CLST Annual Report 2013

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.

In 2013 CLST had its 10th anniversary! This was celebrated on November, 22 with a series of presentations and a mini-musical.

CLST was founded in January 2003. Its objective is to contribute, directly or indirectly, to the development of language and speech technology. CLST is active in fundamental and applied research, application development and consultancy. Ideal projects for CLST are research projects with a link to applications. Whenever possible, CLST publishes about these projects in research journals and conference proceedings. Projects are funded and carried out within national and international research programmes, while contracts from commercial parties are also acquired.

CLST focuses on research in and application of Language and Speech Technology (LST) in the following domains:

- Data mining & knowledge discovery
- Language learning and language teaching
- Communication in Health Care
- Forensics

Our cross-over activities are bundled in the domain:

- Resources and research infrastructure

Particular strengths of CLST are the following:

- Knowledge and expertise in speech technology, particularly in fundamental ASR technology and speech data
- Knowledge and expertise in language technology, particularly in the integration of rule-based and statistical processing of written language
- Combination of knowledge and expertise in the field of language and speech technology
- Combination of fundamental and applied research
- Strongly embedded in an eHumanities context
- Knowledge and expertise in the design, collection and validation of spoken and written language resources, and in the creation and maintenance of linguistic databases
- A firm position in a network of national and international contacts with additional expertise and knowledge

---

1 The programme of the event is still available at [http://www.ru.nl/clst/news/celebration-10-years/](http://www.ru.nl/clst/news/celebration-10-years/)
Board

Board members in 2013 were R. van Hout (chair), A. van den Bosch, D. van Leeuwen, P. Fikkert and N. Oostdijk, J. Oomen (Netherlands Institute for Sound and Vision) R. Smeulders (BrilliantBrains). N. Schröder, senior controller of the Faculty, is advisor of the Board. Final responsibility for CLST lies with the Dean of the Faculty. Executive director is H. van den Heuvel.

Projects

Below we present an overview of the research projects in which CLST was involved in 2013, sorted per research domain.

Data mining & knowledge discovery:
- TM4IP (Text Mining for Intellectual Property) (completed in 2013)
- Speech as metadata II: Speaker diarization for the Netherlands Institute of Sound and Vision
- Automatic detection of threatening tweets

Language learning and language teaching:
- ACMoLA (A Computational Model of Language Acquisition)
- DigLin (webtools for feedback on words and short sentences read by nonnatives who start reading)
- FASOP (Corrective feedback and acquisition of syntax in oral proficiency)
- GOBL (Games Online for Basic Language learning)
- HLT4LL (An interdisciplinary meeting on the contribution of HLT to the future of language learning) (completed in 2013)
- LangTERRA (EC network project)
- LIN (A validated reading level tool for Dutch)

Communication in Health Care:
- ComPoli (LST to improve patients’ communication possibilities with a eHealth webportal)
- INSPIRE (Speech processing in realistic environments, e.g. in hearing aids and cochlear implants)

Forensics:
- BBfor2 (Bayesian Biometrics for forensics)
- COST Action IC1106: Integrating Biometrics and Forensics for the Digital Age
- Automatic detection of threatening tweets (completed in 2013)

Resources, tools and research infrastructure:
- CLAM-Support (Supporting CLAM as interface for tools in a web application)
- DCS (Data Curation Service for CLARIN-NL) (completed in 2013)
- D-LUCEA (Database of the Longitudinal Utrecht Collection of English Accent) (completed in 2013)
- Elftal (Automatic Recognisers for Eleven African Languages) (completed in 2013)

2See for detailed information on all projects: http://www.ru.nl/CLST
- Nederlab (Repository of digitised Dutch texts and search interfaces)
- RemBench (demonstrator project interlinking databases centered around the life of the painter Rembrandt van Rijn)
- VALID (Curation of 5 databases with language pathology data collected in the Netherlands)
- TTNWW (Tools for Dutch LST as web services in a workflow) (completed in 2013)

In this overview we find projects from various funding sources:
- CLARIN-NL projects: CLAM-S, DCS, D-Lucea, RemBench, VALID, TTNWW
- EU projects: BBfor2, COST IC1106, DigLin, GOBL, INSPIRE, LangTERRA
- NWO, STW & ZonMW projects: ACMOLA, ComPoli, FASOP, LIN, Nederlab
- Other: Dutch Government (incl. Dutch Language Union), Netherlands Institute for Sound and Vision. Projects: Elftal, HLT4LL, Speech as metadata, Threatening tweets

CLST is the coordinating partner of two Marie Curie ITN projects, viz. BBfor2 and INSPIRE, and of two LLP projects, viz. DigLin and GOBL.

