

# ru | STEVIN

Dutch-Flemish research programme for  
Dutch Language and Speech Technology

## STEVIN mid-term evaluation Fact File



May, 2008

## STEVIN fact file

### Contents

Title and duration programme	STEVIN: Flemish-Dutch research programme for Dutch Language and Speech Technology (STEVIN: Spraak- en taaltechnologische voorzieningen voor het Nederlands) – 2004 - 2011	
Summary	Summary of the objectives of the DUTCH-Flemish STEVIN Programme	page 3
STEVIN organisation	<ul style="list-style-type: none"> <li>• HLT-Board</li> <li>• STEVIN Programme Committee</li> <li>• International Assessment Panel</li> <li>• STEVIN Programme bureau and some supporting working groups</li> </ul>	page 4
STEVIN Budget and Funding Organisations	<p>STEVIN budget = 11.4 M€: Flanders 3.8 M€, the Netherlands 7.6 M€</p> <ul style="list-style-type: none"> <li>• The Netherlands: Ministry of Education, Culture and Science), Netherlands Organisation for Scientific Research. Ministry for Economic Affairs</li> <li>• Flanders: EWI, Department of Economics, Science and Innovation of the Flemish Government</li> </ul>	page 5
STEVIN Calls	STEVIN Funding instruments – assessment procedures and statistics Distribution funding over Dutch-Flemish/academic-industrail recipients	page 6
STEVIN Funding Instruments (and their max. budget) 2003-2007	<ul style="list-style-type: none"> <li>• 1st Call for Proposals for strategic research proposals and HLT infrastructures (data &amp; tools) (max. budget 2 M€)</li> <li>• 2nd Call for Proposals for strategic research proposals and HLT infrastructures (data &amp; tools) (max. budget 3,8 M€)</li> <li>• Three Calls for tender for specific HLT infrastructures (max budget 1,6 M€)</li> <li>• Call for proposals for applied research (max. budget 2,3 M€)</li> <li>• Three Calls for proposals for demonstration projects (max. budget 1 M€)</li> <li>• Call for proposals for educational projects (max. budget first call € 27.500)</li> </ul>	page 8
STEVIN R&D priorities	Overview STEVIN R&D priorities (as were already included in the original STEVIN programme description) and how they are covered by the STEVIN projects funded in the different funding schemes.	page 12
STEVIN assessment criteria	Standard STEVIN R&D assessmentcriteria (as were already included in the original STEVIN programme description) and code of conduct for the STEVIN IAP and STEVIN PC	page 14
STEVIN project details	Overview of STEVIN projects, including statistics with respect to STEVIN priorities covered and project details as to consortia partners, budget, duration and project summary.	page 17
STEVIN IPR	STEVIN IPR and standards policy, schematic overview HLT actors, leaflet for data providers	page 59
HLT activities	List of HLT activities organised by or financially supported by STEVIN	page 62

## Introduction – Summary STEVIN Objectives

Dutch is ranked as the 40th most widely spoken language of the world's 6,000 languages. Most of the 22 million Dutch native speakers live in the Netherlands and the Flemish region of Belgium. Nevertheless the market for human language technology for Dutch (HLTD) is too limited to attract important investments by industry in HLTD. Therefore, cross border cooperation among governments, businesses and academia has been established, resulting in a Flemish/Dutch HLTD research programme. The programme is called STEVIN, which is a Dutch acronym for 'Essential Speech and Language Technology Resources for Dutch'.

The STEVIN programme for Dutch language and speech technology is a coordinated effort of:

- o de Nederlandse Taalunie;
- o het departement Economie, Wetenschap en Innovatie van de Vlaamse Overheid (EWI);
- o het Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen);
- o het Fonds voor Wetenschappelijk Onderzoek - Vlaanderen (FWO-Vlaanderen);
- o het Nederlandse Ministerie van Onderwijs, Cultuur en Wetenschap (OCW);
- o het Nederlandse Ministerie van Economische Zaken (EZ);
- o het Nederlandse EZ-agentschap SenterNovem;
- o de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO);
- o het NWO-gebied Exacte Wetenschappen (EW);
- o het NWO-gebied Geesteswetenschappen (GW).

This six-year programme aims to contribute to the further progress of HLTD in Flanders and the Netherlands and stimulate innovation in this sector. In addition, it will strengthen the economic and cultural position of the Dutch language in the modern ICT-based society.

The STEVIN-programme was launched in 2005. It is jointly financed by the Flemish (Department of Economy, Science and Innovation) and Dutch governments (Ministry of Education, Culture and Science, Ministry of Economic Affairs and the Netherlands Organisation for Scientific Research).

STEVIN will run until 2010 with a total budget of 11.4 million euros. STEVIN is coordinated by the Dutch Language Union and supervised by a board of representatives of the funding bodies. A programme committee, including both academic and industrial representatives, is responsible for scientific and content related issues. An international assessment panel of highly-respected HLT-experts evaluates the submitted R&D proposals. A programme office, a joint collaboration of the Netherlands Organisation for Scientific Research and the Dutch innovation agency SenterNovem, takes care of the operational matters.

The STEVIN goals are defined in the framework of a stratified innovation system. Each actor of the system needs to be addressed, resulting in a mix of funding instruments and programme activities.

- funding for collaboration between universities and between universities and industry to realise an adequate basic language resources kit
- funding strategic R&D on HLTD
- funding appealing demonstration projects illustrating the feasibility and value of HLTD applications
- funding networking activities to stimulate HLTD knowledge transfer
- organising networking activities and HLTD promotional events

To enable the use and re-use of STEVIN results, a particular IPR-arrangement has been set up. The materials (software, data etc.) must be handed over to the Dutch Language Union so they can be made available to third parties through the Dutch HLT Agency ('TST Centrale' [www.tst.inl.nl](http://www.tst.inl.nl)). The Dutch HLT Agency helps resolve IPR issues, is responsible for the management, maintenance and distribution of HLTD materials, and also acts as a servicedesk. The STEVIN IPR and standards policy is described in detail in one of the annexes to this factfile. More information on the Flemish/Dutch STEVIN programme and the STEVIN projects can be found on: [www.stevin-tst.org](http://www.stevin-tst.org).

## STEVIN organisation – Tasks and Responsibilities

### Programme Board : (composition as of January 2007)

The Board of the STEVIN programme is responsible for the final funding decisions. The Board is also responsible for supervising two activities that are related to the STEVIN programme, i.e. the *TST-centrale* and the *Makel en Schakel* activities carried out by the *Nederlandse Taalunie*.

- o the general secretary of the Nederlandse Taalunie;
- o representatives of the partners financing the STEVIN programme:
  - \* the Flemish Government department Economy, Science and Innovation (EWI)
  - \* the Institute for the Promotion of Innovation by Science and Technology in Flanders
  - \* the National Fund for Scientific Research (Belgium) (FWO)
  - \* the Dutch Ministry of Education, Culture and Science (OCW)
  - \* the Dutch Ministry of Economic Affairs (EZ)
  - \* the Netherlands Organisation for Scientific Research (NWO, NWO-GW, NWO-EW))
- o two senior language and speech technology experts: prof. Dirk Van Compernelle, Leuven University and prof. John Nerbonne, Groningen University.

### Programme Committee (composition as of January 2008)

- o Prof. dr. Jan Odijk (Utrecht University (formerly also Nuance)) – chair
- o Prof. dr. Jean-Piere Martens (Gent University)
- o Prof. dr. Frank van Eynde (Leuven University)
- o Prof. dr. Walter Daelemans (Antwerpen University)
- o Mw. Debbie Kenyon-Jackson (Cumlingua bv Nijmegen)
- o Dr. Arjan van Hessen (Telecats BV / Twente University)
- o Prof. dr. Louis Boves (Radboud University Nijmegen)
- o Drs. Remco van Veenendaal (INL / TST centrale)
- o Dhr. Jan van Sas (The LingWareHouse / Karel de Grote-hogeschool Antwerpen)
- o Dr. ir. Kris Van Bruwaene (VRT)
- o Dr. ir. Ruud Smeulders (Rabobank Groep ICT, IBA)
- o Dr. Leonoor van der Beek, Q-go Amsterdam.

