

language**matters**@UKZN

**University Language Planning and Development Office
Fourth Quarterly Report: 2017**

INSPIRING GREATNESS



Langa Khumalo,

PhD Linguistics.

Director: Language Planning and
Development Office

The major event in this quarter was the Language Research Symposium (LRS), which is hosted by the University Language Planning and Development Office biennially. The second LRS highlighted the need to develop human language technologies (HLTs) in African languages for use as enablers in teaching and learning in higher education institutions. The ULPDO flagship programmes were also enhanced in the period under review. Corpus development, Terminology development, the Bua Le Nna project and the monitoring of HLTs is covered in detail in this report. ULPDO continues to create synergies with institutions that promote the cultivation and advancement of African languages. There is report on synergies on projects of mutual benefit between Edinburgh University, and other Universities in South Africa, which include the University of Cape Town, University of South Africa, and the Cape Peninsula University of Technology.

**Fourth Quarterly Report
October-December 2017.**

Contents

1. Introduction.....	4
2. Terminology Development	5
2.1. Social Work Terminology Consultative Workshop	5
2.2. Information Technology Terminology Consultative Workshop.....	6
2.3. Social Work Terminology Verification Workshop	7
2.4. Social Work Terminology Standardization Workshop.....	8
6. Running Projects.....	9
6.1. Bua Le Nna Sessions.....	9
7. ULPDO HLT's.....	10
7.1. Corpus Collection and cleaning	10
7.2 Term Bank, Spellchecker and the Zulu Lexicon.....	10
8. Workshops and Conference attendance	11
8.1 Director's conference and workshops attendance.....	11
9. ULPDO 2017 Language Research Symposium	12
10. Training workshops	14

1. Introduction

The period under review saw the University Language Planning and Development Office (ULPDO) host the second Language Research Symposium (LRS) in October 2017. The LRS is hosted by ULPDO biennially in order to provide a platform to reflect on the University's responsiveness to its Language Policy and Language Implementation Plan. One of the stated objectives of the University Language Policy is to "*achieve for isiZulu the institutional and academic status of English*" and to "*provide facilities to enable the use of isiZulu as a language of learning, instruction, research and administration.*" (Language Policy of the UKZN, 2014:2). It is in this regard that the LRS provides a platform for academics to confer on the advances and challenges in the intellectualization of isiZulu. The second LRS highlighted the need to develop human language technologies (HLTs) in African languages for use as enablers in teaching and learning in higher education institutions.

The ULPDO flagship programmes were also enhanced in the period under review. The IsiZulu National Corpus (INC) is now the biggest African language corpus with 31 million running words. New disciplines went through the terminology development processes of consultation, verification, and standardization. These include Social Work and Information Technology. These were added onto the isiZulu Term Bank. Visitors to the isiZulu Term also continued to grow, which is indicative of the positive uptake of this dissemination tool.

The multilingual project aimed at recognizing other official languages in order to enhance inclusivity and social cohesion is enabled through the SeSotho for communication project. This project is known as the Bua Le Nna (Let's Talk) project, introduced in 2014 by ULPDO in order to teach SeSotho for communicative purposes.

