

**RESEARCH UNIT FOR EXPERIMENTAL PHONOLOGY**

**UNIVERSITY OF STELLENBOSCH (RUEPUS)**

ANNUAL REPORT: 1998 / 1999

<b>1</b>	<b>BACKGROUND</b>	<b>4</b>
<b>1.1</b>	<b>MISSION AND VISION</b>	<b>4</b>
<b>1.2</b>	<b>LONG TERM GOALS</b>	<b>4</b>
<b>2</b>	<b>PROGRESS REPORT</b>	<b>4</b>
<b>2.1</b>	<b>ACHIEVEMENT OF GOALS</b>	<b>4</b>
<b>2.2</b>	<b>COMPLETED PROJECTS</b>	<b>6</b>
2.2.1	ASPECTS OF PHONETIC AND PHONOLOGICAL INTERFERENCE IN XHOSA SPEECH COMMUNICATION.	6
<b>2.3</b>	<b>CURRENT PROJECTS</b>	<b>6</b>
2.3.1	PROGRAMME 1: LINGUISTIC PROJECTS	7
2.3.1.1	The production and perception of statement and question intonation in Xhosa	7
2.3.1.2	The phonetics and phonology of click articulations in Nguni	7
2.3.2	PROGRAMME 2: RESOURCE DEVELOPMENT AND CAPACITY BUILDING PROJECTS	8
2.3.2.1	Preferred syllable structures in Nguni and Sotho	8
2.3.2.2	SASPEECH	8
2.3.2.3	Development of computer software	9
2.3.2.4	Research capacity building projects	9
2.3.3	PROGRAMME 3: LANGUAGE AND SPEECH TECHNOLOGY PROJECTS	10
2.3.3.1	Modelling the intonation of imperatives in a Xhosa text-to-speech system (TTS)	10
2.3.3.2	Speech recognition in African languages (DACST 1998 proposal)	10
2.3.3.3	Language Engineering (LE): Principles and Practice	10
<b>2.4</b>	<b>PUBLICATIONS AND CONFERENCE PRESENTATIONS</b>	<b>11</b>
2.4.1	PAPERS PUBLISHED IN (OR SUBMITTED TO) ACCREDITED JOURNALS	11
2.4.2	PUBLICATIONS IN NATIONAL AND INTERNATIONAL CONFERENCE PROCEEDINGS	11
2.4.3	CHAPTERS IN BOOKS	11
2.4.4	BOOKS	12
2.4.5	PAPERS READ AT INTERNATIONAL CONFERENCES	12
2.4.6	PAPERS READ AT NATIONAL CONFERENCES	12
<b>2.5</b>	<b>STAFF ACTIVITIES</b>	<b>12</b>
2.5.1	VISITS ABROAD	12
2.5.2	VISITORS	13
2.5.2.1	Prof E Wolff. Afrikanistik, University of Leipzig, Germany	13
<b>2.6</b>	<b>CONFERENCES AND SYMPOSIA PRESENTED</b>	<b>13</b>
2.6.1	TWO DAY WORKSHOP ON LANGUAGE ENGINEERING: 10 <sup>TH</sup> INTERNATIONAL CONFERENCE OF THE AFRICAN LANGUAGE ASSOCIATION (ALASA), UNIVERSITY OF SOUTH AFRICA, PRETORIA, JULY 1999	13
2.6.2	COLLOQUIUM ON EXPERIMENTAL SPEECH RESEARCH, AT THE UNIVERSITY OF THE NORTH, QWA-QWA BRANCH, PHUTHADITJHABA, JUNE 1999.	13
<b>2.7</b>	<b>IMPLEMENTATION OF RESEARCH</b>	<b>13</b>
<b>3</b>	<b>INFRASTRUCTURE</b>	<b>13</b>
<b>3.1</b>	<b>STAFF</b>	<b>13</b>
<b>3.2</b>	<b>RESOURCES</b>	<b>14</b>
<b>3.3</b>	<b>EQUIPMENT</b>	<b>14</b>
<b>4</b>	<b>PLANNING FOR NEXT YEAR (2000/2001 - YEAR 3 OF CYCLE 2)</b>	<b>14</b>
<b>4.1</b>	<b>PROGRAMME 1: LINGUISTIC PROJECTS: THE PHONETICS - PHONOLOGY INTERFACE</b>	<b>15</b>
4.1.1	THE PHONETICS AND PHONOLOGY OF CLICK ARTICULATIONS IN NGUNI (CONTINUATION)	15
<b>4.2</b>	<b>PROGRAMME 2: RESOURCE DEVELOPMENT PROJECTS</b>	<b>15</b>
4.2.1	PREFERRED SYLLABLE STRUCTURES IN NGUNI AND SOTHO (CONTINUATION) - TO BE EXPANDED TO "SALText"	15
4.2.2	SASPEECH (CONTINUATION)	15
4.2.3	TEXT CORPORA: PARSING TOOLS FOR THE AFRICAN LANGUAGES (ON HOLD)	16
<b>4.3</b>	<b>PROGRAMME 3: LANGUAGE AND SPEECH TECHNOLOGY PROJECTS</b>	<b>16</b>
4.3.1	LANGUAGE ENGINEERING: PRINCIPLES AND PRACTICE (CONTINUATION)	16
4.3.2	DACST 2000 (APPENDIX A)	16
<b>4.4</b>	<b>STAFF</b>	<b>18</b>