### STEVIN International Assessment Panel: assessment and ranking (composition per 1/1/2007)

- o Prof. dr. Hans Uszkoreit (DFKI - Germany)
- o Prof. dr. Gábor Prószéky (Morphologic - Hungary)
- o Prof. dr. Roger Moore (Sheffield University – UK)
- o Dhr. Paul Heisterkamp (DaimlerChrysler - Germany)
- o Dr. Gilles Adda (LIMSI - France);
- o Dr. Nicoletta Calzolari (ILC - Italy)
- o Dr. Stelios Piperidis (ILSP - Greece)
- o Prof. dr. Anne Abeillé (Université Paris 7 – France).

### Programme Office

Together the Dutch organisations NWO and SenterNovem have been selected by the *Nederlandse Taalunie* to form the STEVIN Programme Bureau that coordinates the STEVIN activities, including the handling and selection process of applications from both Dutch and Flemish applicants.

- o NWO Alice Dijkstra, Brigit van der Pas
- o SenterNovem Martijn Nuijten
- o Nederlandse Taalunie Peter Spyns, Elisabeth D'Halleweyn

### Furthermore some advisory working groups have been set up:

- o Working Group for STEVIN supporting activities (which includes representatives from non-STEVIN programmes and projects), to coordinate HLT supporting activities in the Netherlands and Flanders.
- o IPR Working Group (led by the Dutch Language Union, includes academic and industrial HLT experts on IPR and legal experts), to co-ordinate and optimize STEVIN IPR practices.

## Summary Dutch-Flemish STEVIN Programme budget

Of the total STEVIN budget 1/3<sup>rd</sup> is funded by the Flemish government and 2/3<sup>rd</sup> is funded by a consortium of Dutch ministries and funding organisations.

Funding by Dutch and Flemish government and funding organisations		
Flanders	€ 3.800.000	*
the Netherlands	€ 7.600.000	**
interest 2.5%	€ 237.000	
	<b>€ 11.637.000</b>	
<p>* Dutch funding provided jointly by the Ministry of Education, Culture and Science), the Netherlands Organisation for Scientific Research (GW, EW, AB) and the Ministry for Economic Affairs</p> <p>** Flemish funding provided by the Department of Economics, Science and Innovation (EWI) of the Flemish Government</p>		
Budget STEVIN funding schemes, supporting activities and management		
R&D projects	€ 8.850.000	<b>76,0%</b>
Demonstration projects	€ 1.009.000	<b>8,7%</b>
Supporting activities	€ 837.000	<b>7,2%</b>
Dutch HLT Archive	€ 300.000	<b>2,6%</b>
STEVIN management	€ 641.000	<b>5,5%</b>
	<b>€ 11.637.000</b>	

As can be seen in the table above: 76% of the budget is spent on R&D projects that were funded in one of the three open calls or in one of the three calls for tender. More than 8.5% was spent for demonstration projects which may stimulate demand for HLT technology. For making sure STEVIN results will be supported and become widely available via the Dutch HLT Archive 2.6% of the total budget, or 3% of the R&D budget is reserved. For the management of the STEVIN programme 5.5% of the programme budget will be spent.

## STEVIN Funding instruments – assessment procedures and statistics

### STEVIN handling agencies

NWO has acted as main handling agency for the open calls for strategic research projects and projects aiming at realizing part of the Dutch basic language resources kit and the calls for tender for i) A speech recognition toolkit for Dutch, ii) A lexical resource for the semantic processing of Dutch and iii) An annotated written Dutch corpus.

SenterNovem has acted as main handling agency for the three calls for demonstration projects and the call for proposals for educational projects.

### STEVIN subsidieregeling

The legal rules applying for the specific Dutch-Flemish granting schemes are laid down in the *Subsidieregeling van de Nederlandse Taalunie tot subsidieverstrekking in het kader van Nederlandse taal- en spraaktechnologie "STEVIN"* (STEVIN-subsidieregeling), which are available from the STEVIN website.

### Selecting the best proposals & managing conflicts of interest

The Nederlandse Taalunie has formally assigned the task of STEVIN Programme Bureau jointly to NWO and SenterNovem. NWO's primary responsibility is to organise the assessment procedure of the open calls and the calls for tenders. For each open call NWO has invited the International Assessment Panel to come to either Amsterdam or Brussels to assess and prioritize all proposals. Proposals were evaluated on the basis of a set of assessment criteria already laid down in the formal STEVIN project description and repeated in the formal call publications. For the calls for tender tender-specific criteria were added. The assessment meetings of the IAP and PC were attended by two representatives from NWO and one from SenterNovem who gave special attention to safeguarding the fairness of the assessment procedure.

One of the main concerns of the Programme Bureau was to deal with conflicts of interest as especially members of the PC would have personal involvement in one of more applications. Considering the size and connectedness of the language and speech technology community in The Netherlands and Flanders, it is like in any other innovation-oriented research area, not realistic to exclude all involvement.

However a number of actions was taken to secure that only the best proposals were selected. One of the main actions take was to have an international panel of experts fly in for the selection process. Furthermore, modelled on the code of conduct used by the European Commission for its Framework Programme, a STEVIN Code of Conduct was formulated for the IAP and PC. All members were required to sign a declaration of conflict of interest and confidentiality and to formally indicate in which – if any – of the submitted project they had any involvement. In doing so, the members committed themselves to strict confidentiality and impartiality concerning their tasks. If a member has a direct or indirect link with the project(s), or any other vested interest, or is in some way connected with the project(s), or has any other allegiance which impairs or threatens to impair his/her impartiality with respect to the project(s), the STEVIN Programme Bureau has ensured that those members did not participate in the review and ranking of the project(s) concerned.

In a two-day meeting, all eligible applications were assessed and ranked by the IAP. The assessment reports were sent to the applicants for a response. Subsequently, on the basis of a) the IAP assessment and ranking, b) the applicants response and c) knowledge of the Dutch and Flemish HLT field, the PC added their remarks to the IAP assessment and also ranked the proposals. In doing so, it was possible to incorporate a Dutch-Flemish perspective in the assessment procedure, which could not be obtained from the international experts.

The final funding decision was made by the HLT Board -- consisting of representatives of the Flemish and Dutch government, along with two senior experts from the field -- on the basis of 1) the IAP assessment and ranking 2) the responses of the applicants, 3) the PC assessment and ranking including a description of way the assessments and ranking were arrived at in the meeting and an explanation for possible differences with the ranking given by the IAP.

## STEVIN Funding instruments

A number of funding instruments (open calls and calls for tender) were implemented:

1. a 1<sup>st</sup> open call (September 2004) – maximum budget 2 M€ - for focussed short term (maximum duration is 2 years) strategic research projects and projects aiming at realizing part of the Dutch basic language resources kit a self-contained result;
2. a 2<sup>nd</sup> open call (in the spring of 2005) – maximum budget 3.4 M€ - for more complex strategic research projects and projects aiming at realizing part of the Dutch basic language resources kit with a longer time frame;
3. a 3<sup>rd</sup> open call (in 2007) – maximum budget 2.3 M€ - for application-oriented research projects;
4. Three calls for tender:
  - A) a call for tender (in 2005) – maximum budget 800 k€ - for i) A speech recognition toolkit for Dutch and ii) A lexical resource for the semantic processing of Dutch
  - B) a call for tender (in 2007) – maximum budget 836 k€ - for An annotated written Dutch corpus;
5. three calls for demonstration projects (in 2005, 2006 and 2007) – maximum budget 1 M€ - for small SME supporting projects stimulating HLT demand;
6. Call for proposals for educational projects (max. budget first call € 27.500).

More details about these calls are given in the next sections.

