CLST Annual Report 2009

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.

Despite the financial recession characterising the general economical climate, CLST can look back at a positive year in which various new projects were acquired and started, and in which our financial situation remained stable.

CLST is active in an impressive number of projects in various domains. In 2009 we acquired projects that open up new horizons in:
- eHumanities
- Forensic speech processing
- Computer Assisted Language Learning

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, text mining).

During 2009, twenty-five researchers were actively involved in CLST projects.

Board

Board members in 2009 were: L. Boves (chairman), A. van Kemenade, D. Iskra, D. Kenyon-Jackson and N. Oostdijk. N. Schröder is advisor of the Board. Final responsibility lies with the Dean of the Faculty. CLST’s executive director is Dr H. van den Heuvel.

Projects

Below we present an overview of the research topics addressed in projects which CLST carried out or participated in in 2009.

eHumanities:
- BATS (Speaker tracking and topic detection in audio archives)
- TM4IP (Text Mining for Intellectual Property)
- LOHW (Living Oral History Workbench – ASR-based access to 250 interviews with veterans & development of a web-based annotation tool)
- VeteranTapes VP (Tools and facilities for making Enhanced Publications based on multidisciplinary re-use of interview data)

Improving Automatic speech recognition:
- Autonomata Too (recognition of names)
- MIDAS (robust speech recognition
- CompanionAble (Robust ASR for independent living of elderly people)

Computational modeling of language acquisition and processing:
- 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)

1 See for detailed information on all projects: http://www.ru.nl/CLST
- TST-AAP (a demosystem for pronunciation training of Dutch)
- FASOP (Corrective feedback and acquisition of syntax in oral proficiency)
- My Pronunciation Coach (Development of a pronunciation trainer for English as second language)

Production & standardisation of language resources (corpora and tools):
- SoNaR (a 500 M words text corpus of written Dutch)
- AAM-LR (Demonstrator web application for annotation of audio recordings)
- Adelheid (demonstrator web application for tagging historical Dutch texts)
- INTER-VIEWs (curation of 250 interviews – see LOHW – and implementation of a search tool)
- TQE (demonstrator web application for automatic quality evaluation of phonetic transcriptions in audio)

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)
- Bbfor2 (Bayesian Biometrics for forensics)

Two projects were completed in 2009:
ACORNS, Veteran Tapes VP

**Acquisition**

CLST acquired funding from the Ministry of Health and Sport (Programme Erfgoed van de Oorlog) for the project “Living Oral History Workbench”. In this project 250 interviews with veterans from World War II and the military missions in New Guinea and Netherlands East Indies are made accessible using ASR-techniques. Further a web based annotation tool is built for these interviews. This is an important project for extending our work in the eHumanities domain.

In the context of e-Humanities, four succesfull project proposals were written for the CLARIN-NL programme\(^2\). These are AAM-LR, Adelheid, INTER-VIEWs and TQE. Short descriptions can be found above.

Our cooperation with the University’s Language and Communication institute Radboud in’to Languages\(^3\) and Mynt Technologies\(^4\) resulted in a successful application for a Valorisation Grant from STW for a feasibility study of a web application for pronunciation training for Dutch learners of English. This application uses speech technology to detect pronunciation errors. Our objective is to write a successful proposal for the Phase 2 Valorisation Grant as well. This would be a great impetus to bringing the product (called My Pronunciation Coach) to the market.

CLST further acquired another Marie Curie ITN project: BBfor2. This project starts in 2010. It 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. International network projects such as Marie Curie ITNs offer excellent opportunities for cooperation between and exchange of researchers in an international European context.

---

\(^2\) [www.clarin.nl](http://www.clarin.nl)

\(^3\) [http://www.ru.nl/radboudintolanguagesuk/](http://www.ru.nl/radboudintolanguagesuk/)

\(^4\) [http://www.mynt-tech.com](http://www.mynt-tech.com)
Participation in new initiatives and awareness

CLST members are involved in 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.5
- 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.6

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.

Personnel

One temporary staff member left CLST in 2009 and found employment elsewhere. Six new employees were added to the team: three PhD students, two junior researchers and one post-doc. At the end of 2009 CLST consisted of three tenure staff members, eight temporary staff members, nine PhD students, and two members with a temporary part-time posting from the Department of Linguistics, summing up to 17.4 fte. The institute’s daily management is in the hands of a director with the support of a secretary. The table below shows a comprehensive overview of the fte distribution in CLST’s personnel.

<table>
<thead>
<tr>
<th></th>
<th>NWO-like projects</th>
<th>Other external funding</th>
<th>Overhead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenure staff</td>
<td>1.8</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>PhD students</td>
<td>5.5</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Other temporary staff</td>
<td>1.8</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>Secondments Dept. Linguistics</td>
<td>0.1</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Director &amp; Secretary</td>
<td></td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>9.2</td>
<td>8.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>

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.

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

5 http://www.clarin.eu
6 http://www.flarenet.eu
**PhD Defenses & Publications**

One CLST project resulted in an academic PhD defense in 2009. The PhD degree was part of the project obligations to which CLST had committed itself.


Publications by CLST employees are listed in the appendix.

**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. In all contributions the affiliation to CLST and the Radboud University is mentioned.

Further PR activities in 2009 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, Levende Talen

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


• Strik, H. (2009). Kilometers maken met AAP. In M. Senten (Ed.), Experiment NL, deel 2 (pp. 190-190). Quest in samenwerking met NWO.