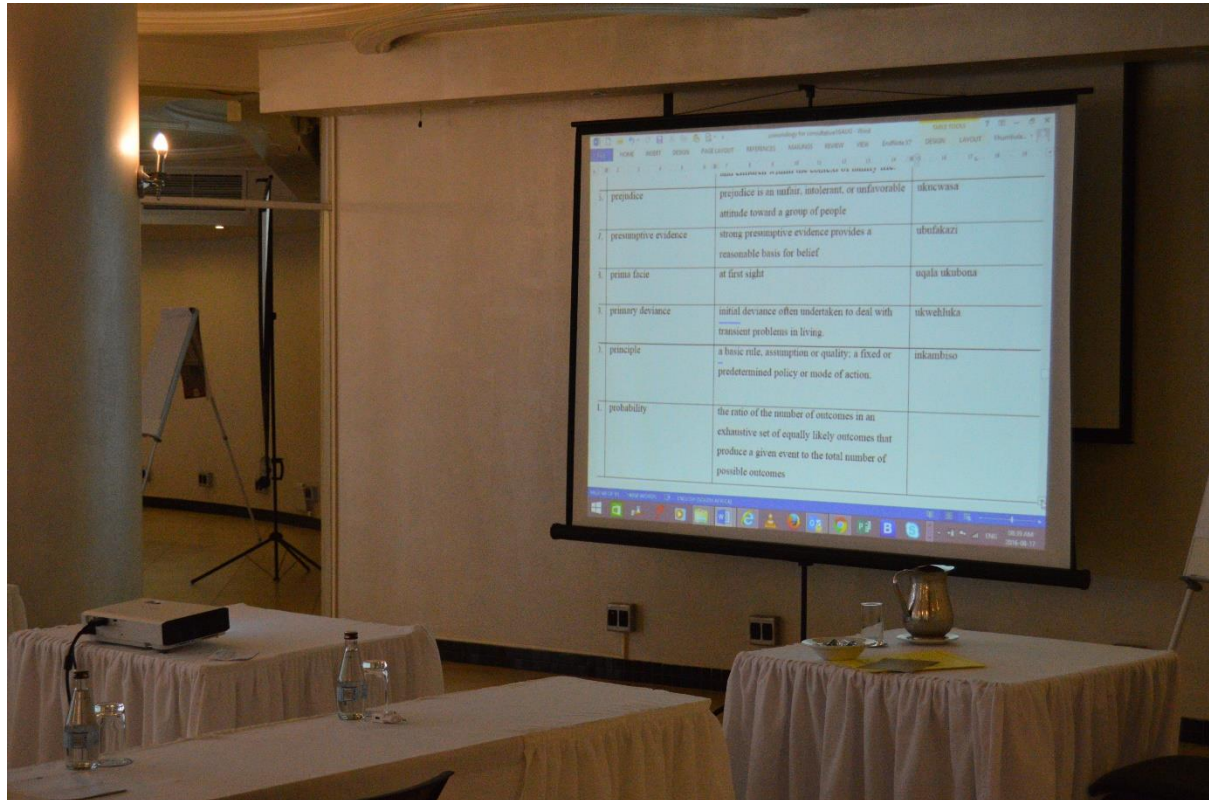
ULPDO continues to create synergies with institutions that promote the cultivation and advancement of African languages. This report details synergies on the MeMat project that is of mutual benefit and brings together Edinburgh University, the University of Cape Town and UKZN. Another project under the auspices of Universities South Africa (Usaf)'s Community of Practice for African Languages (CoPAL) is a terminology project that is coordinated by UKZN, University of South Africa, and the Cape Peninsula University of Technology.

Dr Langa Khumalo
Director: ULPDO

2. Terminology Development

2.1. Social Work Terminology Consultative Workshop

The ULPDO hosted the Social Work Terminology Consultative Workshop on 19-21 September 2017 at the Protea Edward Hotel.



The picture was taken during the workshop. Collaborators for this workshop included, Prof A. M Maphumulo, Dr G Mazibuko, Dr N Sibiyi, Dr B Zungu, Mr. Z. Xala, Mr B Nxumalo, Mr. SW Sibeko, Mrs. M Ngongoma, Ms G Mkhize, Mr B Mdluli and S Gumede, the subject specialists included; Dr S Mathe and Ms N Ngcobo and Mr S Mbambo and the ULPDO staff members.

It is encouraging to see more disciplines partaking in the language development initiatives. This is indicative of UKZN commitment to the intellectualization of the indigenous languages. The consultation workshop is the first stage in the terminology development process. The development of Social Work terminology is in consonant with UKZN's pursuit of the intellectualization of isiZulu language so that it carries the scientific rigors of research, teaching and learning. The terminology will ultimately be made accessible through open source technologies, and thus will be available for researchers, students and government departments.