## Summary statistics STEVIN funding R&D and demonstration projects (in k€)

In the table below an overview is presented of how STEVIN funding awarded to R&D projects and demonstration and education projects was distributed over

- a) Dutch and Flemish partners: 64% - 36%
- b) Academic and industrial partners: 83% - 17%

<i>Distribution STEVIN R&amp;D funding over Dutch-Flemish recipients</i>			
<b>Netherlands</b>			
Universities	4.995		
Industry	1.245		
<b>Total Netherlands</b>		<b>6.240</b>	<b>64%</b>
<b>Flanders</b>			
Universities	3.133		
Industry	450		
<b>Total Flanders</b>		<b>3.583</b>	<b>36%</b>
<b>Total STEVIN R&amp;D funding</b>		<b>9.823</b>	

<i>Distribution STEVIN R&amp;D funding over academic and industrial recipients</i>			
Universities	8.128		<b>83%</b>
Industry	1.695		<b>17%</b>
<b>Total STEVIN R&amp;D funding</b>		<b>9.823</b>	

From the figures above one may conclude that the realizations meet the target percentages set by the Dutch and Flemish funding organisations.

## 1. 1st Call for Proposals for strategic research proposals and HLT infrastructures (data & tools) for focussed short term projects with a self-contained result (maximum duration is 2 years) (max. budget 2 M€)

### Objectives STEVIN 1st open call for proposals

Proposals had to relate to basic linguistic resources (tools and data), fundamental strategic research and applications in the areas of language and speech technology, all of which had to contribute to an appropriate digital language infrastructure for Dutch. Proposals could be submitted both in the area of language technology, and in the area of speech technology, and were preferably relevant to both areas. For cross-border consortiums the standard bench fee was increased by 50%.

### Evaluation procedure full proposals submitted in the 1st open call

All applications were presented to a panel of international experts in language and speech technology. This international panel evaluated the applications based on the assessment criteria mentioned in the call and formulated a set of recommendations for the STEVIN Programme Committee. Due to time constraints, the procedure for this call did not allow the applicants to respond to the panel's recommendations. Based on the applications and the panel's recommendations, the Programme Committee also assessed the applications and determined the order of priority of the eligible proposals. On the basis of both the IAP advice and the Programme Committee's advice, the Board of the STEVIN programme finally determined which projects were funded.

### Time frame STEVIN 1st call – 2004 – length call procedure 3 months

- \* September 15: opening call and brokerage event in Tilburg
- \* November 2: closing date call: 19 proposals were submitted
- \* November 25 and 26: assessment and ranking of all proposals by STEVIN IAP
- \* December 3: assessment and ranking of all proposals by STEVIN PC
- \* December 15: determining short list by Board of the STEVIN programme

### Statistics: Number and percentages submitted and funded proposals, listed by type

	speech	language	speech/language combined	total
Submitted proposals	6 (30%)	9 (50%)	4 (20%)	19 (100%)
Funded proposals	2 (40%)	3 (60%)	0 (0%)	5 (100%)

## 2. 2nd Call for Proposals for strategic research proposals and HLT infrastructures (data & tools) (max. budget 3,8 M€)

### Objectives STEVIN 2nd open call for proposals

Proposals had to relate to basic linguistic resources (tools and data), fundamental strategic research and applications in the areas of language and speech technology, all of which had to contribute to an appropriate digital language infrastructure for Dutch. Proposals could be submitted both in the area of language technology, and in the area of speech technology, and were preferably relevant to both areas. For cross-border consortiums the standard bench fee was increased by 50%.

### Evaluation procedure pre-proposals/full proposals submitted in the 2nd open call

The selection of the pre-proposals was carried out by the Programme Committee. The Programme Committee specifically assessed the expected contribution to the STEVIN aims. Applicants of approx. 15 promising pre-proposals received a recommendation to submit a full proposal according to a specified format. The project leaders of selected pre-proposals were invited by the STEVIN Programme Committee for a short session during which the PC advised them as to the way they might extend their pre-proposal into a full proposal.

All eligible full proposals submitted in the open call were presented to a panel of international experts in language and speech technology. For this call the same experts were asked to serve in the

International Assessment Panel as the ones that acted as such in the first STEVIN Call be it that - to lower the workload - for this call two extra experts were asked to serve in this panel. The composition of the International Assessment Panel was given on the STEVIN website. This international panel evaluated and ranked the eligible applications based on the assessment criteria mentioned in the call and formulated a set of recommendations for the STEVIN Programme Committee.

The International Assessment Panel's assessment was sent to the applicants for comments. Based on the applications, the panel's recommendations and the applicant's response to the panel assessment, the Programme Committee again assessed the applications and determined the order of priority of the eligible proposals. On the basis of the International Assessment Panel's advice and the Programme Committee's advice, the Board of the STEVIN programme determined which projects were funded.

### Time frame STEVIN 2nd call – 2005 – length call procedure 8 months

- \* March 30: opening call and brokerage event in Antwerpen
- \* April 26: closing date call: 34 pre-proposals were submitted
- \* May 23: assessment pre-proposals by STEVIN PC – 18 pre-proposals selected
- \* June 13: interviews with applicants selected pre-proposals
- \* September 2: closing date call: 18 full proposals submitted
- \* October 6 and 7: assessment and ranking of all proposals by STEVIN IAP
- \* October: applicants formulated response to IAP assessment
- \* November 14 and 15: assessment and ranking of all proposals by STEVIN PC
- \* November 29: determining short list by TST Board

**Statistics:** Number and percentages submitted pre-proposals, full proposals and funded proposals, listed by type

	speech	language	total
Submitted pre-proposals	16 (48%)	18 (52%)	34 (100%)
Submitted proposals	6 (33%)	12 (66%)	18 (100%)
Funded proposals	3 (50%)	3 (50%)	6 (100%)

### 3. STEVIN 3rd Open Call for proposals for applied research (max. budget 2,3 M€)

#### Objectives STEVIN 3rd open call for proposals

The STEVIN programme aims to have a balanced programme covering all layers (resources, research and development, technology integration in applications, end users). In the preceding calls, only a few projects focused on technology integration in applications. For this reason in the 3<sup>rd</sup> call especially proposals for *application-oriented research projects* were invited.

#### Evaluation procedure full proposals submitted in the 3rd open call

All eligible full proposals submitted in the open call were presented to a panel of international experts in language and speech technology. For this call the same experts were asked to serve in the International Assessment Panel as the ones that acted as such in the second STEVIN Call. The composition of the International Assessment Panel was available on the STEVIN website. This international panel evaluated and ranked the eligible applications based on the assessment criteria mentioned in the call and formulated a set of recommendations for the STEVIN Programme Committee. The panel's assessment was presented to the applicants for comments. Based on the applications, the panel's recommendations and the applicant's response to the panel assessment, the Programme Committee also assessed the applications and determined the order of priority of the eligible proposals. On the basis of the International Assessment Panel's advice and the Programme Committee's advice, the Board of the STEVIN programme determined which projects were funded.

## Time frame STEVIN 3rd call – 2007 – length call procedure 4 months

- \* April 15: opening call
- \* May 22: closing date call: 15 full proposals were submitted
- \* July 5 and July 6: assessment and ranking of all proposals by STEVIN IAP
- \* July: applicants formulated response to IAP assessment
- \* August 7-8: assessment and ranking of all proposals by STEVIN PC
- \* August 21: determining short list by Board of the STEVIN programme

## Statistics: Number and percentages submitted and funded proposals, listed by type

	speech	language	total
Submitted proposals	5 (33%)	10 (66%)	15 (100%)
Funded proposals	2 (40%)	3 (60%)	5 (100%)

## 4. Three Calls for tender for specific HLT infrastructures (max budget 1,6 M€)

### Objectives STEVIN calls for tender

The realisation of a number of specific top priorities for Dutch HLT

1. A speech recognition toolkit for Dutch
2. A lexical resource for the semantic processing of Dutch
3. An annotated written Dutch corpus

### Evaluation procedure full proposals submitted in the calls for tender

The assessment and ranking of full proposals targeting the specific priorities was carried out by the STEVIN Programme Committee. The proposals and the Programme Committee assessment were subsequently forwarded to the International Assessment Panel for commenting. On the basis of the Programme Committee's advice and the comments of the International Assessment Panel, the Board of the STEVIN programme finally determined which projects were funded.

