

**External Evaluator's Report on
Research Unit for Experimental Phonology**

University of Stellenbosch

December 2001

John Local
Professor of Phonetics and Linguistics
Department of Language and Linguistic Science
University of York, UK

| | | |
|----------|---|----------|
| 1 | PREAMBLE | 3 |
| 2 | SPECIFIC EVALUATION | 3 |
| 2.1 | THE EXTENT TO WHICH THE OBJECTIVES OF THE UNIT HAVE BEEN ATTAINED | 3 |
| 2.2 | THE QUALITY OF THE RESEARCH | 4 |
| 2.2.1 | RELEVANCE | 4 |
| 2.2.2 | INTERNATIONAL STANDARDS | 4 |
| 2.2.3 | SUBSTANTIVE AND METHODOLOGICAL DEVELOPMENT | 5 |
| 2.2.4 | THEORY BUILDING | 5 |
| 2.3 | THE QUANTITY OF THE RESEARCH | 5 |
| 2.3.1 | PUBLICATIONS AND MEDIA DISSEMINATION | 5 |
| 2.3.2 | REPLICATION OF RESEARCH | 5 |
| 2.4 | NATIONAL AND INTERNATIONAL NETWORKING | 6 |
| 2.5 | GRADUATE TRAINING | 6 |
| 2.6 | OTHER RESEARCH CAPACITY BUILDING | 6 |
| 3 | 3. RECOMMENDATIONS | 7 |
| 3.1 | CONTINUATION | 7 |
| 3.2 | TYPE OF RESEARCH | 7 |
| 3.3 | METHODOLOGICAL APPROACH | 8 |
| 3.4 | DISSEMINATION OF FINDINGS | 8 |
| 3.5 | IMPLEMENTATION | 8 |
| 3.6 | PROPOSAL | 9 |
| 4 | CLOSING OBSERVATIONS | 9 |

1 PREAMBLE

The Research Unit for Experimental Phonology was established at the University of Stellenbosch with funding from the CSD in 1994. In 1997 I acted as an External Evaluator for RUEPUS. At that time the Unit was unique in South Africa - a situation that still holds at the present.

I was able to report in 1997 that the research of RUEPUS was of international standard, there was dissemination of results to national and international audiences, progress was being made in capacity building and that overall RUEPUS was providing 'exceedingly good value for money'.

I have been invited again in 2001 to evaluate RUEPUS in the final year of its second cycle. In what follows I provide a formal assessment of the work and progress of the RUEPUS since 1997. I address my comments under the headings provided in the brief of the National Research Foundation (NRF). It is my understanding that due to a change in the policy at the NRF the award for the final year extends only to December 2001.

The initial aims of RUEPUS were primarily linguistic-theoretical and focussed on aspects of the description and analysis of the official languages of South Africa. There was a limited amount of research related to possible applications in speech technology. Since my evaluation in 1997 the focus of RUEPUS has changed significantly. Following recommendations presented to and accepted by the Unit's Advisory Board, the work of the RUEPUS now centres on research and development in the field of language and speech technology applications with specific reference to the indigenous languages of South Africa.

In order to provide this evaluation I have had extensive discussions with the Director and other members of the Unit, reviewed the Annual RUEPUS Reports (97-98, 98-99, 99-00, 00-01) related documents and publications and have seen demonstrations of the current work within the African Speech Technology (AST) project. (The RUEPUS Annual Report for 2000-2001 contains a complete record of activities and is attached as Appendix A.) I have also consulted with colleagues from Europe and elsewhere concerning their perspectives on the work of the Unit particularly in the Speech Technology domain.

My principal conclusions are that

- RUEPUS has delivered outstanding value for money and more than justified its NRF funding;
- The current African Speech Technology project conducted within RUEPUS is making impressive progress and puts the Unit at the forefront of language and speech technology research in SA;
- RUEPUS should continue its nationally important work as a National Resource Centre for Electronic Text and Speech.

2 SPECIFIC EVALUATION

2.1 The extent to which the objectives of the unit have been attained

RUEPUS has been remarkably successful in meeting (and in some cases) exceeding what might have been expected in respect of the achievement of its objectives. In all areas of its work the Unit has been substantively successful. Given the level of funding that the Unit has received it has excelled in continuing to attain the objectives set out in the Annual Reports. It has produced innovative linguistic research, established a strong network of partners and supporters within and beyond South Africa and has made genuine progress in research capacity building.

The work of the RUEPUS is now organised under three programmes (Linguistic projects; Resource development projects; Language and speech technology projects). This has been a sensible and cogent reorganisation that has allowed for development of the Unit's infrastructure and consolidation of its research strengths. In my opinion within each of these programmes the Unit has substantially met its key objectives.

