

Annual report for the Allex project in 1998

NUFU PROJECT NR 28/96

The annual report for The ALLEX Project 1998 was produced jointly in Oslo and Harare by Oddrun Grønvik and Dr. Herbert Chimhundu on the basis of reports, meetings, minutes etc during the year and decisions taken during the ALLEX planning meeting i Oslo 30.11.-4.12.1998. The report was presented for comments to the whole of the ALLEX team in January - February 1999, and all comments are integrated.

A. MAIN ACTIVITIES AND PROGRESS OF THE YEAR

The chief events of the ALLEX Project 1998 were

- a) The 1998 February Workshop (16. - 20.2.1998)
- b) ALLEX Course in computational linguistics at UZ for the UZ ALLEX team 23. - 27.2.1998
- c) Two three-month guest researcher visits to UO 15.4. - 5.7.1998, cf point 4. below
- d) Visit to language centres to Tanzania, South Africa and Malawi by Dr. Jefferies and Mr Ndlovu,
- e) Visit to Afrilex and technical visit to the UZ team (dates) by Dr Ridings
- f) ALLEX UZ internal evaluation seminar 27.-28.7.1998
- g) The first Ph.D. scholar, Edgar Mberi, arrived at UO 15.8.98 to start his sandwich programme.
- h) Technical visit to UZ by Dr. C. E. Ore 17. - 25.10.1998
- i) Planning visit by Dr Herbert Chimhundu to UO 27.11.-4.12.98

For no. a, c, d and f the ALLEX Project has separate reports which are available on request.

1998 was spent in

1. Corpus work
2. Lexicographical work
3. Preparing for lexicography to get institutionalised at UZ
4. Training in lexicography and computational linguistics
5. Competence building
6. Technical development and equipment

1. Corpus Work

Corpus Development and administration:

The size of the corpora has grown steadily in 1998. The importance of constant quality controls at UO under Oddrun Grønvik's direction before the material reaches the corpus administrator, Daniel Ridings in Gothenburg still has to be emphasised.

As yet it is impossible to turn the responsibility for administration over to the ALLEX team in Harare. The tools being used to prepare and organise the corpus require a good understanding of the procedures, and the consequences involved when procedures are deviated from. This takes more training in information technology than the project yet has been able to provide.

In practice this is not a problem hampering project progress, except that corpus material can only be added to the on-line version during visits to Harare by project participants from the Nordic countries. In the near future this problem will be eliminated, as the on-line versions can be updated in Gothenburg or Oslo and still be accessed by team members no matter where they are, since the whole system has been organised around the potential the internet offers. In other words, graduate students and researchers can access their material from Oslo, Harare and Gothenburg on an equal basis.

Certain thresholds in computer literacy must be passed before the practical administration of the system can be turned over to staff in Harare. The aim is to reach this level in Phase 3.

Some advance has been made in the introduction of higher level analysis tools for corpus investigation. A morphological analyser has been built in Gothenburg. Once the level of computer literacy in the UZ team is raised another step, it will be possible to turn over its completion to MA or PhD students. This looks promising and some interest was expressed during the 1998 workshop and reiterated later during the year. To date tangible steps have not been taken.

The corpus organisation and morphological analyser were presented at the AFRILEX conference in July 1998. A request was made for a written version of the presentation from the AFRILEX journal, but was not met. Daniel Ridings feels that it would be more proper to make a concerted effort at documentation together with other team members.

For both corpora the training of editors and M.A. Scholars was done conscientiously, took up a lot of time, but also contributed signally to corpus expansion in 1998. Especially, the work of Ignatius Mabasa at UO and the scanner training offered by Moreblessings Mawema and Karren Dube after their return from UO has given impetus to corpus expansion. The contribution of the M.A.Scholars has been essential to progress for both corpora. Literary contents are listed in Appendix 11.

Shona: 38 books and a number of files of oral material have been added. Some overlooked oral materials were found towards the end of 1998, and are now being encoded. The corpus has been expanded from ca 1,2 mill to 2 mill words in 1998. In 1998 the Shona corpus has come into regular use for the whole team.

