

**RESEARCH UNIT FOR EXPERIMENTAL PHONOLOGY**

**UNIVERSITY OF STELLENBOSCH (RUEPUS)**

ANNUAL REPORT: 2000/2001 (Updated)

<b>1</b>	<b>BACKGROUND</b>	<b>3</b>
1.1	MISSION AND VISION	3
1.2	LONG-TERM GOALS	3
<b>2</b>	<b>PROGRESS REPORT</b>	<b>4</b>
2.1	ACHIEVEMENT OF GOALS	4
2.2	CURRENT PROJECTS	4
2.2.1	PROGRAMME 1: LINGUISTIC PROJECTS	5
2.2.1.1	The production and perception of statement and question intonation in Xhosa	5
2.2.1.2	The phonetics and phonology of click articulations in Nguni	5
2.2.2	PROGRAMME 2: RESOURCE DEVELOPMENT AND CAPACITY BUILDING PROJECTS	6
2.2.2.1	Preferred syllable structures in Nguni and Sotho	6
2.2.2.2	Research capacity building projects	6
2.2.3	PROGRAMME 3: LANGUAGE AND SPEECH TECHNOLOGY PROJECTS	6
2.2.3.1	Language Engineering (LE): Principles and Practice	6
2.2.3.2	African Speech Technology Project (AST)	8
2.3	COMPREHENSIVE SUMMARY OF PUBLICATIONS AND CONFERENCE PRESENTATIONS	8
2.3.1	PAPERS PUBLISHED (ALL IN ACCREDITED JOURNALS) (14 + 7 PAPERS)	8
2.3.2	PAPERS PUBLISHED AS CONFERENCE PROCEEDINGS (27 + 1 PAPER)	10
2.3.3	CONFERENCE / WORKSHOP PARTICIPATION	12
2.3.3.1	International	12
2.3.3.2	National	15
2.4	STAFF ACTIVITIES	16
2.4.1	VISITS ABROAD	16
2.4.2	VISITORS	18
2.5	CONFERENCES AND SYMPOSIA PRESENTED (15)	18
2.6	IMPLEMENTATION AND IMPACT OF RESEARCH	20
<b>3</b>	<b>GRADUATE TRAINING</b>	<b>20</b>
3.1	COMPLETED MA CANDIDATES	20
3.2	IN PROGRESS (MA)	21
3.3	COMPLETED DOCTORAL STUDY (DLITT)	21
3.4	IN PROGRESS (DLITT)	21
<b>4</b>	<b>INFRASTRUCTURE</b>	<b>21</b>
4.1	STAFF	21
4.2	RESOURCES	22
4.3	EQUIPMENT	22
<b>5</b>	<b>FUTURE PLANNING: CREATION OF NATIONAL ELECTRONIC TEXT AND SPEECH RESOURCE CENTRE</b>	<b>22</b>
5.1	REQUEST 1	24
5.2	REQUEST 2	25
5.3	REQUEST 3	25
<b>6</b>	<b>BUDGET</b>	<b>25</b>
6.1	STATEMENT OF INCOME AND EXPENDITURE FOR THE PREVIOUS FINANCIAL YEAR (1998/9)	25
<b>7</b>	<b>GENERAL</b>	<b>25</b>

## 1 BACKGROUND

This progress report covers activities of the **third year (2000/1)**<sup>1</sup> as well as of the **fourth year** and **final year** of the second cycle of the Unit. Due to a change in policy at the National Research Foundation (NRF) the award for the final year (2001/2) only extends to December 2001. This report will therefore focus on activities up to September 2001.<sup>2</sup> In view of the fact that this is probably the last report in the existence of RUEPUS with the financial support of the NRF, it will

- i. Present a concise report on the activities during the above mentioned period in the prescribed format for annual progress reports as specified by the NRF,
- ii. Present a comprehensive overview of all research and related outputs emanating from the Unit since its founding in 1994 and it will assess the impact of the Unit in national and international context, and
- iii. Suggest a strategy for consideration that could, with the possible contribution of the NRF, other parastatal bodies and the private sector, lead to the establishment of a much-needed *National Electronic Text- and Speech Resources Centre* (NETS). This could be a valuable contribution to the development of Human Language Technologies (HLT) in South Africa as espoused in the Foresight (1999) project of national government.

### 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, are to be *developed at technological level* to such an extent that

- These languages will meet the challenges of the Information Society in a technologically driven new 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 new century;
- An *electronic resource facility* making provision for digital speech and text corpora will be developed and maintained as a national asset for further academic and industrial development.

As indicated in the previous report, the vision regarding a *national resource facility* has become one of the main foci of RUEPUS and this vision could become a reality if the concrete suggestions in section 5 (below) were to be supported.

### 1.2 Long-term goals

In order to assess the impact of the Unit it may be necessary to take note of the specific

<sup>1</sup> Up to 2001 the annual funding covered the period April to March of the following year.

<sup>2</sup> The date closest to the annual meeting of the Advisory Board.

long-term goals as supported by the Advisory Board since 1998:

The 1998/1999 Report envisaged the following 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."*

In principle this goal is still pursued, however, the proposed name should be reconsidered and the centre should primarily focus on the creation and maintenance of digital text and speech resources as will be argued in section 5 below.

## 2 PROGRESS REPORT

### 2.1 Achievement of goals

The goals set by RUEPUS for the short, medium and longer term have been achieved to a very large extent. These goals relate to

- **Academic activities**, i.e. conducting research projects and the dissemination of research results through **21 scientific articles** and **27 conference publications**,
- **Networking**, i.e. making contact with colleagues at national and international level, *inter alia* by organising 15 conferences/seminars and participating in **32 international and 22 national conferences/seminars**, and hosting **17 international visitors** for shorter or longer periods of time.
- **Research capacity building**, i.e. at graduate and post graduate levels as part of formal academic courses (**14 completed plus 2 ongoing Masters degrees**, and **2 completed Doctoral degrees** with another two in progress); through 3 dedicated research capacity building projects and different specialised workshops,
- **Fund raising**, i.e. creating a self-sustainable infrastructure conducive to research and development, by securing a **R9.07 million project** from the Innovation Fund.

### 2.2 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)*

## 2.2.1 Programme 1: Linguistic projects

### 2.2.1.1 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 was **completed** in **November 2000** with the attainment of a **DLitt.** degree by Ms Jones. This was an excellent thesis and praised by both internal and external examiners. This study **contributed** to an understanding of the perceptual difference between statements and questions as manifested in reactions of mother tongue speakers/listeners of Xhosa. It made available invaluable data that could be utilised in language technology applications, e.g. the synthesis of Xhosa speech.

The results of this project were disseminated as follows (see also Report 1999/2000):

- Two presentations at national conferences/workshops of the African Language Association of Southern Africa (ALASA).
- Two presentations at international conferences/workshops of ALASA.
- Two articles in an accredited national journal.
- One contribution to a Festschrift.

(Details in summary.)