One key objective, which merits particular mention, is that of fund raising activities to support research. The 1998-1999 RUEPUS Annual Report envisaged

'...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.'

With the prestigious DACST award of R9.08M for the African Speech Technology project RUEPUS has been spectacularly and (for a Unit situated within the Arts and Humanities) unusually successful. The AST project entitled Promoting the development of the official languages of South Africa through language and speech technology applications provides the basis for a significant step forward towards these aims.

2.2 The quality of the research

I concluded in 1997 that RUEPUS produced research output of a high standard. All available evidence shows that since then RUEPUS has continued to be consistently productive and has generated sustained high-quality outputs that have been published in national and international locations. The research has continued to address practical and theoretical issues of significance and has maintained genuine scholarship with accessibility. The dissemination strategy of the Unit is good and members of the RUEPUS are active participants at national and international meetings.

2.2.1 Relevance

The research undertaken by RUEPUS has both national and international relevance. It has the strength of combining theoretical insight and practical applicability. The Unit continues to produce excellent linguistic research and to make an important contribution to theoretical linguistics and to our understanding of the languages of South Africa. Projects have been seen through to successful completion and work within the unit has yielded novel findings regarding the perception of tonal contrasts in Xhosa, challenged conventional wisdom on the production of clicks and provided new and valuable statistical analysis of the structure and distribution of syllable types in the Sotho and Nguni languages. The work conducted on these projects is not simply of academic interest but also has practical application to the ongoing African Speech Technology project. This is especially to be welcomed.

Current work under the AST project is clearly focussed, well-directed to practical implementation and application and looks set to deliver substantive results via the prototype automated telephone hotel booking system.

2.2.2 International standards

The research work of RUEPUS, which is documented in the Annual Reports, is of a high standard and can hold its own with work of other international scholars. I have read a number of the phonetics/phonology publications produced by the Unit and find them to

be of an international standard of excellence. The research shows a solid understanding of contemporary theory and practice and the research methodologies employed evidence best practice. Work under the AST project is undertaken following best international practice. The international standing of Unit's work is evidenced in part by the regular acceptance of papers by members of RUEPUS at international conferences and symposia where serious refereeing of submissions is undertaken.

2.2.3 Substantive and methodological development

The Unit's research encompasses challenging areas of language description and analysis that involve methodologies from scientific and humanities domains. The recent work in the AST project has further engaged with a range of technological and implementation issues that should provide novel opportunities for methodological development

2.2.4 Theory building

Since 1997 RUEPUS has continued to general research results of theoretical relevance in linguistics phonetics and experimental phonology. This work makes a genuine contribution to scholarly understanding and contemporary linguistic theory (particularly in the domain of speech production).

2.3 The quantity of the research

2.3.1 Publications and media dissemination

The Unit has amply delivered in respect of research output and dissemination of research findings. It is very active in publishing its research. Since 1997 members of RUEPUS have had some 28 papers published in journals and conference proceedings (involving 13 different authors). There are a number of articles accepted for publication and yet more still being considered for publications or in progress.

Members of the Unit have been very active in making presentations at conferences. Since 1997 there have been 36 presentations at national and international conferences (full, verifiable information is presented in the RUEPUS Annual Report 2000-2001, Appendix A). Having been present at some of these presentations and having spoken to other senior colleagues who were present I can report that the work has enjoyed a good reception amongst the broader international research community and is recognised to be of importance.

The AST project began officially on January 1 2000 and various members of RUEPUS have presented papers on the ongoing research to international meetings in places as various as China, Denmark France, Finland, Greece, Switzerland. I myself heard a number of excellent presentations by members of the project at PRASA 2001 (the annual meeting of the Pattern Recognition Association of South Africa) at Franschoek this November. These presentations ranged from Database design and implementation, signal processing and dialogue modelling to a demonstration of the prototype of the first ever Xhosa speech synthesis system.

The research work of the Unit has also been given exposure through the medium of newspaper reports and a radio interview. I would expect that as the AST project develops that there would be increasing interest from a variety of media sources both within South Africa and beyond. It is to be hoped that the University of Stellenbosch will engage fully with the promotion of this research on the national and international stage.

2.3.2 Replication of research

Work conducted at RUEPUS is in the public domain and therefore open to replication. The Unit has developed/is developing publicly available software tools and databases that will facilitate other research groups in exploring the results of RUEPUS research.

However, in the context of the current and future work of the AST project there will be delicate issues of possible commercial confidentiality and Intellectual Property to manage.