Ndebele: The Ndebele Corpus was increased from 325 000 words to 1.2 mill words at the end of 1998; it consists of mostly oral material, but also 7 books-

2. Lexicographical Work

Database Development:

In 1998 there has been a further development of the dictionary databases. To keep the maintenance costs as low as possible, the system is developed to meet the needs for both teams. It is, however, possible to indicate the actual language (Shona or Ndebele) to get the correct meta-vocabulary both in the final printouts and on the screen. The ALLEX'98 workshop resulted in a few minor changes. During the visit of the guest researchers in April to June the database system was adjusted and streamlined towards the headword selection process. This work was continued in close co-operation with the database administrator of the Shona team during the technical visit in October (17/10 - 25/10). The visit was successful and the key members of the teams had a good command of the use of the system. In addition a CD-burner was installed and the database administrator of the Shona team was trained in using this device. The possibility of making CDs has made it possible to make secure backups at frequent and regular intervals. It also has eased the data exchange between Oslo and Harare. The internet connection provided by Samara can be used to send small sized emails, but is useless as a channel for sending data to and from Oslo/Gothenburg and Harare. It will hopefully be replaced by a better with an acceptable band width.

During the visit of dr. Chimhundu in December the preservation of the audio tape with the oral material was discussed. Ore suggested that the tapes should be transferred to CDs. The CD-medium is much more stable than the magnetic tapes and the conversion to digital form make it easier to store, retrieve and study the oral material . The idea was discussed and it was decided that the conversion should be tested in practice during 1999.

Advanced Shona Dictionary (ASD)

Headword collection: The database has been expanded from 16 000 to 27 500 entries. Headwords only are entered for the last 11 500 words added.

Editorial tools: The Style Manual was greatly expanded, and Metalanguage List were revised in 1998.

Editorial formats: A number of trial entries (ca 100) from all word classes have been edited, with a view to expanding the format register from DRC (compounds, idioms, multi-word expressions etc), and the formats have been added to the Style Manual.

Definitions: On the basis of the corpus, a number of new meanings have been added to the entries from DRC.

General Ndebele Dictionary (GND)

Headword collection: 18800 headwords had been entered by the end of 1998. Grammatical information is entered for 12 000.

Editorial tools: Style manual revised and expanded, particularly for defining formats. Metalanguage list completed.

Editorial formats: Roughly 150 trial entries from all important word classes were written by K. Dube during her guest researchership at UO, and discussed in the whole team. These formed the basis of the editorial formats described in the Style Manual for GND.

Definitions: The letters A, B, part of D, I and Z were defined on paper (not entered into the database). Some semantic sets were finished (letters of the alphabet, days of the week, months etc).

3. Preparing for institutionalising lexicography at UZ

Visit to the Norwegian Language Council: The Guest researchers visited the Norwegian Language Council before going back to UZ, to gain information on the administration of language policy issues in Norway, and reported on their findings to the UZ team.

Visit to language centres in Malawi etc: Dr. Jefferies and Mr Ndlovu visited language centres in Malawi, Tanzania and South Africa in May 1998 and gathered information which lead to important recommendations for planning institutionalisation of lexicography in Zimbabwe. A solid report from these visits has proved invaluable in preparing the institutionalisation proposal.

Internal evaluation seminar 27. – 28.7.98: This seminar was organised in UZ campus for the UZ ALLEX teams. It focussed on

- work efficiency within and between the UZ teams,
- adjusting work plans to planned absences because of leave
- institutionalisation (with the Ndlovu/Jefferies report from other language centres in Africa)
- State of affairs at UZ re the ALLEX Project

There is agreement that these internal seminars for the UZ team are important and valuable, as it often is difficult to meet and keep threads together between workshops. After institutionalisation, this kind of gathering will continue to be an important element in ALLEX life.

Proposal for institutionalising Lexicography at UZ: A proposal drafted by the UZ ALLEX team was brought to Oslo for the administration week 30.11 – 4.12.1998. It was then discussed and extensively revised in connection with the 1999 Workshop, and will now be processed through the UZ administrative system with a view to completing the institutionalisation process by the end of 1999. cf Appendix 9.

Impact of ALLEX Project

As an outcome of the ALLEX Project in Zimbabwe, and especially the acclaim that greeted the DRC, a *National Language Advisory Panel* was appointed in 1997 by the Ministry of Education and Culture. Dr. Chimhundu was elected chairman. In 1998, the panel produced *Report on the formulation of a National Language Policy*, which embodies a plan for creating adequate language planning instruments for the indigenous languages of Zimbabwe. In this plan, institutionalisation lexicography at UZ plays a major role.

In 1998, Dr. Chimhundu was also appointed head of the reconstituted *National Arts` Council*, a decision that was greeted with acclaim in the national press.

4. Training in lexicography and computational linguistics

1998 Workshop: 1998 was an important year in ALLEX history, in that the training in lexicography since 1992 really began to bear fruit on a general level. Plainly put, the Scandinavian participants for the first time felt fully confident that what was said in lectures, was understood and responded to as they would expect at their own institutions.