The workshop was attended by collaborators together with subject specialists who worked tirelessly to coin and improve isiZulu term equivalents. A total of 494 terms were workshopped and are now ready for the verification stage.

2.2. Information Technology Terminology Consultative Workshop

The ULPDO hosted the Information Technology Terminology Consultative Workshop on 21-23 November 2017 at the Protea Edward Hotel.



In the picture are the collaborators for this workshop included, Prof. A. M. Maphumulo, Dr G. Mazibuko, Dr N. Sibiya, Dr B. Zungu, Mr. Z. Xala, Mr. B. Nxumalo, Mr. S. W. Sibeko, Mrs. M. Ngongoma, Ms G. Mkhize, Mr. B. Mdluli and S. Gumede, the subject specialists included; Mr S. Dlamini and the ULPDO staff members.

As alluded to above the consultative workshop forms the first crucial stage in the terminology development process. This initial stage is very important in order to consider a broad-based stakeholder view before the terms are considered for verification and later authentication. The information technology terminology is the second batch of terms in this discipline that is being put through the terminology development process. It is encouraging that this discipline is growing the scope of its technical stock in the isiZulu language. This development is an important precursor to the intellectualization

of isiZulu in order to aid the creation of novel teaching and learning pedagogies, which will improve access to epistemologies and student success.

The workshop was attended by collaborators together with subject specialists, who worked tirelessly to create appropriate isiZulu terminology for this discipline. A total of 536 terms were workshoped and are now ready for the verification stage.

2.3. Social Work Terminology Verification Workshop

The ULPDO hosted the Social Work Terminology Verification Workshop on the 7th-9th of November 2017 at the Coastlands on The Ridge Hotel.



In the picture are the collaborators for this workshop included, Prof A. M Maphumulo, Dr N Sibiya, Mr. Z. Xala, Mr B Nxumalo, Ms V Masuku, the subject specialists included; Dr S Mathe PanSALB included Mr J Simelane and the ULPDO staff members.

The verification workshop is the second stage in the terminology development processes. The workshop is led by a technical committee appointed by the Pan South African Language Board (PanSALB) whose remit is to refine the terms submitted as the output of the consultation workshop. This is the penultimate stage that prepares terms for the last stage, which is authentication (sometimes

referred to as standardization). Subject specialists continue to be a crucial part of these workshops to offer their technical expertise on the subject, usage and application of the terms in their context to help linguists verify them appropriately. A total of 521 terms were verified and are ready for the authentication.

2.4. Social Work Terminology Standardization Workshop

The ULPDO hosted the Social Work Standardization Workshop on 05-07 December 2017 at the Protea Edward Hotel.



In the picture are the collaborators for this workshop included, Prof. A. M. Maphumulo, Dr. N. Sibiya, Dr. B. Zungu, Dr. G. Mazibuko, Dr. S. Ntuli, Dr. L. Makhubu, Dr. P. Dlamini, Mr. Z. Xala, Mr. B. Nxumalo, Mr. B. Mdluli, Ms. D. Msibi, Mr. Makhoba, the subject specialists included Dr. S. Mathe; and the ULPDO staff members.

The standardization workshop forms the last stage in the terminology development processes. The PanSALB national language body (UMZUKAZWE) meets to consider and authenticate the terms submitted by the verification committee. The NLB met to authenticate terminology for Social Work. The authenticated terminology will be stored for dissemination in our two language technology platforms, the isiZulu Term Bank and the Mobile Application. The NLB, sitting with the subject specialists, spent three days thoroughly engaging these terms in a meticulous process that saw 520 terms standardized and ready for dissemination and use.

6. Running Projects

6.1. Bua Le Nna Sessions

The ULPDO successfully held four Bua Le Nna sessions in October, November and December 2017 respectively.