2.4 National and international networking

RUEPUS has an excellent record in networking lead by the exemplary work of the Director. There is ample evidence of effective network building within South Africa and beyond with both formal and informal links. Good links exist within the University of Stellenbosch itself between RUEPUS and other departments. There is tangible evidence of active networking in the collaboration of the Unit with the University of Pretoria and Transkei that has resulted in the award of the funding for the AST project. Further evidence comes from the international links established with the Universities of Hamburg, Leipzig and Stuttgart. The link with Stuttgart, in particular, has provided a successful research collaboration that resulted in a number of joint publications and the appointment post-doctoral researcher (supported by funds from a Feodor Lynen Stipendium from the Alexander von Humboldt Stiftung and from the University of Stellenbosch). RUEPUS has also been effective in facilitating interaction of its staff with other academics and strengthened its academic network of contacts by welcoming a number of overseas visitors.

Members of the Unit have also been actively engaged in network building through their vigorous participation at national and international venues. Viewed from my position in Europe, the international profile and visibility of RUEPUS has been significantly strengthened over the period by this activity and by the Director's appointment to the editorial boards of *Afrika und Übersee* (Hamburg) and the *Journal of the International Phonetic Association* (JIPA) (Los Angeles/London), and his appointment to the management group of the International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (COCOSDA).

2.5 Graduate Training

For a Unit of the size of RUEPUS there has been good, solid delivery in respect of graduate training. A substantial number of higher degrees have been completed - 10 Masters and 2 Doctorates with others still in progress. (Full information is provided in the Annual Report 2000-2001.)

As well as delivering good quality training through graduate programmes, RUEPUS has seriously addressed its responsibility for research capacity building some more successful than others. The Unit has worked with the University of Transkei, the University of the North (Qwa-Qwa branch) and the University of Venda to provide training and research development - only the last of these has failed to be genuinely productive. The lack of success with the University of Venda appears to be in part a consequence of the lack of an appropriate infrastructure support at Venda itself.

These collaborations have produced a small core of researchers and skilled mother tongue speakers of isiXhosa and Sesotho who are currently all actively involved in the AST project. In addition to this it is clear from my discussions with the Director of the Unit that specific capacity building related to the training of a number of graduate students involved in projects is taking place on a continuous basis.

2.6 Other research capacity building

A second strand of research capacity building is being effectively pursued by RUEPUS through the development and collection of speech and text databases for isiXhosa, Zulu, Sesotho, Afrikaans and SA English (es).

Researchers in RUEPUS have also created publicly available software resources for speech

and language processing (documented in Annual Report 1997-1998). All these resources will provide indispensable infrastructure support for a developing Language Technology economy. They will also provide a reusable resource for other researchers engaged in the analysis and description of these languages.

It is impressive to record that the AST project involves 82 people at different levels and across different institutions (including the Universities of the North, Natal, Port Elizabeth, Pretoria, Potchefstroom and the Western Cape) of which more than 50% are drawn from the Historically Disadvantaged Communities. It is anticipated that apart from the capacity building consequences of the AST project the results that could be delivered could afford enormous opportunities for mother-tongue speakers to participate in future research projects and employment opportunities. The AST project has a significant contribution to make in this respect in 'bridging the digital divide' in South Africa.

3 3. RECOMMENDATIONS

3.1 Continuation

RUEPUS has established itself as a true international centre of excellence with talented and able staff in key positions. It is difficult to over-estimate the importance of the work of RUEPUS within the South Africa Language Technology context. It is the single most important grouping of individuals with great potential. It has developed a strong infrastructure and efficient and effective lines of management. The Unit has already established significant international links with key players in Language Technology and has good links with other key players (e.g. HE institutions, commercial sector interests and Governmental bodies) within South Africa.

In my opinion it is vital to the future development and success of Language Technology in South Africa that the work of RUEPUS be continued in the context of Speech and Language Technology. The most obvious and cost effective way of doing this would be to reconfigure RUEPUS as a National Resource Centre for Electronic Text and Speech.

I endorse the aim (Annual RUEPUS REPORT 2000-2001 §5) of creating such a Centre concur with the description given in the RUEPUS Annual Report for 2000-2001 of the key roles to be played by such a National Resource Centre for Electronic Text and Speech

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

I would add

- v. Define and implement standards for the collection, annotation and representation of electronic resources and deal with legal aspects of their access and distribution;
- vi. Manage and co-ordinate the strategic planning and delivery electronic language resources and liaise with key players engaged in resource development.

3.2 Type of Research

The redirection of RUEPUS towards work in Language Technology has been timely, well motivated and effective. This work should be continued. In particular the Unit should focus on the collection of electronic resources in speech and text for the official languages of South Africa. The Director should continue with the preparation of appropriate plans for the establishment of a National Resource Centre for Electronic Text and Speech that could consolidate this research activity.

3.3 Methodological Approach

The Unit should maintain and develop its successful interdisciplinary approach. The evolving core research agenda, manifest in the AST project, absolutely requires cross-disciplinary expertise. The AST project requires basic research, applied research, linguistic and technological expertise in order to implement knowledge and deliver management.