## Time frame STEVIN Call for tender 1 and 2 – 2005 – length call procedure 8 months

- \* March 30: opening call for tender 1 and 2
- \* May 9: closing date call: 4 proposals were submitted (1x tender 1; 3x tender 2)
- \* May 23: assessment proposals by STEVIN PC – more info requested from consortia
- \* June 30: second discussion extended proposals by STEVIN PC
- \* August: written assessment and ranking of all tender proposals by STEVIN IAP
- \* November 29: funding decision by Board of the STEVIN programme

## Time frame STEVIN Call for tender 3 – 2007 – length call procedure 9 months

- \* December 1: opening call for tender 3
- \* February 28: closing date call: 1 proposal was submitted
- \* March/April: assessment proposal by STEVIN IAP – more info requested from consortium
- \* August 7: assessment tender proposal by STEVIN PC
- \* August 21: funding decision by Board of the STEVIN programme

### Statistics

The statistics for this call are rather straightforward: for tenders 1 and 3 the HLT field formed broad Dutch/Flemish consortia containing all essential actors in the field that submitted a joint proposal. For tender 2, three proposals were submitted, one of which was selected.

## 5. Three Calls for proposals for demonstration projects (max. budget 1 M€)

### **Objectives STEVIN calls for demonstration systems**

The objective of the STEVIN calls for demonstration systems was to try and stimulate demand for HLT technology by funding short-term (maximum length 15 months) projects for building demonstration projects using proven HLT technologies. Demonstrators could target to open new markets or new domains. Project consortia had to be lead by a Dutch or Flemish HLT SME and could consist of both industrial and academic partners. Maximum size of a demonstration project is € 100.000. Three calls were opened in respectively 2005, 2006 and 2006. The total budget of the three calls was € 1.000.000.

### **Evaluation procedure full proposals submitted in the STEVIN calls for demonstration projects**

All applications were assessed by a committee consisting of the two senior managers of the STEVIN Programmebureau, the STEVIN coordinator at the Dutch Language Union, an ICT expert from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT) and an ICT expert from SenterNovem, the Dutch Ministry for Economic Affairs agency for innovation and sustainable development. As of the 2<sup>nd</sup> call all proposals shortlisted by the assessment committee were sent for a sanity check to the STEVIN Programme Committee. On the basis of the advise from the assessment committee and the sanity judgment of the PC the HLT Board made the final funding decision.

### **Time frame STEVIN STEVIN calls for demonstration projects 2005, 2006 and 2007 – length call procedure 5 months**

- \* July 2005: opening 1<sup>st</sup> call
- \* October 15 2005: closing date 1<sup>st</sup> call: 8 proposals were submitted
- \* November 2005: assessment and ranking of all proposals by assessment committee
- \* December 2005: 3 proposals funded by Board of the STEVIN programme
- \* July 2006: opening 2<sup>nd</sup> call
- \* October 15 2006: closing date 2<sup>nd</sup> call: 19 proposals were submitted
- \* November 2006: assessment and ranking of all proposals by assessment committee
- \* December 2006: sanity check by STEVIN programme committee of shortlisted proposals
- \* December 2006: 6 proposals funded by Board of the STEVIN programme
- \* July 2007: opening 3<sup>rd</sup> call
- \* October 15 2007: closing date 3<sup>rd</sup> call: 13 proposals were submitted
- \* November 2007: assessment and ranking of all proposals by assessment committee
- \* December 2007: sanity check by STEVIN programme committee of shortlisted proposals
- \* December 2007: 5 proposals funded by Board of the STEVIN programme

### **Statistics: Number and percentages funded proposals, listed by type**

	speech	language	speech/language combined	total
Funded proposals	4 (30%)	7 (50%)	3 (20%)	14 (100%)

### **Call for proposals for educational projects (max. budget first call 27.500 €)**

A first call for educational projects was openen in 2007. Projects should aim to make students between age 15-20 in educational settings (school, museams, etc) aware of the possibilitis of language and speech technologies. Two calls remain in 2008 and 2009. The maximum budget for each of these calls is € 55.000. Proposals submitted in the calls are assessed by a panel of Dutch and Flemish educational experts. Their advise is sent for a sanity check to the STEVIN Programme Committee. On the basis of the advise from the assessment panel and the sanity judgment of the PC the HLT Board makes the final funding decision. In the 1<sup>st</sup> call only one eligible proposal was submitted and funded.

## The STEVIN Priorities (as formulated in the original STEVIN project description)

Proposals can relate to basic linguistic resources (tools and data), fundamental strategic research and applications in the areas of language and speech technology, all of which have to contribute to an appropriate digital language infrastructure for Dutch. Proposals can be submitted both in the area of language technology, and in the area of speech technology, and are preferably relevant to both areas. For cross-border consortiums the standard bench fee (see 'eligible costs' on page 4) will be increased by 50%. Examples of language and speech applications which can be targeted are presented after the priorities for speech technology resources and research and those for language technology resources and research.

For speech technology, the priorities are:

*for resources:*

- speech and multimodal corpora for:
  - applications such as CALL (Computer Assisted Language Learning);
  - applications in which names and addresses play an important role;
  - CCQA applications (questions and answers in call centres), educational applications;
- multimodal corpora for applications of broadcast news transcription or person identification;
- text corpora for the development of stochastic language models;
- tools and data for the development of:
  - robust speech recognition;
  - automatic annotation of corpora;
  - speech synthesis;

*for research:*

- robustness of speech recognition;
- output treatment (inverse text normalization);
- confidence measures;
- adaptation;
- lattices.

For language technology, the priorities are:

*for resources:*

- richly annotated monolingual Dutch corpora;
- electronic lexicons;
- aligned parallel corpora;

*for research:*

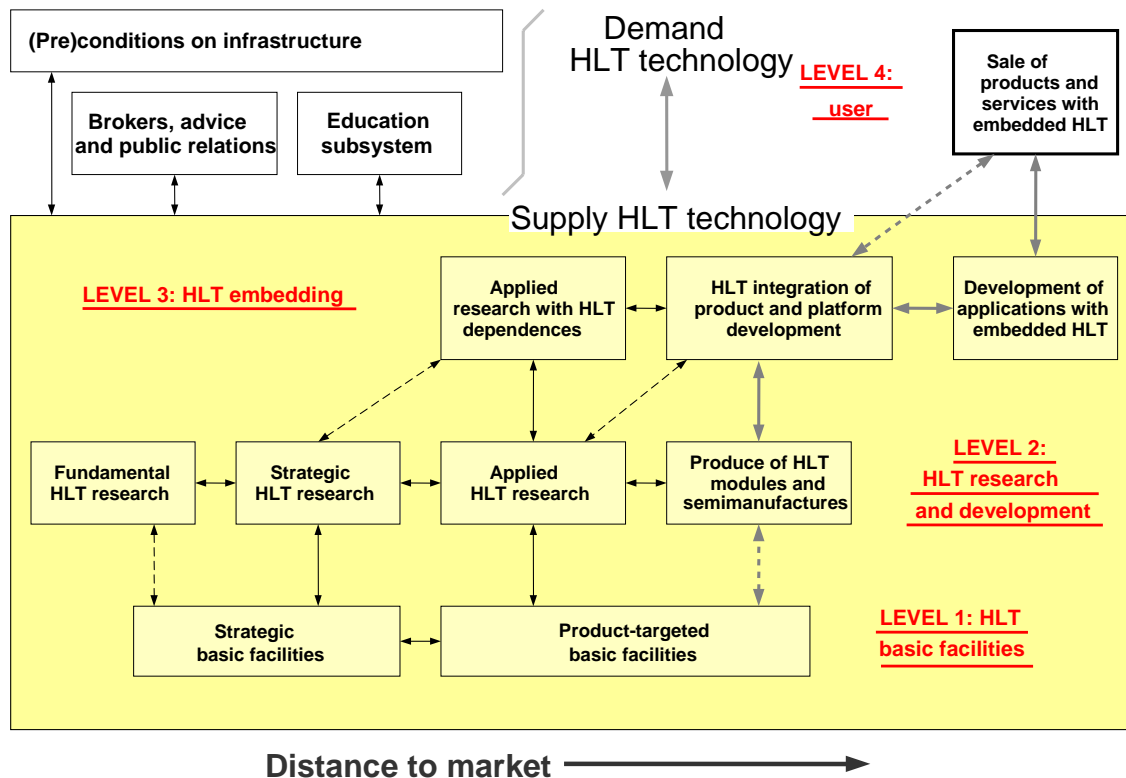
- semantic analysis, including semantic tagging and integrating morphological, syntactic and semantic modules;
- text pre-processing;
- morphological analysis;
- syntactic analysis (robust parsing).