Last but not least, CLST was engaged in contract research with companies. We have set up collaborations in language technology with Davinci Products and Revisely. For CMG and ABN-AMRO we have carried out several batches of speech transcriptions. With X-MI Cross Media Interaction we collaborate in automatic speech recognition services.

**Acquisition**

In the new ambitious NWO programme Creative Industry, Helmer Strik and Toni Rietveld acquired a project named CHASING (CHAllenging Speech training In Neurological patients by interactive Gaming). The aim of the project is to study to what extent serious games and games’ principles motivate and support neurological patients in speech training, and to develop and test an interactive, intuitive, user adaptive game for speech training that makes use of advanced speech technology and is compatible with mobile platforms. The project duration is from 2014 to 2018 and involves a Phd student and a postdoc for CLST. Project partners are the Sint Maartenskliniek and Waag Society.

CLST started collaboration with the University of Groningen in building a writing assessment tool for the College voor Examens (the organization that is responsible for central exams for Dutch secondary schools).

We prolonged our collaboration with the Netherlands Institute for Sound and Vision with a new project on improving speaker recognition and exploiting speaker information for user searches.

With pride we mention the grant of a patent applied for by Helmer Strik and Catia Cucchiarini Automated system for training oral language proficiency (WO 2013172707 A2), 21 Nov 2013. This patent is also important for our spin-off company Novolanguage®, in which we will formally start participating in the course of 2014.

**Participation in new initiatives and awareness**

At the national level CLST keeps track of new trends and developments. To this end, CLST’s director participates at Board level (Treasurer) in NOTaS (Nederlandse Organisatie voor Taal-en Spraaktechnologie), an organization in which national companies and knowledge centers cooperate to stimulate the development and application of language and speech technology in the Dutch language. In addition, he is member of the ELRA Board. ELRA’s missions are to promote language resources for the Human Language Technology (HLT) sector, and to evaluate language engineering technologies³.

⁴ [http://www.elra.info](http://www.elra.info)
CLST is involved in CLARIN-NL, which is a large-scale national collaborative effort to create, coordinate and make language resources and technology available and readily usable in the Humanities and Social Sciences.\(^5\) In CLARIN-NL Antal van den Bosch and Nelleke Oostdijk are members of the national advisory panel. They are also member of the technical advise committee of Taalportaal\(^6\) and of the editorial board of CLIN\(^7\). In Nederlab, Antal van den Bosch is member of the Steering Committee as chair of the technical track. CLST is represented in the International Speech Communication Association Special Interest Group (ISCA SIG) on Speech and Language Technology in Education (SLaTE\(^8\)) by H. Strik (vice-president) and C. Cucchiarini.

CLST members are also involved in European initiatives pertaining to the language resources infrastructure. Involvement is considered of some importance as the infrastructure impacts on the possibilities for research in the area of language and speech technology:

- CLARIN-ERIC\(^9\)
- 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.\(^10\)
- Meta-Net which is a Network of Excellence dedicated to fostering the technological foundations of a multilingual European information society. META-NET is building the Multilingual Europe Technology Alliance (META). Bringing together researchers, commercial technology providers, private and corporate language technology users, language professionals and other information society stakeholders.\(^11\)
- CLARIAH, which aims to design, construct, and exploit a facility for eHumanities research.\(^12\)
- COST IC-1106\(^13\)

CLST is member of the LDC and of ELRA, and as such an important entry for the acquisition of language resources by CLS researchers.

ICT infrastructure

Together with the Department of Linguistics and the Department of Business Communication (Funds Antal van den Bosch), CLST set-up a powerful computing and storage infrastructure at the Faculty of Science. The equipment was expanded in 2013 and now consists of seven 16-core computing units and over 175 TB disk storage. CLST values this infrastructure as vital substitute of our computing and storage capacities. Co-financed by CLS, we have employed an ICT-specialist and programmer for system maintenance and user support.

In 2013 we formulated and published a policy for software maintenance for CLST-employees.