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

**JC Roux, G Dogil & M Jessen**

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

This project with international colleagues may officially be regarded as **completed**. The results of this project were disseminated as follows (also see Report 1999/2000):

- Presentations at five international and three national conferences
- Five publications as proceedings of international conferences.
- Two publications in an accredited national journal.
- One pending publication in the prestigious international *Journal of Phonetics*. (To be published in Vol 4: 2001)

(Details in summary.)

The single most important **contribution** of this research was the use of the so-called Wigner Distribution Analysis technique that clearly refuted the "affricative nature of clicks" claim as proposed by Traill and Ladefoged and other contemporaries. The

acoustic signal of most click types indicate a temporal order of frication and then plosivisation and not vice-versa.

## **2.2.2 Programme 2: Resource development and capacity building projects**

### **2.2.2.1 Preferred syllable structures in Nguni and Sotho**

**JC Roux, PW Lewis, V Macingwane, Sibongile Madi, B Wiese.**

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

This project is **to be completed in December 2001**. The project aims at determining statistically the nature of syllable structures in the Sotho and Nguni languages and a possible relationship between "preferred" syllables and the structure of loan words. It is corpus based and utilises a new software pattern analysis programme (*Patana 2001*), which was developed for this purpose.<sup>3</sup> It furthermore attempts to present explanations in the area of speech production for attested "preferred syllables".

The results of this project have been disseminated as follows (see Report 1999/2000):

- i. One presentation at an international conference.
- ii. Two presentations at national workshops of ALASA.

#### **In preparation:**

Two publications for presentation to the *South African Journal for African Languages*. (The previously reported intended submission of Roux & Louw was postponed in view of new data that had become available in the research process.)

(Details in summary.)

### **2.2.2.2 Research capacity building projects**

As indicated in Report 1999/2000 (p.6) research capacity building actions were to be taken care of within the *African Speech Technology* (AST) project.

## **2.2.3 Programme 3: Language and Speech Technology Projects**

### **2.2.3.1 Language Engineering (LE): Principles and Practice**

**JC Roux, J Masalesa, Maritza van den Heuvel**

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

This project has been making extremely good progress in assessing the nature of Human Language Technology (HLT) development worldwide. This project is interpretative in nature and the material gathered serves to assist in language planning with respect to the development of Human Language Technologies (HLT) in South Africa in particular.

In the previous report (1999/2000: p.7) two specific events directly and indirectly

---

<sup>3</sup> This programme also has applications in the AST project.

related to this project were mentioned. For perspective these two points are restated below:

### **Contribution to the formulation of language policy in South Africa**

The Director was invited to host a workshop at the Second National Language Indaba of the Department of Arts, Culture, Science and Technology (DACST) in Durban in May 2000 regarding the role of Language and Speech Technology in language policy in South Africa. He also presented a plenary lecture on this topic at the same conference. From both these events a number of specific recommendations followed, most of which eventually found their way into the *Final Draft Language Policy and Plan for South Africa* (5 June 2000) of the Advisory Panel on Language Policy which was presented to the Minister of Arts, Culture, Science and Technology and which is to be tabled in Parliament during this year.

### **Contribution to the formulation of a national strategy for Human Language Technologies in South Africa**

The Director was invited to join a national Steering Committee on Language and Information Technology set up by DACST and the Pan South African Language Board (PANSALB). He was eventually responsible for the drafting of a strategic planning document entitled *The Development of Human Language Technologies in South Africa: Strategic Planning*, which is currently under consideration by DACST and PANSALB. This strategic plan, if accepted, has far reaching implications for the development of National Lexicography Units (NLUs) and for the future of RUEPUS as well."

With respect to the last topic it may be reported that significant progress has been made albeit very slowly. In October 2000 PANSALB accepted the above-mentioned strategic document, and together with some proposals forwarded it to DACST for consideration. In August 2001, the minister for Arts, Culture Science and Technology, Dr Ben Ngubane appointed a small *Advisory Panel for the Implementation of Human Language Technology in South Africa* for the period August 2001 to August 2002 to advise him on further actions in this regard. Two members associated with RUEPUS have been included in this panel of six persons, i.e. Dr Wynand Coetzer (Advisory Board Member) as well as the Director who was requested to avail himself as co-ordinator of the panel. This development is indeed significant as it is seen as another step by National Government to implement recommendations of, *inter alia*, the Foresight (1999) project that emphasised the development of HLTs in South Africa.

In **October 2000** the Director presented an overview of HLT development activities in South Africa in his capacity as member of the management committee of the *International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (COCOSDA)*, at the **International Workshop on Language Resources and Technology**, Beijing, China.

In **May 2001** the Director attended a three-day training workshop in **XML Mark-up Language** presented at the University of Edinburgh. This was mainly sponsored by internal publication funds of the department of African languages.

In **July 2001** DACST presented the Director with an international travelling grant to present the Strategic Plan to an international audience at a *Workshop on Human Language Technology and Knowledge Management* at the **39<sup>th</sup> Annual Conference of the Association of Computational Linguistics** in Toulouse France. The reaction was extremely positive whilst the magnitude of the task (accommodating eleven languages) was frequently mentioned.

The Director was asked to address a number of decision makers in **Parliament** on **7 August 2001** on the topic *The promotion of multilingualism through HLT applications*. The audience also included members from the Provincial government, including the Premier of the Western Cape, Mr G Morkel. The presentation was well received with a number of requests for further future contact.

This project and the AST project also host the Africa-rapporteur's web page for COCOSDA reporting on HLT activities in Africa, cf. [http://www.ast.sun.ac.za/Cocosda\\_index/htm](http://www.ast.sun.ac.za/Cocosda_index/htm). (This is a new web page and some parts are still under construction.)

### 2.2.3.2 African Speech Technology Project (AST)

*African Speech Technology* (AST) is the working title of the DACST Innovation Fund Project (Round 2) awarded to RUEPUS in 1999 that officially commenced on January 1, 2000. The approved title of the project is *Promoting the development of the official languages of South Africa through language and speech technology applications*. This project is overseen by a Steering Committee comprising representatives of the consortium responsible for the project. It has a prescribed mode of report, notably the submission of quarterly progress reports. A website (<http://www.ast.sun.ac.za/>) has been created for this project with links to and from the RUEPUS website.

This has indeed become a major project involving at this stage, eighty two (82) co-workers at different levels of expertise. The project is making good progress at all levels. Progress is monitored by two Progress Monitors appointed by DACST. The original proposal made provision for the development of an automated telephone based information retrieval system in three languages (SA English, Afrikaans and Xhosa); however, the team was requested to add Zulu and Sesotho as well. The additional languages are now proving to impact on the expected delivery date (December 2002), and a request has been put to DACST for an extension of the project (without any additional costs involved) by ten months (i.e. to end October 2003).

In **September 2001** the Director presented an overview of AST development activities in South Africa in his capacity as member of the management committee of the *International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (COCOSDA)*, at its annual **International Workshop on Language Resources and Technology**, Aalborg, Denmark.

**Appendix A** represents **Quarterly Report 6** (June 2001) explicating the activities of the project in more detail. According to the current schedule this represents the half way mark of the project.

The negotiations with the SAIL Trust of Belgium did not come to fruition due to internal problems experienced by the mother company, Lernhout and Hauspie.