In the area of applications (both for speech & language technology), examples to be targeted on are:

- information extraction from audio-transcripts created by speech recognizers;
- speaker accent and identity detection;
- monolingual or multilingual information extraction;
- semantic web;
- dialogue systems and Q&A solutions, especially in multimodal domains;
- automatic summarization and text generation applications;
- machine translation;
- educational systems.

## Coverage of STEVIN Priorities by the projects funded within the STEVIN programme

STEVIN priorities together address different aspects of the stratified innovation system (as depicted below). STEVIN advocates in integrated approach all layers in the stratified system are addressed, i.e. development of language and speech resources and tools, stimulating innovative fundamental and strategic research, stimulating application oriented research, promote HLT embedding in existing applications and services, stimulate HLT demand via demonstrator projects and encourage cooperation and knowledge transfer between academia and industry.



In the table below an overview is given of how the projects funded in the various STEVIN funding schemes cover the STEVIN priorities and the layers of the innovation model.

Percentage of STEVIN funding per STEVIN priority		
• Speech technology resources	21,5%	
• Language technology resources	29,0%	
<b>% STEVIN funding for basic resources</b>		<b>50,5%</b>
• Speech technology research	14,5%	
• Language technology research	9,1%	
<b>% STEVIN funding for basic research</b>		<b>23,6%</b>
• Speech technology application-oriented projects	7,4%	
• Language technology application-oriented projects	8,2%	
<b>% HLT Application-oriented projects</b>		<b>15,6%</b>
Speech technology demonstration projects	3,7%	
Language technology demonstration projects	6,6%	
<b>% HLT Demonstration projectss</b>		<b>10,3%</b>
<b>% STEVIN funding for speech technology</b>		<b>47,1%</b>
<b>% STEVIN funding for language technology</b>		<b>52,9%</b>

## Standard STEVIN Assessment criteria

### Quality and innovative character of the proposal

- Clarity in problem definition and innovative power of the project.
- Suitability and effectiveness of the research design and methodology. In particular, an explicit component of evaluation, or in the case of linguistic resources, an explicit validation plan must be included in the proposal.
- Impact of the project on a wide range of applications and its importance to applications that are relevant to the industry.
- Competence of the participating groups (including past performance).
- Feasibility of the goals.
- The goal is to have a balanced programme covering all layers (basic resources, research and development, technology integration, end users) in the chain approach and properly integrating them. The contribution of individual projects to this overall programme goal will therefore be a criterion in their evaluation.
- Balanced cooperation and task division within the project.
- Availability of the required infrastructure.

### Economic aspects of the project proposal

- Is there cooperation with or support by companies?
- What are the prospects for spin-offs and/or other new developments?
- Opportunities for applying the results in industry and/or society.

### Contribution to the STEVIN-programme

- Conformity to the focus of the programme and fit in with the priorities set. The project must focus on the Dutch language and must contribute to improving or at least securing the position of the Dutch language in the modern information and communication society.
- Perspectives on knowledge transfer and network creation. In particular it is to the advantage of a project proposal if the expertise of Dutch and Flemish groups or companies are combined, if research institutes and companies jointly make a proposal, or if the proposal relates both to language and to speech technology.

### IPR, avoiding duplication and standards

- The proposal must contain a clear plan for the proper treatment of intellectual property rights (IPR), both for the resources provided by third parties and for the results of the project. The working principle must be that the data, tools and other practical spin-offs resulting from the STEVIN-projects are made available in a non-discriminative way in the TST-centrale.
- The proposal must prove that the applicants have a precise and up-to-date picture of what is already available in terms of basic resources. Preferably the resources to be developed in the project do not exist yet. If it is known or can be presupposed that the resources exist but are not generally accessible, the proposal should contain a plan to avoid disturbance of the market, i.e. unfair competition must be avoided.
- The R&D Community must be able to access, use and exploit the basic resources resulting from the STEVIN-projects on non-discriminate terms. The applicants have to declare themselves willing to negotiate on this with the TST-centrale and to sketch the conditions that apply. Conclusion of a contract on the IPR-arrangements is a necessary condition for awarding of funding.
- The proposal must fit in with existing standards and apply these where possible, or cooperate on the development of new standards so that a maximum reuse of the basic resources developed is guaranteed.

Some specific criteria defining application-oriented proposals were added for the 3<sup>rd</sup> call where this type of proposals was specifically invited.

## **Code of Conduct for Independent Experts appointed in the international STEVIN Assessment Panel (IAP)**

1. The task of an expert is to participate in a confidential, fair and equitable review of project(s) according to any programme-specific review documents. He/she must use his/her best endeavours to achieve this, follow any instructions given by the STEVIN Programme Bureau to this end and deliver a constant and high quality of work.
2. The reviewer works as an independent person. He/she is deemed to work in a personal capacity and, in performing the work, does not represent any organisation.
3. The independent expert must sign a declaration of conflict of interest and confidentiality before starting the work, by which he/she accepts the present Code of Conduct. Invited independent experts who do not sign the declaration will not be allowed to work as a reviewer.
4. In doing so, the independent expert commits him/herself to strict confidentiality and impartiality concerning his/her tasks. If a reviewer has a direct or indirect link with the project(s), or any other vested interest, or is in some way connected with the project(s), or has any other allegiance which impairs or threatens to impair his/her impartiality with respect to the project(s), he/she must declare such facts to the responsible STEVIN Programme Bureau official as soon as he/she becomes aware of this. The STEVIN Programme Bureau ensures that, where the nature of any link is such that it could threaten the impartiality of the reviewer, he/she does not participate in the review of the project(s) concerned.
5. Reviewers may not discuss any project details with others, including other reviewers or STEVIN Programme Bureau officials not directly involved in the review of the project, except during the formal review session moderated by or with the knowledge of the responsible STEVIN Programme Bureau official.
6. Where it has been decided that project details and/or project deliverables are to be posted or made available electronically to reviewers, who then work from their own or other suitable premises, the reviewer will be held personally responsible for maintaining the confidentiality of any documents or electronic files sent and returning or destroying all confidential documents or files upon completing the review as instructed. Reviewers may seek further information (for example through the internet, specialised databases, etc.) in order to allow them to complete their examination of the project details and/or deliverables, provided that the obtaining of such information respects the overall rules for confidentiality and impartiality. Reviewers may not show the contents of the deliverables or information on the project(s) to third parties (e.g. colleagues, students, etc.) without the express written approval of the STEVIN Programme Bureau. It is forbidden for reviewers to make direct contact with the project participants.
7. Reviewers are required at all times to comply strictly with any rules defined by the STEVIN Programme Bureau for ensuring the confidentiality of the review process and its outcomes. Failure to comply with these rules may result in exclusion from the immediate and future reviews, without prejudice to penalties that may derive from other applicable Regulations.