In the picture are the students who participated in this session. The language champions and ULPDO staff members are not in the picture.

The Bua Le Nna (Sesotho for Let's Talk) programme has been proceeding very well since the beginning of 2017, and students continue to embrace it. The office noted the growing excitement from the students in learning the Sesotho language in order to affirm multilingualism as pronounced in the Institutional language policy. This initiative is responsive to the imperative to enhance social cohesion and inclusivity at UKZN, which is important in uniting the university community. This project is one of the flagship projects of the ULPDO, working in close collaboration with student services. Participating students managed to learn conversational skills in Sesotho. Evaluation forms together with registers have indicated a steady increase in attendances at each Bua Le Nna sessions that are held fortnightly at student residences. It is hoped that in 2018 more students will be interested to participate in project as we attempt to spread the project to other campuses.

7. ULPDO HLT's

7.1. Corpus Collection and cleaning

The isiZulu National Corpus (INC) grew substantially in this quarter. The INC increased by 5 735 962 running words from 26 million running words to break the 30 million word mark to a new record size of 31 775 804 tokens. The figure 1 below shows a screenshot of the current INC size.

Figure 1. The isiZulu National Corpus. 31.7 Million words.

	Overall	0649	0650	0650	0651	0652	0653	0654	0655	0656	0657	0658	0659	0660	0661	0662
text file																
file size	257 651 312 113 922 118 340	75 712	97	136	182	80	34	126	133	98	86	95	193	71		
tokens (running words) in text	31 775 804	13 109	13 542	8 734	11	16	22	9 093	4 165	15	15	11	10	10	23	8 404
tokens used for word list	29 893 420	12 333	12 748	8 287	10	15	20	8 655	3 999	14	14	10	9 472	10	21	7 789
sum of entries																
types (distinct words)	3 970 000	7 224	7 471	5 112	6 542	8 923	8 465	5 609	2 584	8 425	7 957	6 511	5 552	6 073	9 122	4 991
type/token ratio (TTR)	13,28	58,57	58,61	61,89	61,57	57,54	42,02	64,81	64,62	57,78	55,33	60,81	58,61	58,20	43,08	64,08
standardised TTR	69,36	72,09	73,65	72,88	73,09	75,78	53,70	75,71	73,57	73,49	72,07	73,58	69,35	72,50	55,42	73,30
STTR std. dev.	30,83	24,91	24,56	24,27	25,25	22,94	37,90	21,96	20,72	25,29	26,20	23,99	28,63	24,54	38,90	23,76
STTR basis	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000
mean word length (in characters)	6,35	6,89	6,87	6,90	6,89	6,82	6,67	7,07	6,56	6,60	6,98	6,93	6,49	7,01	6,66	6,73
word length std. dev.	3,37	3,29	3,20	3,17	3,24	3,09	3,28	3,15	3,23	3,16	3,25	3,16	3,34	3,19	3,26	3,34
sentences	2 322 360	845	840	578	711	1 038	1 489	609	260	890	1 002	743	621	753	1 833	522
mean (in words)	13,05	14,60	15,18	14,34	14,95	14,94	13,53	14,21	15,38	16,38	14,35	14,41	15,25	13,86	11,55	14,92
std. dev.	203,74	12,24	13,84	12,41	13,36	11,37	10,87	12,72	13,28	13,59	12,13	11,45	16,14	11,63	10,30	12,81
paragraphs	240 280	110	82	62	73	71	35	61	19	55	166	43	196	92	31	112
mean (in words)	124,41	112,12	155,46	133,66	14,62	14,95	14,94	13,53	14,21	15,38	16,38	14,35	14,41	15,25	13,86	11,55
std. dev.	1 345,41	268,98	293,43	239,75	28,00	48,00	6,00	1,00	27,00	6,00	29,00	6,00	9,00	22,00	54,00	2,00
headings																
mean (in words)																
std. dev.																
sections	2 385	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
mean (in words)	12 533,93	12	12	8	10	15	20	8	3	14	14	10	9	10	21	7
std. dev.	20 297,20															

The INC is currently the biggest African language corpus, which is an important resource in the development of isiZulu human language technologies. It has been a critical resource in the development of the isiZulu spellchecker version 1 and version 1.1. In order to continue to grow this crucial resource ULPDO continues to unlock new relationships in order to access new corpus materials. To this end the ULPDO is currently engaging the Mazisi Kunene Foundation to create a synergistic relationship that will allow ULPDO to access the isiZulu materials.