(Details of publications, conference activities etc are presented in the Summary).

## 2.3 Comprehensive summary of Publications and conference presentations

### 2.3.1 Papers published (All in accredited journals) (14 + 7 papers)

- 1995 Roux, JC. Prosodic data and phonological analyses in Zulu and Xhosa. *South African Journal of African Languages*, 15.1: 19-28.



- 1995 Roux, JC. On the perception and production of tone in Xhosa. *South African Journal of African Languages*, 15.4: 196-204.
- 1996 Roux, JC & Ntlabezo, BN: Phonetic motivation for phonological processes: Labio-dental assimilations in Xhosa. *South African Journal of Linguistics*, 31:139-151.
- 1996 Lewis PW & Roux, JC. A phonological process analysis of the acquisition and loss of clicks in Xhosa. *South African Journal of African Languages*, 16.1: 1-9.
- 1996 Wissing, DP & Zonneveld, W. Final devoicing as a robust phenomenon in second language acquisition: Tswana, English and Afrikaans. *South African Journal of Linguistics*, Supplement 33, pp 4-55.
- 1996 Wissing, DP & Coetzee, AW. Die stemlose eksplosief van Afrikaans: 'n eerste verkenning. *South African Journal of Linguistics*, Supplement 33, pp 1-19.
- 1996 Van Rooy, AJ & Wissing, DP. Degrees of neutralisation during syllable-final devoicing: evidence from second language phonetics. *South African Journal of Linguistics*, Supplement 33, pp 77-98.
- 1996 Nortjé, Sylvia & Wissing, DP. Accepting accent: attitude of high school pupils towards language varieties. . *South African Journal of Linguistics*, Supplement 33, pp 139-162.
- 1998 Roux, JC, Wissing, DP & R Pretorius. Front vowel raising in Tswana. *South African Journal of Linguistics*, Supplement 36, 93-103.
- 1998 Lewis, PW & Roux, JC. Xhosa L2 speech intelligibility. *South African Journal of Linguistics*, Supplement 36, 51-63.
- 1998 Jones, J, Louw, J & Roux, JC. Queclaratives in Xhosa: An acoustical analysis. *South African Journal of Linguistics*, Supplement 36, 3-18.
- 1998 Jones, J, Louw, J & Roux, JC. Perceptual experiments on queclaratives in Xhosa. *South African Journal of Linguistics*, Supplement 36, 19-32.
- 1998 Roux, JC. Xhosa: A tone or pitch accent language? *South African Journal of Linguistics*, Supplement 36, 33- 50.
- 2000 Roux, JC & Louw, Philippa H. Black South African English (BSAE) and Speech Technology Applications. *South African Journal of Linguistics*, Supplement 38:4-13
- 2001\* Naidoo, Shamila. Distinctive feature theory: From linear to non-linear: An application to isiZulu. Accepted for publication in *South African Journal for African Languages* (SAJAL)
- 2001\* Jessen, M & Roux, JC. Voice characteristics associated with stops and clicks in Xhosa: An acoustic investigation. Accepted for publication in *Journal of Phonetics* (provisionally Vol 4)
- 2001\* Jessen, M An acoustic study of distinctive stop types and click accompaniments in Xhosa. Submitted for publication to an international journal: *Journal of Phonetics* (June, 1999), no feedback yet.

- (2001\*) Lewis, PW, Roux, JC. On the syllable structure of Zulu and Xhosa. For presentation to the *South African Journal for African Languages*. (Completed, to be submitted.)
- (2001\*) Lewis, PW, Roux, JC. On the syllable structure of Zulu and Xhosa. For presentation to the *South African Journal for African Languages*. (In progress)
- (2001\*) Jones, J, Roux, JC. Acoustic and perceptual qualities of queclaratives in Xhosa. For presentation to the *South African Journal for African Languages*. (In progress)
- (2001\*) Louw, PH, Roux, JC. A new algorithm for grapheme to phoneme conversions in African languages. For presentation to the *South African Journal for African Languages*. (In progress)

### 2.3.2 Papers published as Conference Proceedings (27 + 1 paper)

- 1995 Scharf, G, Hertrich, I, Roux, JC & Dogil, G. An articulatory description of clicks by means of electromagnetic articulography. *Proceedings of the 13<sup>th</sup> International Congress of Phonetic Sciences, ICPhS 95*, Stockholm, Sweden. Vol 1: 378-379.
- 1995 Roux, JC, Dogil, G & Wokurek, W. Click articulations in Xhosa: New perspectives through Wigner distribution analysis. *Proceedings of the 13<sup>th</sup> International Congress of Phonetic Sciences, ICPhS 95*, Stockholm, Sweden. Vol. 1: 502-505.
- 1995 Wissing, DP & Roux, JC. The interrelationship between VOT and voice assimilation. *Proceedings of the 13<sup>th</sup> International Congress of Phonetic Sciences, ICPhS 95*, Stockholm, Sweden. Vol. 1: 50-53.
- 1996 Wissing DP & Roux, JC. Voice assimilation in (Tswana) Afrikaans: A phonetic solution for a phonological problem. *SA Linguistics 1995*, Ed EF Kotzé, University of Port Elizabeth Publication Series B23, 262-268.
- 1996 Van Rooy, AJ. Word-final devoicing by Tswana and Afrikaans speakers of English: Second language interference or universal tendency? *SA Linguistics 1995*, Ed EF Kotzé, University of Port Elizabeth Publication Series B23, 214-227.
- 1996 Wissing, DP & Roux, JC. Voice assimilation in (Tswana) Afrikaans: A phonetic solution for a phonological problem. *SA Linguistics 1995*, Ed EF Kotzé, University of Port Elizabeth Publication Series B23, 262-268. (Reported in 1996)
- 1996 Dogil, G, Mayer, J, Mavimbela, P, Roux, JC & Tuomi, S. Speech coding and apraxia of speech: Evidence from Xhosa. in *Clinical Phonetics and Linguistics. Proceedings of the 5<sup>th</sup> Annual International Conference of the ICPLA*, (eds W Ziegler & K Deger) München, Germany.
- 1996 Dogil, G & Roux, JC. Notes on unencoded speech: Clicks and their accompaniments in Xhosa. *Proceedings of the Sixth Australian International Conference on Speech Science and Technology, SST-96*. (Ed P McCormack & A Russel), 55-60.
- 1998 Dogil, G, Mayer, J & Roux, JC. Syllables and unencoded speech: Clicks and their accompaniments in Xhosa. In *PHONOLOGICA 96: Syllables?! Eds. J Rennison & K Kuhnhammer*. The Hague: Holland Academic Graphics. 49-60.

