

Annual report for 1997  
the Alex project

NUFU PROJECT NR 28/96

The annual report for The ALLEX Project 1996 was produced jointly in Oslo and Harare by Oddrun Grønvik and Dr. Herbert Chimhundu on the basis of reports, minutes etc during the year and decisions taken during the ALLEX planning meeting i Oslo 20-24.11.97.

#### MAIN ACTIVITIES AND PROGRESS OF THE YEAR

1997 was spent in

- a) creating viable corpora for Shona and Ndebele
- b) starting work on the Advanced Shona Dictionary (ASD) and the General Ndebele Dictionary (GND)
- c) preparing for lexicography to get institutionalized at UZ

- a) Creating viable corpora for Shona and Ndebele

During 1997 the corpora for Shona and Ndebele were converted to TEI standard, which is the current international standard for text exchange over the network. This involved a great deal of learning and hard work for both teams, but also means that the corpora now have a high standard and can, when the time is ready, be made accessible internationally over the network. This should give a big boost to research in Shona and Ndebele not only in Zimbabwe, but in all places where Languages from Southern Africa are studied.

#### SHONA:

All of the oral material collected in 1993 - is now encoded and tagged. This material is unique, irreplaceable and extremely valuable. Proper safeguarding of this national cultural resource has to be a UZ task in 1998. About twenty books, half novels, half textbooks, have been scanned and tagged and added to the corpus.

#### NDEBELE:

Extensive oral collections were carried out in 1997, despite practical problems for the fieldworkers (floods, malaria etc). These materials are now being encoded and tagged.

Five books (one play, the rest novels) have been scanned and tagged.

In all, the corpora now have a size of

This is enough to start editing the two planned dictionaries from, though the corpora are to be added to continuously.

- b) Starting work on the Advanced Shona Dictionary (ASD) and the General Ndebele Dictionary (GND)

It must be mentioned, however, that the administrative burden on the UiO Academic co-ordinator is far in excess of what was expected, and leaves very little time within the time allotted to the Project for lexicography proper. This is unsatisfactory and will have to be dealt with, cf point G below.

- c) Preparing the lexicographical work to be done in phase 2

The ALLEX team has now been expanded with an Ndebele editorial team.

The Ndebele team have started preparing their field work (excursion to Bulawayo and surrounding area September 1996) and the editorial apparatus (style manual etc) for the General Ndebele Dictionary. They receive constant support from the Shona team, a member of which is present for consultation at all Ndebele team meetings.

There is a feeling that the Ndebele team is not getting off the ground with the same panache that the Shona team did in 1992. One reason could be a present shortage of trained linguists in the Ndebele team. We are dealing with this in recruiting amongst the M.A. students.

The Shona editorial team spent the autumn of 1996 getting their corpus ready for conversion to international standards (TEI). They have started preparing the style manual and project design for the Advanced Shona dictionary. They have also corrected the database for Duramazwi reChiShona, so that it now contains all corrections made in during proof-reading in the spring of 1996.

Both teams have received intensive instruction in the use of computers, in order to free the project from the restriction of needing secretaries (copy typists). The administrative staff now assists in project management and follow-up.

Mr. S. Chibanda of the Economics department at the UZ taught the computer courses in the autumn of 1996, and has agreed to be local IT consultant to the UZ team for the time being, hopefully to the end of phase 2.

d) Guest researchers in 1996 (completion of phase 1)

Moreblessings Chitauro 15.1.-9.3,  
Francis Matambirofa 15.1.-5.3

e) Training of technical staff

Training of project assistants by Dr. Daniel Ridings during June workshop (17.-21.6.) in addition to running contact via email throughout the year between Harare team and Dr C. E. Ore and Dr. D. Ridings.

f) Technical development and equipment

Machinery for a small project network (15 IP pcs and a UNIX server) was bought in the autumn of 1996, and the team moved on to the best of current commercial software for work and communications.

