CLST Annual Report 2008

CLST is the “Centre for Language and Speech Technology”. CLST operates as a separate unit within the Faculty of Arts of the Radboud University Nijmegen.

The year 2008 was a positive year for CLST.

CLST is active in an impressive number of projects in various domains. In 2008 we acquired projects that open up new horizons in:
- eHumanities
- Forensic speech processing

The projects are funded through grants from the EU, NWO (the Netherlands Organisation for Scientific Research), The Dutch language Union (STEVIN-programme), the Ministry of Economic affairs (IOP). A number of projects are carried out for companies (speech transcriptions, validation of resources, text mining).

At the end of 2008, eighteen researchers were actively involved in CLST projects.

The financial situation of CLST is stable.

Projects

Below we present an overview of the research topics addressed in projects\(^1\) which CLST carried out or participated in 2008.

eHumanities:
- A Propos (Pro-active text editing for researchers)
- BATS (Speaker tracking and topic detection in audio archives)
- D-kwadraat (A search interface for dialect databases on the internet)
- TM4IP (Text Mining for Intellectual Property)

Improving Automatic speech recognition:
- Autonomata Too (recognition of names)
- MIDAS (robust speech recognition)
- N-BEST (benchmark evaluation of speech recognizers of Dutch)
- NORISC (new ASR techniques to capture long-span speech processes of speech
- SPRAAK (a speech recognizer for Dutch)

Language acquisition and speech technology:
- ACORNS (Acquisition of Communication and Recognition Skills)
- Veni-project Odette Scharenborg (Avoiding the ham in hamster)
- ACMoLA (A Computational Model of Language Acquisition)

Computer assisted second language learning and speech technology:
- DISCO (ASR-based CALL for training oral proficiency In Dutch)
- TST-AAP (a demosystem for pronunciation training of Dutch)
- Repetitor (pronunciation error detection using ASR for L2 learners of Dutch)

Production of language resources:
- D-Coi (Pilot 50 M words text corpus of written Dutch)

\(^1\) See for detailed information on all projects: http://www.ru.nl/CLST
- SoNaR (a 500 M words text corpus of written Dutch)

Validation of language resources (by SPLEX):
- LILA (speech databases for voice driven telephony services in Asia and the Pacific region)
- LC-STAR II (lexicons for domain specific speech-to-text and text-to-speech)

A number of these projects were completed in 2008:
A Propos, D-COI, D-kwadraat, LILA, N-BEST, NO-RISC, SPRAAK

CLST also participates in Training Networks funded by the EC in the Marie Curie Programme:
- S2S (exploring how humans and computers understand speech using fine phonetic details)
- SCALE (Speech Communication with Adaptive Learning)

**Acquisition**

CLST acquired funding from the SURF foundation for the project “Veteran Tapes VP: Enriched publication based on multidisciplinary re-use of interview data”. This is an important project for extending our work in the eHumanities domain.

CLST further acquired another Marie Curie project: BBfor2. This project will create a training programme that will both address the fundamental biometric detection technologies issues and the typical forensic usage and requirements of these biometrics. These international network projects offer excellent opportunities for cooperation between and exchange of researchers in an international European context.

In the framework of the NWO Small Programmes a project was acquired entitled “Corrective feedback and the acquisition of syntax in oral proficiency”. This is an important acquisition for maintaining our front position in the domain of Computer assisted second language learning using language and speech technology.

With the LC-STAR consortium a new project for the validation of lexicons in various languages for speech technology purposes is scheduled. It is expected to start in 2009.

With Logica a series of new speech transcription tasks is planned for 2009.

**Participation in new initiatives and awareness**

CLST members are involved in a series of European initiatives with potential impact on the landscape of language and speech technology:

- CLARIN, which is a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily usable in the Humanities and Social Sciences.\(^2\)
- FLaReNet, which aims at developing a common vision of the area of language resources and language technologies, and fostering a European strategy for consolidating the sector, thus enhancing competitiveness at EU level and worldwide.\(^3\)

At national level CLST keeps track of new trends and developments as well. To that end, CLST participates at Board level in NOTaS. This is a foundation in which national companies and knowledge centers cooperate to stimulate the development and application of language and speech technology in the Dutch language.

\(^2\) http://www.clarin.eu
\(^3\) http://www.flaarenet.eu
**Personnel**

Five temporary staff members left CLST during 2008 due to project completion. Four of them have found employment elsewhere. Two new PhD members were employed. At the end of 2008 CLST consisted of three tenure staff members, six temporary staff members, six PhD students, and three members with a temporary part-time posting from the Department of Linguistics, summing up to 13 fte. The institute’s daily management is in the hands of a director with the support of a secretary.

Annual performance interviews were held with all staff members. These interviews serve to achieve a match between personal ambitions, job contents and career opportunities of individual employees.

For training and education of its personnel, CLST has a budget of 6 kEUR. Most staff members followed one or more courses, most of which were offered by the Radboud University as part of its career policy.

**Customer Satisfaction**

For most of its projects CLST is obliged to write progress reports. These progress reports follow the templates offered by the funding bodies (STEVIN, EU). Customer satisfaction in other projects (typically with companies) is monitored by e-mail or telephone. Customer responses are almost without exception very positive.

**Public Relations**

Every CLST research employee visited at least one international conference or workshop. Attendance of workshops is, as a rule, restricted to conferences to which the employee contributes an accepted paper.

Further PR activities in 2008 were:
- Website updates and improvement of its position in rank orders of search engines
- Folders and brochures (Dutch, English)
- Publicity in newspapers and periodicals (such as DIXIT, Kennislink, Onze Taal)

Henk van den Heuvel, Director CLST
H.vandenHeuvel@let.ru.nl
http://www.ru.nl/CLST