- 1998 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.
- 1998 Roux, JC. 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.
- 1999 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.
- 2000 Louw, PH, Roux, JC, & Botha EC. Synthesizing prosody for commands in a Xhosa TTS system. *Proceedings of the 6<sup>th</sup> International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China, Vol 3: 263-266.
- 2000 Louw, PH, Roux, JC, & Botha EC. Synthesizing prosody for commands in a Xhosa TTS system. *Proceedings of the 11<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*, 7-10. Johannesburg: University of the Witwatersrand (Reprint)
- 2000 Roux, JC, Botha, EC, & Du Preez, JA. Developing a multilingual telephone based information system in African languages. *Proceedings of the Second International Conference on Language Resources and Evaluation*. Athens. Vol 2. pp 975-980. Also in CD-ROM format
- 2000 Roux, JC. Language and Technology. *2<sup>nd</sup> Language Indaba Proceedings*. Durban. National Language Service. Pp 80-85.
- 2000 Roux, JC, Swart, PH. Black South African English (BSAE) and speech technology applications, *Workshop on Black South African English, International Conference on Linguistics in South Africa*, 46-51, University of Cape Town.
- 2001 Roux, JC. Zulu Tonology and its relationship to other Nguni languages: A reply to Kisseberth and Cassimjee - The empirical status of tonal data. *Proceedings of the Second International Symposium on Cross linguistic studies of Tonal Phenomena*, (Ed S Kaji) Tokyo University of Foreign Studies, Japan, pp 361-368.
- 2001 Roux, JC. Human Language Technology (HLT) development in an African context: Planning for the next decade in South Africa, *Proceedings of the Workshop on Human Language Technologies and Knowledge Management, 39<sup>th</sup> Annual meeting of the Association of Computational Linguistics*, Toulouse, France, pp 25-32.
- 2001 Roux, JC. The African Speech Technology Project: Challenges for the collection of speech databases. *COCOSDA International Workshop on Language Resources*, Aalborg, Denmark pp 14-15. (Also full presentation at <http://www.slt.atr.co.jp/cocosda>)
- 2001 Louw, PH, Botha, EC & Roux, JC **Developing** multilingual speech corpora in African languages. *Proceedings of Eurospeech 2001*, Aalborg, Denmark. Vol 3: 522-525.

- 2001 PH Louw, JC Roux & EC Botha. African speech technology (AST) telephone speech databases: Corpus design, contents and early validation statistics. Proceedings of the 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001), pp 115-119.
- 2001 E de Villiers and JA du Preez. The advantage of using higher order HMM's for segmenting acoustic files. Proceedings of the 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001), Franschhoek, pp 120-122.
- 2001 HA Engelbrecht and JA du Preez The interplay of signal analysis and phoneme modelling techniques on phoneme recognition. Proceedings of the 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001), Franschhoek, pp 123-127.
- 2001 MW van den Heuvel and EC Botha. Dialogue development for the African Speech Technology hotel booking and enquiry system. Proceedings of the 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001). Franschhoek, pp 128-134.
- 2001 TR Niesler and JC Roux. Natural language understanding in the DACST-AST dialogue system. Proceedings of the 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001) Franschhoek, pp 134-137.
- 2001 A Visagie and JA du Preez. Sinusoidal modelling in speech synthesis, a survey. Proceedings of the 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001). Franschhoek, pp 138-142.
- (2001) Roux, JC, Bosch, SE. HLT Initiatives in South Africa - implications for human development, empowerment and democratisation. *Proceedings of the International Symposium on Text in context: African languages between orality and scriptuality*, University of Zurich, Switzerland. (In print) (October 2001)

### 2.3.3 Conference / Workshop participation

#### 2.3.3.1 International

(Name of presenter underscored, if more than one author)

- 1995 JC Roux. A perceptual approach to the study of tone. Laboratory workshop conducted as part of a two-day workshop at the 8<sup>th</sup> *International Biennial Conference of the African Language Association of Southern Africa (ALASA 95)*, University of Stellenbosch, Stellenbosch.
- 1995 Scharf, G, Hertrich, I, Roux, JC & Dogil, G. An articulatory description of clicks by means of electromagnetic articulography. Poster session conducted by Roux & Scharf at the 13<sup>th</sup> *International Congress of Phonetic Sciences, ICPhS 95*, Stockholm, Sweden.
- 1995 Roux, JC, Dogil, G & Wokurek, W. Click articulations in Xhosa: New perspectives through Wigner distribution analysis. Paper presented at the 13<sup>th</sup> *International Congress of Phonetic Sciences, ICPhS 95*, Stockholm, Sweden.
- 1995 Wissing, DP & Roux, JC. The interrelationship between VOT and voice assimilation. Paper read at the 13<sup>th</sup> *International Congress of Phonetic Sciences*,

*ICPhS 95*, Stockholm, Sweden.

- 1996 G Dogil & JC Roux. Syllables and unencoded speech: Clicks and their accompaniments in Xhosa. *The 8<sup>th</sup> International Phonology Meeting*, University of Vienna, Vienna.
- 1996 JC Roux. Noisy clicks in Nguni. One-day Workshop on *Phonetics and Phonology*, Hungarian Academy of Science, Budapest, Invited opening paper. (Organiser: Prof M Gosy)
- 1996 JC Roux & G Dogil. Notes on unencoded speech: Clicks and their accompaniments in Xhosa *The 6<sup>th</sup> International Conference on Speech Science and Technology*, Adelaide, Australia.
- 1996 G Dogil, J Mayer, P Mavimbela, S Tuomi & JC Roux. Speech coding and apraxia of speech: Evidence from Xhosa. *The 5<sup>th</sup> International Conference of Clinical Phonetics and Linguistics*, München, Germany.
- 1996 JC Roux, PW Lewis & J Louw. Preferred syllable structures in Xhosa. *8<sup>th</sup> International Phonology Meeting*, University of Vienna, Vienna.
- 1996 JC Roux. Human-machine communication in a developing country. *International symposium on culture, communication and development*, HSRC & UNESCO, Pretoria.
- 1996 JC Roux. Developing speech-based communication systems in the indigenous languages of South Africa. *The Third International Conference of the Third World Science, Technology and Development Forum*. University of Natal, Pietermaritzburg.
- 1997 JC Roux. Phonetic explanations for phonological phenomena: Labiodental assimilations in Xhosa. *2<sup>nd</sup> World Congress of African Linguists (WOCAL 2)*, Leipzig, Germany.
- 1997 JC Roux. On the phonetic representation of clicks: Some experimental phonetic considerations. *16<sup>th</sup> International Congress of Linguists*, Paris, France.
- 1998 JC Roux. SASPEECH - Establishing speech resources for the languages of South Africa Presentation at the *First International Conference on Language Resources and Evaluation*. Granada, Spain.
- 1999 JC Roux. *10<sup>th</sup> Biennial International Conference of the African Language Association of Southern Africa (ALASA 99)*. UNISA, Pretoria.
- Keynote speaker on invitation: *Language and Technology: Challenges for African languages in the next millennium*.
- 1999 Swart, PH. Prosodic features of Imperatives in Xhosa: A perceptual experiment. Presentation at *10<sup>th</sup> International Conference of the African Language Association of Southern Africa*.
- 1999 JC Roux **On** Xhosa L2 speech and intelligibility: Ejectives, implosives clicks. Presentation at the *14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, USA.
- 2000 Roux, JC & Swart, PH. Workshop and poster presentation entitled Black South

African English and Speech Technology Applications. *International Conference on Linguistics in Southern Africa*. University of Cape Town.