7.2 Term Bank, Spellchecker and the Zulu Lexicon

The office is engaged in continuous monitoring and evaluation of the use of technologies that were launch in November 2016. The language technologies are open source, which means they are currently available for free. The

records are indicating that the uptake by end users is indeed growing. So far, 5001 users have visited the isiZulu terminology bank. Over a 1000 and 1500 people have downloaded the ZuluLexicon Mobile application, and the isiZulu spellchecker.

This is indeed evidence that UKZN has become the center of isiZulu development and intellectualization.

Figure 2

People accessing the Term Bank

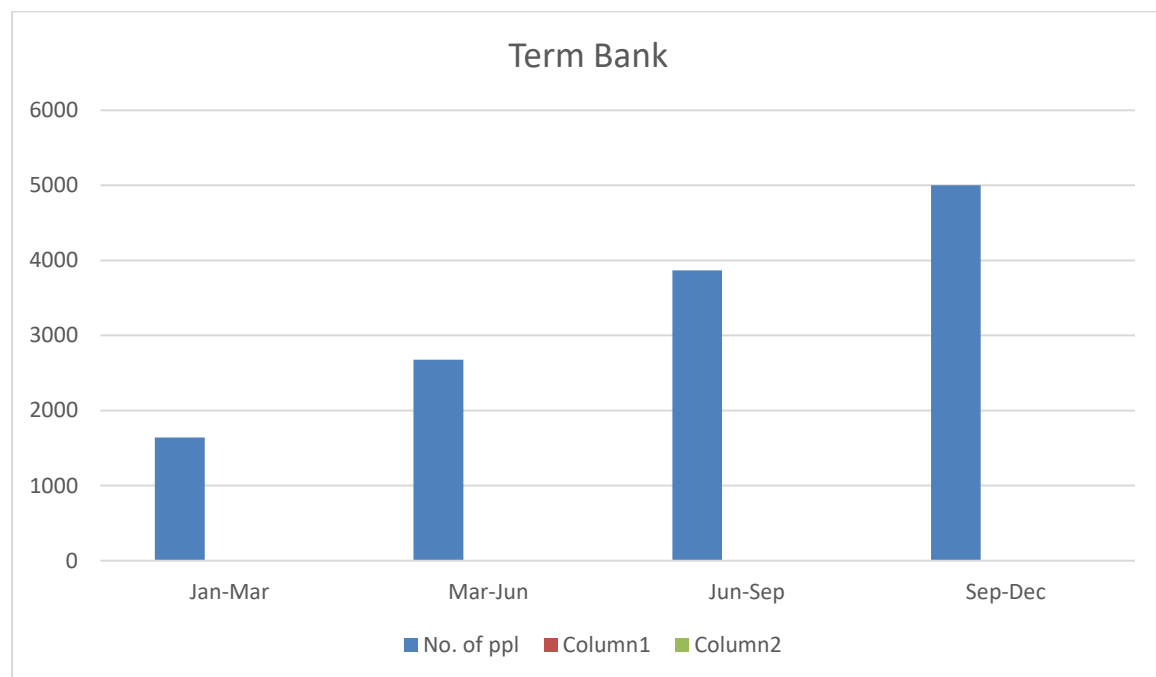


Figure 2 shows that in the first quarter of 2017 only 1642 people had visited the term bank. The number of visitors grew in the second quarter with more than 1000 visitors, which resulted to a total of 2678 people accessing the terms back. The third quarter saw the number grow by a further 1100 to 3869 visitors to the term bank. In the period under review, the number grew by more than 1130 visitors to 5001. This was an impressive positive ending to the year and the expectations are high for 2018.

