, editing, or compiling dictionaries’. In order to write, edit, or to compile dictionaries, there must be a language or languages and words to be involved in the compilation of the dictionary. There must be rules or control mechanisms with regard to the writing of these words in order to have uniformity in the form of spelling, orthography, etc. which are not necessarily the work of a lexicographer but that of the standardizing bodies or the Language Boards. This means that a balance is needed between orthography development, language standardization and lexicography in order to be able to write, edit and compile dictionaries effectively, especially with reference to foreign acquisition and the newly acquired lexical items.

Wikipedia (2016) defines language standardization as the process by which conventional forms of a language are established and maintained. Standardization may occur as a natural development of a language in a speech community or as an effort by members of a community to impose one dialect or variety as a standard. With reference to the South African indigenous languages, standardization can be regarded as the reshaping, the purifying and the accreditation of languages, especially the terminology of the language. This is usually done by choosing one variety, or combining the varieties of the language to form one acceptable language which is henceforth regarded as a standard variety. In most cases these standard varieties are used as official languages for official communications. The system of standardising languages is usually done by language bodies which are regarded as having authority over the language and its development. Previously, in South Africa, the language standardizing bodies were known as Language Boards. After 1995 these Language Boards were phased out and most of their functions were transferred to the National Language Bodies (NLB’s) which were formed as one of the structures of the Pan South African Language Board (PanSALB), in accordance with the Pan South African Language Board Act, number 59 of 1995, as Amended.

It is the constitutional mandate of the National Language bodies to insure that all new words which originate, inter alia, from developments of new concepts, dialectal varieties and foreign acquisitions, are standardized to form part of the languages. Recently, in South Africa the standardization of terminology is very slow. This is due to several reasons varying from lack of enough fund allocations for the standardising bodies, mal-administrations and lack of political wills to support language developments and even purism. As a result, we have many words which are used by the various communities differently or written differently by the communities because the spelling and orthography rules are still outdated to accommodate newly formed words. This is the situation in which the lexicographers found themselves, especially the indigenous languages lexicographers, because there are mostly confusions regarding the choice of spelling or word structures to be followed due to lack of sufficient directives in the form of spelling and orthography rules. The responsibility of a lexicographer is to compile dictionaries, and not to prescribe on issues like spellings and orthographies. These are all mandatory functions of the National Language Bodies and their sub committees. The malfunctioning of the systems of standardization is a huge stumbling block (is bad news) to lexicography because, for the sake of progress, the lexicographers will continue to
lemmatize words which the Language Bodies did not have chance to give accreditation or will never get that chance due to various administrative shortfalls. The ultimate result is that the same words which are defined by different lexicographers will always have different spellings, which is a serious irregularity when considering the fact that dictionaries are supposed to be authorities and accurate reservoirs of accredited language. All confusions, arguments and differences pertaining to spellings and orthographies are solved by referring to the dictionary. The following are examples of words which need the attention of the standardizing bodies to give accreditation so that the lexicographer can write these in the dictionary:

- Stokisi and Setokisi ‘prison’
- Jele and Tšele ‘gaol’, ‘prison’
- Mmankhoço and Mmankhotlo ‘owl’
- Tjheke and Tšheke ‘cheque’
- Bara and ‘Para’ ‘bar’, ‘ladies bar’

Even though the compilations of comprehensive dictionaries do not necessarily require the use of the standard languages and the standard terminologies, lack of proper control and proper balance between standardization and dictionary compilations cause serious confusions and setbacks in the development of lexicography in the South African indigenous languages. These confusions lead to, inter alia, the following consequences:

- There will be no uniformity with regard to the spellings of the newly acquired words
- There will be no guidance with regard to orthography, e.g. tjheke or tšheke
- The abovementioned irregularities will lead to the publications of dictionaries which will not command authority in the language.

References


PanSALB Act Number 59 of 1995, as Amended.
The lexicographic treatment of isiZulu negation
Sonja BOSCH (boschse@unisa.ac.za)
Department of African Languages, University of South Africa
Gertrud FAASS (faassg@uni-hildesheim.de)
Department of Information Science and Natural Language Processing, University of Hildesheim

Negation is described by Crystal (1994:231) as “A process or construction in GRAMMATICAL and SEMANTIC analysis which typically expresses the contradiction of some or all of a sentence’s meaning.” As an important instrument of language use, one would therefore expect negation to be dealt with in dictionaries. However, as Dahl (1979) states, negation phenomena appear to be at the border of lexicon and grammar, thus, one could argue that grammatical issues are not a matter for lexicography. Electronic dictionaries, on the other hand, are nowadays seen rather as language information tools, that is, they are to contain and to present extra-lexicographic data about a language as well (cf. Prinsloo et al. 2012), so negation again comes into play. However, to the knowledge of the authors, only very few works are available on the lexicographic treatment of negation of which none concerns isiZulu.

There are a number of isiZulu negation strategies of which the most common is verbal negation which appears in inflected verbs, see (1) showing diadic negation and (2) displaying polyadic negation.

(1) a) ngi- -hamb- -a
   SC1p VRoot VEnd
   1st person Sg. go
   ‘I go’

   b) a- ngi- -hamb- -i
   neg SC1p VRoot VEnd (neg)
   not 1st person Sg. go
   ‘I do not go’

(2) a) u- -ya- -hamb- -a
   SC01 long present tense VRoot VEnd
   3rd person sg go
   he/she goes
   ‘he/she goes/is going’

   b) a- ka- -hamb- -i
   neg SC01_neg VRoot VEnd (neg)
   not 3rd person sg. go
   ‘he/she does not go / is not going’

In isiZulu we also find several lexicalized negation word forms used in the imperative as shown in (3), but not all are described by existing dictionaries.

(3) a) mus- a- uku- -hamb- -a
   VRoot VEnd SC15 VRoot VEnd
   do not (imperative) class 15 = infinitive go
   ‘Do not go!’ (semantically stronger than simple negation)
b) yek- -a uku- -hamb- -a  
VRoot VEnd SC15 VRoot VEnd  
stop (imperative) class 15 = infinitive go  
‘Do not go!!’ (semantically stronger than 3 a)

With the regular negated constructions as shown in (1) to (3), we also find a number of additional rules for specific verbs or verb forms. Passive verbs, for example, may not use the negative verbal ending -i when being negated, but retain the positive -a. There are also so-called defective verb forms such as –sho ‘say’ which take irregular negative suffixes, for example –ongo instead of the regular –anga.

(4) a) ba- -sh- -o  
SC02 VRoot VEnd  
3rd person pl. go  
‘They said’

b) a- ba- -sh- -ongo  
neg SC02 VRoot VEnd (neg)  
not 3rd person pl. go  
‘They did not say’

We will first describe the negation phenomena of isiZulu that a dictionary is expected to present; and then we will show how the bilingual dictionary of Doke and Vilakazi (1964), the bilingual learners’ dictionaries of Dent & Nyembezi (1969) and of De Schryver (2010), and the monolingual general dictionaries of Nyembezi (1992) and Mbatha (2006) deal with these, providing suggestions on enhancing their presentations, for example by making more extensive use of textboxes illustrating the linguistic phenomena in question (cf. Gouws and Prinsloo, 2014). Lastly, we will discuss ways and methods (and also problems) to present these issues in two kinds of electronic dictionaries: isiZulu.net and a learners’ dictionary which also has to take didactic issues into account. As Prinsloo et al. (2012) rightly state: “there are numerous complex situations where users need more detailed support than currently available in e-dictionaries, to make valid and correct choices”. The proposal of Kovarikova et al. (2012) to interconnect affirmative and negative forms individually via referencing tools in e-dictionaries will also be considered.

References
eZulu for learners: http://10.1.7.145/eZulu-mobile/index.php (Unisa internal only).
isiZulu.net https://isizulu.net/ (accessed on 02/03/2016).


*****

Demonstration of the Pilot Project for a new online Woordeboek van die Afrikaanse Taal

Willem F. BOTHA (wfb@sun.ac.za)
Bureau of the WAT, Stellenbosch University

Hanelle FOURIE BLAIR (hfourie@sun.ac.za)
Bureau of the WAT, Stellenbosch University

Work on the WAT was started in the pre-theoretic era of lexicography. In 1946 Dr PC Schoonees, Editor-in-Chief of the WAT desperately searched for a lexicographic manual of some kind in Europe, but without success. Not even the South African Academy for Science and Arts could be persuaded to assist in the compilation of the dictionary.

Consequently, what had been started as a Standard Afrikaans dictionary modelled on the Dutch Van Dale Groot Woordenboek der Nederlandse Taal gradually developed into a comprehensive Afrikaans dictionary. By the end of the second millennium around 2001 the WAT realised that electronic dictionaries were the future of lexicography. The advantages of electronic dictionaries over printed dictionaries had become evident.

By 2003 the WAT was available as a CD and as an online dictionary. This was no mean feat as like most other comprehensive dictionaries, the WAT was not compiled electronically. In a very expensive and labour intensive project, in which the contractor lost a considerable amount of money, the printed dictionary was digitalised. A digitalised dictionary can, however, not offer all the progressive state of the art functions of a modern online dictionary.

In 2015 it was therefore decided to redesign the online WAT platform after a thorough study of the latest lexicographic theories on online dictionaries and the state of the art Dutch, English and German online dictionaries.

Because of the fact that earlier parts of the online dictionary consist of digitalised texts, created before the era of electronic compilation, it means that there are greater constraints on the amount of manipulation or customised presentation of content that can be done at this stage. The older texts are not properly tagged or marked and do not in fact appear in specific searchable fields, but only appear as if they do because all formatting was done by hand as surface mark-up only when the text was first prepared for a CD ROM version of the WAT. Therefore, the proposed changes will only be applied to a pilot project consisting of S – SKOOI (volume XIV of the WAT) for the moment, and will eventually include more text, the lemmas and their articles, as more volumes are completed.

The major redesign features of the proposed pilot project focus on both presentation and the overall look and feel of the online dictionary as well as certain elements of the content of the online dictionary.
In the interest of greater user friendliness and improved on-screen information retrieval, it was decided to streamline the default article content in certain ways, and also to allow greater user interaction in the sense that users will be allowed to determine which data types will appear on the screen. One example of this is the option to turn citations and other usage examples on and off. This could potentially save a lot of space on the screen and eliminate the need for excessive scrolling in the case of longer polysemous lemmas with lots of usage examples.

Other changes include, amongst others:

1. the use of different font sizes to clearly indicate the lemma's position at the head of the article, for example;
2. the decision to henceforth provide no phonetical transcription of pronunciation, but to rather indicate syllabification and stress;
3. writing out the full polysemy marker number for each separate subcomment on semantics in the case of polysemous and polyfunctional lemmas (as opposed to the current elliptical numbering);
4. the use of a few new non-typographical structural indicators for both synonym and variant cross references.

One of the great advantages of working with text or manuscript that has been entered and tagged in the correct digital data fields, is not only that it makes it possible for the user to ultimately search in specific fields only (e.g. to limit searches to look for words as they appear in definitions, or in example material, etc.), but it also makes it possible to label or give headings to certain data types so that they may appear as much more discrete units on the screen and become both easily identifiable and navigable by the user, which in turn leads to greater ease of navigation within any given article in general. The pilot project proposes the use of lightly coloured strips or bands as a background to certain headings, for idioms and undefined compounds, for example, to clearly demarcate these data types.

As a result of the extensive redesigned features of the pilot project, a thorough user guide or help section will be also be compiled and will contain screen shots of actual pages. Separate non-typographical structural indicators will make the relevant help sections easily accessible without requiring the user to navigate away from the page on which they are.

The presentation will consist of a series of page mock-ups of the proposed pilot project to illustrate the points being discussed.

*****

Are we there yet? An investigation of the VivA-app as Afrikaans’ answer to the mobile dictionary market
André H. DU PLESSIS (ahdp@sun.ac.za)
Department of Afrikaans and Dutch, Stellenbosch University

Botha (2014) and Heid (2014) mention that there is still limited research available on mobile dictionaries when compared to research on online or print dictionaries. In concordance Du Plessis (2015:16) also states that undervaluing mobile lexicographic tools across languages is detrimental to lexicographic progress and innovation. The same applies to the Afrikaans dictionary environment. In 2014 no other Afrikaans dictionary marketed as a mobile dictionary besides the mobile WAT was available (Du Plessis, 2015:14). This changed at in the middle of 2015 when a new mobile dictionary application appeared. The “Virtuele instituut vir Afrikaans” (VivA), a non-profit research organization and service provider aimed at developing digital language tools and portals for Afrikaans, released an Afrikaans mobile dictionary portal called the Woordeboekportaal-app van die Virtuele instituut vir Afrikaans or
simply the VivA-app (Coetzee, 2015). The VivA-app seems to incorporate all the needed modern functionality of current mobile dictionaries produced by the larger and better funded publishers such as Merriam-Webster or Oxford. At first glance the VivA-app makes use of standard web and mobile practices such as swipe functions, interactive layouts and advanced search options. These functional structural and technological elements speak to the modern users’ need for fast, easily accessible information as well as to these users’ current expectations of mobile applications (Du Plessis, 2015:38). However when this product is critically and lexicographically evaluated, it becomes clear that the VivA-app has some worrying deficiencies that will impede its target users.