- 2000 Roux, JC, Botha, EC, & Du Preez, JA. Developing a multilingual telephone based information system in African languages. *Second International Conference on Language Resources and Evaluation*. European Language Resources Association (ELRA), Athens, Greece.
- 2000 Masalesa, J & Roux, JC Language Engineering in Southern Africa: a challenge for the new millennium. *Interim Conference of the African Language Association of Southern Africa Gaborone*, Botswana.
- 2000 Roux, JC. Report on Language Resources in Africa. *COCOSDA international workshop on language resources and technology evaluation: International and regional perspectives*, Beijing, China. (Session chair and presentation of paper.)
- 2000 Louw, PH, Roux, JC, & Botha EC. Synthesizing prosody for commands in a Xhosa TTS system. *Proceedings of the 6<sup>th</sup> International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China.
- 2000 Schwardt L & Du Preez, J. Efficient mixed-order hidden Markov model inference *ICSLP 2000: 6th International Conference of Spoken Language Processing*, Beijing, China.
- 2000 Schwardt L & Du Preez, J. Automatic language identification using mixed-order HMMs and untranscribed corpora. *ICSLP 2000: 6th International Conference of Spoken Language Processing*, Beijing, China.
- 2000 Roux, JC. Zulu Tonology and its relationship to other Nguni languages: A reply to Kisseberth and Cassimjee - The empirical status of tonal data *International Symposium on Cross-linguistic studies of Tonal Phenomena*, Tokyo University of Foreign Studies (Presentation of paper and panellist.)
- 2001 Roux, JC. Human Language Technology (HLT) development in an African context: Planning for the next decade in South Africa *Workshop on Human Language Technology and Knowledge Management, 39<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Toulouse, France.
- 2001 Jones, J. Roux, JC. Acoustic and perceptual qualities of queclaratives in Xhosa. *11<sup>th</sup> International Biennial Conference of ALASA*, University of Port Elizabeth.
- 2001 Roux, JC. Computational aspects of tone in Nguni. *Workshop on computational description of tone in African languages*. University of Helsinki, Finland. (Presentation on invitation)
- 2001 Roux, JC. The African Speech Technology Project: Challenges for the collection of speech databases. *COCOSDA Workshop: Language Resources and Technology Evaluation*, Aalborg, Denmark. (Session Chair and presentation).
- 2001 Louw, PH, Botha EC & Roux JC. Developing multilingual speech corpora in African languages. *Eurospeech 2001*, Aalborg, Denmark. (Poster presentation and programme demonstration).
- 2001 Roux, JC. Human Language Technologies and the promotion of multilingualism in South Africa. The *2<sup>nd</sup> South African Human-Computer Interaction Conference*,

University of Pretoria, Pretoria. Presentation of paper. (September 2001)

- 2001 Roux, JC & Bosch SE. HLT Initiatives in South Africa - implications for human development, empowerment and democratisation. *International Symposium: Text in context: African languages between orality and scriptuality*. Zurich, Switzerland. Co-author of paper. (October)

### 2.3.3.2 National

- 1995 Wissing, DP & Roux, JC. Voice assimilation in (Tswana) Afrikaans: A phonetic solution for a phonological problem. Paper read at the *31<sup>st</sup> Annual Congress of the Linguistic Society of Southern Africa (LSSA)*, University of Port Elizabeth.
- 1995 Van Rooy, AJ. Word-final devoicing by Tswana and Afrikaans speakers of English: Second language interference or universal tendency? Paper read at the *31<sup>st</sup> Annual Congress of the Linguistic Society of Southern Africa (LSSA)*, University of Port Elizabeth.
- 1996 Roux, JC Laboratory notes on click articulations in Xhosa. Paper read at the *Regional Conference of the African Language Association of Southern Africa*, Bellville. University of the Western Cape.
- 1996 Jones, J & Roux JC. Queclaratives in Xhosa. National Conference of ALASA, Rand Afrikaans University, Johannesburg.
- 1996 Jones, J & Roux JC. Demonstration of resynthesis techniques and perceptual testing strategies. *Regional Workshop of ALASA*, at UNISA, Pretoria.
- 1996 Roux, JC, Laboratory notes on click articulations in Xhosa. Paper read at the *Regional Conference of the African Language Association of Southern Africa (ALASA)*, Bellville. University of the Western Cape.
- 1996 Roux, JC. Noisy, nasal and voiced clicks in Nguni: Some experimental observations. *Annual National Conference of ALASA*. Rand Afrikaans University, Johannesburg.
- 1997 Roux, JC. Human-machine interaction in South Africa. *National Workshop on Speech, Text and Technology*, University of Cape Town.
- 1998 Roux, JC, Wissing, DP & Pretorius, R. Problems regarding front vowel variation in Setswana. 1998. Presentation at the *National Conference of the Linguistic Society of South Africa*, University of Potchefstroom.
- 1998 Roux, JC. On vowel raising and perception. Presentation at *National Conference of ALASA*, Potchefstroom University, Potchefstroom.
- 1999 Roux, JC. Introducing Translation Technology to the South African scene. Invited plenary address at *National Seminar of the South African Translators' Institute*. ESKOM. Midrand.
- 2000 Roux, JC Human Language Technologies and its development in South Africa. Workshop at *Second National Language Indaba*. Organised by DACST and PANSALB, Holiday Inn, Durban
- 2000 Roux, JC Language and Technology: Challenges for African languages in the

new millennium. Plenary paper at *Second National Language Indaba*. Organised by DACST and PANSALB, Holiday Inn, Durban

- 2000 Schwardt L & Du Preez, J. Mixed-order HMMs for language recognition. *IEEE COMSIG-2000*, Stellenbosch.
- 2000 Louw, PH, Roux, JC, & Botha EC. Synthesizing prosody for commands in a Xhosa TTS system. 11<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA).
- 2001 Roux, JC. Networking Human Language Technologies. *HCI-SA and ALASA Special Interest Group on HLT Workshop*, UNISA, Pretoria. Presentation of paper.
- 2001 PH Louw, JC Roux & EC Botha. African speech technology (AST) telephone speech databases: Corpus design, contents and early validation statistics. 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001)
- 2001 E de Villiers and JA du Preez. The advantage of using higher order HMM's for segmenting acoustic files. 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001), Franschhoek.
- 2001 HA Engelbrecht and JA du Preez The interplay of signal analysis and phoneme modelling techniques on phoneme recognition. 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001), Franschhoek.
- 2001 MW van den Heuvel and EC Botha. Dialogue development for the African Speech Technology hotel booking and enquiry system. 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001). Franschhoek.
- 2001 TR Niesler and JC Roux. Natural language understanding in the DACST-AST dialogue system. 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001) Franschhoek.
- 2001 A Visagie and JA du Preez. Sinusoidal modelling in speech synthesis, a survey. 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001). Franschhoek.