8. Workshops and Conference attendance

8.1 Director's conference and workshops attendance

The Director of the ULPDO attended the Community of Practice for African Languages (CoPAL) sub-committee meeting responsible for Terminology, Translation and Interpreting. The meeting was held from 2 -3 November 2017 at the University of South

Africa, in Pretoria. The subcommittee is made up of Professor S Madonsela (UNISA), Mr Theo Rodrigues (CPUT) and Dr L Khumalo (UKZN). The committee is responsible for developing portals for sharing multilingual disciplinary terminologies to aid access of these and other digital resources.

The Director also attended the Medical Machine Translation (MeMat) Workshop, from 4 – 8 December 2017 at the University of Cape Town (UCT), in Cape Town. MeMat is a collaborative project between Edinburgh University, UCT and the University of KwaZulu-Natal. The project seeks to develop medical text corpora for the development of automated machine learning in order to ultimately aid patient-doctor interaction.

Some of the delegates at the MeMat Workshop at UCT



9. ULPDO 2017 Language Research Symposium

UKZN's Language Planning and Development Office hosted the second bi-annual Language Research Symposium at Howard College campus on the 25th of October 2017. The theme for this symposium was 'Human Language Technologies and Computer Assisted Language Learning Tools as enablers in the intellectualization of isiZulu at UKZN'. The Symposium provided a platform for academics involved in various language projects across all UKZN to review progress on the implementation of the University's Language Policy and Plan.

Prof Sonja Bosch delivering the first keynote address



Professor Sonja Bosch from the University of South Africa (UNISA) delivered the first keynote address on the intellectualisation of African languages. She emphasized the importance of developing technologies to aid the language intellectualization programme.

Prof Monwabisi Ralarala delivering a keynote address



Professor Monwabisi Knowledge Ralarala, who is the Director for the Fundani Centre for Higher Education Development at Cape Peninsula University of Technology, delivered the second keynote address. He also underscored the importance of developing technological tools to advance the development of African Languages. He made practical examples on tools that have hitherto been developed showing how they have aided the process of intellectualizing the African languages.



A wide range of papers were presented at the Symposium by UKZN academics including five ULPDO staff members. Different media houses attended this event including Ukhozi FM, iLanga newspaper etc.

The symposium was a huge success. This was gleaned from the university media platforms and articulated by the delegates who completed evaluations forms to provide us with feedback.

10. Training workshops

As part of its developmental strand, ULPDO held a successful UKZN/UKHOZI training workshop on the 11th of October 2017 at the UNITE building.

In the picture below are Ukhozi FM participant, Language Experts and ULPDO staff members.



The objective of the workshop was to forge synergistic relations between UKZN and Ukhozi FM particularly on matters of mutual interests. In particular to train Ukhozi FM employees on the proper utilization of isiZulu language, as well as to ensure that the standardized terminology is disseminated through their media, and reaches their large audience. The workshop was attended by the entire staff of uKhozi FM including their senior management team.

The training was facilitated by three leading isiZulu language experts, Dr Mpumelelo Mbatha, Mr Zweli Xala, and Mr Nhlanhla Mtaka. Dr Mbatha focused on the accuracy, richness and history of isiZulu language, Mr Xala dealt with translation techniques from other languages into isiZulu, and Mr Mtaka addressed challenges emanating from code switching and translanguaging between isiZulu and English.

The workshop was evaluated as a huge success. Evaluation conducted by the office at the end of the workshop indicate that delegates really gained a lot from this training and suggested that ULPDO hosts more of such informative workshops. The dissemination of specialized disciplinary terms have already began, since the broadcasters are using the terms and applying the language techniques that they were taught at the workshop.