Unfortunately, this dictionary application is hampered by a lack of lexicographic planning. This is something that, according Kwary (2013: 69), happens far too often with mobile dictionaries. Due to the fact that this dictionary is relatively new on the market, this paper will focus on an initial lexicographic analysis of the application, i.e. an overview of the possible lexicographic shortcomings will be presented with specific focus on user experience. This can however not be done without a theoretical framework. Du Plessis (2015) formulated an integrated theoretical framework for the analysis of the mobile WAT which can also be used for evaluating and analysing other mobile dictionaries from a lexicographic and usability perspective. This integrated theoretical framework encompasses the function theory (Bergenholtz and Tarp, 2003), the general theory of lexicography (Wiegand 1991; 1996 and 1998) and usability standards, such as ISO 9241, and theories such as interaction design (International Organization of Standardization, 2010; Sharp, Rogers and Preece, 2007). This framework allows for a holistic and critical analysis of almost any mobile dictionary as it pertains to the user(s), the dictionary structure and the human-computer interaction and technological standards. With the aid of this theoretical framework the abovementioned analysis of the VivA-app is thus possible.

From the preliminary investigation the following, amongst other, shortcomings can be highlighted. This dictionary portal application provides users with access to data from seven Afrikaans dictionaries or language resources such as the Afrikaanse Woordelys en Spelreëls, the Kuberwoordeboek/Cyber dictionary, the Entrepreneurskapwoordeboek, the Voorsetselwoordeboek and a range of resources from CTexT (Coetzee, 2015). Interestingly this dictionary does not include the WAT as a resource even though it is available for use on VivA’s online portal. One can speculate that the VivA-team does not regard the mobile WAT as a usable mobile dictionary and therefore did not include it. The relevancy or integration of the other dictionaries into an application are also questionable, for example the Kuberwoordeboek/Cyber dictionary is now 10 years old and aimed at semi-experts or experts and, as Swart and Du Plessis (2015:144) mention, this dictionary’s data could be seen as inconsistent and outdated. This is most certainly a problem for the target user group of this dictionary application – high school pupils and students – as this will lead to a situation where users’ needs cannot be satisfied, especially since this dictionary is primarily aimed at satisfying the communicative function.

The VivA-app has some structural issues too. For example, the users are left to their own devices when it comes to disseminating the provided data given when a search is done. Although it is clear in which dictionary or language resources the user can extract information, the way in which the relevant articles are presented needs to be re-evaluated with a lexicographic mind-set as the given article structure and microstructure can be confusing and difficult to use. Positives of this dictionary are the functional access structure it employs, as well as the use of standard web and mobile practices such as mentioned.
The VivA-app is a step in the right direction for Afrikaans mobile dictionaries, but as this preliminary investigation will show, it is not yet the answer to this language group’s need for a fully functional mobile lexicographic resource.

References
Sentrum vir Tekstegnologie. 2015. Woordeboekportaal-app van die Virtuele instituut vir Afrikaans (VivA). [Version 1.0.0]

Redesigning the DSAE’s presentation and navigation features: adapting a historical dictionary for the modern online user
André H. DU PLESSIS (ahdp@sun.ac.za)
Department of Afrikaans and Dutch, Stellenbosch University
Tim VAN NIEKERK (T.vanNiekerk@ru.ac.za)
Dictionary Unit for South African English, Rhodes University

During a special session at the 20th AFRILEX International Conference the pilot online version of A Dictionary of South African English on Historical Principles (DSAEHist) was presented and critically discussed (cf. Van Niekerk, 2015). During this session practical and theoretical obstacles and shortcomings of the online DSAEHist, especially with regard to
general usability, layout and functionality, were highlighted. After this conference, a collaboration between the Dictionary Unit for South African English (DSAE) and the University of Hildesheim (UH) was formalised and undertaken to “address the need for a thorough adaptation of the dictionary text to support publication on multiple electronic platforms” (Van Niekerk, 2015). Another collaboration has also since been established with Stellenbosch University (SU) to address the dictionary structure, layout and functionality. This paper will focus only on work undertaken during the initial stages of the DSAE-SU collaboration. Please note that while these are new enhancements, they in some cases represent the refinement and implementation of prior adaptations and prototypes developed during the DSAE-HU collaboration (for parallel enhancements see Van Niekerk et al., 2016, forthcoming).

As with other print dictionaries that enter the electronic medium, the pilot version of the online DSAEHist still felt and looked like a print dictionary albeit with a somewhat more user-friendly access structure. Another difficulty that the DSAE encounters is the fact that this dictionary is based on historical principles and needs to showcase large synchronic and diachronic datasets. This means that the article and microstructure, as well as the access structure, need to be adapted. Apart from this, the general layout of the dictionary also needs to be re-evaluated and redesigned to make it more usable on online platforms. For all of the aforementioned to be achieved, a theoretical framework with the target user and usability as focus areas needs to be adopted. Therefore, the theoretical and practical work done by Van Niekerk et al. (2016, forthcoming) as well as Du Plessis (2015) is used to generate a conceptual version of the new online DSAEHist where the layout, interactivity and presentation of the data are addressed.

The first step in modernising the DSAEHist was to rethink the general layout. Here the screenshot structure has been adapted so that search routes, user guidance, the access structure and user interactivity become visible and prominent features of the dictionary. Design changes have also been considered with new fonts, colours and possible images being incorporated. A key part of this step has been adapting the landing page into a more engaging and user-friendly layout that provides easily-accessible data and links. Some preliminary ideas involve shifting and relabelling links to different parts of the dictionary, moving the alphabet bar (as part of the macrostructure design), increasing the search bar size and functionality, and providing user-interactive features such as images of words found in the dictionary, a quotation capturing option, etc.

With the above-mentioned in mind, it is also necessary to adapt the article structure and microstructure. The current article structure is a legacy of the print edition and its space-saving editorial policies, augmented with basic hyperlinks for cross-references between articles and to the bibliography of quotation sources. In order to make the dictionary articles more user-friendly and to showcase the historical data, a repositioning and labelling of data categories was considered as well as including a more multi-layered or dynamic approach to the presentation of content. Again styling changes have also been employed in the actual dictionary articles. Thus far the headword, part-of-speech and pronunciation have been moved and restyled, and data categories such as forms, plurals and origin have been changed from inline to expandable block-level elements. The paraphrases of meaning have been placed in a text box which is foregrounded while the quotations and other data types have been repositioned and remodelled to be expandable. This is not all that has been done and is also only the initial phase of restructuring the entire dictionary. There are however challenges in adapting some data types which require further exploration.

As shown, there are many areas in the DSAE that need attention before it can be regarded as a fully adapted and functional online dictionary, but due to lack of funding and
personnel, progress is slow. Nonetheless, the preliminary work done is a step in the right direction in bringing a valuable cultural language resource into the digital era. This paper hopes to reignite the discussion originally presented by Van Niekerk in 2015 and engage the lexicographical community on questions relating to the requirements and possibilities of online dictionary interfaces.

References

Integrating online reference sources into a comprehensive dictionary culture
Rufus H. GOUWS (rhg@sun.ac.za)
Department of Afrikaans and Dutch, Stellenbosch University

The formulation of current lexicographic theories has primarily occurred in this era and within the ambit of printed dictionaries. Topics like the different structures of dictionaries, access to data in dictionaries and the presentation and treatment to ensure optimal user-friendliness were directed at printed dictionaries. Albeit that some of these issues are not medium-specific and are also relevant to online dictionaries, the emergence of the digital era compels metalexicographers to re-appraise, adapt and expand their theories in order to make provision for new approaches in the planning and compilation of online dictionaries. The digital era has progressed much faster than digital lexicographic products and lexicography yet again has become a follower instead of a leader. A real change in lexicographic theory and its applications should see that theoretical lexicography not only has followed a contemplative approach of merely analysing and discussing existing dictionaries, but also activates a transformative approach that gives guidance with regard to the planning and compilation of innovative lexicographic products. Lexicographic theory needs to be a leader in the development and production of both printed and online dictionaries and not only a follower.

Practical lexicographers should take cognizance of new work in the field of theoretical lexicography to enhance the quality of their printed and online dictionaries, but they also need to be aware of the changing needs and reference skills of the envisaged users of new lexicographic and other reference works. In this regard it is important not only to identify the reference needs but also the reference culture of those members of society that represent the target audience. Lexicography should look ahead and plan for the future and this implies that the default target user should no longer be regarded as older people, many of whom belong to lexicographically lost generations, but rather those people who constitute the digital community, the netizens, i.e. net citizens, that belong to Generation Z. These people and their reference needs and skills differ completely from the needs and skills of preceding
generations. A Baby Boomer or even a Generation X or Millennial approach to dictionaries and other reference works will not suffice the new generation of potential dictionary users.

The transfer from a paper to an online reference world demands a change in the prevailing dictionary culture. Users need to be made aware of new ways to access data, with traditional dictionaries only one of the instruments by means of which this access is possible. The advantages, restrictions and possibilities of the digital era need to be integrated into a comprehensive dictionary culture that changes the approach of theoretical lexicographers and their works as well as the attitude of dictionary users.

This paper gives a brief account of some aspects of the development from paper to online lexicography and the implications for the formulation of lexicographic theories. The focus then switches to Generation Z as a preferred target group of future dictionaries. Attention is given to aspects like the relation between Generation Z and digital devices, their overall reliance on the internet, the need to be connected and their FOMO, i.e. their fear of missing out. It is shown that lexicography should take cognizance of this potential target user group and dictionaries should be planned accordingly. In order to establish a significant and interactive relation between the netizens and lexicography it is important that children should be exposed to online reference works, including dictionaries, from an early age. This could initiate a life-long lexicographic relationship.

The paper suggests ways in which lexicographers can introduce new types of dictionaries and new approaches to the transfer of data. Dictionaries should not be planned and used as isolated products but need to be regarded as part of a more comprehensive reference family. Links between dictionaries and other online sources will increase the interactive relation that strengthens the need for connectivity that characterises members of Generation Z.

The paper argues in favour of a completely expanded interpretation and application of the concept of a dictionary culture.

*****

The lemmatisation of Ideophones and Interjections in Cuenod’s Tsonga–English Dictionary
Madala C. HLUNGWANI (Crous.hlungwane@univen.ac.za)
MER Mathivha Centre for African Languages, University of Venda
Memory MPAPA (Memory.mpapa@yahoo.com)
Department of African Languages and Literature, Great Zimbabwe University

The paper analyses Cuenod’s Tsonga-English Dictionary with a purpose to determine the extent to which linguistic information has been considered in the lemmatisation of ideophones and interjections. Linguistic information involves aspects of linguistics such as phonology, phonetics, morphology, syntax and semantics. Four types of ideophones will be examined in this study, namely, manner ideophones such as gaa! (lie on one’s back), Sound emitting ideophones, for example dloomu! (sound of something like a stone falling into the water), colour ideophones such as juu! (be red), and smell ideophones such as mbvhee! (sink). Whereas the interjections that are investigated in this study are classified into three types, namely, interjections associated with surprise, for example, e! (interjection of fear, e.g. of something fearful that occurs unexpectedly), interjections associated with fear, such as manano! (mother! Interjection of fear), interjections of happiness, for example, ahee! (interjection of approval), and interjections associated with anger, for example, ase!
With respect to phonology and phonetics which deals with the speech sounds, the study examines whether the sound of the lemmas entered under the two mentioned parts of speech above have been marked or not, and to determine whether consistencies or inconsistencies do exist in the marking of such sounds, if any. Since morphology deals with the structure of words, including specifying the parts of speech, it is the aim of the study to establish whether Cuenod did specify the parts of speech of the lemmas entered under the ideophones and interjections. On syntax, a subfield that is concerned with the structure of sentences, the study examines whether the lemmas that have been entered under the two parts of speech under consideration have been used in example sentences in order to demonstrate how such lemmas may be used in contexts. When it comes to semantics, the aspect that deals with meanings, the aim is to determine whether the meaning that has been provided of the lemma under consideration is adequate or not, that is to say whether such a meaning may be viewed as a description of the lemma under consideration or the equivalent translation of the English lemma thereof, taking into account the fact that the dictionary under investigation is a bilingual text from English to Xitsonga. The study highlights those aspects of linguistic information that have been considered and those that have not been considered. On aspects that have been considered, it is the aim of the study to determine those aspects that have been lemmatised correctly and those that have not been lemmatised correctly, if any. The study also establishes the consistencies and the inconsistencies that exist, if any, in the lemmatisation of the two parts of speech under investigation when it comes to the linguistic information. The study employs the sampling method, that is, a reasonable number of lemmas representing the two parts of speech is extracted from the dictionary. The theory that is adopted in this study is the modular theory, the theory that takes into account all the aspects of language in the lemmatisation of dictionaries. It is the aim of the study to point out the shortfalls that exist in the lemmatisation of Cuenod’s dictionary, and to also come up with ways in which such shortfalls can be overcome in the compilation of future dictionaries of any other language in generally and of Xitsonga specifically.

*****

Intellectualization through terminology development
Langa KHUMALO (khumalol@ukzn.ac.za)
Linguistics Program, School of Arts, University of KwaZulu-Natal

Intellectualization is a term originally used by Havranek (1932), a linguist from the Prague School, to characterize a process that a language undergoes in its advancement. It has been famously associated with the development of Tagalog in the Philippines, which involved its lexical enrichment through terminology to enable its use in academia. This paper critiques the terminology development processes designed and implemented as a strategy for the intellectualization of isiZulu (cf. Khumalo, 2016) at the University of KwaZulu-Natal (henceforth UKZN).