New, tailor-made (corpus and database) software was developed and adapted in 1996 by Dr. Ore and Dr. Ridings for phase 2, and is now being introduced. We will report on its efficacy in the annual report for 1997. This software is essential for the success of phase 2, and will also benefit the lexicographical units at UiO and GU.

## B. DEVIATIONS BETWEEN THE PLANNED ACTIVITIES (ANNUAL WORKPLAN) AND PROGRESS OF THE YEAR

The workplan 1996-99 is going to be adhered to as concerns all main objectives. Everything is delayed by six months or so in comparison with the workplans put forward in the application, in which we supposed we would be able to start phase 2 in earnest from mid 1996.

## C. MAJOR INVESTMENTS IN 1996

11 Intel Pentium pc-s , one of which is a UNIX server , linked in a project network within ALL/UZ.

## D. PROPOSAL TO NUFU OF MAJOR CHANGES

The ALLEX Project have no proposals of major changes.

## E. INFORMATION ABOUT FUNDING FROM OTHER SOURCES

The ALLEX project has not received funding from other sources in 1996.

## F. MAJOR OUTSIDE FACTORS AND INPUTS

We planned phase 2 of the ALLEX project on the assumption that the UZ and especially the UZ Research Board would continue to support activities not directly involved in but supportive of and essential to the ALLEX Project. In 1996, no such support has been received owing to financial and administrative difficulties at the UZ. Decision-taking on the long-term future for lexicography at the African languages department has also been slowed down for this reason, but the proper procedures will be started in 1997.

## G. ORGANISATIONAL OR ADMINISTRATIVE CHANGES IN THE ALLEX PROJECT

The only changes of importance are those following from the introduction of the Programme Agreement. However, those are considerable.

While the project partners now have better control and can run things more efficiently, the amount of administrative work has increased enormously. Indeed, after the new administrative arrangements for NUFU projects were implemented in 1996, project administration turns out to take up so much time for the Norwegian academic co-ordinator that it seriously affects the good will towards the programme in the Department of Scandinavian Studies at the University of Oslo. Our experience is that the administrative support granted in phase 1 has been largely withdrawn, while the project has grown in size and complexity. The Department of Scandinavian Studies deals with the ALLEX account and everything to do with it, but cannot otherwise free administrative staff to step into the breach.

This could turn out to become a serious problem later in phase 2.

## I. WORK PLAN AND SCHEDULE FOR 1997

Technical visit from UiO/GU in autumn of 1997, if necessary.  
Technical training of UZ team member (preferably Sihle Moyo) in autumn of 1997.

### a) Work plan Shona team 1997

#### LEARN

- \* to use Filemaker Pro, the corpus programme, Windows 95, Word 7
- \* to convert and polish up corpus
- \* to add to the corpus
- \* to use the corpus as lexicographical tool (source of headwords, forms, meanings, examples ...)
- \* more Ndebele

#### TASKS

Make style manual ASD, cf comments on editing format

Decide roughly on dictionary size

- \* in pages: not much more than 1000, if one volume
- \* in number of entries: aim for 60 000 (add 45 000 from left out ms to GSD, Hannan, corpus)
- \* in text: current format ca 35 letters x 50 lines x 2 columns x 500 pages = 1,75 mill points.

NOB:  $65 \times 75 \times 2 \times 1000 = \text{ca } 9,5 \text{ mill}$

Collins Cb  $50 \times 70 \times 2 \times 1000 = 7 \text{ mill}$

$9\,500\,000 : 60\,000 = 160$      $9\,500\,000 : 40\,000 = 240 \text{ (ca)}$

$7\,000\,000 : 60\,000 = 120 \text{ (ca)}$

$1\,750\,000 : 16\,000 = 110 \text{ (ca)}$

Expand corpus (how much?)

Shona oral material - how much done, in what state, how much left?