The annual ALLEX Workshops have always been important fora for teaching lexicography and linguistics while focussing on project progress. In 1998 the corpora had reached an acceptable size for concordancing, and the amount of computers and computer training at ALL made the techniques of computational linguistics accessible to all team members. The 1998 ALLEX workshop focussed heavily on basic analysis of concordance materials and criteria for extracting and differentiating lemma and lexemes.

Course in Corpus Linguistics: The Workshop was followed by a concentrated course in corpus linguistics for the UZ team members. The course content corresponded to what UO or GU would offer in one term, with a preset syllabus of about 300 pages, and with exercises and presentation of results by ALLEX UZ team members at the end. The Ph.D. candidates who had registered at the UZ by the end of 1997, got an opportunity to present their thesis projects and demonstrate some newly learnt techniques.

The course in corpus linguistics (taught by Pernilla Danielsson and Daniel Ridings, GU) was very successful. They were asked whether they would be prepared to hold the same course under a longer period of time, and at a slower pace, in the future. This was answered to the affirmative but would require some advance planning to follow through.

5. Competence building

The ALLEX Project in 1998 had 2 guest researchers, 3 Ph.D. scholars, one of whom started his sandwich course at UO, and three M.A. scholars at UZ.

Guest researchers: Both women, one from the Shona team and one from the Ndebele team, spent 15.4.-5.7.1998 at INL/UO as guest researchers in 1998. Both reported in writing on the work done and the quality of their stay.

Moreblessings Mawema :

Corpus work: parsing and correcting 95 files of oral material, and scanning two books, all added to the corpus

Database: working regularly with database and developing format in co-operation with Dr. Ore.

Lexicography and semantics: headword selection (ca 3000 words added), classification, description and defining of words, compounds and multi-word units. Review of editorial tools (Style manual, metalanguage list).

Karren Dube:

Corpus work: 20 000 words oral material prepared and added to corpus; scanning and tagging several books

Database: adding grammatical information to all headwords (ca 14 000 at the time), clearing it of inconsistencies, practising use of database with Dr. Ore;

Lexicography and semantics: training in analysing language for isolating and describing meaning in definitions and settling definition formats for GND. 150 trial definitions completed.

Edgar Mberi, the first of three Ph.D. scholars (all registered at UZ in 1997) arrived in Oslo to start his sandwich course 15.8.98. He was well received and settled in very quickly. He presented his project at a staff seminar at INL/UO 16.10.1998 and did well. Although the Arts Faculty UO does not offer courses at the Ph.D. level in English, the Faculty Linguistics seminar for the spring term of 1999 has lexicography and linguistics as a theme and will as far as possible be taught in English.

M. A. Students: One of the 3 1997-98 M.A. scholars interrupted his training at the end of 1997. The other two,

Gift Mheta (Shona) and Finex Ndlovu (Ndebele) completed their theses and course work as planned, did well, and were also of great assistance to the ALLEX UZ team through their part time work on the corpus.

6. Technical development

In addition to corpus and database development, the following developments took place:

The internet: In 1998 The ALLEX Project at UZ were allocated a phone line and signed up for internet services from Samara. Samara's follow up was not good enough and in addition there were considerable IT problems affecting the whole of UZ. Although local email was a great advance, we were not able to transfer files of any size. For 1999, a better service provider will be sought, in contact with the UZ computer centre.

ALLEX homepage: In 1998, the ALLEX project established a home page from UO (<http://www.dokpro.uio.no/allex/allex.html>) which in time will have all information in all languages involved. From it one can access the Duramazwi reChiShona and search in it (<http://www.dokpro.uio.no/allex/gsd.html>). The Corpora are accessible to team members over the net (<http://ldb20.svenska.gu.se/>). Information on the net about the corpora has been presented by Dr Ridings (<http://svenska.gu.se/~ridings/allex.html>).

B. DEVIATIONS BETWEEN THE PLANNED ACTIVITIES (ANNUAL WORK PLAN) AND PROGRESS OF THE YEAR

The work plan 1996-99 has been adhered to in all essentials, including the planned extension into the year 2000.

C. STATUS AND DEVIATIONS IN RELATION TO THE CO-OPERATION'S MAIN AND SPECIFIC OBJECTIVES

There are no deviations under this point.

D. PROPOSAL TO NUFU OF MAJOR CHANGES

The ALLEX Project has no proposals of major changes.

E. MAJOR INVESTMENTS IN 1998

There were no major investments in 1998.