<b>5</b>	<b>BUDGET</b>	<b>18</b>
<b>5.1</b>	<b>STATEMENT OF INCOME AND EXPENDITURE FOR THE PREVIOUS FINANCIAL YEAR (1998/9)</b>	<b>18</b>
<b>5.2</b>	<b>BUDGET FOR NEXT FINANCIAL YEAR (2000/2001)</b>	<b>19</b>
<b>6</b>	<b>GENERAL</b>	<b>19</b>

## 1 BACKGROUND

This progress report covers activities of the **first year** of the second cycle of the Unit, i.e. activities that have taken place in the financial year ending **March 1999**. For the sake of a wider perspective, especially in view of the landing of a major project that will affect all future activities dramatically, activities related to the period April 1999 to October 1999 will also be mentioned. The report will be presented in the prescribed format for annual progress reports as specified by the CSD.

Following specific suggestions presented to and accepted by the Advisory Board at its annual meeting in 1997 as well as in 1998 the focus of RUEPUS has been redirected towards research and development in the field of language and speech technology applications with specific reference to the indigenous languages of South Africa. (Please consult previous reports in this regard.) This led to the acceptance of a new strategic plan that is taken into account in the drafting of this report.

### 1.1 Mission and vision

The **mission** of RUEPUS remains unchanged as described in the 1997/8 report, i.e. as the active promotion of the official languages of South Africa at **technological level** through research, facilitation and development activities with a focus on the African languages and the speakers of these languages.

The **vision** upheld by RUEPUS is that all languages spoken in South Africa, and more specifically the African languages, be *developed at technological level* to such an extent that

- They will meet the challenges of a technologically driven next millennium;
- A *new generation of language engineers* will be created which will play a central role in the development of language based applications which will infiltrate all walks of life in the next century;
- A technology driven *national resource facility* making provision for speech and text corpora will be developed and maintained. (See pp 1-2 of 1997/98 Report).

### 1.2 Long term goals

The strategic plan discussed and accepted by the Advisory Board in 1998 envisages that through focused activities RUEPUS will be functional in setting up a new, financially self sustaining institution with exclusive focus on **language engineering** (LE) within the African languages. RUEPUS should eventually become part of this new institution in the final year of its second cycle (i.e. in 2002). This new centre, the **(South) African Language Engineering Centre ((S)ALEC)**, will operate on a national scale to the benefit of all potential researchers in LE in the (South) African languages bringing together resources from the academic community with financial input from business and industry and government according to a model similar to that implemented by the European Union.

## 2 PROGRESS REPORT

### 2.1 Achievement of goals

The goals set by RUEPUS reflect on

- **Academic activities** (the running of research projects and the dissemination of research results through publications and conference presentations),
- **Networking** (making contact with colleagues at national and international level, *inter alia* by organising conferences/seminars and/or participating in other conferences/seminars)
- **Research capacity building** (at graduate and post graduate levels as part of formal academic courses; through dedicated research capacity building projects; through specialised workshops)
- **Fund raising** to create a self-sustainable infrastructure conducive to research and development.

In assessing the achievement of the **academic aims** it may be stated that a substantial number of publications (nine plus one in press) have seen the light during this period, and research results have been disseminated through five presentations at international conferences and three at national conferences as specified below.

As far as **networking** is concerned the overall visibility of RUEPUS at national as well as at international level was maintained through conference participation and international projects. This allowed for good interaction between RUEPUS staff members and other academics of whom a large number hailed from historically disadvantaged communities (at national conferences). International research co-operation was strengthened with the award of a Feodor Lynen Stipend of the Alexander von Humboldt Stiftung to Dr Michael Jessen of the Institute for Computational Linguistics at the University of Stuttgart, which enabled him to spend nine months with RUEPUS during this period.

Good progress has been made in the field of **research capacity building (RCB)** as a new initiative with the University of the North (Qwa-Qwa branch) was launched, as well as the presentation of a workshop at an international conference. (Reports in the appropriate sections below).