## **Code of Conduct for members of the STEVIN Programme Committee**

1. The task of an expert is to participate in a confidential, fair and equitable review of project(s) according to any programme-specific review documents. He/she must use his/her best endeavours to achieve this, follow any instructions given by the STEVIN Programme Bureau to this end and deliver a constant and high quality of work.
2. The PC member works as an independent person. He/she is deemed to work in a personal capacity and, in performing the work, does not represent any organisation.
3. The PC member must sign a declaration of conflict of interest and confidentiality before starting the work, by which he/she accepts the present Code of Conduct. PC members who do not sign the declaration will not be allowed to be present at the assessment meeting of the PC.
4. In doing so, the PC member commits him/herself to strict confidentiality and impartiality concerning his/her tasks. If a reviewer has a direct or indirect link with the project(s), or any other vested interest, or is in some way connected with the project(s), or has any other allegiance which impairs or threatens to impair his/her impartiality with respect to the project(s), he/she must declare such facts to the responsible STEVIN Programme Bureau official as soon as he/she becomes aware of this. The STEVIN Programme Bureau ensures that, where the nature of any link is such that it could threaten the impartiality of the reviewer, he/she does not participate in the review of the project(s) concerned.
5. PC members may not discuss any project details with others, including other PC members or STEVIN Programme Bureau officials not directly involved in the review of the project, except during the formal assessment meeting moderated by or with the knowledge of the responsible STEVIN Programme Bureau official.
6. Where it has been decided that project details and/or project deliverables are to be posted or made available electronically to PC members, who then work from their own or other suitable premises, the PC members will be held personally responsible for maintaining the confidentiality of any documents or electronic files sent and returning or destroying all confidential documents or files upon completing the review as instructed. PC members may seek further information (for example through the internet, specialised databases, etc.) in order to allow them to complete their examination of the project details and/or deliverables, provided that the obtaining of such information respects the overall rules for confidentiality and impartiality. PC members may not show the contents of the deliverables or information on the project(s) to third parties (e.g. colleagues, students, etc.) without the express written approval of the STEVIN Programme Bureau. It is forbidden for PC members to make direct contact with the project participants.
7. PC members are required at all times to comply strictly with any rules defined by the STEVIN Programme Bureau for ensuring the confidentiality of the review process and its outcomes. Failure to comply with these rules may result in exclusion from the immediate and future reviews, without prejudice to penalties that may derive from other applicable Regulations.

## Overview of STEVIN projects

1st Call for Proposals for strategic research proposals and HLT infrastructures (data & tools) (max. budget 2 M€) - 2004

- Automata for deriving phoneme transcriptions of Dutch and Flemish names (AUTONOMATA)
- Coreference Resolution for Extracting Answers (COREA)
- Dutch Language Corpus Initiative (D-coi)
- Identification and Representation of Multi-word Expressions (IRME)
- Extension of CGN with speech of children, non-natives, elderly and human-machine interaction (JASMIN-CGN)

2nd Call for Proposals for strategic research proposals and HLT infrastructures (data & tools) (max. budget 3,8 M€) - 2005

- Detecting and Exploiting Semantic Overlap (DAESO)
- Dutch Parallel Corpus (DPC)
- Large Scale Syntactic Annotation of written Dutch (Lassy)
- Missing Data Solutions (Midas)
- Northern and Southern Dutch Benchmark Evaluation of Speech recognition Technology (NBest)
- STEVIN can PRAAT

Call for proposals for applied research (max. budget 2,3 M€) - 2007

- Autonomata, Transfer of Output (Autonomata TOO)
- Dutch ILanguage Investigation of Summarization technology
- Development and Integration of Speech technology into COurseware for language learning (DISCO)
- Dutch Online Media Analysis (DuOMAn)
- Parse and Corpus based Machine Translation (PaCo-MT)

Three Calls for tender for specific HLT infrastructures (max budget 1,6 M€) – 2005/2007

- Speech Processing, Recognition & Automatic Annotation Kit (Spraak)
- Combinatorial and Relational Network as Toolkit for Dutch Language Technology (Cornetto)
- Stevin Nederlandstalig Referentiecorpus (SoNaR)

Three Calls for proposals for demonstration projects (max. budget 1 M€) – 2005/2006/2007

- Rechtsorde
- GemeenteConnect!
- Spraakgestuurde Nummerbord Retrieval Tool
- Audiokrant
- Spelling- en grammaticacontrole voor dyslectische gebruikers
- Rechtspraakherkenning
- Klinkende Taal
- SpelSpiek
- Voice Assess
- Alfabetisering Anderstaligen Plan (AAP)
- Esay Info
- Hulp bij Auditieve Training na Cochleaire Implantatie (HATCI)
- Nederlandstalige Ondertiteling (Neon)
- Sprekende zelfcorrigerende woordvoorspeller voor dyslectische gebruikers (WooDy)

Call for proposals for educational projects (max. budget first call € 27.500) - 2007

- Vooronderzoek TST in het voortgezet onderwijs
- Taal en spraaktechnologie op Kennislink

**Overview proposals funded in the 1st Call for Proposals for strategic research proposals and HLT infrastructures (data & tools)  
(max. budget 2 M€)**

<i>acronym</i>	<i>coordinating institute and other academic partners</i>	<i>industrial partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed (subject)</i>	<i>planned duration</i>	<i>funding</i>
AUTONOMATA	Ghent University (Jean-Pierre Martens)  Radboud Univ. Nijmegen Utrecht University	TeleAtlas Scansoft	VL/NL	Speech resources  (speech synthesis)	24 mnths	€ 322.848
COREA	Groningen University (Gosse Bouma)  Antwerpen University	Language and Computing	NL/VL	Language resources  Language research  (semantic annotation)	24 mnths	€ 353.875
D-coi	Radboud Univ. Nijmegen-CLST (Nelleke Oostdijk)  Tilburg University Antwerpen University Twente University Utrecht University Groningen University Leuven University	Polderland	NL/VL	Language resources  Speech resources  (Corpus written Dutch protocols)	14 mnths	€ 566.531
IRME	Utrecht University (Jan Odijk)  Groningen University	Van Dale Lexicografie	NL	Language resources  Language research  (semantic and syntactic annotation)	24 mnths	€ 389.500
JASMIN-CGN	Radboud Univ. Nijmegen- CLST (Catia Cucchiarini)  Leuven University	TalkingHome	NL/VL	Speech resources  (speech corpus)	24 mnths	€ 419.471

## Automata for deriving phoneme transcriptions of Dutch and Flemish names (AUTONOMATA)

### Project co-ordinator

Prof. dr. ir. J.-P. Martens  
Gent University  
ELIS Speech Lab  
Sint-Pietersnieuwstraat 41  
B-9000 Gent, België  
Telephone: +32 9 264 33 95  
E-mail: Jean-Pierre.Martens@elis.ugent.be  
URL: [www.elis.ugent.be](http://www.elis.ugent.be)

### Project consortium

1. Prof. dr. ir. J.-P. Martens (Universiteit Gent, ELIS Speech Lab)
2. Dr. H. van den Heuvel (Radboud Universiteit Nijmegen, Centre for Language and Speech Technology - CLST)
3. Dr. ir. G. Bloothoof (Universiteit Utrecht, Utrecht institute of Linguistics - UiL-OTS)
4. Ir. L. Peirlinckx (TeleAtlas)
5. Dr. ir. J. Verhasselt (Nuance Communications International)

**STEVIN funding: 322.848 euro**

**Duration: 01/06/2005 – 31/05/2007**

### Project summary

This project aims to build two resources: (1) a grapheme-to-phoneme (g2p) conversion tool set for creating good phonetic transcriptions for TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) applications with a focus on phonetic transcriptions of names, and (2) a corpus of spoken name utterances for supporting more research towards better automatic name recognition.

Since all presently available g2p converters perform poorly on names, the project will create and make available to third parties, dedicated name g2p converters (for Dutch and Flemish) that will be designed to produce high quality canonical name transcriptions of person names and address items. The machine learning tools that will be used to design these converters will be made available to third parties as well. This way they can be applied to develop dedicated g2p converters for name categories that are not handled in this project.

It is acknowledged that the deployment of LST applications involving ASR of Dutch and Flemish could be raised significantly if (among other things) one would succeed in surpassing the present state-of-the-art in name recognition. This will first of all require tools for creating good canonical transcriptions of these names, as envisaged in this project, but on top of that it will also call for new methods for predicting the kind of variations of these pronunciations one is likely going to encounter in spoken name utterances of native and non-native speakers of Dutch and Flemish. For the development of such methods, one needs a substantial corpus of spoken name utterances. Such a corpus is presently not available for Dutch nor Flemish, and this project proposes to create one.