South African Higher Education is increasingly waking up to the imperative of transmuting African languages to be the kernel of the academy. It is in this vein that the
UKZN language policy and plan (2006 revised 2014) recognizes the prominent role of language in teaching and learning. Through its language policy and plan, UKZN has taken an initiative to cultivate isiZulu so that it becomes a vehicle in knowledge production and knowledge dissemination. It has been persuasively argued that in order for African languages to be used in education as languages of instruction, innovation, science, mathematics and logic, there has to be a clear, conscious and careful process of the intellectualization of these languages. According to Finlayson & Madiba (2002), in the South African context intellectualization is a meticulous procedure aimed at accelerating the growth and development of hitherto underdeveloped African languages to augment their capacity to effectively interface with modern developments, theories and concepts.

It is our contention that germane to this process is the development of discipline specific terminology. The paucity of such specialized terminology is often cited as the reason why African languages cannot be used as languages of teaching and learning, hence their discernment as shallow and inadequate (cf. Shizha, 2012). The paper critically evaluates the UKZN terminology development model (see Fig 1.) used to harvest, consult and authenticate isiZulu terminology for Administration, Architecture, Anatomy, Computer Science, Environmental Science, Law, Physics, Psychology, and Nursing. Outflow platforms for the terminology are loosely listed as the ‘database’ and ‘development platform’ but there is no clear end user platform for students, who seem to be the main end-user-targets of the whole terminology development initiative. The paper will propose an improved model to cater for AnyTime Access, which is convenient for student needs between lectures, and improve the harvesting mechanism in the existing model.

Figure 1. UKZN Terminology Development Model (cf. Khumalo 2016)
The paper will thus argue that terminology development is typically a top-down and selective process involving workshops often with a few discipline experts, terminologists, lexicographers and linguists, which is very resource intensive. After explicating stage 1 and 2 in the model in Fig. 1, an argument will be advanced for crowdsourcing (cf. Keet & Barbour, 2014) as bottom-up alternative to terminology harvesting including usurping certain functions of the consultative workshops in order to improve efficiency and reach while cutting back on expenditure. It will also be emphasized that intellectualization is a gradual process, which culminates in a scientific language capable of reflecting the rigor of objective thinking “[…] in which the terms approximate concepts and the sentences approximate logical judgements.” (Nekvapil 2006:2228). It is in this light that we argue that terminology development in itself and by itself cannot result in intellectualization. Natural Language Generation (NLG) in African languages will be proposed as a complimentary computational process that can enhance the verbalization of logical theories in isiZulu.

References

*****
Term Creation through Translation: An Evaluation of Shona Phonetics and Phonology Terminology at GZU
Penniah MABASO-SHAMANO (mabasopeniah@gmail.com)
Department of African Languages and Literatures, Great Zimbabwe University

The paper evaluates Shona terms that were created to teach phonetics and phonology courses at Great Zimbabwe University (GZU). It analyses the techniques that were used to create the terminology with the aim of proposing alternative translation techniques that can be used in creating terms in African languages. Most phonetics and phonology terms are alien to Shona and as a result, there are no suitable Shona equivalents. In practice, however, the translator of phonetics and phonology terms soon learns that very often he has to create equivalence rather than merely look for it (Chimhundu 1996). The phonetics and phonology terms to be discussed in this paper were created using different translation processes and strategies such as borrowing, neologising, compounding, transliteration, circumlocution among many other processes. Chimhundu’s (1996) Scan and Balance theory will be used to evaluate the created terms. Among all universities in Zimbabwe that offer African languages, GZU is the only university that uses indigenous languages as medium of instruction when teaching African languages courses. The main reason the other universities prefer English to indigenous languages as medium of instruction in teaching African languages is due to the lack of adequate terminology in the latter for specialised fields such as linguistics. This lack of adequate terminology can be partly attributed to the language use policy in Zimbabwe whereby African languages were not used in specialised fields such as law, education, science and technology. Shona and Ndebele, the two indigenous languages taught in schools, teacher training colleges and universities were taught in the medium of English from Advanced level up to tertiary level. GZU offers Shona, Ndebele, Shangani and Venda at bachelor’s and master’s levels. However, as aforementioned, all these languages lack adequate terminology to teach especially linguistic courses. The lecturers and students for these courses had a mammoth task of creating terminology to teach the different modules offered in the above mentioned indigenous languages. Most linguistic reference books are written in English, so lecturers and students have to translate information from English to Shona, which has proved to be too difficult for them. Term creation workshops were held at GZU to try to deal with the problem of lack of terminology in indigenous languages. Different indigenous languages practitioners from different tertiary institutions convened for a two-day workshop at GZU. Due to the ‘specialised’ nature of phonetics and phonology, it was too difficult to come up with the indigenous terms. Chimhundu and Chabata’s (2007) Duramazwi reDudziramutauro neUvaranomwe (A Dictionary of Linguistics and Literary Terms) was also used as a source for terms. In their dictionary, Chimhundu and Chabata used different translation strategies to come up with the terms. Most of the Shona phonetics and phonology terms in Chimhundu and Chabata’s dictionary and those created at the workshop will be evaluated. The strategies used in developing the new terms and the terms themselves have shortcomings. Most Shona terms used in phonetics and phonology are neologisms, transliterated terms, circumlocuted terms which are usually longer than the English terms, compound neologisms and blended terms and in some instances English words were translated to Shona through a whole statement. These shortcomings hinder the effective use of Shona in the teaching of phonetics and phonology courses. It will show that there is overuse of transliterated terms due to the lack of Shona equivalents for English terms. Most single English words were translated into compound neologisms or phrases after attempts to reduce them to one word terms failed. In other instances, circumlocution led to the problem of creating longer terms than the original which are not user-friendly. This paper discusses and evaluates the different phonetics and phonology terms created and the different strategies
and processes used in creating them. The paper will also suggest possible ways to come up with user friendly terms and ways to standardize them.

References

*****

Reflections towards the revision of Dikixinari ya Xitsonga ya ka Lingua Franca
Ximbani E. MABASO (mabasxe@unisa.ac.za)
University of South Africa
Mangalani J. HLONGWANE (hlongwane@limpopoleg.gov.za)
University of South Africa

Dikixinari ya Xitsonga ya ka Lingua Franca is the first monolingual dictionary to be published in Xitsonga since this language was reduced to writing. While the publication has been received with great enthusiasm by end-users, a number of shortcomings have been observed in it. The aim of this paper is to highlight issues of concern in the dictionary with the purpose of alerting the author (The Xitsonga National Lexicography Unit), the publisher (Lingua Franca Publishers) and the end-user about some of the problematic issues that need attention. Questions pertaining to lexicographic approaches adopted in the compilation of the dictionary and lexicographic treatment of some dictionary texts and entries are raised. For instance, do front-, middle- and back-matter texts serve any functional purposes? Do lemmatisation strategies and cross-referencing employed in this dictionary facilitate the access of data and information retrieval in a user-friendly manner? Language matters including those dealing with grammar are also considered: Is word-formation and syntactic structure presented in the dictionary compliant with current and standard Xitsonga rules? For example, are words such as ‘mukhanselara’ < ‘councillor’ acceptable in Xitsonga? Furthermore, how is typography (the syllabification of words such as ‘nhlamu-selo’ and ‘ngo-pfu’); punctuation (are punctuation marks employed in the dictionary? For example the abbreviation of names of languages such as ‘xizulu’ and ‘xinghezi’ with an initial lower case
letter); spelling and orthography (for instance the difference between ‘nwa’/‘n’wa’ and ‘nomo’/‘nomu’) and addressing inconsistencies such as ‘nhlapfi’ / ‘hlampfi’ / ‘nhlampfi’; tone marking (is it necessary or not and why?) and the use of international phonetic script (for example the incorrect rendering of the nasal sound ‘n’ as [?] instead of [ŋ] in the word ‘nkolo’; the use of shortenings (such as inconsistency between ‘rienc.’ and ‘rence.’ for ‘riencisi’) and other related lexicographic issues. All these aspects will be assessed in order to determine whether they are factually and grammatically correct and whether they make the dictionary user-friendly. Since the dictionary is in circulation, the presenters of this paper hope that the author and publisher will note the concerns and find a way to intervene by any acceptable corrective measures (for example by way of including errata in the copies that are still in the shelves in bookstores, and the media to reach out to the users who have already bought the copies with wrong information) for the benefit of the end-user while author and the publisher prepare for revision and publication of a second edition. The end-users will also not adopt the incorrect forms once these will have been marked as such for them.

The paper will be grounded on theories of lexicography and terminography within the language standardisation paradigm and the development of monolingual dictionaries for general purposes in developing languages. These theories will provide internationally recognised benchmarks against which Xitsonga monolingual dictionaries can be compared and assessed. The publication will also be measured against the current acceptable Xitsonga standardised language rules on spelling and orthography to determine whether it complies.

Some of the presenters’ preliminary conclusions will include the finding that (1) the work was not properly edited because it is teeming with typographical errors; (2) the work has a lot of grammatical errors such as the wrong formation of the past tense and incorrect concordial agreement; (3) there are many inconsistencies such as the pronoun ‘yeyeleyo’ vs ‘yoyoleyo’; (4) inconsistent and wrong punctuation as in the abbreviation ‘n.sw.’ vs ‘n. sw.’ and the use of lower case instead of upper case and vice versa (e.g. ‘xitsonga’); (5) no cross-reference between related lexical entries such as ‘bakatswa’, ‘mbakatsi’ and ‘mbhakatsi’; in other instances the user is referred to an entry that does not exist in the dictionary (e.g. ‘riqingho’ cross-referenced from ‘foni’); (6) incorrect labelling of lemmata such as ‘ndza’ and ‘ndzi’ which are labelled as ‘ris’ < risivi (pronoun) instead of ‘xitw.’ < xitwananisi (concord/agreement morpheme); (7) there is no table of content in the dictionary; (8) there is no page numbering in the front matter section); (9) there are incorrect definitions of some lexical entries such as ‘akhironimi’ (acronym).

In view of all the above mentioned factors, it is recommended that the dictionary be reviewed and thoroughly edited and proofread urgently before it is revised and published as a second edition. If publications that are marred with errors/mistakes and wrong information are allowed to be used by the public, there is a danger that Xitsonga dictionaries in general will lose credibility among the end-users and will cause more confusion for standardisation efforts.

References
Xitsonga Dictionaries a boom or doom for the standardisation of Xitsonga: a linguistic introspection
Ximbani E. MABASO (mabasxe@unisa.ac.za)
University of South Africa

One of the main functions of a dictionary is to serve as a standard reference when there is a uncertainty among the users of a language relating to the correct spelling and orthography of a word. In this respect, the dictionary is always consulted as a source and absolute authority to verify the correct form, including tone notation and international phonetic script. With regards to this honourable status of the dictionary, this paper intends to probe to what extent Xitsonga dictionaries uphold their authority in this respect and determine whether a dictionary should be allowed to have incorrect spelling and orthography at all, and to what extent such incorrectness should be tolerated if it does exist. If such errors/mistakes cannot be tolerated in a language, the study will seek answers as to what corrective measures are taken by relevant authorities. The study will use the standardisation theoretical framework and employ content analysis as its methodology which involves literature review.

The discussion will take the format of a review of two existing Xitsonga monolingual dictionaries, namely Dikixinari ya Xitsonga ya ka Lingua Franca (2014) and Tihlungu ta Rixaka (Dikixinari ya Ririmi ra Xitsonga (2016); four bilingual dictionaries, namely Tsonga-English Dictionary (1967), Dictionary of Basic English-Tsonga Across the Curriculum (1983), English-Tsonga: Tsonga-English Pocket Dictionary (1974), Dikixinari ya Xitsonga/English (2005); three publications on Xitsonga standard orthography and spelling rules, namely Terminology and Orthography No.3 (1976), A unified standard orthography for Xitsonga / Xichangana (South Africa and Mozambique) – (2003), Milawu ya Mapeletelo na Matsalelo ya Xitsonga (2008); the South African Multilingual Dictionary (2014). Each dictionary will be reviewed alone to check for internal consistency, authenticity and reliability and compliance with prescribed standard orthography at the time of its publication before a cross-comparison is drawn against the rest of them. Since language is dynamic, the study will track developmental trends of Xitsonga spelling and orthography rules and assess their implementation in the dictionaries because the two resource types should be complementary. For purposes of this presentation I have decided to focus on dictionaries for general purposes. I am aware of the existence of dictionaries for specific purposes like the dictionary of Xitsonga/Changana surnames, namely Swivongo swa Machangana (1995).

Discrepancies will be highlighted and relevant conclusions will be drawn from them. Some of the preliminary conclusions are the following: (a) a dictionary should strive to be faultless from the word-go, it should not be assumed that this is impossible to do; (b) the first edition of a dictionary may be allowed to have very few (specified number) of spelling and orthography and other typographical errors/mistakes according to prevailing rules at the time of its publication. For instance there are times when the language authorities themselves are still under-developed and have no tangible solutions at the time and therefore need more time for further research and rule formulation (e.g. the case of mikova/minkova and xihari/xiharhi); (c) a dictionary should not be allowed to include variant spelling and orthography prior to approval by the relevant standardising authority; (d) where dialectal
variation is allowed, the dictionary should specify the name of the specific dialect involved instead of simply assuming that a variant form belongs to some unspecified dialect; (e) there should be a prescription for the number of years that a faulty publication should be allowed in the public domain (for example, a maximum of five years). The Tsonga/English Dictionary was never revised since 1967 and it is 49 years old today without any prospect for revision; (f) newly formulated rules should be incorporated in all existing dictionaries as a matter of urgency and such dictionaries should not be prescribed in schools before they have been revised; (g) a solution should be sought to resolve problems where the copyright owner of a dictionary no longer exists; (h) National Language Bodies should be empowered to recall and remove from bookshelves those dictionaries that do not comply with standardised rules and regulations for dictionary compilation.