**Fund raising** for research and development purposes has been extremely successful following the Board's 1998 recommendation that the Director could be allowed to spend more time on this issue. At that point in time the Board took note of the fact that a proposal to the **Innovation Fund** of the Department of Arts, Culture, Science and Technology (DACST) by the University of Pretoria (Dept of Electronic Engineering with RUEPUS as a consortium member) had been unsuccessful and that no funds would be forthcoming from that source. When the second round for submission of proposals to the Innovation Fund (1999-2002) was announced, the Director took it upon him to redraft the proposal, to focus more on language issues, to set up a new consortium (including a historically disadvantaged university, and two new members from the private sector) and to submit the proposal in March 1999. In September 1999 it was announced that this proposal had been successful and that an amount of R9.08 million had been awarded to the project entitled *Promoting the development of the official languages of South Africa through language and speech technology applications*.<sup>1</sup>

This project, which will be described in more detail below, (4.3.2 and Appendix A) will obviously have a profound effect on the nature of the projects and activities of RUEPUS in its final years up till 2002.

---

<sup>1</sup> Board members are please requested to treat this information as confidential until an official announcement has been made by the Minister of Arts, Science, Culture and Technology in November 1999. This project will informally be referred to as the **DACST 2000 project**.

## 2.2 Completed Projects

### 2.2.1 Aspects of phonetic and phonological interference in Xhosa speech communication.

**PW Lewis, JC Roux & G Dogil**

For a full description of this project - please see the Annual Report 1996/7

This project makes a significant contribution to an understanding of the so-called "difficult" and "easy" sounds of a language such as Xhosa, both from the views of non-mother tongue speakers and mother tongue listeners. The results have implications for intercultural communication, language learning and the teaching of phonetics, as well as for speech technology as sets of very valuable acoustic and perceptual data have been made available for technology development.

#### Dissemination of results:

##### Doctoral thesis:

- **Lewis, P 1998.** *Phonetic aspects of Xhosa second language speech communication*, DLitt, University of Stellenbosch, pp 380

##### Article in accredited journal

- **Lewis, P & Roux, JC 1998** On Xhosa L2 speech and intelligibility: vowels and clicks. *South African Journal of Linguistics*, Supplement 36, 51-63.

##### Published international conference proceedings

- **Roux, JC & Lewis, PW 1999.** On Xhosa L2 speech and intelligibility: Ejectives, implosives and clicks. *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, USA; Regents of the University of California. CD-ROM format.

## 2.3 Current projects

Projects are conducted within the following programmes:

- **Programme 1:** Linguistic projects  
*(The phonetics-phonology interface)*
- **Programme 2:** Resource development projects.  
*(Resource development in African languages and capacity building)*
- **Programme 3:** Language and speech technology projects  
*(Human-machine interaction through African languages)*

### 2.3.1 Programme 1: Linguistic projects

#### 2.3.1.1 The production and perception of statement and question intonation in Xhosa

**Ms Jackie Jones & JC Roux**

For a full description of this project - please see the Annual Report 1996/7

This project is nearing completion with expected date of completion early in 2000. The researcher is making extremely good progress having completed a vast array of acoustic and perceptual analyses. An evaluation of the implications of these results for phonological and psycholinguistic theories as well as for its relevance towards human-machine interfaces is currently in progress.

##### **Dissemination of results:**

##### **Previously reported:**

- i. One presentation at national conference of ALASA, Johannesburg: Jones (1996)
- ii. One presentation at regional workshop of ALASA, Pretoria: Jones (1996)
- iii. One presentation at an international workshop of ALASA, Durban (1997)

##### **New:**

##### **Articles in accredited journal**

- **Jones, J, Louw, J & Roux, JC. 1998.** Queclaratives in Xhosa: An acoustical analysis. *South African Journal of Linguistics*, Supplement 36, 3-18.
- **Jones, J, Louw, J & Roux, JC. 1998.** Perceptual experiments on queclaratives in Xhosa. *South African Journal of Linguistics*, Supplement 36, 19-32

##### **Contribution to Festschrift**

- **Jones, JJ & Roux, JC. 1999.** Queclaratives in Xhosa in *African Mosaic: A Festschrift for JA Louw*, ed R Finlayson, Pretoria, UNISA, pp 164 - 172

#### 2.3.1.2 The phonetics and phonology of click articulations in Nguni

**JC Roux & G Dogil**

For a full description of this project - please see the Annual Report 1996/7

This is an **ongoing study** without any specific date of completion due to the complexity of the topic. Dr Michael Jessen, a member of the Institute for Computational Linguistics (IMS) at the University of Stuttgart joined in on this project as a postdoctoral researcher in 1998. The project is progressing well whilst making interesting and new contributions to our knowledge on this sound type.

##### **Dissemination of results:**

##### **Previously reported:**

- i. Presentations at international conferences in, respectively, Vienna (Dogil, 1996)<sup>2</sup>, Budapest (Roux, 1996), München (Dogil, 1996), Adelaide (Roux, 1996), Paris (Roux & Dogil, 1997) as well as at national and regional conferences in South Africa: (Roux, 1996a, 1996b).
- ii. Three publications as proceedings of international conferences: Dogil et al (1996), Dogil & Roux (1996), Dogil, Mayer & Roux (1998)
- iii. One publication in an accredited journal: Roux & Lewis (1996)

**New:****Proceedings of international conferences**

- **Roux, JC & Dogil G. 1998:** On the phonetic representation of clicks: Some experimental phonetic considerations *Proceedings of the 16<sup>th</sup> International Conference of Linguists*. (Ed B Caron), CNRS, Paris. CD-ROM format.