Personnel

Eight temporary staff members left CLST in 2013\(^14\) and found employment elsewhere. Four new employees were added to the staff, two junior researchers\(^15\), one senior researcher\(^16\) and one ICT-specialist\(^17\). During 2013 the staff employed by CLST consisted of five tenure staff members, nine

\(^5\) [http://www.clarin.nl](http://www.clarin.nl), As of Feb. 29\(^{th}\), 2012, CLARIN is also an official ERIC
\(^6\) [http://www.taalportaal.org](http://www.taalportaal.org)
\(^7\) [http://hmi.ewi.utwente.nl/clin2013/](http://hmi.ewi.utwente.nl/clin2013/)
\(^8\) [http://www.sigslate.org](http://www.sigslate.org)
\(^9\) [http://www.clarin.eu](http://www.clarin.eu)
\(^10\) [http://www.flarenet.eu](http://www.flarenet.eu)
\(^12\) [http://www.clariah.nl](http://www.clariah.nl)
\(^13\) [http://www.cost.eu/domains_actions/ict/Actions/IC1106](http://www.cost.eu/domains_actions/ict/Actions/IC1106)
\(^14\) Christina Bergmann, Polina Drozdova, Michele Gubian, Miranti Indar Mandasari, Rahim Saedi, Yang Sun, Bouke Versteegh, Maarten Versteegh
\(^15\) Polina Drozdova, Vanja de Lint
\(^16\) Martin Reynaert
\(^17\) Wessel Stoop
temporary staff members (of which five postdocs), nine PhD students, and two members with a temporary part-time posting from the Department of Linguistics, summing up to 13.1 fte (excluding 0.3 ICT support). The institute’s daily management is in the hands of the director with the support of a secretary (together 0.5 fte). The Table below shows the distribution of the CLST staff (in fte) over various types of contracts and financing. Student assistants are excluded from this overview.

<table>
<thead>
<tr>
<th></th>
<th>NWO-like projects, incl. CLARIN-NL</th>
<th>Other external funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenure staff</td>
<td>2.2</td>
<td>0.8</td>
</tr>
<tr>
<td>PhD students</td>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Other temporary staff:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>1.0</td>
<td>0.75</td>
</tr>
<tr>
<td>Postdoc</td>
<td>1.1</td>
<td>1.25</td>
</tr>
<tr>
<td>Secondments Dept. Linguistics</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Sum</td>
<td>7.3</td>
<td>5.8</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. However, staff members preferably follow courses offered for free by Radboud University to its employees as part of the University’s career policy.

**Internships**

This year have started extra efforts in training and educating students in LST-related research and development. Apart from attending fairs with internships offerings for students, we maintain a webportal where proposals for internships are displayed[^18], both at Bachelor and at Master level.

**Customer Satisfaction**

For most of its projects CLST is obliged to write progress reports. These progress reports follow the templates provided by the funding bodies (NWO, EU, CLARIN-NL). Customer satisfaction in other projects (typically with companies) is monitored on a regular basis (by personally contacting the customer e-mail or telephone). Business customers such as Davinci and Revisely articulate that they value our work and would like to extend and intensify collaboration.

**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. Where applicable the affiliation to CLST and the Radboud University is mentioned.

Further PR activities in 2013 were:

- Website demo updates[^19] now including:
  - Demo of the Oral History Annotation tool from the INTER-VIEWS project
  - Data search in video recordings of popular TV series: Boer zoekt vrouw (BATS project)
  - Spelling correction for Dutch (Valkuil) and English (Fowlt)
  - Courses on the spelling of Limburgian dialects.


**PhD Defenses & Publications**

One CLST project resulted in an academic PhD defense in 2013:


In CLST projects with PhD students, supervision of these students requires a substantial time investment by faculty staff members. By way of compensation CLST can provide extra funding for PhD students to complete their theses when faculty funds are exhausted. This is decided on a case by case basis by the CLST-director. In 2012 such a contribution resulted in a successful PhD defense by Folkert de Vriend. For 2014 such a provision has been made for Steve Bodnar.

Publications by CLST employees or others involved in CLST projects are listed in the Appendix.

---

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

---


Publications 2013

Journal papers:


- http://dx.doi.org/10.3389/fpsyg.2013.00676 (Metis)

- http://dx.doi.org/10.1162/COLI_a_00149 (Metis)

- http://www.clinjournal.org/node/29 (Metis)


- http://dx.doi.org/10.1007/s10579-013-9236-1 (Metis)


- http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6584746&queryText%3Dmandasari (Metis)


Dissertation:

Book chapters:


Proceeding papers:


- http://dx.doi.org/10.1109/ICB.2013.6613025 (Metis)


Radboud Universiteit Nijmegen

9


**Patent:**