Finally, the paper intends to present these recommendations to the National Language Bodies and National Lexicography Units under the auspices of the Pan South African Language Board for consideration.

References
Xitsonga National Lexicography Unit. 2014. Dikixinari ya Xitsonga ya ka Lingua Franca. Giyani: Lingua Franca.

*****

A discussion of isiNdebele Terminology
Katjie S. MAHLANGU (sponono.mahlangu@up.ac.za)
IZiko lesiHlathululi-mezwi sesiNdebele, University of Pretoria

The paper explores the extent to which terminologies created by terminologists are understood and used by speech communities with a focus on isiNdebele. The isiNdebele language is one of the Nguni languages that is more under-resourced than the other official languages in the Republic of South Africa. The article will mainly examine isiNdebele terminology and specifically the terms coined by the Language Board in its publication entitled The isiNdebele Terminology and Orthography No.1 of 2001. It will also examine some of the terms found in Information and Communication Terms (2003), the Multilingual Mathematics Dictionary (2005) and the Multilingual Soccer Terminology (2009). In coining these terms the Board seems to have had a problem in its application of the international guidelines for term formulation as well as the creation strategies for coining terms. The
argument is that terminology must be taken seriously because it plays an important role in any language service. The Constitution of South Africa (1996) recognises isiNdebele as one of the eleven official languages and the Department of Arts, Culture, Science and Technology (DACST), as a public service provider, has the duty of realising fully Section 6 of the Constitution, that provides the legal requirements for language usage in South Africa. However, the ISO Technical Committee 37 Terminology (ISO/TC 37) (Principles and coordination) is tasked with establishing the principles and methods for terminology, and its specific aim is to standardise the methods for creating, compiling and coordinating terminologies. For developing and protecting language rights, PanSALB has created thirteen National Language Bodies (NLB’s) as advisory structures to take care of standardisation (e.g., spelling or orthography rules), terminology development, verification and authentication of terminologies and also to assist with the standardisation of terms among other things. In the past, the National Terminology Services (NTS) used to work in collaboration with the old Language Boards and currently the terminologists of the National Language Service (NLS) work in consultation with the National Language Bodies. After a specific terminology list has been finalised, it is then taken to the language bodies for verification and authentication of terminology, so that they can also assist with the standardisation and stabilisation of terms as well as with popularising terms. Since 1994, the Department of Arts, Culture, Science and Technology (DACST) has employed African language terminologists to develop and document African language terminology in a variety of subject fields (isiNdebele included). Because of the manner in which technology is growing globally, neologisms i.e., new terms, have to be coined for new inventions, situations and environments. Some of the multilingual terminology lists that were compiled are: weather terms, basic health terms, HIV/AIDS terms, building terms, election terms, banking terms, commercial and financial terms, computer terms, mathematical terms, natural science terms, soccer terms, water and sewerage terms and the like. However, some of the challenges experienced during the terminology development in isiNdebele give rise to the inconsistencies in terminology itself. The paper further highlights the problems faced in terminology development in isiNdebele. The concept and practice of terminology are poorly understood by the speakers and in some cases the new terms are not even known.

The purpose of this paper is, therefore, to provide some input on amendments regarding aspects of these terms that were coined in isiNdebele terminologies. Furthermore, these terminologies normally cause controversy in relation to the orthography or spelling of language. The input is made in the hope that it will make a contribution toward the development of isiNdebele terminology as one of the official languages.

*****

The use of a corpus in the development of an electronic monolingual Sesotho school learners’ dictionary
Malefu J. MAHLOANE (475894@students.wits.ac.za / mahloane.malefuro@gmail.com)
University of Witwatersrand

The main output of the project reported here is a monolingual electronic learners’ dictionary for South Sotho. Its specific objectives included the expansion of Sesotho practical and
theoretical lexicography, the development of a corpus and its use in developing the monolingual dictionary, and also, the annotation of the corpus for further lexicographic use. This project has used frameworks from the disciplines of general lexicography (Wiegand, 1983; Neubauer, 1983; Gouws & Bergenholtz, 2010), Bantu languages lexicography (Kiang, 1992; de Schryver, 2009; Prinsloo & de Schryver, 2001; Nkomo & Wababa, 2013; Otlogetswe; 2013), corpus linguistics (Graeme, 1998; Biber et al., 1998; McEnery & Wilson, 2001) as well as Natural Language Processing models (Trausan-Matu, 2000). A number of questions were asked in the beginning of the project and throughout the project, but the primary question which was also identified as a problem that the project could attempt to solve was how a monolingual dictionary can serve as a language learning resources to the L1 speakers of that very language, as well as being an aide in second language acquisition. This question arose following the hypothesis that L1 speakers of Sesotho from rural and peri-urban areas present poor academic results in high school due to a lack of language learning resources that can help them regarding the Medium of Instruction at school, which is English in South Africa. It could be argued that there are English dictionaries and other language resources for English however, because English is not the language through which the learners conceptualise the world in and is merely a language they encounter in their school books, it is not sufficiently rooted in their cognition, as a result, school learners may learn English concepts but remain without an understanding of the concepts and thus, the failure to apply themselves in English. In addition to this, it was also found that the learners have an additional struggle concerning the expected knowledge of their own language. According to Bloom’s Taxonomy of learning, there are certain expected output skills for learners at different stages, and from a small observatory study and focus group discussion with groups of Grade 10 high school learners, the research discovered that the learners did not have the expected comprehension skills of Sesotho, and they also did not know the connotations of lots of words presented in their Sesotho activity books and literatures. For such a reason, a monolingual learners’ dictionary is seen as a necessary language resource, since there is not one available at present. The research project has aimed to produce a monolingual electronic dictionary alongside the final report. Because this is an electronic dictionary, a number of dictionary-making software were experimented with until a decision was made to employ TshwaneLex dictionary making software. The lexicon was harvested from the corpus that was also compiled for this project and words, their connotations, and usage examples were observed from the corpus. Due to the conventions that guide corpus compilation, the corpus was further manually annotated with metaphor properties using Lakoff and Johnsen’s (2003) taxonomy of metaphor, which is also done as a step towards representing Sesotho in Computational Intelligence (Mahloane & Trausan-Matu, 2015) and to support the significance of an annotated corpus for lexicography. For the current presentation, the processes of using a corpus for a monolingual dictionary will be reported on. This will answer the research question on what are the low-cost approaches for developing Bantu languages lexicography further, as the snail-progression of Bantu languages in lexicography in the 21st century has been reported to be a result of financial factors. The advantages and disadvantages of a corpus in lexicography will also be discussed as observed from the project procedures. The project will also demonstrate how a monolingual dictionary can be useful in
creating more efficient bilingual Sesotho dictionaries in the long run, given the numerous inaccuracies found in the currently available Sesotho bi/multilingual dictionaries.

References

*****

Powers and problems of language modernisation in lexicography: A review of terminological approaches in Northern Sotho dictionaries
Kegaugetjwe M. MAKGOPA (makgopakm@tut.ac.za)
Tshwane University of Technology

Do dictionaries still retain their purpose of sustaining the legacy of native languages? According to Gouws and Bergenholtz, (2012:33) “lexicography is divided into two related components viz. practical lexicography which is the art or craft of compiling, writing and editing dictionaries and theoretical lexicography which is the scholarly discipline of analysing and describing the semantic, syntagmatic and paradigmatic relationship within the lexicon (vocabulary) of a language, developing theories of dictionary components and structures linking data in dictionaries, the needs for information by users in specific types of situation, and how users may best access the data incorporated in printed and electronic dictionaries.” Therefore, Northern Sotho dictionary should preserve the original language of Northern Sotho as it calls itself “Seshego sa Polelo.” which translates to “a storage of language for future i.e. generation to generation.” Lexicographers should endeavour to preserve the language as dictionaries are resources of language.
The purpose of the study is to explore the present newly developed words in dictionaries. Dictionaries are precisely instruments to accumulate, list and store vocabulary (words). The writer is not anti-modernisation, but rather concerned about the death of Northern Sotho language.

In the new dispensation technology takes a lead in communication. It comes with new invented concepts that enable effective communication to take place. It is inevitable to develop and align with the new terminology as technology emerges at a rapid rate. If attention is not paid, modernisation may lead to the death of Northern Sotho and this will be like harnessing a language towards its death.

Phillipson (1992:27) advocates that all African writers should be capable of writing in African languages instead of promoting English at the expense of their own languages. According to him, multilingualism is possible only if speakers of African languages are ready to write in their own languages. He also argues that the choice of a former colonial language rather than African languages should be seen in the light of the fact that 90% of the population in Africa today speak only African languages. It is a fact that language is dynamic and changes with time, but this does not mean that African languages should be compromised. Furthermore, Phillipson (1992:27-29) claims that Africans’ thought is imprisoned in foreign languages and that Africans ignore the fact that anything in English is reaching only one percent of African population. The implication is that an elderly person who has never been at school may not understand what the words means for an example: antropolotši, metholotši, sociolotši, bolumu and AIDS. Northern Sotho dictionaries will be studied in terms of their inclusion and treatment of vocabulary associated with modernity. Adopted words from the existing Northern Sotho dictionaries will be identified and juxtaposed with the existing words that are replaced or compromised to further support the argument. The principles of terminography and lexicography presented by Alberts (2001:71) will be applied in the analysis of the data that will be obtained from dictionaries.

The study will bring awareness to Northern Sotho lexicographers and terminographers to take into consideration that dictionaries are the storage of language. However elderly people are gurus of a language and need not be left behind when developing a language. Dictionaries should restore words that are user-friendly to the indigenous speakers of the language, elderly people inclusive. This approach will benefit both learners and elderly people as it will identify concepts by their use.

References

Efficacy of bilingual dictionaries as translator’s aid: A case of Sesotho sa Leboa and isiXhosa
Kwena MASHAMAITE (mashakj@unisa.ac.za)
Committee Services Section, University of South Africa
Feziwe SHOBA (shobafm@unisa.ac.za)
Department of Linguistics and Modern Languages, University of South Africa

Translation profession has grown tremendously in South Africa in the new dispensation. This is due to globalisation and the promotion of multilingualism as espoused in the South African
Use of Languages Act of 2012 (Act No.12 of 2012). One of the principles of this act, in particular principle 4.6, recognises the community-based approach where each governing structure is to identify and promote its working languages to promote multilingualism. It is in this context that translators need adequate and efficient bilingual dictionaries to translate texts in various African languages. The translators are required to translate a number of documents and texts in order to facilitate communication. Whenever translators face a terminological and translation problem, the first stop is a dictionary, particularly a bilingual dictionary. Newmark (1998: 29) posits that “the bilingual dictionary is the translator’s single, first and most important aid and a translator who does not consult one in doubt is arrogant or ignorant or both.”

One of the primary tasks in translation is choosing a translation equivalent from among a set of semantically related words. It is posited in this paper as a problem statement that the majority of bilingual dictionaries in African languages do not provide adequate translation equivalents in the target language from which translators need to choose during their translation activities. The situation is exacerbated in some cases by the lack of equivalents between the source and the target language. This viewpoint supports that of Gauton’s (2008: 106): “This non-equivalence between languages is also the root cause of the difficulties with which the translator or user of the bilingual dictionary has to contend.”

This paper assesses the efficacy of bilingual dictionaries of Sesotho sa Leboa and isiXhosa to check whether they really help or support translators of African languages during the process of translation. Gonzalez-Jover (2005) argues that a dictionary should serve as a problem-solving tool and any dictionary including the translation-oriented dictionaries, should take into consideration the user who consults it to solve a specific problem. A translator’s problem is mainly that of finding a suitable translation equivalent in the target language. Consequently, bilingual dictionaries of African languages should be designed to serve as tools to solve the specific problems of translators as well. Such problems include finding the appropriate equivalent in a specific context. In other words, there should be sufficient context-based equivalents to choose from. For example, the word ‘school’ should denote an institution for learning, pupils, building, a process of being educated, teaching activities, a college or group of related departments in a university or a group of people with similar views.

A purposive sampling technique will be employed to select some bilingual dictionaries of Sesotho sa Leboa and isiXhosa as sources of data for the investigation. Purposive sampling is a technique that is used in qualitative research to select specific characteristics of a population which will enable the researcher to best answer the research question. Data gleaned from the two sources will be analysed using the documentation analysis method. In the analysis of the content, the researchers will critically examine the data categories and determine whether or not they assist the translators with their translation tasks. The extracts or examples from the selected bilingual dictionaries will be used as the source text for analysis.