**Article in accredited journal (submitted for publication)**

- **Jessen, M & Roux, JC (\*) Voice** characteristics associated with stops and clicks in Xhosa: An acoustic investigation. Submitted for publication to an international journal: *Journal of Phonetics* (June, 1999)

**2.3.2 Programme 2: Resource development and capacity building projects****2.3.2.1 Preferred syllable structures in Nguni and Sotho****JC Roux, A Jones, J Louw, Dorothy Calata & Bongiwe Fingo**

For a full description of this project - please see the Annual Report 1996/7

This **ongoing study** with an initially expected completion date of **December 1999** is making good progress. This project however will need to continue and link on to the new DACST 2000 project.

**Dissemination of results:****Previously reported**

- i. One presentations at an international conference: Vienna (Roux, 1996)
- ii. One presentation at national workshop on Speech, Text and Technology, UCT, (A Jones, 1997)
- iii. One presentation at workshop at 9<sup>th</sup> International conference, of ALASA, Durban (A Jones, 1997)

Nothing new to be reported.

**2.3.2.2 SASPEECH**

This project may be regarded as pivotal for any real significant development within

---

<sup>2</sup> Full details regarding these "older" references may be found in the 1997/8 Report.

RUEPUS in view of its declared change in focus and movement towards language and speech technology applications. For this reason extremely much attention was devoted by the Director to promote this project and to apply in December 1998 to PANSALB for the funding of a sub-project entitled *The development of multi-purpose spoken language databases for the languages of South Africa*. Response from PANSALB was rather disappointing: not only was the proposal rejected, but judged on the nature of the feedback it is clear that the organisation has no real idea of what is taking place in the field of language and speech technology at international level and its implications for the South African scene.

The SASPEECH project was however brought to the attention of the international community at a presentation of the first international conference of language and speech resources (details below). This was the only presentation from an African country.

This project will exist next to the DACST 2000 project and provision will be made in the NRF (CSD) budget as specified later.

### **Dissemination:**

Conference presentation and publication in conference proceedings

- **Roux, JC. 1998:** SASPEECH: Establishing speech resources for the indigenous languages of South Africa. *Proceedings of the First International Conference on Language Resources and Evaluation*. (Eds A Rubio et al). Granada, Spain. European Language Resources Association - ELRA. Vol 1, 343-350.

### **2.3.2.3 Development of computer software**

Apart from the reported eight software programmes that have been developed at RUEPUS (see 1998 Report) a new Windows NT based programme is under construction for automated phonetic transcriptions and following from that a series of "information mining" algorithms. This will enable researchers to access any text or speech database to recover any type of information on phonetic, phonological, morphological, and eventually syntactic strings.

### **2.3.2.4 Research capacity building projects**

#### **Previously reported**

Two RCB projects with respectively the University of Venda, and University of Transkei.

#### **New:**

#### **Research capacity building with the Department of Sesotho, University of the North, Qwa-Qwa.**

This project was designed and implemented along the same lines as the two previous projects (see 1998 Report). The person in charge of the newly installed phonetic workstation is Mr K Selebeleng as acting head of department. As part of the staff training programme the following colloquium took place:

- **Colloquium** on experimental speech research, at the University of the North, Qwa-Qwa branch, Phuthaditjhaba, June 1999

**JC Roux & M Jessen:** *Speech acoustics and the sound system of Sesotho*

### 2.3.3 Programme 3: Language and Speech Technology Projects

#### 2.3.3.1 Modelling the intonation of imperatives in a Xhosa text-to-speech system (TTS)

**P Swart, J Louw, L Botha, & JC Roux**

For a full description of this project - please see the Annual Report 1996/7

This project is nearing completion and will be disseminated in a MA thesis at the end of 1999. Apart from setting specific guidelines of how to develop a TTS system in an African tone language, an important methodological point has been addressed regarding the nature of tagged speech when used as a basis for concatenation in speech synthesis.

#### **Dissemination of results:**

#### **Previously reported:**

- i. Presentation at two day workshop at 9<sup>th</sup> International ALASA conference, Durban (Swart, 1997)

#### **New:**

- **Swart, PH.** 1999. Presentation at 10<sup>th</sup> International Conference of the African Language Association of Southern Africa entitled **Prosodic features of Imperatives in Xhosa: A perceptual experiment.**

#### 2.3.3.2 Speech recognition in African languages (DACST 1998 proposal)

RUEPUS entered into a consortium proposal agreement with the departments of Electric and Electronic Engineering at respectively, the Universities of Pretoria and Stellenbosch which was presented as an Innovation Fund proposal to the Departments of Arts, Culture, Science and Technology. The topic of this project was *Development of African Language Speech Recognition Software and Voice based Information System Prototypes*.