## 2.4 Staff activities

### 2.4.1 Visits abroad

During 2000/2001 the Director attended and participated in the following events:

- i. The *COCOSDA international workshop on language resources and technology evaluation: International and regional perspectives*, Beijing, China. (Session chair and presentation of paper.) (October 2000)
- ii. The 6<sup>th</sup> *International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China. (Co-presenter of a poster) (October 2000)
- iii. Series of invited guest lectures at the *Institut für Afrikanistik* at the University of Leipzig as part of centenary celebrations. (November 2000) (On invitation, dues paid by host.)



- iv. The *International Symposium on Cross-linguistic studies of Tonal Phenomena*, Tokyo University of Foreign Studies Presentation of paper and panellist during plenary discussion. (December 2000. On invitation - all dues paid by host).
- v. The *Training Workshop on XML Mark-up Language for Linguistic Applications*, University of Edinburgh. (May, 2001)
- vi. The *Workshop on Human Language Technology and Knowledge Management, 39<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Toulouse, France. Oral and poster presentation (On request by DACST - no cost involved). (July 2001)
- vii. The *COCOSDA international workshop on language resources and technology evaluation: International and regional perspectives*, Aalborg, Denmark. (Session chair and presentation of paper.) (September 2001)
- viii. The *Eurospeech 2001 International Conference*, Aalborg, Denmark. (Co-presenter of a poster.) (September 2001)
- ix. The *2<sup>nd</sup> South African Human-Computer Interaction Conference*, University of Pretoria, Pretoria. Presentation of paper. (September 2001)
- x. The *HCI-SA and ALASA Special Interest Group on HLT Workshop*, UNISA, Pretoria. Presentation of paper. (September 2001)
- xi. The 12<sup>th</sup> Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2001), Franschhoek. Co-presenter of two papers. (November 2001).

The Director has been invited by an international steering committee under the convenorship of Prof Antonio Zampolli (Pisa, Italy), to become a founding member of the *International Federation for Language Resources*, as a representative of Africa. This foundation meeting is to take place in Madrid on 17 September 2001. Due to the tragic events of 11 September 2001, this meeting has been postponed.

In order to build high level capacity, two of the (RUEPUS)/AST **staff members** were sent to **specialised training courses**<sup>4</sup> in the creation and maintenance of, respectively, text and speech databases:

- i. Mrs Philippa Louw participated in the *Text and Speech Corpora* ELSNET Summer School in Prague, Czech Republic, 16-23 July 2001.
- ii. Mrs Maritza van den Heuvel attended EUROLAN 2001, Summer Institute of *Creation and exploitation of annotated language resources* from 30 July to 11 August 2001 in Romania.

At this point in time it is probably safe to say that these two persons can hardly be replaced for the specialised work that they are doing. This expertise must be retained in one way or the other after the termination of RUEPUS and the AST project. These two persons form the epicentre of a proposal regarding the establishment of a National Electronic Text and Speech Resource Centre as envisaged in Report 1999/2000 (p 14) and as set as a future goal for the Unit.

<sup>4</sup> Attendance of these courses was sponsored in part by the AST project and through University research publication funds.

### 2.4.2 Visitors

Over the past seven and a half years the following visitors spent longer or shorter periods of time at RUEPUS, all contributing to capacity building in one form or the other. Previously unreported additions to this list appear in shadowed format.

**Prof. Greg Dogil**, Chair of the Institute for Computational Linguistics (IMS), University of Stuttgart (1996).

**Prof Tore Janson**, University of Göteborg, Sweden. (1997)

**Prof Ryohei Kagaya**, Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies (1997)

**Prof Seiji Niimi**, Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo. (1997)

**Prof John Ohala**, Department of Linguistics, University of California, Berkeley. (1997)

**Prof John Local**, Department of Linguistics, University of York, United Kingdom (1997)

**Prof Ludwig Gerhardt**, Department of Linguistics, University of Hamburg, Germany (1997)

**Dr Anne Cutler**, Scientific Director, Max Planck Institute, Nijmegen, The Netherlands (1997)

**Dr James McQueen**, Max Planck Institute, Nijmegen, The Netherlands. (1997)

**Dr Michael Jessen**, (Feodor Lynen Post-doc) Institute for Computational Linguistics (IMS), University of Stuttgart (1999).

**Prof John Rennison**, Department of Linguistics, University of the Witwatersrand (1999)

**Prof Ekkehard Wolff**, Institut für Afrikanistik, University of Leipzig, Germany. (1998 & 1999)

**Prof Arvi Hurskainen**, Institute for Asian and African Studies, University of Helsinki, Finland. (1999)

**Prof. Wim van Dommelen**, National Technical University of Trondheim, Norway. (2000)

**Prof Katherine Demuth**, Dept. of Cognitive & Linguistic Sciences, Brown University, Providence, USA. (2000)

**Prof Mark Johnson**, Dept. of Cognitive & Linguistic Sciences, Brown University, Providence, USA. (2000)

**Prof James Duplessis Emejulu**, Omar Bongo University, Gabon. (2000 & 2001).

### 2.5 Conferences and symposia presented (15)

Workshops/ Colloquia organised by RUEPUS since its inception. Previously unreported appearances on this list appear in shadowed format.

1995      **Workshop on The nature of phonetic and phonological interference** at

- the University of Potchefstroom. Presenters: Profs Monica Jacobs, Daan Wissing, and JC Roux.
- 1995 **Workshop** (2 days) on **Explanation in Phonology** as part of the *8th Biennial International Conference of the African Language Association of Southern Africa (ALASA 95)* in Stellenbosch. Presenters Profs Chuck Kisseberth and Farida Cassimjee (University of Illinois), JSM Khumalo (Wits), Dr J Claughton (Rhodes), and JC Roux.
- 1996 **Workshop** on **Experimental Phonetics and Speech Technology** at the University of Potchefstroom on 21 February 1996. Presenters: Prof G Dogil, Stuttgart, and JC Roux.
- 1996 **Workshop** on **Experimental Phonetics and Speech Technology** at the University of South Africa on 22 February 1996. . Guest: Prof G Dogil, Stuttgart, and JC Roux
- 1997 **Colloquium: The Phonetics of Phonology**, University of Stellenbosch. Presenters Profs John Ohala (UCL at Berkeley), JC Roux
- 1997 **Colloquium: The Phonetics of Phonology**, University of South Africa. Presenters Profs John Ohala (UCL at Berkeley), JC Roux
- 1997 **Colloquium: The Phonetics of Phonology**, University of Venda, and Thohoyandou. Presenters Profs John Ohala (UCL at Berkeley), JC Roux
- 1997 **Colloquium on the value of cross-linguistic research in psycholinguistics**, Stellenbosch. Presenters: Prof A Cutler and Dr J McQueen, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands & JC Roux
- 1997 **National Workshop on Speech, Text and Technology**, UCT. In co-operation with Prof C Brink Department of Applied Mathematics, UCT (Wide range of presenters.)
- 1997 **Two day workshop on 'The teaching of phonetics'**. *9<sup>th</sup> International Conference of the African Language Association (ALASA)*, University of Natal, Durban. Presenter: Ms P Swart & JC Roux
- 1999 **Colloquium** on **Experimental Speech Research**, at the University of the North, Qwa-Qwa branch, Phuthaditjhaba. Presenters: Dr M Jessen & JC Roux.
- 1999 **Workshop** on **African Language Engineering in the New Millennium**. At the *10<sup>th</sup> Biennial International Conference of the African Language Association of Southern Africa (ALASA 99)*. UNISA, Pretoria. Presenter: JC Roux
- 2000 **Workshop** on **Tagging and Glossing Sesotho**. Stellenbosch. Presenters: Profs. Mark Johnson and Katherine Demuth (Brown University, USA).
- 2001 **Workshop** on **Human Language Technologies**, UNISA, Pretoria. 12 & 13 September. Joint venture between ALASA-SIG for Language and Speech Technology Development and the African Speech Technology project.
- 2001 **Workshop** on **Human Language Technologies for e-governance in a multilingual society**. Stellenbosch 25/26 October. (Including three international speakers). AST-project organised by the SA Foundation for Language and Speech Technology Development (SAFLAST). Aimed at decision

makers and IT HODs in national government service. Due to the tragic events of 11 September 2001, which prohibited keynote speakers to attend, this meeting was postponed to a later, as yet undecided date.