The significance of the study is to contribute into the body of knowledge in the field of bilingual lexicography with regard to the future design of bilingual dictionaries to include the user needs of translators. It is further recommended that in view of the shortcomings of the current bilingual dictionaries in African languages in serving the needs of translators, future bilingual dictionaries should be polyfunctional if their broad goal is to be problem-solving tools. Bilingual dictionaries should not only provide equivalents in the target language. They should also provide the meaning of words and example sentences indicating the various contexts in which the headword can be used. In this manner, the bilingual dictionary will
assist the translator to solve the problem of choosing the correct equivalent looking at the various contexts of use.

References

*****

The role of translation in lexicography with special reference to Tshivenda-English dictionaries
Mashudu MAŢHABI (Mashudu.nthambeleni@univen.ac.za)
University of Venda

According to Mongwe (2008), dictionaries are indicators or mirrors of a country’s social, cultural, scientific and technological development. Inadequate translations of lemmas in the dictionary lead to poor translation whilst appropriate translation brings about effective communication. It is important therefore to evaluate the role that translation plays in dictionary compilation. Languages such as Tshivenda have presently been given an official status in South Africa (Constitution of the Republic of South Africa, 1996). It is thus important that dictionaries of good standard should be compiled for the convenience of people in all aspects of life. Gouws and Prinsloo (1998:46) point out that “dictionaries are the instruments of linguistic and communicative empowerment, therefore, lexicographers have to make sure that their intended target users receive an optional linguistic presentation and aimed at the specific needs and reference skills of well-defined users”. The emphasis here is that the role of dictionaries is to present the contents to readers in such a way that these will gain understanding of the words that they are looking for. In addition, dictionaries help to promote language study. A dictionary is a book (or a website) which lists a large number of words of a language and gives their meaning and sometimes gives examples of their correct usage. Pei (1996:69) describes a dictionary as “a list of words of a language usually in alphabetical order with their meaning, often their derivations, and occasionally their histories”. A lack of properly translated Tshivenda-English dictionaries is a great challenge to dictionary users. In addition, it is without a doubt that the language communities of all the official languages in South Africa need dictionaries of good quality not only to bridge the communication gap, but also to document and preserve the rich variety of South African languages. In the light of this information, it is thus crucial to conduct a study on whether translation in the compilation of Tshivenda dictionaries has been performed properly or not. The aim of this paper is to examine the role of translation in lexicography with special reference to selected Tshivenda-English dictionaries (bilingual dictionaries). The selected dictionaries are: Tshikota (2006) Ṭhalusamaipfi Tshivenda/English Dictionary, and Van
Warmelo (1989) Venđa Dictionary. 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 Tshivenđa bilingual dictionaries?

This paper will utilise a qualitative method and will adopt a descriptive research design. The reason for using this method is that the qualitative research method is largely descriptive and it assists the researcher to answer why things are the way they are. The researcher will use primary and secondary research approaches to gather relevant information about translation in dictionaries. Interviews will be conducted with lexicographers, university lecturers, language practitioners, educators and court interpreters which will be randomly selected. Data will be analysed based on different answers from different questions posed to different respondents. Recommendations will be drawn based on the answers from the respondents. If this study conducted well, lexicographers will be able to compile user-friendly dictionaries. Recommendations will be drawn based on the answers given by different respondents.

References

Problems and perspectives in Civili lexicography
Paul A. MAVOUNGOU (moudika2@yahoo.fr)
Centre de Recherches en Etudes Germaniques et Interculturelles, Université Omar Bongo
Hugues S. NDINGA-KOU MBA-BINZA (hugues.ndingakoumbabinza@nwu.ac.za)
Centre for Text Technology, North-West University

The present paper comes within the framework of overhauling early reference works of Gabonese languages. It focuses on the lexicography of Civili. The latter is a developing Bantu language spoken in Gabon, in the Republic of Congo, in the Democratic Republic of the Congo, in the whole enclave of Cabinda as well as in the north of Angola. From linguistic point of view, Civili is part of the Kongo group (H10) of the Bantu linguistic branch. According to Guthrie’s classification, Civili is referred to as H12a (Guthrie 1948).

The literature review conducted by Mavoungou & Ndinga-Koumba-Binza (2010a) indicated that all the literature about Civili that is available deals with linguistics, anthropology, lexicography and religion. Concerning anthropology, studies have been devoted to the ethnic group of Bavili, which comes from the former Loango kingdom, a part of the Great Kongo kingdom, formed in the 14th century (cf. Mavoungou & Ndinga-Koumba-Binza 2010a & 2010b, Mavoungou Pambo 1999, Merlet 1991, Hagenbuecher-Sacripanti 1973 & Martin 1972). In relation to lexicography, some missionaries have compiled word lists, lexica as well as dictionaries in Civili (e.g. Marichelle, 1902 & 1912). These missionaries’
works have been followed by more recent publications including a lexicon (Tchimina 2016), a dictionary (ILALOK 2008) and a textbook comprising an abundant bilingual glossary (Loembe 2005).

This paper lays stress on the tradition of lexicography in the Gabonese languages. Although it aims at addressing specific issues in Civili, metalexicographic proposals in the other African languages will be necessarily brought to the fore. On this line of argumentation, it has been variously shown (cf. Prinsloo 1992, 1994, 2009, 2014; Prinsloo & De Schryver 1999) that African languages share certain common lexicographic problems which are completely unknown to the English or Afrikaans lexicographic situation, such as tonal indication, handling of singular versus plural forms of nouns, selection inclusion versus omission of derivatives of the verb.

Therefore, suggestions will be made along the lines of what already exists as far as the lexicographic traditions in the African languages are concerned. In this paper, a specific attention is essentially paid on problems faced in current Civili lexicography as well as on possible lookouts for Civili lexicography expansion. The paper makes a repertoire of major issues in Civili lexicography and draws guidelines for the development of Civili practical lexicography.

Lexicographic issues in the Civili language concern not only lemmatization procedures (e.g. lemma selection processes, lemma ordering, etc.) but as well as orthographic representations (e.g. orthographic alphabet, word division, etc.) and tonal indications (especially in view of distinguishing between homophones). This paper presents each issue and suggests proposals to deal with each in the attempt to give an answer to each of the following questions:

(i) What are the lemmatization processes in Civili dictionaries?
(ii) What are the orthographic representations practices in Civili dictionaries?
(iii) What is the tonal indications consideration in Civili dictionaries?

According to the word tradition, lexical items are entered in their complete forms, i.e. prefix plus stem, while in the stem tradition lexemes are lemmatized under the stem without their prefixes. Both these traditions represent a problematic theoretical issue in linguistics and lexicography. Therefore, they are critically evaluated in this paper with special reference to Civili. Prior to a discussion of the above-mentioned points, some information about the Civili language will be given.

The paper concludes with agreeing on the fact that the arrangement of lemmata remains a bone of contention among various lexicographers working on Bantu languages. It is also viewed in this paper that in keeping with the user-perspective, the word tradition should be preferred to the stem tradition. With regard to lemma-clustering ordering, this study emphasizes that nested and niched arrangements usually add to the problem in multilingual dictionaries.

References


*****

**Electronic dictionaries and lexicographical theory: A case study of the *Dictionnaire Intercatif pour Jeunes Apprenants***

Ludwine MBINDI ANINGA (suncine9208@gmail.com)

Institut de Recherches en Sciences Humaines (IRSH/ CENAREST)

Centre de Recherches en Etudes Germaniques et Interculturelles (CREGI/UOB)

Reference books have been compiled for Gabon’s languages since the early 1840s by both Gabonese and non-Gabonese people. In that regard, there are major dictionaries and other lexicographic works done by colonial explorers and religious missionaries before the nation’s independence. The current expansion of lexicographical works has brought quite a number of lexica, dictionaries and other reference books into the Gabonese market. A number of old dictionaries have been revised and new ones have been compiled as well. Each major linguistic community in Gabon now has at least two dictionaries, and there is now at least one lexicon for each minority linguistic group. This new wave of lexicographical works does not leave the field of electronic and online lexicography behind. More and more websites offer data on Gabonese languages, and several electronic materials such as CD-ROM, online lexica exist. However, all these works are produced mostly by non-trained lexicographers and very few of these works involve sound theoretical substance. A question is also posited on the usefulness of these works which have yet to be used in education, in popular literature or elsewhere.
Meanwhile, Human Language Technologies (HLT’s) research and development and the outbreak of Information Communication Technologies (ICT’s) have nowadays allowed not only a quicker information access at one’s choice but also an easier way for human-machine communication and interaction in both written and spoken language. Experiences in countries such as South Africa and Japan have shown that the development of computational linguistics and language technology industries contribute significantly to language promotion. Technological products from computational linguistic research (e.g. electronic spellcheckers, grammar checkers, machine translation software, speech synthesisers, etc.) have consequently enhanced further research on languages with limited resources.

Both HLT’s and ICT’s have been deemed as quick and efficient ways for endangered-language promotion and revitalisation. Based on UNESCO’s criterion for endangered languages, most Gabonese languages have been regarded as endangered languages. In fact, Gabon is known as a multilingual nation with languages that are, for most of them, classified as endangered by UNESCO. Although several linguistic research works have been successfully conducted, there is still a gap between research results and their use for community development. One of the reasons for such a situation is the fact that research products are not accessible to common users. It is a fact that Gabonese languages revitalisation and promotion are duly dependable on a better visibility of scientific outcomes about those languages, both at national and international levels. The inaccuracy on the number of existing Gabonese languages despite the numerous classification and inventory works on these languages is the perfect illustration. Hence more, not only the number of languages is not well determined, but also the ones that have been classified are not even encouraged for vernacular use, nor are even being promoted and included in the education system.

Therefore, the move to computational linguistics and/or computational lexicography for language research in Gabon would be an advantage for Gabonese linguistic communities in addition to the economic impact such research could have. With its capacity of large data processing, computational research saves time and allows better research results dissemination. The choice for computational linguistic research has led to the compilation of an electronic dictionary intended for users between 4 and 16 years old: the Dictionnaire Interactif pour jeunes Apprenants (DIA). The DIA is a multilingual onomasiological dictionary with Gabonese languages, English, French, and German as treated languages. It comes within a bigger project which is aimed at compiling corpora for Gabonese major native languages.

The present paper seeks to demonstrate the importance of a solid theoretical background while dealing with the compilation of an electronic dictionary such as the DIA. It will look at the compilation process of the dictionary, as well as its framework, its content and to an extent, at the impact on the consultation process. The paper will also describe the application of such an electronic dictionary in the education system. It will demonstrate the benefit of using such a dictionary to reverse the extinction process of Gabonese languages.

*****

Developing an LSP bilingual glossary to support international students at the University of Johannesburg
Guy-Roger MIHINDOU (gmihindou@uj.ac.za)
Academic Development and Innovation (ADI), University of Johannesburg
Mapelo TLOWANE (mapelot@uj.ac.za)
Academic Development and Innovation (ADI), University of Johannesburg
South Africa is emerging as one of the world's most exciting study destinations. According to IIE (2013) - an annual publication of IEASA, the number of international students has “grown dramatically” since 1994, from 12,600 to 72,875 in 2012. Since 2007, the national average growth rate for international students has been estimated at 5.47%, higher than the international rate of 4.4% per annum. South Africa is among the top ten study destinations in the world.

Lee and Sehoole (2015) found that, among others, affordable fees, government subsidies for students from the region, proximity to home and cost of living, the strong reputation of higher education and currency of its qualifications, are significant pull factors for international students want to study in South Africa.

These “pull” factors are mentioned by nearly 1,700 students that Lee and Sehoole (2015) surveyed, while other aspects such as accommodation difficulties, difficulties in understanding South African languages, lack of funding opportunities, support and adjustment challenges, lack of South African friends and sometimes xenophobic attitudes towards African students were among obstacles international students articulated.

The majority of these students originate from non-English speaking countries and others from countries where English is either 3rd, 4th language, or is not the language of instruction in institutions of higher learning. Finding themselves in South Africa, such students are confronted with and challenged by the fact that English is used as a language of communication and as a language of teaching and learning at South African universities.

In response to the severity of the language problems faced by international students both upon registration and during their academic endeavours, the Division of Internationalization (DI) and the Academic Development Centre (ADC) of the University of Johannesburg initiated the creation of University of Johannesburg English Language Programme (UJELP).

UJELP’s focus is on the development of English for communication and for academic purposes. International students who require support in English language and Academic Literacies enrol in the programme, which comprises two components. Firstly, tuitions in the English language, and secondly, the English for specific purpose in the form of Academic Literacies (AL) as a domain or “new language”, intended to assist students (both local and international) in integrating academia. UJELP’s AL programme makes it possible to equip students with academic vocabulary and language for specific disciplines such as English for Law (EFL), Language for Economic Sciences (LES) Language and Skills for Science (LSS).

In this study the plan is to create an LSP glossary in English and Sesotho sa Leboa with the intention of assisting enrolled international students to learn an additional South African language. The glossary will comprise of the vocabulary of the most useful expressions in Sesotho sa Leboa and some basic terms and phrases from specific disciplines.

International students will be afforded an opportunity to learn a new South African language in an LSP context that will help them fit into the South African society. On the other hand, international students, known for their great academic performance, would be able to help some local students if they can communicate with them in the South African vernacular languages, which will then facilitate code-switch, and peer learning.