The successes of RUEPUS have been achieved through excellent leadership and effective management of the Unit. The management of RUEPUS is outstandingly good and evidences good practice. It is efficient and has promoted positive attitudes amongst staff and made astute appointments. The management style has clearly encouraged researchers in their pursuit of excellent work and has facilitated good interdisciplinary relations between the Unit and other bodies. It is clear that it is this quality management that has allowed the Unit to respond well and effectively to advice and changing circumstances.

If the vision of a National Resource Centre for Electronic Text and Speech is achieved it will be important that the management structures and skills developed under RUEPUS are appropriately implemented. In particular the current Director of RUEPUS (Prof. J. Roux) should continue as the Director of the new Centre with the speech database manager (Ms P. Louw) and text resources manager (Ms M. van den Heuvel) retaining their current roles. In addition it will be necessary to provide appropriate levels of administrative support for the successful running of such as Centre.

3.4 Dissemination of Findings

The policy of dissemination of results through publishing of research and scholarly papers, attending national and international conferences and running workshops/symposia should continue as part of any development. As the AST project progresses it is to be hoped that use is made of the full range of media outputs so that lay-accessible dissemination of the work can be pursued.

It is also hoped, too, that there will be an opportunity in future to revive the symposium planned for October 25 which had to be cancelled following the withdrawal of international speakers post the September 11 events in America. The symposium entitled Human Language Technologies for E-governance in a Multilingual Society, involving leaders of the AST project, members of the South African Foundation for Language and Speech Technology Development, would have provided opportunity for the important dissemination of ideas to governmental and other key players in South Africa.

3.5 Implementation

Results from the Resource Development Project have already contributed significantly to policy formulation on the development of Human Language Technologies in South Africa. A plenary presentation by the Director at the Second Language Indaba of National Government in Durban in May 2000 led to the recognition and eventual inclusion of different aspects of language technology in the final Language Policy and Plan for South Africa that is currently being debated in Parliament. These results also found their way into a strategic planning document entitled The development of Human Language Technologies in South Africa: Strategic planning (June 2000) of the Department of Arts, Culture, Science

and Technology (DACST) and the Pan South African Language Board (PANSALB). The acceptance by DACST and PANSALB of this plan, which arose directly from RUEPUS activities, has led to questions in National Parliament and eventually to the establishment of a Ministerial Advisory Panel on the Implementation of Human Language Technologies in South Africa. The Minister nominated the Director of RUEPUS to this panel.

The development of a National Resource Centre for Electronic Text and Speech (as proposed in the 1998-1999 and 2000-2001 Annual Reports of RUEPUS and accepted by the NRF-based Advisory Body) stands at the core of the above-mentioned implementation-plan. Given the origins of this plan it would be entirely appropriate for the University of Stellenbosch and the National Research Foundation to pro-actively seek to secure the establishment of such a Centre on the Stellenbosch campus. Such a centre, based inter alia on the track record of RUEPUS and its existing expertise, could continue as a driving force at national level and, with its international connections, could serve as a hub for language and speech technology developments in the rest of Africa.

3.6 Proposal

In order to take advantage of its already established leadership role in the field it is suggested that the University of Stellenbosch (e.g. Rector and Vice Rector research) and the National Research Foundation (e.g. President and Manager of the ICT focus area) engage in high level negotiations in a serious attempt to present and establish Stellenbosch as the host institution for a National Resource Centre for Electronic Text and Speech. I suggest that the University of Stellenbosch and the NRF should expeditiously formulate a joint proposal to be presented to the Ministerial Advisory Panel early in 2002. This would facilitate the smooth and timely continuation of research, expertise and staffing and the maintenance of a true 'centre of excellence'. Such a proposal might be expected to detail the infrastructure provided by the University of Stellenbosch as well the financial mechanisms for long-term commitment to the Centre.

4 CLOSING OBSERVATIONS

There is much to celebrate in the distinctive contributions that RUEPUS has made and continues to make. Since its inception it has produced quality research and made its findings known to the wider research community. It has promoted work on the indigenous languages of South Africa and recently, with the AST project, has consolidated this research in a technological framework. The Director of the Unit has worked with commitment, energy and dedication and has provided exemplary leadership that has enabled RUEPUS to be a success. Recent developments via the African Speech Technology project promise to deliver equally good results.

The NRF (and the University of Stellenbosch) is to be commended for funding this research unit. The investment has been handsomely repaid. The funding has not only allowed RUEPUS to conduct high-quality research but has allowed the Unit to establish the infrastructure and expertise which enabled the successful application to the Innovation fund of DACST for the African Speech Technology project.

I wish the Unit continued success in what would hopefully be its continuation as the National Resource Centre for Electronic Text and Speech in South Africa.

John Local
2 December 2001