F. MAJOR OUTSIDE FACTORS AND INPUTS

The situation at UZ in 1998: The UZ was closed to B. A. students from May 1998 and until the end of the year. M.A. students and Ph. D. candidates were allowed to carry on with their regular activities, as was the UZ academic staff. This situation was in the short term beneficial to the ALLEX Project because it gave the UZ team members time and peace to work on the project. However, it means that 1999 will be a very difficult working year, with a double intake of students.

Communication (email, telephone etc): Communication was not improved in the ways we hoped in 1998, partly because of poor internet service, but mainly because of very poor general communication facilities at UZ, in addition to short term difficulties.

The situation at UO in 1998: In 1998, UO has had to cut its budget and plan for tighter budgets in the future. This endangers project co-operation such as NUFU because the institutional cost (overhead) at department or faculty level is not compensated. These problems will have to be addressed in preparation for phase 3 of the NUFU programme.

In 1998, the Arts Faculty at UO hesitated to register the ALLEX Ph.D. candidates because of the lack of a research training programme in English. This matter was satisfactorily addressed early in 1999.

G. ORGANISATIONAL OR ADMINISTRATIVE CHANGES IN THE ALLEX PROJECT

Institutionalising co-operation with GU: Throughout the ALLEX Project, two Nordic universities have been active in the co-operation, UO and GU. A co-operation agreement has been entered into by the Arts Faculties of UO and GU, an important factor in it being the co-operation within the ALLEX Project.

In the ALLEX planning meeting 1998, an Amendment to the NUFU Programme Agreement was drawn up and later signed by all three co-operating departments (cf appendix 10). It formalises current practice in the co-operation, namely that there are three equal partners behind the ALLEX Project for the rest of Phase 2. A corresponding arrangement is planned for Phase 3.

As part of the formalisation process, Oddrun Grønvik and Dr. Ore were present during a contact visit by a delegation from GU headed by their Dean of Arts Faculty Dr. xx to the Dean to Arts Faculty at UO Dr. Even Hovdhaugen, 22.12.98.

Administrative assistance at UO: In 1998 M.A. scholar Ignatius Mabasa assisted the UO coordinator with general administration and with corpus expansion and quality control. He also started constructing an ALLEX home page for Internet. Mr. Mabasa proved a great asset to the ALLEX project in every respect.

INL/UO has made available an office for the ALLEX Project throughout 1998 and provided administrative services, especially for finance administration. Relations have been very good. Co-operation with the International department at UO has also been excellent, and the support provided invaluable to project efficiency.

UZ support: Has been as before, i.e. a reasonable amount of support granted the difficult situation at UZ.

The NUFU coordinator at UZ: UO employs Mr Alexio Gumi at the UZ to look after the NUFU Projects and assist the academic co-ordinators at UZ. While Mr Gumi has always been friendly and helpful, 1998 was a breakthrough year in relation to the ALLEX project. He saw to it that the team was warned when funds were running out, helped trace misplaced transfers and assisted the ALLEX team signally in keeping exact accounts and preparing a realistic 1999 budget. Thanks to him, project administration was significantly improved in 1999.

The ALLEX team at UZ understood right from the start the importance of building a project administration within the department, and of documenting all work and decisions through proper reporting procedures. Without this administrative effort, the team could not have planned for institutionalisation by the end of 1999.

H. FUNDING FROM OTHER SOURCES

The UZ Research Council: Although the UZ co-ordinator again applied to the UZ Research Council for additional funding, the application was turned down. It is argued that the ALLEX Project already is well off. The argument that UZ is meant to contribute, and that NUFU does not offer full coverage, is not regarded as decisive.

The Quota Programme at UO: Ph.D. Candidate Edgar Mberi was included in UO Quota programme for the academic year 1998-99. He will have additional funding from the ALLEX programme.

GU participation: Although this arrangement was formalised only at the end of 1998, Dr. Ridings and other GU team members (Pernilla Danielsson, Christian Sjögren) have been part of the ALLEX Project in the whole of phase 2, with their own funding from SIDA. The GU team members have their own account with SIDA.

I. COMMENTS TO FINDINGS IN NUFU EVALUATION

Along with other NUFU Projects in Zimbabwe and Botswana, the ALLEX Project was evaluated in the summer of 1998, cf NUFU reports covering the UiN side and the UiD side separately. The Report covering UiN does not say much about individual projects. The summary is fair enough, but the report does not help much towards an understanding of what it takes to plan and implement a good project.

The Report covering NUFU projects in Botswana and Zimbabwe discusses each project in depth. The UZ team took the evaluation process very seriously and prepared for giving as correct a picture of the ALLEX Project as they could. They feel that the project was perceived and appreciated on its merits, and are grateful for the work done by the reporters.

J. ANY OTHER FACTORS

All factors of relevance have been mentioned above.