Sesotho sa Leboa is one of the most used languages at the University of Johannesburg and the city as a whole. The city of Johannesburg is cosmopolitan and therefore the glossary will assist the students in learning about and integrating the community through the language. The present study intends to (i) compile a bilingual glossary for international students, (ii) make it accessible to both groups and facilitate its use (iii), as well as monitor and evaluate its impact on the lives of its users. The glossary will be compiled in 3 phases: a) the conception, including its structures, b) the data collection and, C) the compilation per se.
As mentioned earlier, the challenges of studying in South Africa listed by international students included accommodation, financial constraints, language, materials support and adjustment challenges, and xenophobia. As a result of these factors, some international students tend to only socialise with other international students. By creating this bilingual glossary, we are promoting interaction amongst different cultures through language. The objective of this project is to ensure that the end product, which is a bilingual and monodirectional glossary, becomes a dynamic practical tool for its end-users in that it can be stored, retrieved, modified and enriched according to their needs.

References

*****

The treatment of synonym paradigm members and plural nouns in the macrostructure of Thanodi ya Setswana and strategies to enhance its reference quality
Nickey S. MOKONE (nickeym@raf.co.za)
Language Unit, Road Accident Fund, South Africa

Ineffective treatment of lexical items and inadequate information in the macrostructure reduce the value of the reference quality of Thanodi ya Setswana. This dictionary is not consistent in its use of structural reference symbols e.g. the use of labels on some headwords to mark vital lexicographic information and reference addresses symbols which result in shortcomings with regard to the use macrostructure as a reference tool to assist users to decode lexicographic meaning and information.

The exclusion of certain plural nouns in the macrostructure of this monolingual descriptive dictionary as part of the lemmatisation procedure put an inexperienced user at a disadvantage especially in cases where these plural forms are more frequently used compared to their singular nouns counterparts. This poor macrostructural treatment of such lemmas creates a long and winding searching procedure which may result in user being frustrated and abandoning the search without getting what he/she was looking for.

The absence of members of a synonym paradigm in the treatment of lexical items in the macrostructure of Thanodi ya Setswana is also a cause for concern. There are certain lexical items which are given as part of the definiens of a lemma and also form part of the synonym paradigm in this dictionary’s article but are not lemmatised in the macrostructure. Although they are given as synonyms paradigm, the exclusion of synonyms paradigm members in the macrostructure of Thanodi ya Setswana creates a problem for the user to access information; it makes it difficult for the user to decode lexicographic information when a lemma which is referred to in a dictionary article for further information and semantic relationship does not link with the lemma that was not used within the synonym paradigm and within the reference position.

44
The macrostructure as a part of the dictionary is instrumental component to provide lexicographic information has to be used in such a way that information retrieval becomes easy. According to Gouws (1993:32), the macrostructural ordering of information should be done in such a way that a user can find answers to most of his her/his questions in the convenient manner. Lexical items presented in the microstructure as part of the synonym paradigm should be lemmatised in the macrostructure because of their occurrence as members of the lexicon. The macrostructure should reflect this lexicon to emphasise lexical cohesion and show semantic relation among these synonym paradigm members, these lemmas should appear in the central list of this dictionary.

Compilers of dictionaries have to decide on a system of lemmatisation. In choosing a certain lexicographic tradition they should consider the practicality and nature of the language; and the possibilities to enhance the reference quality of a dictionary. There is a lexicographic tradition adopted in Thanodi ya Setswana which never considered the practicality and nature of the possibilities to enhance the reference quality of a dictionary. Gouws (1991: 75) states that to ensure a sound treatment of all lexical items found in a dictionary, a traditional word-based approach should be replaced by a broader lexicon-based approach that offers a more comprehensive reflection of the lexicon. It will be better to reconsider this principle and ensure that regular plural nouns are lemmatised to enhance reference quality of a dictionary. This poor macrostructural treatment of lemmas in Thanodi ya Setswana has resulted in suggested strategies in the treatment of lemmas to enhance reference quality for this monolingual descriptive dictionary.

A suggested strategy to lemmatise plural nouns in Thanodi ya Setswana should be done in such a way that the user can easily find these lexical items in this dictionary’s macrostructure, and access the information associated with them. This type of lemmatisation, that includes plural nouns, is not sufficiently presented under the treatment of singular nouns. The inclusion of plural nouns in the macrostructure will improve the accessibility, reliability and user-friendliness of the dictionary, which are most likely to be searched for by the dictionary user. This shows a modern trend in lexicography to ensure that user-friendliness will be accomplished.

The lemmatisation of lexical items and treatment has a bearing on the user-friendliness of this dictionary. The compilation of a dictionary using practical lexicographic theories has to ensure that user’s needs are met. The macrostructure as part of the dictionary’s instrumental component has to be used in such a way that information retrieval becomes easy.

The lexicographic theories of ordering lemmas in a dictionary’s macrostructure need to be reviewed to suit the nature of the language and user it serves. The lemmatisation in the macrostructure should reflect the lexicon of the language in question. There is no doubt that the best dictionary is one which enables the user to find the required information without difficulty in the first place where he/she looks. Should the suggested treatment procedures be adopted in a descriptive monolingual dictionary like Thanodi ya Setswana, the dictionary’s reference quality will be significantly improved.

References
Pronunciation in African language dictionaries
Hugues S. NDINGA-KOUMBA-BINZA (hugues.ndingakoumbabinza@nwu.ac.za)
Centre for Text Technology, North-West University

In general, pronunciation can be defined as the way certain speech sounds are uttered (cf. Richards & Schmidt 2002: 429; Crystal 199: 274-275). In dictionaries, pronunciation is rather the way a certain word should be articulated according the orthographic alphabet used to represent the phonetic and/or phonological units (segments and suprasegments) of the language or languages for which a particular dictionary is made. In dictionaries of most major world’s languages (e.g. Dutch, English, French, German, etc.) pronunciation is one of the key components of the dictionary microstructure.

The present paper results from an examination of various published dictionaries and a number of planned dictionaries in African languages. The languages concerned in this study are languages of the Bantu phylum, comprising of Nguni and Sotho languages of South Africa (namely Sepedi, Sotho, Tswana, Xhosa and Zulu) and of a few native languages of Gabon (namely Civili, Geviya, Fang, Pove, Yilumbu and Yipunu). A total of forty dictionaries and fifteen metalexicographical dictionary plans were examined.

The initial aim of the examination was to verify whether pronunciation is indicated in African languages dictionaries. The intention of this investigation was motivated by the assumption that most African languages, especially Bantu languages, have orthographic alphabets essentially based on their phonetic-phonological systems (cf. Hombert 1990a & 1990b, Ndinga-Koumba-Binza 2010). As such, words that may make the lemmata of these languages’ dictionaries should be easy to pronounce – at least at segmental level – and thus may not need any particular indication for pronunciation.

The subsequent aim of the investigation was to figure out how pronunciation is indicated, if it is, in the African language dictionaries. This leads to the question of what would be the necessity for pronunciation information in dictionaries of languages which have one of the simplest orthographic alphabets. An answer to this question is multi-faceted in a number of metalexicographical studies for dictionary compilation. One common view of these studies is that pronunciation is part of the linguistic information to include in a dictionary article at microstructural level.

A number of other issues were also considered as the dictionaries and metalexicographical plans that constitute the subject matter of this study were being checked. Such issues include: How adequately is pronunciation indication in terms of the phonetics-phonology dichotomy? In fact, pronunciation can be related to both phonetics and phonology, which are theoretically two different disciplines with two different formal representations for sound transcription. In phonetics, language sound is transcribed with square brackets while in phonology language is transcribed within oblique bars.

Another issue investigated was how adequately pronunciation information relates to the language orthographic alphabet. This question is relevant when one notes that a number of African languages lack standardised spelling rules in the orthographic alphabet. Even when an orthographic alphabet does exist, it is not necessarily clear whether it is a phonetically-based alphabet (i.e. based on the entire phonetic system of the language) or a phonologically-based alphabet (i.e. based only on the phonemic system of the language).

Findings of this investigation show two perspectives to mention. First, there exists a correlation between dictionary typology and pronunciation indication. This is to say that not all dictionary types deem it necessary to indicate pronunciation. For a specialised dictionary, for instance, Bergenholtz & Tarp (1995) signalled that information on pronunciation can be left out when “a specialized dictionary is being designed with a view to writers and problems of writing only” (Bergenholtz & Tarp 1995: 134).
This paper will contain three sections. Section 1 will give a general account of pronunciation indication in dictionary compilation. Section 2 will be concerned with the current investigation methodological approach. Section 3 will present the examination results and the research findings discussion. The discussion will include a contribution to categorising dictionaries according to the necessity for pronunciation information in the microstructure.

References

Lexicography as a discipline and profession in Zimbabwe: A critical appraisal

Eventhough NDLOVU (eventthough@yahoo.co.uk)
Unit for Language Facilitation and Empowerment, University of the Free State

“Language policy… can influence the number and types of reference works produced and used Hartman and James (1998: 51).”

This paper is an appraisal of the status of lexicography as a discipline and profession in Zimbabwe. A review of the research output of the African Languages Research Institute (henceforth: ALRI), Zimbabwe’s lexicography unit, shows that ALRI’s main thrust is language research and development as well as dictionary making. The unit’s dictionary making activities seem to only focus on IsiNdebele and ChiShona, which were the national languages and the only local languages taught up to tertiary level in the country before the Zimbabwean Constitution Amendment (No. 20) Act. In terms of output, ALRI has compiled; Isichazamazwi SesiNdebele (2001) the first Zimbabwean Ndebele monolingual general purpose dictionary, followed by Isichazamazwi SezoMculo (2006), a trilingual musical terms dictionary, for the Ndebele language. Apart from these two dictionaries, there is also the bilingual Ndebele Practical Dictionary by Pelling published in 1965, which however is not ALRI’s research output. In total, only 3 dictionaries have been compiled in IsiNdebele. In ChiShona a total of 5 dictionaries have been compiled, namely Duramazwi reChiShona (1996), Duramazwi Guru reChiShona (2001), Duramazwi reUrapi neUtano (2004), Duramazwi reMimhanzi (2005), Duramazwi reDudziramutauro noUvaranomwe (2007) and the soon-to-be printed Shona Children's Dictionary. Apart from these works compiled by ALRI, Hannan (1959; 1974) compiled the Standard Shona Dictionary and the Shona Standard Dictionary respectively while Dale (1981) compiled the Duramazwi: A Basic Shona-English Dictionary giving us a total of 8 dictionaries in Shona. It is therefore clear,
then, that lexicography as a discipline and profession in Zimbabwe is still at its nascent stages, especially when compared with South Africa, which has lexicography units for each official language. As such, a lot of work lies ahead in terms of dictionary making in Zimbabwe, especially in the wake of the promulgations of Section 6 subsections 1-4 of the Zimbabwean Constitution Amendment (No. 20) Act, which states that Zimbabwe has 16 officially recognised languages. This section obliges the state, its institutions and all government agencies at all levels, ALRI included, to ensure that these languages are treated equitably and that their use and development are promoted and advanced. ALRI’s research output therefore presents an interesting case study where the researcher seeks to account for this state of affairs in Zimbabwean lexicography with the view to address the following key questions; Why the bias towards ChiShona and IsiNdebele in lexicography? Why the strong bias towards ChiShona? To what extent can the language policy account for the state of affairs in Zimbabwean lexicography? Given the intricate relationship between lexicography and language policy in terms of the general language policy and the language-in-education policy, a critical discourse analysis of the policy documents that enshrine the Zimbabwean language policy and language-in-education policy reveals that the lack of an enabling language policy environment in Zimbabwe accounts for the underdevelopment of lexicography as a discipline and profession in the country. It also explains why the bias towards ChiShona and IsiNdebele. Semi-structured interviews with language practitioners and language experts, publishers, policy critics and makers also showed that past and present promulgations of the language policy and language-in-education policy have not provided and do not provide a conducive environment for the growth and development of lexicography as a discipline and profession in Zimbabwe. It was noted that the marginalisation of most the local languages, their constricted functional space and the low status accorded to them, particularly in the education sector, legal fraternity and in information dissemination constrains the growth and development of lexicography. A close analysis of the curriculum policies of selected state universities in Zimbabwe reveals that lexicography is only taught as a semester-long course at undergraduate level, revealing its low status as a discipline and profession. Interactions with seasoned translators and interpreters showed that the lack of professionalisation of translation and interpretation, and their lack of recognition could also explain why lexicography has not developed significantly as a discipline and profession, especially given that these disciplines and professions permeate each other so fully that it is difficult to separate them. Dictionaries are key tools for translators and interpreters and where multilingual service provision is promoted by the language policies, dictionaries in the concerned languages are compiled to support the work of translators and interpreters. As a result, this will see the growth and development of lexicography and inevitably that of translation and interpreting.