**AUTONOMATA website:** <http://speech.elis.ugent.be/autonomata/>

## Coreference Resolution for Extracting Answers (COREA)

### Project co-ordinator

Dr. G. Bouma  
Rijksuniversiteit Groningen  
Faculteit der Letteren, Informatiekunde Alfa-Informatica)  
Oude Kijk in 't Jatstraat 26  
Postbus 716  
NL-9700 AS Groningen  
Telephone: +31 50 363 59 37  
E-mail: [gosse@let.rug.nl](mailto:gosse@let.rug.nl)  
URL: [www.rug.nl/let](http://www.rug.nl/let)

### Project consortium

1. Dr. G. Bouma (Rijksuniversiteit Groningen, Alfa-informatica)
2. Prof. dr. W. Daelemans (Universiteit Antwerpen, Centrum voor Nederlandse Taal and Spraak - CNTS, en Universiteit Tilburg, Induction of Linguistic Knowledge - ILK)
3. J.-L. Verschelde (Language and Computing NV)

**STEVIN funding: 353.875 euro**

**Duration: 01/05/2005 – 30/04/2007**

### Project summary

Co reference resolution is a key ingredient for the automatic interpretation of text. It has been studied mainly from a linguistic perspective, with an emphasis on establishing potential antecedents for pronouns. Practical applications, such as Information Extraction (IE), summarization and Question Answering (QA), require accurate identification of co reference relations between noun phrases in general. Computational systems for assigning such relations automatically, require the availability of a sufficient amount of annotated data for training and testing. For Dutch, annotated data is scarce and co reference resolution systems are lacking.

In this project, we aim to develop a robust system for assigning such relations automatically, and we will investigate the effect of making co reference relations explicit on the accuracy of systems for IE and QA. We will annotate a limited amount of application-specific corpus material, which is required for the evaluation of the co reference resolution system in the context of IE and QA. The project contributes to the goals of STEVIN by providing a robust co reference resolution system which is applicable in a range of applications for Dutch, such as information extraction, question answering and summarization. In addition, general guidelines for co reference annotation will become available and a tool will be developed to support the annotation of co reference in text. Finally, a limited amount of data annotated with co referential information, including spoken language data, will be produced.

**COREA website:** <http://www.cnts.ua.ac.be/~hoste/corea.html>

## Dutch Language Corpus Initiative (D-coi)

### Project co-ordinator

Dr. N. Oostdijk  
Radboud Universiteit Nijmegen  
Faculteit der Letteren  
Centre for Language and Speech Technology (CLST)  
Postbus 9103  
NL-6500 HD Nijmegen  
Telephone: +31 24 361 27 65  
E-mail: [n.oostdijk@let.ru.nl](mailto:n.oostdijk@let.ru.nl)  
URL: [www.let.ru.nl](http://www.let.ru.nl)

### Project consortium

1. Dr. N. Oostdijk (Radboud Universiteit Nijmegen, Centre for Language and Speech Technology - CLST)
2. Dr. A. van den Bosch (Universiteit Tilburg, Induction of Linguistic Knowledge - ILK)
3. Drs. Th. van den Heuvel (Polderland Language and Speech Technology BV)
4. Prof. dr. F. de Jong (Universiteit Twente, Human Media Interaction - HMI)
5. Dr. P. Monachesi (Universiteit Utrecht, Utrecht institute of Linguistics - UiL-OTS)
6. Dr. G. van Noord (Rijksuniversiteit Groningen, Alfa-informatica)
7. Prof. dr. F. Van Eynde (Katholieke Universiteit Leuven, Centrum voor Computerlinguïstiek - CCL)

**STEVIN funding: 566.531 euro**

**Duration: 01/06/2005 – 31/12/2006**

### Project summary

The project proposed here can be characterized as a preparatory project and aims to produce a blueprint for the construction of a 500-million-word corpus of contemporary written Dutch. This will entail the design of the corpus and the development (or adaptation) of protocols, procedures and tools that are needed for sampling data, cleaning up, converting file formats, marking up, annotating, post editing, and validating the data. In order to support these developments, a 50-million-word pilot corpus will be compiled, parts of which will be enriched with linguistic annotations. The pilot corpus is intended to demonstrate the feasibility of the approach. It will provide the necessary testing ground on the basis of which feedback can be obtained about the adequacy and practicability of various annotation schemes and procedures, and the level of success with which tools can be applied. Moreover, it will serve to establish the usefulness of this type of resource and annotations for different types of HLT research and the development of applications. The Danish Center for Sprogteknologi (CST) will undertake the evaluation of the protocols and procedures. At the end of the project, the pilot corpus together with all other results obtained within the project will be handed over to the Dutch Language Union and be made available through the Flemish-Dutch HLT Agency (TST-centrale).

**D-coi website:** <http://lands.let.ru.nl/projects/d-coi/>

## Identification and Representation of Multi-word Expressions (IRME)

### Project co-ordinator

Prof. Dr. J. Odijk  
Universiteit Utrecht  
Faculteit der Letteren  
Utrecht institute of Linguistics OTS (UiL-OTS)  
Janskerkhof 13  
NL-3512 JK Utrecht  
Telephone: +31 30 253 60 76  
E-mail: [jan.odijk@let.uu.nl](mailto:jan.odijk@let.uu.nl)  
URL: [www-uilots.let.uu.nl](http://www-uilots.let.uu.nl)

### Project consortium

1. Prof. dr. J. Odijk (Universiteit Utrecht, Utrecht institute of Linguistics OTS - UiL-OTS)
2. Dr. G van Noord (Rijksuniversiteit Groningen, Alfa-Informatica)
3. Dr. G. Bouma (Rijksuniversiteit Groningen, Alfa-Informatica)
4. Dr. J. Zuidema (Van Dale Lexicografie BV)

**STEVIN funding: 389.500 euro**

**Duration: 01/06/2005 – 31/08/2007**

### Project summary

The central problems that the project addresses are (i) the lack of large and rich formalized lexicons for multi-word expressions for use in NLP; (ii) the lack of proper methods and tools to extend the lexicon of an NLP-system for multi-word expressions given a text corpus in a maximally automated manner. Therefore, the project aims to develop innovative methods and tools for the automatic identification and lexical representation of multi-word expressions. Concomitantly, a 5.000 entry corpus-based multi-word expression lexical database for Dutch will be developed. The database will be externally validated, and its usability will be evaluated in two independent NLP-systems for Dutch. The project contributes to the development of electronic lexicons, in particular for Dutch. The MWE database to be developed fills a gap in existing lexical resources for Dutch. The project carries out strategic research into generic methods and tools for MWE identification and lexical representation, focusing on Dutch, but these tools will be largely language-independent and can also be used for other languages, new domains, and beyond this project. In this way the project contributes directly to strengthening the digital infrastructure for Dutch.

**IRME website:** <http://www-uilots.let.uu.nl/irme/>

## Extension of CGN with speech of children, non-natives, elderly and human-machine interaction (JASMIN-CGN)

### Project co-ordinator

Dr. C. Cucchiari  
Radboud Universiteit Nijmegen  
Faculteit der Letteren  
Centre for Language and Speech Technology (CLST)  
Postbus 9103  
NL-6500 HD Nijmegen  
Telephone: +31 24 361 57 85  
E-mail: [C.Cucchiari@let.ru.nl](mailto:C.Cucchiari@let.ru.nl)  
URL: [www.let.ru.nl](http://www.let.ru.nl)

### Project consortium

1. Dr. C. Cucchiari (Radboud Universiteit Nijmegen, Centre for Language and Speech Technology - CLST)
2. Prof. dr. H. Van hamme (Katholieke Universiteit Leuven, ESAT/PSI Speech Group)
3. Dr. ir. F.M.A. Smits (TalkingHome)

**STEVIN funding: 419.471 euro**

**Duration: 01/04/2005 – 30/09/2007**

### Project summary

Large speech corpora (LSC) constitute an indispensable resource for conducting research in speech processing and for developing real-life speech applications. In 2004 the Spoken Dutch Corpus (Corpus Gesproken Nederlands - CGN) became available, which constitutes a plausible sample of standard Dutch as spoken by adult natives in the Netherlands and Flanders. Owing to budget constraints, CGN does not include speech of children, non-natives, elderly people and recordings of speech produced in human-machine interactions.