## 2.6 Implementation and impact of research

**Programme 1 (Linguistic projects)** has demonstrated the intricate relationship between phonetics and phonology and the role of primary data in the process of the construction of linguistic theories. New insights have been presented regarding

- i. The nature of tonal perception in the communication process through various acoustic and perceptual experiments, and
- ii. The phonetic nature of click articulations in the Nguni languages through the use of innovative experimental techniques (e.g. Wigner Distribution analysis),
- iii. The nature and distribution of syllables in the Nguni and Sotho languages through quantitative analyses of text corpora.

Research results obtained in **Programme 2 (Resource development projects)** are finding their way not only into aspects of legislation on language issues, but also into the development of appropriate strategies to implement Human Language Technologies in South Africa. International networking has been a key feature of this programme.

**Programme 3 (Language and speech technology projects)** is setting the scene for a wide range of natural language applications in the South African context through the AST project. Capacity building and the acquisition of reusable electronic text and speech data are additional features of this programme.

## 3 GRADUATE TRAINING

Graduate training took place within different projects within the Unit or through facilitation by the Unit through the use of specialised equipment. The following students graduated since the inception of RUEPUS:

### 3.1 Completed MA candidates

- 1994** PW Lewis. Aspects of the phonological acquisition of clicks in Xhosa.
- 1994** TM Maxela. The phonology and phonetics of Xhosa demonstratives: A non-linear approach. (Cum laude)
- 1994** JN Mdekazi. An Autosegmental analysis of Xhosa nouns.
- 1996** KZ Selebeleng. Phonetic and phonological aspects of vowel raising in Sesotho (cum laude)
- 1997** PH Nkuna. Homorganic nasal assimilation in Tsonga: A Feature Geometry approach.
- 1998** BN Ntlabezo. Nasal assimilations in Xhosa: An experimental analysis.
- 1999** PH Swart. The prosodics of Imperatives in Xhosa: Implications for the development of a text-to-speech system. (Cum laude)

- 2001** I Olivier. Duurverskynsels by enkelvoudige nasale en nasaalverbindinge in Xhosa.
- 2001** FS Shoba. The representation of phonetic-phonological information in Nguni dictionaries. (Cum laude).
- 2001** M Kgasago. Palatalization in Northern Sotho.
- 2001** P Maepa. Theory and data: Nasal strengthening in Northern Sotho.
- (2002)** J Masalesa. Phonetic aspects of style variation in Sesotho. (March 2002)
- (2002)** PN Nemakhavhani. Aspects of assimilation processes in Venda. (March 2002)
- (2002)** N Rhadazhadhi. Linear and non-linear analyses of palatalisation in Venda. (March 2002)

### 3.2 In progress (MA)

P Madigoe. (December 2002)

N Lumadi. (December 2002)

### 3.3 Completed Doctoral Study (DLitt)

**1998** PW Lewis: Phonetic and phonological interference in Xhosa speech communication.

**2001** J Jones: Queclaratives in Xhosa: An acoustic-perceptual investigation.

### 3.4 In progress (DLitt)

S Naidoo: **Aspects of Zulu phonology: A feature geometry approach.**

OK Chikane The **interaction between morphology and phonology in Setswana.**

## 4 INFRASTRUCTURE

### 4.1 Staff

The following persons were appointed for the periods specified:

Ms Philippa Louw (MA)	Ad hoc 1/4/2000 -
Ms Maritza van den Heuvel (BA Hons)	Ad hoc 1/4/2000 -
Mr R Rhadzhadzhi (BA Hons)	Ad hoc 1/4/2001 -
Mr K Nemakhavhani (BA Hons)	Ad hoc 1/4/2001 -
Mr KP Kgasago (BA Hons)	Ad hoc 1/4/2000 -
Mr M Maepa (BA Hons)	Ad hoc 1/4/2000 -
Mr Vuyani Macingwane (BA Hons)	1/2/2001 - 30/11/2001

Ms S'bongile Madi (MA)	1/2/2001 - 31/12/2001
Mr James Masalesa (BA Hons)	1/2/2001 - 31/12/2001
Ms S du Plessis	1/4/2000 - 31/10/2001
Dr PW Lewis (DLitt)	1/2/2001 - 31/10/2001

#### 4.2 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. Electronic language resources are being developed through the AST project.

#### 4.3 Equipment

The AST project has contributed to a significant growth in technical infrastructure. Special arrangements will need to be negotiated at the end of the AST project with government on the use of this equipment.

### 5 FUTURE PLANNING: CREATION OF NATIONAL ELECTRONIC TEXT AND SPEECH RESOURCE CENTRE

A particular vision as well as a long-term goal of the Unit has always been the desire to be instrumental in the founding of a *National Electronic Text and Speech Resource Centre* (see previous reports as well as 1.1 and 1.2 of this report). One of the key aspects within the modern field of Human Language Technologies (HLT) is that of *re-usable resources*. No serious development in the field of HLT can take place without well-developed sets of electronic text and speech databases. These databases are typically used for academic research, or for the development of particular language based systems for a wide range of applications by developers.

National language resource centres typically

- i. Acquire text and speech data in electronic format from as many fields and modes as possible,
- ii. Perform first level annotation / mark up of the data in a preferred mark up language (currently XML), and develop automated annotation software,
- iii. Develop software tools (programmes) to manipulate and/or extract information from large databases
- iv. Serve as a distribution point ('clearing house') of data to typical users, which could be academics, dictionary units, developers etc.

As sets of annotated data and services presented by such a centre are commodities that may be sold, these centres eventually function as self-contained units. At the same time these centres act as living text and speech archives of the languages concerned. Over and above national language resource centres in different countries there are currently two major institutions in the world active in this field, i.e. the Linguistic Data Consortium (LDC) resident in Penn State University in the USA (<http://www ldc.org/>) and the European Language Resource Association (ELRA) in Paris (<http://www elra.org/>). These institutions focus on the main languages of the world with virtually no data on the languages of Africa.

The founding of such a centre in Africa will pave the way for serious work in the field of HLT on the continent.