This project application was unsuccessful, however, the new DACST 2000 project is to replace this project. (See Section 4: Planning for next year.)

#### 2.3.3.3 Language Engineering (LE): Principles and Practice

**JC Roux, J Louw, PH Swart, J Masalesa**

A detail description of this project is found in the 1997/8 Report.

The CSD awarded an Internship to RUEPUS to fund the position of Mr Masalesa. The project is making good progress in creating, *inter alia*, an awareness in the country towards the need for LE development. A pre-conference seminar of two days was presented at the 10<sup>th</sup> International Conference of ALASA in July 1999, as well as a plenary lecture at the same conference as detailed below. Resulting from the workshop was the founding of the *ALASA Special Interest Group on Language and Speech*

*Technology Development*, set up with specific aims to further the cause of LE in SA.

### International conference participation

**Roux, JC 1999: 10<sup>th</sup> Biennial International Conference of the African Language Association of Southern Africa (ALASA 99).** UNISA, Pretoria.

- Organiser and presentation of workshop: *African Language Engineering in the New Millennium*.
- Keynote speaker on invitation: *Language and Technology: Challenges for African languages in the next millennium*.

## 2.4 Publications and conference presentations

### 2.4.1 Papers published in (or submitted to) accredited journals<sup>3</sup>

- **Roux, JC, Wissing, DP & R Pretorius. 1998** Front vowel raising in Tswana. *South African Journal of Linguistics*, Supplement 36, 93-103
- **Lewis, PW & Roux, JC. 1998.** Xhosa L2 speech intelligibility. *South African Journal of Linguistics*, Supplement 36, 51-63
- **Jones, J, Louw, J & Roux. JC. 1998.** Queclaratives in Xhosa: An acoustical analysis. *South African Journal of Linguistics*, Supplement 36, 3-18.
- **Jones, J, Louw, J & Roux. JC. 1998.** Perceptual experiments on queclaratives in Xhosa. *South African Journal of Linguistics*, Supplement 36, 19-32.
- **Jessen, M & Roux, JC (\*)** Voice characteristics associated with stops and clicks in Xhosa: An acoustic investigation. Submitted for publication to an international journal: *Journal of Phonetics* (June, 1999)

### 2.4.2 Publications in national and international conference proceedings

- **Roux, JC & Dogil G. 1998:** On the phonetic representation of clicks: Some experimental phonetic considerations *Proceedings of the 16<sup>th</sup> International Conference of Linguists*. (Ed B Caron), CNRS, Paris. CD-ROM format.
- **Roux, JC. 1998:** SASPEECH: Establishing speech resources for the indigenous languages of South Africa. *Proceedings of the First International Conference on Language Resources and Evaluation*. (Eds A Rubio et al). Granada, Spain. European Language Resources Association - ELRA. Vol 1, 343-350.
- **Roux, JC & Lewis, PW 1999.** On Xhosa L2 speech and intelligibility: Ejectives, implosives and clicks. *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, USA; Regents of the University of California. CD-ROM format.

### 2.4.3 Chapters in books

- **Jones, JJ & Roux, JC. 1999.** Queclaratives in Xhosa in *African Mosaic: A*

<sup>3</sup> In the previous report some of these articles were reported as having been submitted only.

*Festschrift for JA Louw*, ed R Finlayson, Pretoria, UNISA, pp 164 - 172

#### 2.4.4 Books

- **Gerhardt, L & Roux, JC. 1998. Otto Dempwolff: Induktiver Aufbau des Urbantu.** Archiv afrikanistischer Manuskripte, Nr 3. R.Köppe Verlag, Köln.

#### 2.4.5 Papers read at international conferences

- **JC Roux. 1998.** Presentation at the *First International Conference on Language Resources and Evaluation*. Granada, Spain. Entitled **SASPEECH - Establishing speech resources for the languages of South Africa**
- **JC Roux. 1999.** *10<sup>th</sup> Biennial International Conference of the African Language Association of Southern Africa (ALASA 99)*. UNISA, Pretoria.
- **Organiser and presentation of workshop:** *African Language Engineering in the New Millennium*.
- **Keynote speaker** on invitation: *Language and Technology: Challenges for African languages in the next millennium*.
- **JC Roux 1999.** Presentation at the *14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, USA. Entitled **On Xhosa L2 speech and intelligibility: Ejectives, implosives clicks**.
- **Swart, PH. 1999.** Presentation at *10<sup>th</sup> International Conference of the African Language Association of Southern Africa* entitled **Prosodic features of Imperatives in Xhosa: A perceptual experiment**.