Locate sources of additional vocabulary

Corpus sources:

- \* ZBC (on disk)
  - \* Weekly paper in Shona computerised from October 97
  - \* Publishers (on disk) etc
  - \* Own input file for each editor?
  - \* Scanning books that were selected earlier and that have not been dealt with Get sources (on paper) and enter into corpus (who and how much)
- (cf appointment with Longman's 15.2.97 to come and get Shona materials on disks - contact Mrs Nyamakura, tel 621661/7)

Select additional headwords (put criteria in Style Manual)

Enter additional headwords with grammatical information into database

45 000: 6 = 7500 new headwords to be selected and entered pr editor. With 40 working weeks, that is ca 200 headwords (including grammatical information) pr editor pr week.

## ORGANIZE WORK AND ALLOCATE TASKS

The databases:

There will be three, one each for GSD (Duramazwi reChiShona), ASD and GND

One or two database administrators? Preferably one. Engage Stan Chibanda until Sihle Moyo comes back?  
Others? The team as a whole must choose/appoint a database administrator

The corpora:

There will be two, one for Shona and one for Ndebele. Editors must expect to take part in working on corpus expansion (especially proofreading).

The team must have (make?) system for keeping manuscripts, tapes, books and disks with corpus content in order and safely stored. One or two keepers of non-electronic materials? All additions must be recorded, and the record should be accessible to all (file on the network?)

The team as a whole must make system for :

- \* taking out only own bit of files for working on
- \* checking on each others' manuscripts from database (make notes and let editor in question do actual corrections)
- \* organizing meetings, excursions etc (from editorial meetings and brief visits to workshops and field trips for many people)
- \* making and keeping records of activities
- \* writing reports (progress, annual etc - only to the extent they benefit the project)
- \* planning work in relation to personnel absences (staff on research and technical visits)
- \* participation and presentations at conferences
- \* distribution of information (to team here and abroad, and other contacts, such as professional readers, field contacts etc)

Personell:

The ALLEX team as a whole must settle what the project assistants are to do, so that both editorial teams get an equal amount of support. Emphasis on tasks of joint importance? Administration of corpora and databases, other general administration. Reach agreement within total ALLEX team and within department.

Who should liaise with Stan Chibanda?

2-3 additional MAs from March 97 - both Shona and Ndebele users. Think about candidates.

Arrange for Miss Mpofo and Mr. Chapata (present ALLEX M.A. scholars) to stay on as trainee editors from June 1997 onwards.

## TIME SCHEDULE

The ASD Style Manual will be reworked by 28.2.97

The editors will start selecting headwords by 1.3.97

The editors will start entering headwords + grammatical information by 1.5.97

Status pr month - list up in ALLEX room?

All headwords should be entered with grammatical information by the end of 1997.

Additions to corpus:

Show status on list in ALLEX room every month (titles/oral text signatures + rough number of words)

The corpus should increase by ca 80-100 000 words pr month (= 3 average novels of 200 pages each)

The corpus should have reached a size of minimum 2 mill words (originally the aim for phase 1), preferably twice as much. The emphasis should be on printed material, preferably in electronic form, but it still has to be tagged and checked (if text quality is high, the proofreading is less important).

Editorial workshop (out of UZ campus) in the autumn, to assess progress with selecting headwords, discuss linguistic problems arising out of this task and polish Style Manual ASD. Likewise prepare more differentiated definition formats.

Guest Researcher for 3-month stay at AL/UiO: Esau Mangoya

D.Phil. Candidate: Edgar Mberi (cf form)

b) Work plan Ndebele team 1997

## LEARN

- \* to use Filemaker Pro, the corpus programme, Windows 95, Word 7
- \* to convert and polish up corpus
- \* to add to the corpus

\* to use the corpus as lexicographical tool (source of headwords, forms, meanings, examples ..)

## TASKS

Make style manual GND, cf comments on editing format

Decide roughly on dictionary size

\* in pages: not much more than 500, + front and back matter

\* in number of entries: aim for 25 000 and trim to 20 000

\* in text: current format ca 35 letters x 50 lines x 2 columns x 500 pages = 1,75 mill points.

1 750 000 : 20 000 = 90 (ca), i. e. 3 lines pr entry on average

Build corpus (how much - aim at 2 mill words?)