References
S'ncamtho lexico-semantics and terminology development in Ndebele: Implications on lexicography and standardisation
Sambulo NDLOVU (matsilaneg@gmail.com)
Department of African Languages and Literatures, Great Zimbabwe University

S'ncamtho is a popular Ndebele/Zulu-based urban youth language in Bulawayo. The language has had influences on the Ndebele language over time; however, most of the influences are not welcome by Ndebele language purists and gate keepers. This youth variety, like others of its type, has an extensive use of metaphor that affects the meaning of some Ndebele terms to create new meanings and in the process add to Ndebele terminology. All Ndebele-based slangs are feeder varieties to S'ncamtho and some of these slangs operate in specific environments of use and have specialised registers that build on to Ndebele scientific terminology in areas such as mining, especially gold panning, and education. S'ncamtho borrows from English, Afrikaans and Zulu to create a lexical stock that is widely used in Ndebele-speaking areas and spreads across age and gender boundaries influencing Ndebele terminology. This paper argues that Ndebele benefits from S'ncamtho its urban youth variety in terms of terminology although efforts at linguistic purism often compromise this contribution. There is terminology borrowing in Ndebele that is facilitated by S'ncamtho and the words operate within Ndebele, especially the spoken form which, according to this paper, is the real language as the written form is an artificial representation of the language. The paper avers that S'ncamtho terminology creates synonyms in some cases whereby a S'ncamtho term or expression becomes as popular as the standard Ndebele counterpart and in some cases the S'ncamtho version can even be more popular. By creating synonyms S'ncamtho is impacting on Ndebele terminology and lexical stock. Apart from synonym creations S'ncamtho has created terminology and expressions that are specialised in some disciplines and domains so much that these terms and expressions help in covering the terminology gap in Ndebele specialised terminology. The paper also argues that S'ncamtho has helped and continues to help in euphemistic term creation in Ndebele especially in the area of reproductive health. The euphemisms help Ndebele speakers to express themselves through S'ncamtho terminology. Structural and semantic manipulations in S'ncamtho create a terminology development platform for Ndebele as there is in some cases wide spread usage of S'ncamtho terminology in Ndebele aggregates. The study collects S'ncamtho terminology from Ndebele youths in Bulawayo, and S'ncamtho specialised terminology is collected from the taxi industry in Bulawayo, gold panners in Bulawayo peri-urban and euphemisms especially reproductive health euphemisms are taken from the youth in the city of Bulawayo. Collected terminology is tested for knowledge and usage across age and gender variables amongst Ndebele speakers to measure the influence of these S'ncamtho terms in the Ndebele language. The research is guided by the assumption that S'ncamtho has spread beyond Bulawayo male youth and is impacting standard forms of Ndebele. The paper goes on to evaluate the treatment of popular S'ncamtho terminology in the Ndebele dictionary Isichazamazwi seSiNdebele and gives recommendations on standardising some S'ncamtho
terminology in Ndebele. Due to popular usage of S'ncamtho Ndebele terminology is laden with S'ncamtho terminology to such an extent that even the education sector is affected. The paper is motivated in part by the debate that arose after the S'ncamtho term for prostitute umahotsha was included in a grade seven (primary school) examination. The inclusion of the term raised a lot of dust with some purists arguing that it is slang and slang cannot be used in examinations however, looking closely at S'ncamtho and Ndebele it is not linguistically and socially possible for S'ncamtho and Ndebele to share space and speakers and remain independent of each other’s influences on the other. The study is conducted through interviews of knowledge testing and observations of usage, interviews on the acceptance of these terms are also conducted. A ‘S'ncamtho analysis’ of the Ndebele dictionary is also a research methodology employed in the paper. While some terms in S'ncamtho border on the anti-language typology some are useful to the Ndebele language as essential terms for expression and gap filling.

*****

Lexicography and language revitalisation for the Tjwao language in Zimbabwe
Shumirai NYOTA (shumirai.nyota@gmail.com)
Great Zimbabwe University
Zvinashe MAMVURA (zvinashe.mamvura@gmail.com)
African Languages Research Institute, University of Zimbabwe

Language documentation for a language threatened with extinction like the Tjwao language of Zimbabwe is critical as a way of revitalising the language. This paper considers the roles of lexicography in the revitalisation of Tjwao, a critically endangered language spoken in Zimbabwe. Tjwao is a moribund language spoken by less than 10 individuals aged between 60 and 90 in the Tsholotsho and Plumtree districts of western Zimbabwe. The Tjwao community itself amounts to approximately 2,500-3,000 members, most of whom have shifted to the local Bantu languages, Ndebele and Kalanga which are in contact with Tjwao. A survey undertaken by Fehn and Phiri in 2015 suggests that the language shift began about two generations back and was accelerated by the resettlement of thousands of people during the establishment of what is now Hwange National Park. Despite its endangered state, Tjwao obtained an official status in 2013, but little has been done to develop or merely safeguard the language. The researchers observed a critical need for a Tjwao dictionary as a way of documenting, preserving and revitalising this endangered language. This paper focuses on the functions of a community-focused dictionary. A corpus-based dictionary on the language has to be compiled urgently before the last speaker dies. There has been no inter-generational transmission of Tjwao, and in that case, the proposed dictionary shall be a Tjwao-Ndebele bilingual dictionary since the younger generation has acquired Ndebele as their first language. In addition, Ndebele is the ‘lingua franca’ for the area in which the Tjwao people are found. The dictionary articles shall give information on part of speech, translation equivalents, sense distinction, examples of usage in source and target languages. All this shall be done in a very user-friendly way. The back matter of the dictionary shall have a glossary on flora and fauna terms, human body parts etc. Plural nouns shall be lemmatised with cross-references to their singular forms. The research team for the dictionary project comprises of six members. Five are linguists, two of whom are first language Ndebele speakers while the sixth one is a language activist, Dave Ndlovu who is also the Chairperson of the Tsholotsho Development Trust, an organisation which works for the empowerment of the speech
community. He is intimately connected with the Tjwao community and is very critical for 
mediating between the community and the researchers or to provide links between the 
researchers and the community. The team will engage with the members of the speech 
community in order to help with capacity building and transference of skills for the younger 
generation which has attended school to be trained as “community linguists” who will work 
in collaboration with the research team. The literate members of the speech community are 
not active speakers. Active speakers shall work with the research team as ‘language 
specialists’ (Beier 2009: 4). The dictionary is aimed at enhancing conversational proficiency 
in the speech community for the generations that have shifted to Ndebele and Kalanga. It 
will be linked to language memory tests and language-learning exercises with audio files. The 
proposed dictionary will be done within the framework of the Master-Apprentice Program, 
originally devised by Leanne Hinton (1997, 2001), Nancy Richardson, and Mary Bates 
Abbott for revitalization of Californian languages. The research team will bring the active 
Tjwao speakers and the other Tjwao people who intend to learn the language together. 
Through instituting this method of compiling dictionaries, the research team will bring 
together active speakers who will be acting as the ‘masters’ and the learners who become the 
‘apprentices’ together. Following Ogilvie (2011: 397-8), the process will be guided by the 
following principles: (1) the master and the lexicographer must not speak together in the 
dominant language (the language which is replacing the endangered language, Ndebele 
and/or Kalanga in this case); (2) only oral (not written) language must be transmitted; (3) the 
lexicographer must be at least as active as the master in deciding what is to be learned and in 
keeping communication going in the language; (4) learning must take place in real-life 
situations and traditional activities for example collecting food, going hunting, cooking, and 
doing crafts; (5) it must all be audio- and audio-recorded for later analysis and use in the 
dictionary. All the efforts to revitalise a language squarely rests on the attitudes that speakers 
of the language have towards their language. The Tjwao community has shown interest in 
learning their language and are passionate for seeing the inter-generational transmission of 
their language. The dictionary will be compiled using the working orthography developed by 
CASAS in 2015 (Sibanda et al. 2015). Some members of the research team have produced 
learning and reading materials in the Tjwao language using the developed orthography.

References
Beier, C. 2009. Implementing Community Involvement in Collaborative Language 
Documentation: The Iquito Case. First International Conference on Language 
Documentation and Conservation.


eds.). The Green Book of Language Revitalization in Practice San Diego: Academic 

Ogilvie, S. 2010. Lexicography and Endangered Languages: What Can Europe Learn from 

Sibanda, M., Mpofu, M., Sibanda, M., Moyo, N. Msindo, B. & Vundla, M.K. 2015. A 
Unified Standard Orthography For Tjoao Languages: Angola; Botswana; Namibia; 
South Africa & Zimbabwe. The Centre for Advanced Studies of African Society 
(CASAS) Monograph Series 264. Cape Town: CASAS.
How much specialised data can and should be incorporated into meaning explanations in general language dictionaries? A discussion based on sports terms

Heidi A.P. PEDERSEN (heap@bcom.au.dk)
Centre for Lexicography, Aarhus University

In my paper I argue that there is a problematic treatment of field specific words in both general language dictionaries as well as specialised dictionaries, using the sports field as my point of departure and focusing on meaning explanations (Agerbo 2015). The treatment in general language dictionaries in particular results in many cases of inadequate solutions that cannot sufficiently accommodate the dictionary users’ needs if we consider these from the viewpoint of the function theory, i.e. that a meaning explanation (and all other data types) is supposed to help a specific user type who has an information need resulting from a specific type of extra-lexicographical situation (Bergenholtz & Tarp 2003). In my study, I investigated a large number of general and specialised information tools in English and Danish. In the first part of the investigation, I analysed the introductions to determine the intended purpose and users of each tool. In the second part, I analysed a selected number of sports lemmas in the dictionaries. The content of the dictionary articles in each dictionary was compared to the stated purpose of the dictionary, i.e. compared to the intended user and user situation. The findings from these analyses showed that a lot of these information tools either do not clearly specify their functions or cannot live up to their intended functions – sometimes they contain too much data, and sometimes they lack relevant data. As an example, the investigated general language dictionaries typically focus on the meaning of sports terms in football and a few more of the most popular sports in those countries, often excluding the use of the terms in the less famous sports (compare the use of red card in football to the use in race walking and badminton – a red card in football means that the player is excluded from the rest of the game, but this is not the situation in the two other sports; regardless of this, it is the football meaning that is explained in all the analysed dictionaries). As a second example, some of the explanations are either incorrect or so simplified that they do not help the user understand what a sports term actually means (the definitions of goalkeeper do not mention the essential meaning element that special rules apply to this person’s contact with the ball – their definitions could be applied to all defenders). And as a third example, in many cases, the dictionaries do not successfully distinguish between different dictionary functions; they mainly focus on reception and in some cases also production, i.e. communicative functions, but the data presented in many of the articles are not all incorporated to support these functions (e.g. a synonym definition is not helpful if a person has read an unfamiliar word in a text and wants to understand what this word means).

I suggest how such identified problems (narrowness, incorrectness or simplicity, and function) should be treated in specialised information tools in the form of a new concept for a number of specialised e-tools on sport, fitness and biomechanics based on the function theory (Bergenholtz & Agerbo 2015). These tools will all be derived from the same database and will be presented to the users as online dictionaries, which means that I can take a much more dynamic approach to the distribution of data and the structuring of these information tools. I also suggest how the problems mentioned above and other identified problems in my investigation could be treated in general language dictionaries. These suggestions will be related to the title of this paper: How much specialised data can and should be incorporated into meaning explanations in general language dictionaries? For the first example above, I suggest that for reception, non-interested laymen will need if not all, then the uses in the most popular sports, possibly described as one general meaning, whereas interested laymen and semi-experts only need the infrequent uses of the terms, possibly combined with a general
description of the term in the form of a phrase or simple sentence. For the second example, I argue that the three main principles for writing meaning explanations should be: relevance, clarity, order and correctness, and that in terms of relevance, the lexicographer must make sure that the data included in a meaning explanation (and also for the other data types) should depend on the dictionary function – nothing more and nothing less than what is necessary for a specific function should be included. For the third example, I provide different suggestions of dictionary articles that are based on the intended user and user situation, among other things demonstrating what a definition of red card would look like for a layman for reception and for a semi-expert for reception, and what it would look like for a layman for interpretation.

References

*****

The Sepedihelper as a support tool for text production in a dictionary use situation
Daniel J. PRINSLOO (dprinsloo@gmail.com)
Entelect
J.V. PRINSLOO (jvprinsloo@gmail.com)
Aerosud
Danie J. PRINSLOO (danie.prinsloo@up.ac.za)
Department of African Languages, University of Pretoria

It has been stated repeatedly in the literature that electronic dictionaries have been slow to supersede paper dictionaries and that the initial electronic dictionaries leave much to be desired. The first generation of electronic dictionaries were little more than ‘paper dictionaries on computer’ not reflecting what De Schryver (2010: 585) calls “the true power of the digital age”. Heid (2009: 1) says that EDs “could do better” and Rundell (2009: 9) refers to “game changing developments that have expanded the scope of what dictionaries can do and (in some respects) changed our view of what dictionaries are for”. Rundell (2012:78) acknowledges the user support tools designed by Prinsloo, Heid, Bothma and Faaß (2011) “which are designed to guide users’ lexical and grammatical choices in text-production mode” as a current trend in dictionary development.

The authors of this paper strongly believe that it is important that the e-dictionary should remain the starting point of choice for information retrieval but it should be strengthened by a variety of linked information and support tools especially for text production purposes. This paper introduces new functionalities in the Sepedihelper, a writing tool for text production in Sepedi (www.sepedihelper.co.za). We describe a dynamic lightweight tool designed for user-involved text production of Sepedi phrases in a dictionary use situation. The user combines his/her existing knowledge with the generation power of the writing tool. It represents a full implementation as follow-up on work in respect of user support tools for isolated complicated grammatical structures such as copulatives, kinship terminology and verbal relative constructions as presented at eLex 2011 (Prinsloo et al.
The tool will be demonstrated for a number of nominal and verbal constructions and it will be shown that the tool can provide support to dictionary users for a multitude of combinations and permutations for all the nominal classes and verbal moods. The paper consists of five parts, i.e. (a) a linguistic abstraction / preparation algorithm which means formulating the linguistic core of the Sepedi constructions to be interpretable by computer programs and software programmers, (b) illustrating the functionality of the tool as electronic dictionary support system, (c) briefly describing the computational pragmatic approaches to the compilation of the tool and (e) reporting on user-feedback done by first year students at the University of Pretoria.