#### 2.4.6 Papers read at national conferences

- **Roux, JC. 1998** Co-author of presentation at the *National Conference of the Linguistic Society of South Africa*, University of Potchefstroom entitled **Problems regarding front vowel variation in Setswana**.
- **Roux, JC. 1998** Presentation at *National Conference of ALASA*, Potchefstroom University, Potchefstroom entitled **On vowel raising and perception**
- **Roux, JC. 1999** Invited plenary address at **National Seminar of the South African Translators` Institute**. ESKOM. Midrand. Title: **Introducing Translation Technology to the South African scene**.

### 2.5 Staff activities

#### 2.5.1 Visits abroad

The Director made a visit in May 1998 to Granada, Spain in order to participate in the *1<sup>st</sup> International Conference on Language Resources and Evaluation*, and again to the United States of America in August 1999 in order to participate in the *14<sup>th</sup> International Conference of Phonetic Sciences*. During this visit contact was also made with speech technology developers in Boston, i.e. Dragon Systems, Speechworks and Lernhout and Hauspie. These visits were of great value for the DACST 2000 project, which materialised soon afterwards.

## 2.5.2 Visitors

### 2.5.2.1 Prof E Wolff. Afrikanistik, University of Leipzig, Germany

Although Prof Wolff primarily visited the Department of African Languages, he had lengthy discussions with staff of RUEPUS and with the Director in view of some specific needs in the area of phonetics and phonology. A tentative arrangement has been reached in terms of which the Director and/or staff members of RUEPUS will assist in the development of a new phonetics laboratory in Leipzig in 2000, and simultaneously supply them with available acoustic speech databases for the study of the South African Bantu languages at phonetic level. These arrangements from 1998 were reconfirmed.

## 2.6 Conferences and symposia presented

### 2.6.1 Two day workshop on Language Engineering: 10<sup>th</sup> International Conference of the African Language Association (ALASA), University of South Africa, Pretoria, July 1999

See 2.3.3.3

### 2.6.2 Colloquium on experimental speech research, at the University of the North, Qwa-Qwa branch, Phuthaditjhaba, June 1999.

See 2.3.4

## 2.7 Implementation of research

Up till now research results have been disseminated as described above. Real world implementation of these results will take place with the development of an interactive multilingual information system.

## 3 INFRASTRUCTURE

### 3.1 Staff

The following persons were appointed for the periods specified:

Mr Jan Louw (M Eng.)	Ad hoc (1998) then 1/10/1998 - 30/11/1998
Ms Pippa Swart (BA Hons)	1/4/1998 - 31/3/1999
Ms D Calata	Ad hoc 1998/9
Ms B Fingo	Ad hoc 1998
Ms M Mdemka (BA)	Ad hoc 1998-9
Ms S du Plessis	1/6/1998 - 31/3/1999
Ms E van der Vyver (BA Hons)	1/6/1998 - 31/3/1999
Mr J Masalesa (BA Hons)	1/1/1999 - 31/3/1999 (31/12/1999)

RUEPUS was awarded an Internship from the CSD that made provision for the position of

Mr Masalesa. He will be more fully employed in RUEPUS following this internship.

### 3.2 Resources

Resources for the study of both technology and phonetics are readily available in the different libraries on campus. RUEPUS furthermore has adequate access to the Internet.

### 3.3 Equipment

RUEPUS had to upgrade some of its computer equipment as routine practice. This is reflected in the financial report. At this point in time the laboratory is adequately equipped for running its own projects. Any new project sponsored from an external source will have to include equipment in its budget - this is indeed the case with the DACST 2000 project where provision has been made for additional hardware and software.

## 4 PLANNING FOR NEXT YEAR (2000/2001 - YEAR 3 OF CYCLE 2)

The strategic plan accepted in 1998 will have to be adapted in view of the DACST 2000 project:

CYCLE 2	ACTIVITIES	STAFF	COSTS
<b>Year 1 1998/1999</b>	<ul style="list-style-type: none"> <li>Research in linguistic processing and technology development</li> <li>Research capacity building projects with HDIs</li> <li>Setting up networks at national and international levels</li> <li>Initial fund raising initiatives</li> </ul>	1 researcher (ling) 1 researcher (techn) Half day secretary	
<b>Year 2 1999/2000</b>	<ul style="list-style-type: none"> <li>Research in linguistic processing and technology development</li> <li>Research capacity building projects with HDIs</li> <li>Networking at national and international levels</li> <li>Special awareness campaigns with major fund raising initiatives</li> <li>Solicitation of projects</li> <li>DACST (1998) proposal on speech recognition</li> <li>SASPEECH</li> </ul>	1 researcher (ling) 1 researcher (techn) Half day secretary	
<b>Year 3 2000/2001</b>	<ul style="list-style-type: none"> <li>ALEC partially in place (with RUEPUS projects)</li> <li><b>DACST 2000: (1/1/2000 - 31/12/2001)</b></li> <li>Research in linguistic processing and technology development</li> <li>LE training / capacity building</li> <li><b>RUEPUS:</b> <ul style="list-style-type: none"> <li><b>Programme 1:</b> Phonetics/phonology interface</li> <li><b>Programme 2:</b> Resource and software development Speech and text corpora</li> <li><b>Programme 3:</b> LE principles and practice</li> </ul> </li> </ul>	<b>Own Staff:</b> 30 Co-workers (nationally) 5 Co-workers (internationally) 1 Researcher (Linguistics) 1 Programmer Ad hoc subjects / co-workers	<b>For 2000:</b> R2.9 million  <b>For 2001:</b> R2.9 million As per budget
<b>Year 4 2001/2002</b> (Final year for NRF funding)	<ul style="list-style-type: none"> <li>ALEC in place with final RUEPUS projects               <ul style="list-style-type: none"> <li>Change of name at end of cycle in March 2002</li> <li>Host to DACST 2000 till end of 2002.</li> </ul> </li> <li><b>DACST 2000: (1/1/2002 - 31/12/2002)</b></li> <li>Research in linguistic processing and technology development</li> <li>LE training / capacity building</li> <li><b>RUEPUS / (S)ALEC<sup>4</sup>:</b> <ul style="list-style-type: none"> <li><b>Programme 1:</b> Resource and software development Speech and text corpora</li> </ul> </li> </ul>	<b>Own Staff:</b> 30 Co-workers (nationally) 5 Co-workers (internationally) 1 Researcher (Linguistics) 1 Programmer (ad hoc) Ad hoc subjects / co-workers	<b>For 2001:</b> R3.1 million As per budget