Complete field work. 30 students with in all 450 tapes in the field. Plan for keying in transcribed tapes? and key in. Learn from Shona team. Use new corpus format

Locate sources of vocabulary

Corpus sources:

\* former Ndebele dictionaries

\* Oral material from field work

\* ZBC (on disk)

\* Publishers (on paper and disk) etc

\* Scanning selected books

\* Own input file for each editor?

Get sources (on paper, disk tape etc) and enter into corpus (who and how much)

Select headwords (put criteria in Style Manual)

Enter headwords with grammatical information into database

25 000: 6 ? = 4000 headwords to be selected and entered pr editor. With 40 working weeks, that is ca 100 headwords (including grammatical information) pr editor pr week.

## ORGANIZE WORK AND ALLOCATE TASKS

The databases:

There will be three, one each for GSD (Duramazwi reChiShona), ASD and GND

One or two database administrators? Preferably one. Engage Stan Chibanda until Sihle Moyo comes back?

Others? The ALLEX team as a whole must choose/appoint a database administrator

The corpora:

There will be two, one for Shona and one for Ndebele. Editors must expect to take part in working on corpus expansion (especially proofreading).

The team must have (make?) system for keeping manuscripts, tapes, books and disks with corpus content in order and safely stored. One or two keepers of non-electronic materials? All additions must be recorded, and the record should be accessible to all (file on the network?)

:

The team as a whole must make and learn and practice a system for :

\* taking out only own bit of files for working on

\* checking on each others' manuscripts from database (make notes and let editor in question do actual corrections)

\* organizing meetings, excursions etc (from editorial meetings and brief visits to workshops and field trips for many people)

\* making and keeping records of activities

\* writing reports (progress, annual etc - only to the extent they benefit the project)

\* planning work in relation to personnel absences (staff on research and technical visits)

\* participation and presentations at conferences

\* distribution of information (to team here and abroad, and other contacts, such as professional readers, field contacts etc)

Personell:

The team as a whole must settle what the project assistants are to do, so that both editorial teams get an equal amount of support. Emphasis on tasks of joint importance? Administration of corpora and databases, other general administration. Reach agreement within total ALLEX team and within department.

Who should liaise with Stan Chibanda?

2-3 additional MAs from March 97 - both Shona and Ndebele users. Think about candidates.

Arrange for Mr Kholomu to join team (stand in for Mr. Hadebe + trainee editorship) from March 1997.

## TIME SCHEDULE

The GND Style Manual will be reworked by 28.2.97

The editors will start selecting headwords by 1.3.97

The editors will start entering headwords + grammatical information by 1.5.97

Status pr month - list up in ALLEX room?

All headwords should be entered with grammatical information by the end of 1997.

Additions to corpus:

Show status on list in ALLEX room every month (titles/oral text signatures + rough number of words)

The corpus should increase by ca 200 000 words pr month (= 7 average novels of 200 pages each)

The corpus should reach a size of 2 mill words. The emphasis should be on getting a fair sample of the oral material entered. Printed material, in electronic form or scanned, is less work, but it still has to be tagged and checked (if text quality is high, the proofreading is less important).

Editorial workshop (out of UZ campus) in the autumn, to assess progress with selecting headwords, discuss linguistic problems arising out of this task and polish Style Manual GND. Likewise prepare definition formats.

Guest Researcher for 3-month stay at AL/UiO: S. Hadebe.

Publications, presentations etc:

For publications, etc from the UZ team members see attached copies of page 11.

Oddrun Grønvik, UiO, has been interviewed a number of times by papers, NRK and has presented the ALLEX project in a television programme about Ivar Aasen (by NRK MøRo, sent 29.9.96).

Lectures:

Studentmållaget i Oslo 14.2.96 Ivar Aasen går att i Zimbabwe

Sommarleiren til Norsk Målungdom i Ørsta 5.8.96 Ivar Aasen i og morsmålsleksikografi i Zimbabwe

Chaired session at the international Ivar Aasen symposium 16.11.96, where Dr. Chimhundu presented a paper on language standardization in Zimbabwe.