The concept of a such a national resource centre has already been accepted by PANSALB (please see the attached full colour diagram indicating the position of such a centre within developments in South Africa (Appendix B). The Minister of Arts, Science and Technology has just appointed an Advisory Panel to determine a strategy for the implementation of this proposal (see also 2.3.3.1). The Director and a member of the RUEPUS Advisory Board are actively involved in this process.

As previously envisaged, there are probably enough grounds to transform RUEPUS into such a centre if the necessary financial support could be found. As such the Unit has the infrastructure to host such a venture as well as the expertise (*inter alia*, built up with the AST project.) The Director is in a position to form international networks (even with possible sponsors) as he is member of the Management Board of COCOSDA (the **International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques**) and has recently been invited to become a founding member of the **International Foundation for Language Resources** which is to be formed in September 2001 in Madrid, with the support of different European Union agencies.

The following tentative arguments may be forwarded for the development of a *National Electronic Text and Speech Resource Centre* (NETS) in South Africa, and at the University of Stellenbosch in particular:

- It will give content to the concept **Reusable Resources**: South Africa is at the brink of developing electronic resources of different kinds which would be of extreme importance for the academic communities as well as potential developers of HLT applications. In particular such a centre would be beneficial for
  - The National Lexicographic Units
  - The planning and development of various university based corpora (see applications to the NRF on ICT)
  - The development and distribution of private corpora

In order for data of this kind to be utilised effectively, i.e. being of such a nature that it may be mutually exchangeable, certain joint conventions and norms should be adhered to.

- It will **prevent the loss** of valuable electronic text and speech data derived from government sources, e.g. at national, provincial and local levels are being lost due to ignorance regarding the value of such data, or due to mismanagement.
- It will allow for an **infrastructure** to be created to allow for the co-ordination and distribution of annotated text and speech data for academic as well as commercial applications for instance through agreements with a body such as the SABC which has enormous resources (acquired *inter alia*, through tax payers' contributions).
- It will be in a **position to liaise** with potential consortium members such as the printed media which, by and large have an interest in the eventual development of text based information retrieval system, automatic summarisation etc. The media could not only benefit from such a centre, but could also make a significant contribution in supplying data in all languages in electronic format to be enhanced (annotated) in different ways for eventual product development.

- Such a centre will become a **living archive** of the nature of the difference language (both written and spoken) in South Africa. As was previously reported SA has lost an extremely valuable asset with the inexplicable demise of the SA Speech Archives. This has been brought to the attention of the relevant Government Subcommittee on Archives, however, nothing yet has transpired.
- The need for such a Centre is on the same level as that of five other **National Centres**<sup>5</sup>, which are directly sponsored by government through the NRF. All of these centres are in the fields of either natural and/or biological sciences. In view of the emphasis that has been placed on the development of HLTs as a priority area in SA (cf. Foresight, 1999 and Policy documents of the Department of Communications, as well as Public Administration) a good case could be made for the founding of such a centre to act as a catalyst for the whole movement.
- This centre could become a **centre for Africa**. A centre of this nature could initially focus on the SA context and at a later phase expand towards West and East Africa and the languages spoken in these areas. This could give some further support to the African Renaissance notion. The University of Stellenbosch has for instance, an exchange programme running with the Omar Bongo University in Gabon and students from various countries have been involved in the MPhil course in Computer Lexicography. We have a network of international colleagues (European, African and Asian) working in this area and who are extremely keen to co-operate in such a venture).
- This **international network** includes the following institutions: Universities in Germany: Stuttgart (Prof Dogil), Bielefeld (Prof D Gibbon) and Leipzig (Prof E Wolff), Finland (Prof A Hurskainen), Sweden - KTH Stockholm (Dr K Gustafsonn, Dr D House), Japan - Institute for Languages and Cultures of Asia and Africa (ILCAA) Tokyo (Prof S Kaji). All of these people are working in the field of data collection, data mining etc and support this initiative. Their expertise in the field could be acquired at any point in time.
- Such a centre could include much needed **pathological language and speech data** covering a wide range of pathologies. These data, which is of extreme importance for the medical sector and a venture such as this, could be co-sponsored by the Medical Council.
- The Centre could also act as a national resource centre for **Sign Languages**. One of the most foremost researchers in this field is from the University of Stellenbosch as well.
- We have **international support** from the *International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (COCOSDA)* as well as from the *European Language Resource Association (ELRA)* - official letters to follow.

## 5.1 Request 1

In view of the fact that

- NRF financial support for the Unit is to be terminated in December 2001, whilst the Unit will remain in place to host the *African Speech Technology* project until

<sup>5</sup> These 'National Facilities' are Hartebeesthoek Radio Astronomy Observatory, JLB Smith Institute for Ichthyology, National Accelerator Centre, and South African Astronomical Observatory.



December 2002, and

in order to

- Utilise the existing infrastructure to the benefit of South African linguists in general, and to
- Retain expertise in the field of electronic text and speech resource development,

it is **requested** that

- The Director be afforded a chance (with the support of the NRF) to devise a **strategic plan** through high level negotiations with relevant decision makers in academia, government, organised media and the private sector during 2002 in view of creating a sustainable, self contained resource centre, i.e. the *South African National Electronic Text and Speech Centre (SANETS.)*

Given the high priority assigned to the development of HLT in South Africa by, *inter alia*, the Foresight (1999) project and numerous official green/white papers of national and provincial governments, this centre could become a National Facility in the same vein as those currently being supported by the NRF (which are all in the field of natural science).

## 5.2 Request 2

- i. That any existing funds remaining after the completion of the activities in December 2001 be retained, and
- ii. That, if necessary, the Director be allowed to apply for *seed money* from the NRF at any point in time during 2002 in order to further the above mentioned aim.

## 5.3 Request 3

That the NRF takes note of this development and through liaison integrate it into any strategic plan that it may have, or wish to develop within the **focus area** Information and Communication Technology and the Information Society in South Africa.

# 6 BUDGET

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

Please see **APPENDIX C**

# 7 GENERAL

During the reported period the Unit has had some national exposure through the *Research Files of the NRF* (available on the NRF website), through newspaper reports in *Die Burger*, and through a radio interview of *Radio sonder Grense* with the Director.

The success of this Unit would not have been possible without the generous continuing financial support of, initially the **Centre for Science Development**, the **National Research Foundation** and the **Research Committee of the University of Stellenbosch**. This support was obviously greatly complemented by the contribution of the **Innovation Fund** of the Department of Arts, Culture, Science and Technology. This support is greatly appreciated. We (all co-workers to a variety of projects over the years)

hope to have demonstrated that a relatively small initial investment could create valuable academic as well as infrastructural benefits in the long run. The fact that the Unit was placed under to the focus area **ICT and Information Society in South Africa** with the amalgamation of the CSD and the NRF was very appropriate.

**The single greatest challenge that now remains is to ensure that the infrastructure and human expertise developed over eight years are appropriately secured in order for the University of Stellenbosch to maintain its leading role in this field.**

**PROF. JC ROUX**

**STELLENBOSCH**

**DIRECTOR: RUEPUS**

**5 DECEMBER 2001**