It will be shown how the Sepedi linguist should abstract the grammar rules for the compilation of nominal constructions such as pronouns and concords in the compilation of e.g. possessive, adjective and nominal relative constructions and for verbs in all of the verbal moods, tenses and aspects as well as for intransitive, transitive and double transitive verbs with and without direct and/or indirect objects. Consider, for example, the functionality of the tool for a double transitive verb in the past tense positive with two objects, one pronominalized by means of an object concord and the other by an emphatic pronoun in Figure 1.

Figure 1: A complex sentence compiled by the Sepedihelper
Step-by-step phrase building is performed in a horizontal line, current buildup (towards the final Sepedi phrase) as well as through stacking the relevant building blocks in a vertical layout generated from user input and required concords generated by the system in a number of steps. For each of the steps there is a clickable icon “?” for getting more cognitive information and typical examples. It also provides for a lookup procedure in cases where the user does not know the Sepedi words and finds it by typing in the English word.

The programming aspects will be outlined in terms of the required database, mark-up of input data, best practices etc. The software uses a machine readable English/Sepedi dictionary. These dictionary entries are marked-up for part of speech. The syntax and components of the grammatical constructions are hard-coded. The Sepedihelper was developed using AngularJS and Bootstrap on the front-end. The back-end was developed in PHP that uses a SqlLite database. An effort was made to use best practices to split out logic and display and input from output to allow maximum flexibility. This allows easier improvement, maintenance and extension of the application.

Finally, user feedback will be given especially in respect of user satisfaction and problems encountered in text production of Sepedi verbal and nominal phrases.

References


55
Lexicographic approach to spoken versus written corpora for Setswana dictionaries
Fannie SEBOLELA (fannie.s01@curroholdings.co.za)
Curro Academy Soshanguve
Danie J. PRINSLOO (danie.prinsloo@up.ac.za)
Department of African languages, University of Pretoria

In the past few decades the consideration of corpus data in the compilation of dictionaries became common practice for most dictionaries. Corpus data and in particular frequency counts were initially used for the benefit of the lexicographer, e.g. to decide on inclusion versus omission of lemmas in the dictionary. Putting information on frequency in the dictionary, albeit not as raw counts but by means of symbols, make frequency considerations visible to the user. Such symbols are, for example, filled versus unfilled diamonds in Collins Cobuild, stars in MED or a combination of letters and numbers in LDOCE. LDOCE distinguishes between frequency counts obtained from a spoken corpus, indicated by the symbol S and the numbers 1, 2 or 3, and frequency counts obtained from a written corpus, indicated by the symbol W and numbers 1, 2, or 3. Both these symbols are stratified into three levels, the top 1,000, the top 2,000 and the top 3,000, indicated by S1/W1, S2/W2 and S3/W3 respectively. The lemma *leg* is marked as S1W1 which means that it is most frequently used in both oral conversation and in written texts. In the case of the lemma *legal*, marked as S3/W1, the W1 indicates that it is most frequently used in written texts but somewhat less in the spoken language and simply means that it belongs to the top 3,000 in spoken English and to the top 1,000 in written English.

The aim of this study is (a) to determine if such explicit labelling of lemmas in Setswana dictionaries in terms of their use in spoken versus written Setswana should be done by the lexicographer and (b) to determine from a case study to what extent spoken Setswana differs from written Setswana.

It is unfortunate that most corpora around the world lack sufficient data from spoken sources. The reason for this is that there are many logistical problems and ethical factors involved in the collection of spoken data and the collection process is much more time consuming and expensive. It is estimated that it is about ten times more expensive in terms of resources to compile an oral corpus of, say, one million tokens (running words) in comparison to compiling a written corpus of the same size.

For this study the oral data was drawn from 60-minute tape recordings of individual interviews of 13 randomly selected persons, male and female, professional and non-professional adults ranging from 20 years to 60 years of age. The themes included Setswana poems and praises, wedding and birthday celebrations, and radio broadcasts. The semi-structured interview of each candidate included the following basic questions: *Aneela ka bokhutswane ka tsa botshelo jwa gago.* (Briefly tell about your life history.) *Bapisa botshelo jwa gago le botshelo jwa segompieno.* (Compare your life style with the present life style.) *Ke eng se o eletsang go bona se fedisiwa kgotsa o sa se rate?* (In your own opinion, which things do you consider most irrelevant or you dislike in the modern way of life?) *A puo le setso ya Setswana di a somarelwa kgotsa di a nyelela?* (Do you consider the language and culture of Setswana conserved or destroyed?)

The oral collections will briefly be analysed with comments on certain lexical items used in the oral communication as well as highlighting peculiarities of oral communication in comparison to the written corpus.

As a second step the ranks of types in the oral corpus will be compared to the ranks of types in the written corpus in order to determine whether differences between the use of Setswana words in written versus spoken language justify explicit labelling in Setswana dictionaries e.g. as has been done for English lemmas in LDOCE.
Studying oral corpora in comparison to written corpora can pinpoint so-called key words. A word can be regarded as ‘key’ if, in a specific corpus, it is used much more (positive key) or much less (negative key) than it is expected in terms of a more general (usually bigger) corpus of the language. Positive and negative key words that will be presented include kereya (find), mara (but), aowa (no), thoma (start), nkebe (maybe), nyaka (want), endene (and then), nate (father), tjelete (money), feisi (fist), gagwe (his/hers), jwa (belong to), gago (yours), jaaka (like), jang (how), motho (person), omongwe (somebody), Mopitlwe (March), etc.

Preliminary results suggest that generally speaking the difference between spoken and written Setswana is not big enough to justify specific labelling/marking of spoken versus written use on macrostructural level. The lexicographer, however, should take cognisance of the differences and account for such differences in the treatment of the lemma, i.e. on microstructural level. A few microstructural strategies will be suggested.

References

*****

Lost and found: Assessing the value of an unknown bilingual dictionary towards the development of Ndau
Emmanuel SITHOLE (emmanesu@gmail.com)
School of Languages: African Language Studies, Rhodes University

Ndau is a language that straddles the south-eastern Zimbabwe-Mozambique border, boasting a combined total population of almost 1.6 million speakers (Instituto Nacional de Estatíesco, 2010; Zimbabwe National Statistics Agency; 2002; 2012). Despite its constitutional recognition as an official language in both countries, the language is still patently marginalised in higher-status domains. Its invidious sociolinguistic positioning and attached deprivations in society could, perhaps, be a result of a number of factors including but not limited to the following: indifferent language policies, negative speaker attitudes, elite closure and language internal deficits. These can be summarised as legal (constitutional), linguistic and socio-political problems engulfing Ndau which could be expeditiously addressed by means of appropriate interventions to succeed in revalorising it in society.

Without negating the impact of other challenges, the present paper focuses on the linguistic dimension and asserts that utilising “lost and found” materials could effectively bolster and enhance contemporary scholarly efforts to document, preserve, maintain and intellectualise the language. The phrase “lost and found” materials evolves from a tacit acknowledgement of the roles that could be played by missionaries’ early works on Ndau, some of which were destroyed by fire alongside a printing plant at Chikore Mission in 1916 while others were never disseminated well and therefore unknown to justifiably merit the “lost” description (Doke, 1931a:132). To give credence to this claim, the century-old bilingual ChiNdau-English and English-ChiNdau Vocabulary with Grammatical Notes produced by the American Board in 1915 (hereafter NEV) is evidently unknown to mother-
tongue speakers, the clergy and researchers. This therefore builds the central argument in this paper that the NEV (1915) can be used as a pedestal to support current and future lexicographical practices in Ndau. This appears to be the norm rather than the exception in lexicography and elsewhere (Hadebe, 2006; Viriri, 2014; Mahlangu, 2014; Nthambeleni and Musehane, 2014).

While conceding that the dictionary’s main twin purposes were to fulfil the principal goals of evangelisation and civilisation thereby abetting the project of colonialism (Fortune, 1979:45), it could now execute a somewhat different but fulcrum role in the development of Ndau lexicography. The NEV (1915) contains 4,000 headwords from all the five dialects of Ndau while also including vocabulary from languages that it was historically and currently in contact with such as Zulu, Afrikaans, English and Shona. By successfully balancing internal linguistic variation in Ndau and handling also other languages’ influence, NEV demonstrated its recognition of the wider sociolinguistic realities in Ndau speaking communities which admittedly is still the case today. This strategically positions it to function as a key source of reference to plan strategies and procedures for the development of new lexicographical works. For instance, its core Ndau vocabulary coverage is useful in planning the compilation of a monolingual dictionary. Being the first and only substantial dictionary, the NEV (1915) can also effectively direct and guide other language services such as bi-directional translation and interpretation which demand skills in bilingualism. Its inclusion of grammatical information is an essential attribute that was in keeping with its goal of transmitting Ndau learning to non-native speakers a century ago. In the present day, this merit could still be capitalised on to assist new foreign learners and mother-tongue speakers who want to learn and perfect their Ndau and English skills respectively. NEV (1915)’s virtuous principles on dialect representation, defining and exemplification can be built on in the crafting and designing of general purpose monolingual and bilingual dictionaries in Ndau.

Besides being imbued with the positive and constructive traits to facilitate the development of lexicography and other corpus planning activities, the NEV (1915) is perhaps, not a sufficient yardstick. This is because it ignores “most derivatives, abstract nouns and illustrative sentences” (American Board Mission, 1915:3) which significantly further truncate its lexical coverage which, as already mentioned, is compartmentalised by its overtly pedagogic scope. This makes the dictionary depart from its traditional and normative task of comprehensively presenting, describing, documenting and preserving a language’s vocabulary as a general dictionary (Alberts, 2011:27). As a dictionary steeped in the traditional approach and method, NEV falls short on a number of key aspects such as orthography, lemmatisation, definitions among other structural deficiencies which adversely impact on its ability to fulfil its function. In view of these normal blemishes, the current paper insists that the dictionary should still be used as a foundational stone in the compilation of future general monolingual and bilingual dictionaries as well as technical dictionaries in Ndau. It is therefore pertinent to restate that the NEV (1915) could be an invaluable contribution in shaping the development of new lexicographical debates and practices in Ndau if the strengths it envisages are fully utilised as will be shown in this paper.

References


*****

Advantages and disadvantages in the use of the internet as a corpus. The case of the online dictionaries of Spanish “University of Valladolid”

Sven TARP (st@bcom.au.dk)

Centre for Lexicography, School of Business and Social Sciences, Aarhus University

There is no doubt that big electronic text corpora can be of great value to lexicographers when performing a series of tasks in connection with the compilation of dictionaries. This has been documented by various scholars engaged in practical lexicography, among them Bergenholtz (1996), Atkins & Rundell (2008), and Hanks (2012). Various high-quality dictionaries have been compiled based upon this empirical basis (see e.g. Sinclair 1997).

Today, corpora composed by texts containing hundreds of millions of words are available to the compilers of dictionaries, and it seems that they never stop growing. In this respect, big data is already a reality, but it should not be forgotten that no corpus, however big, can stand up to the enormous collection of texts and words which can be accessed through the Internet. The development of methods allowing for the use of this almost unlimited empirical basis constitutes undoubtedly a challenge more and more relevant to lexicography, for which reason it is the topic of this paper.

The paper will start with a brief discussion of the complex relationships existing between different classes of empirical basis which can be used in a lexicographical project, i.e. introspection, multispection, corpora, the Internet, existing dictionaries, textbooks or other information sources, or a combination of these.

After that, the paper will discuss some of the most important advantages and disadvantages of using the Internet as a corpus, in comparison with the “traditional” corpora in which the texts have been selected according to specific criteria relevant to the tasks that have to be accomplished. Among the disadvantages, the paper will discuss the problems related to the dubious origin and quality of many texts; and among the advantages, it will discuss the time factor, the big amount of texts available, and the continuous actualization with the most recent words and expressions. In spite of the undeniable disadvantages, the paper will conclude that it is perfectly possible, and even beneficial, to use the Internet as the main empirical source, without resorting to the “traditional” corpora”, when the objective is the production of dictionaries of still higher quality.

As an example, the paper will then show how the Internet has been used as the main empirical source in order to select lemmata and meaning units (senses) in a Spanish online
project, i.e. the Online Dictionaries of Spanish “University of Valladolid” (Diccionarios Valladolid-Uva), a project which is currently in an advanced phase of compilation.

The various methods applied to accomplish the two tasks mentioned will be discussed and some examples (translated in to English) of dictionary articles will be presented, especially focusing on the selected lemmata and senses which are already stored in the database sustaining the seven dictionaries included in the project. Furthermore, a comparison will be made between these lexicographical data and those appearing in other general online dictionaries of Spanish (among them, the one edited by the Royal Spanish Academy), and some examples will be provided which shows how the chosen methodology frequently permits the selection of a bigger number of lexicographical data relevant to the foreseen user group.

Finally, the paper will provide some general conclusions as well as some recommendations and suggestions for future lexicographical projects.

References

Thanks are due to the Spanish Ministry of Economy and Competitiveness for funding the project “La Teoría Funcional de la Lexicografía: Diseño y Construcción de Diccionarios de Internet” (Ref. FFI2014-52462-P).