<sup>4</sup> It is envisaged that (S)ALEC will continue after termination of NRF-funds at that point in time, but with independent funds originating from industry.

	o <b>Programme 2: LE principles and practice</b>		
--	--	--	--

The planning for the third year (2000/2001) below is in accordance with the proposed strategic plan as presented above:

#### **4.1 Programme 1: Linguistic projects: The phonetics - phonology interface**

##### **4.1.1 The phonetics and phonology of click articulations in Nguni (Continuation)**

**JC Roux, G Dogil, M Jessen and M Jokweni**

This is an ongoing study with estimated date of completion in 2001. Dr Michael Jessen, a member of the Institute for Computational Linguistics (IMS) and Dr Mbulelo Jokweni of UNITRA joined the team. The scope of the project has been expanded to include *inter alia* the study of ejectives and implosives in more detail. The relationship between phonetics and phonology will be explored with reference to these sound types as well as the implications of these sound types for automatic speech recognition and speech synthesis.

This project may only require new software - this is reflected in the budget.

#### **4.2 Programme 2: Resource development projects**

##### **4.2.1 Preferred syllable structures in Nguni and Sotho (Continuation) - to be expanded to "SALText"**

**JC Roux, J Louw, James Masalesa & F Olivier**

This project will be adapted to include so called "information mining" strategies in order to analyse the morphophonotactic structures of these languages in a systematic way. This will necessitate innovative software design to create computer-based tools for use by linguists. Mr Frank Olivier (a computer science undergraduate) has already proven his computational skills and will be responsible for these programmes.

The new acronym "SALText" refers to "South African Languages Text" corpora. Although the focus will exclusively remain on the Sotho and Nguni languages, it is envisaged that this project will eventually be expanded to all South African languages and that multilingual re-useable text resources and computer based access tools will be available for linguistic and technological applications.

Provision is made in the budget for

- i. language informants (subjects)
- ii. database labellers
- iii. computer programmer.

##### **4.2.2 SASPEECH (Continuation)**

**JC Roux, J Louw, P Swart, M Jokweni, K Qamata, A Sotashe**

This project will be continued with co-researchers from Unitra focussing on Xhosa.

Provision is made in the budget for language informants (subjects)

#### 4.2.3 Text Corpora: Parsing Tools for the African Languages (On hold)

**JC Roux, Marianna Visser, Kathy Demuth (Brown University, USA) and Annie Zaenen (Xerox, France)**

This project, which was announced in 1998, is still on hold depending on high-level negotiations with Xerox (France).

#### 4.3 Programme 3: Language and Speech Technology Projects

##### 4.3.1 Language Engineering: Principles and Practice (Continuation)

**JC Roux, P Swart & J Masalesa**

This project entails a continuous study of developments regarding LE technologies as well as of models of language and speech processing, and principles and policies underlying implementation.

No financial implications for the budget except for the position of Mr Masalesa.

##### 4.3.2 DACST 2000 (Appendix A)<sup>5</sup>

**Project co-ordinator: JC Roux**

This project entitled *Promoting the development of the official languages of South Africa through language and speech technology applications*, is the result of a consortium proposal to the **Innovation Fund** of the Department of Arts, Culture, Science and Technology and amounts up to a total of R9 077 620,00 spread over a period of three years.

This can be seen as a major breakthrough towards the sensitising of Government for the role it has to play in developing the languages of the nation at technological level. A total of 35 co-workers will be involved, thirty of which will be South African citizens of which at least 15 will be mother tongue speakers of an African language. The activities will take place at various campuses across the country with RUEPUS as base.

This project is of further importance to RUEPUS in that it actually provides the financial backing for the transformation process of RUEPUS to ALEC as envisaged in the Strategic Plan accepted by the Advisory Board.

Detail may be found in Appendix A. The description below will be used by DACST as a press release when officially announcing the new successful projects:

---

**University of Stellenbosch**

**Research Unit for Experimental Phonology -**

**Department of African Languages**

---

**Promoting the development of the official languages of South Africa through**

---

<sup>5</sup> Please see footnote 1 and the request for confidentiality.

**language and speech technology applications.**

---

**Proposal number:** 21213

---

**Focal area:** Information Society

---

**Project duration:** 3years

---

**Total funding:** R 9,077,620

---

Funding year 1: R 2,983,750

---

Funding year 2: R 2,903,590

---

Funding year 3: R 3,190,280

---

This project aims to introduce the languages spoken in South Africa into the domain of interactive human-machine communication as the world moves into a technologically driven next millennium. In order not to become marginalized in any way it is necessary that the official languages spoken in this country be developed at technological level to meet the challenges of the information society.

It has become clear within developed countries that interactive voice enabled information systems functioning in particular languages are of the most powerful tools to impart information and knowledge on a broad basis within and across multilingual societies. These systems that invoke **voice recognition, linguistic processing** and **speech generation** fall within the domain of Language and Speech Technology (LST) and find applications in all spheres of life. The single most important value of voice based LST applications is the fact that these systems are accessible to all people regardless their state of literacy as communication takes place in natural spoken language.

Apart from the implementation of an awareness campaign directed towards potential high profile end users of these systems, a **multilingual query and booking system** will be developed. This fully automated system will accept telephonic queries submitted in four languages, i.e. in either **South African English, isiXhosa, Sesotho** or **Afrikaans** after which it will access an information database and eventually supply the caller with the necessary information in good quality synthetic speech. Although the domain of the queries is initially restricted to **the hotel industry**, provision will be made for later "plug-in" modules, e.g. information on travel schedules, cinema and theatre programmes, insurance products, banking rates, automobile products, weather

conditions etc.

The project will be conducted by a consortium of three interdisciplinary teams from respectively the universities of Stellenbosch (African Languages and Electronic Engineering) Pretoria (Electronic Engineering), Transkei (African Languages), and West Technology Holdings with co-workers from the universities of Potchefstroom, Port Elizabeth and the University of the North (Qwa-Qwa branch).

**Project Co-ordinator:** Prof. Justus Roux; **Phone:** +27 21 808 2017

[jcr@maties.sun.ac.za](mailto:jcr@maties.sun.ac.za)

### **National and international conferences**

As in the past provision is made in the budget for two national and one international conference attendance by staff members of RUEPUS in order to disseminate research results and to be involved in capacity building programmes.

## **4.4 Staff**

The services of a young **linguistic researcher** is required to assist in theoretical work related to the linguistic project in Programme 1. It is important to keep abreast with developments in phonological theory and its language specific applications and in this regard the services of an MA student in African linguistics is urgently needed. Provision is made in the budget for such a person (Item 6). Two mother tongue speakers of African languages have indicated that they may be available for this position.

The Unit must have access to the services of a skilled computer scientist to develop software programmes for projects on Programmes 2 and 3. Mr Frank Olivier, a final year student in Computer Science have been involved in programme development within RUEPUS and the Department of African Languages during 1999. In order to retain his services for ad hoc programming tasks in 2000 (when he will be enrolled for an Honours degree), provision is made accordingly in the budget (Item 7)

**Mr James Masalesa** is a current Intern in the Unit with great research potential. In order to retain his services, provision is made for an amount of R50 000 in the budget (Item 8), which is on par with the amount he currently receives from the NRF.

Provision is also made for ad hoc co-workers / labellers / language assistants who have up to now fulfilled an extremely important task in the gathering of authentic speech material and/or in the perceptual testing of this data in experiments. These are normally undergraduate African students or Matriculation scholars from Khayamandi. It is envisaged that data gathering will need to be increased dramatically next year, therefore the anticipated need for 200 hours` participation (Item 9).

## **5 BUDGET**

### **5.1 Statement of income and expenditure for the previous financial year (1998/9)**

Please see **APPENDIX B**

## **5.2 Budget for next financial year (2000/2001)**

Please see **APPENDIX C**

## **6 GENERAL**

Thank you very much to the benefactors of RUEPUS, i.e. the NRF (CSD) and the Research Committee of the University of Stellenbosch, for continuous support over the years. It has enabled RUEPUS to grow in many ways to the point of recognition by Government in entrusting us with this DACST 2000 project from the Innovation Fund. Without your support, as well as that of the Advisory Board and the dedicated work of all staff members this would not have been possible.

**PROF. JC ROUX**

**STELLENBOSCH**

**DIRECTOR: RUEPUS**

**29 OCTOBER 1999**