Since such recordings would be extremely useful for conducting research and for developing HLT applications for these specific groups of speakers of Dutch, the present proposal aims at extending CGN in three dimensions. First, by collecting a corpus of contemporary Dutch as spoken by children of different age groups, non-natives with different mother tongues and elderly people in the Netherlands and Flanders (JASMIN-CGN), we aim at an extension along the age and mother tongue dimensions. In addition, we intend to collect speech material in a communication setting that was not envisaged in CGN: human-machine interaction. Therefore, in this project part of the speech material from the three speaker groups will be collected in a setting of human-machine communication. We expect that the knowledge gathered from these data can be generalized to developing appropriate systems also for other speaker groups (i.e. adult natives). One third of the data will be collected in Flanders and two thirds in the Netherlands.

**JASMIN-CGN website:** <http://www.esat.kuleuven.be/psi/spraak/projects/JASMIN/>

Overview proposals funded in the 2nd Call for Proposals for strategic research proposals and HLT infrastructures (data & tools)  
(max. budget 3,8 M€)

<i>acronym</i>	<i>coordinating institute and other academic partners</i>	<i>industrial partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed (subject)</i>	<i>planned duration</i>	<i>funding</i>
DAESO	Tilburg University (Emiel Kraemer)  Antwerpen University Universiteit van Amsterdam	Textkernel	NL/VL	Language research  Language resources  (Semantic / discourse annotation)	36 mnths	€ 487.000
DPC	KU Leuven (Piet Desmet)  Hogeschool Gent		VL	Language resources  (Multilingual corpora / translational equivalents)	34 mnths	€ 498.000
LASSY	Groningen University (Gertjan van Noord)  KU Leuven		NL/VL	Language resources  (Syntactic treebank)	36 mnths	€ 496.000
MIDAS	KU Leuven (Hugo Van hamme)  Radboud Univ. Nijmegen	Nuance	VL/NL	Speech research  (Robust ASR)	48 mnths	€ 499.000
NBest	TNO-TM (David van Leeuwen)  KU Leuven, Twente University, Radboud Univ. Nijmegen, Ghent University, SPEX, TU Delft		NL/VL	Speech resources  (ASR benchmarks for evaluation)	29 mnths	€ 470.000
STEVIN can PRAAT	Universiteit van Amsterdam (Paul Boersma)  Leiden University SPEX	Speech-Minded	NL	Speech resources  (ASR, annotation tool)	24 mnths	€ 114.000

## Detecting and Exploiting Semantic Overlap (Daeso)

### Project co-ordinator

dr. E. Kraemer  
Tilburg University  
Faculteit Communicatie en Cultuur  
Taal en Informatica  
Warandelaan 2  
5037 AB Tilburg  
Telephone: +31 13-466 25 68  
E-mail: [e.j.kraemer@uvt.nl](mailto:e.j.kraemer@uvt.nl)  
URL: <http://let.uvt.nl/research/ti>

### Project consortium

1. dr. E. Kraemer (Tilburg University)
2. prof. dr. W. Daelemans (Antwerp University)
3. prof. dr. M. de Rijke (University of Amsterdam)
4. drs. J. Zavrel (Textkernel)

**STEVIN funding: 487.000 euro**

**Duration: 01/10/2006 – 30/09/2009**

### Project summary

The well-known fact that similar information can be expressed in many different ways is one of the major challenges in building robust NLP applications. It is commonly assumed that such applications can be improved with knowledge of how natural language expressions relate to each other, for instance in terms of paraphrases (same semantic content, different wording) or entailments (one expression implied by the other). DAESO investigates the detection of semantic overlap between Dutch sentences and the exploitation of this knowledge in a range of NLP applications. For this purpose, tools will be developed for the automatic alignment and classification of semantic relations (between words, phrases and sentences) for Dutch, as well as for a Dutch text-to-text generation application which fuses related sentences into a single grammatical sentence, which may be a generalization, a specification or a reformulation of the input sentences. To facilitate development and testing of these tools, an annotated monolingual Dutch parallel/comparable corpus of 1M words will be developed, consisting of pairs of texts that express comparable information. The utility of the resources and tools will be demonstrated in the context of three applications: (1) question-answering systems (improved recall, more complete answers), (2) information extraction (improved recall), and (3) summarization (beyond extraction: sentence compression, sentence fusion, anaphora resolution).

**Daeso website:** <http://daeso.uvt.nl/>

## Dutch Parallel Corpus (DPC)

### Project co-ordinator

Prof. dr. Piet Desmet  
Katholieke Universiteit Leuven Campus Kortrijk  
Etienne Sabbelaan 53  
B-8500 Kortrijk  
Telephone: +32 (0) 56 24 61 85  
E-mail: [Piet.Desmet@kuleuven-kortrijk.be](mailto:Piet.Desmet@kuleuven-kortrijk.be)  
URL: [http://www.ling.arts.kuleuven.ac.be/franling\\_n/pdesmet](http://www.ling.arts.kuleuven.ac.be/franling_n/pdesmet)

### Project consortium

1. Prof. Dr. Piet Desmet (Katholieke Universiteit Leuven Campus Kortrijk)
2. Prof. Dr. Willy Vandeweghe (Hogeschool Gent, School of Translation Studies)
3. Dr. Hans Paulussen (Katholieke Universiteit Leuven Campus Kortrijk)
4. Dra. Lieve Macken (Hogeschool Gent, School of Translation Studies)

**STEVIN funding: 498.000 euro**

**Duration: 01/05/2006 – 28/02/2009**

### Project summary

Gealigneerde parallelle corpora vormen noodzakelijk bronmateriaal voor een groot aantal multitalige toepassingen, zoals machinevertaling (in het bijzonder corpusgebaseerde MT zoals statische en example-based MT), computer-ondersteunde vertaaltools, informatie-extractie, multilinguale terminologie-extractie, en computer-ondersteund talenonderwijs.

Op dit ogenblik zijn er slechts weinig kwaliteitsvolle parallelle corpora met Nederlands als centrale taal beschikbaar, en hun beschikbaarheid wordt voor de onderzoeksgemeenschap bemoeilijkt door auteursrechtelijke restricties. Daarom is de aanmaak van een parallel corpus een van de prioriteiten van het STEVIN-programma.

In dit project beogen we een kwaliteitsvol zinsgealigneerd parallel corpus van 10 miljoen woorden aan te maken voor de talenparen Nederlands-Engels en Nederlands-Frans. Het corpus zal bidirectioneel zijn (Nederlands als brontaal en doeltaal), zodat het kan gebruikt worden als een comparable corpus (waarbij oorspronkelijk in het Nederlands geschreven teksten kunnen vergeleken worden met teksten vertaald naar het Nederlands). Een gedeelte van het corpus zal drietalig zijn, waarbij Nederlandse teksten vertalingen hebben naar het Engels én het Frans. Het corpus wordt verrijkt met taalkundige annotaties.

Om de kwaliteit van het corpus te waarborgen, evenals de multilinguale beschikbaarheid voor de gehele onderzoeksgemeenschap, zal iedere stap in de aanmaak, het structureren en het annoteren gevalideerd worden door een gebruikersgroep van specialisten in de taalkunde en taaltechnologie. Aangezien het Nederlands de scharniertaal is van het corpus, zullen we nauw samenwerken met de onderzoekers van het D-COI project die een pilootcorpus aanmaken van 50 miljoen woorden van hedendaags geschreven Nederlands.

Om het corpus beschikbaar te stellen van de hele onderzoeksgemeenschap, wordt voor de auteursrechten nauw samengewerkt met de TST-centrale die instaat voor de distributie van het corpus.

**DPC website:** <http://www.kuleuven-kortrijk.be/DPC>

## Large Scale Syntactic Annotation of written Dutch (Lassy)

### Project co-ordinator

dr. G.J.M. van Noord  
Rijksuniversiteit Groningen  
Faculteit der Letteren - Alfa-informatica  
Oude Kijk in 't Jatstraat 26  
Postbus 716  
9700 AS Groningen  
Telephone: +31-50-3637811  
E-mail: [vannoord@let.rug.nl](mailto:vannoord@let.rug.nl)  
URL: <http://www.rug.nl/let/onderzoek/onderzoekinstututen/clcg/onderzoek/compuling>

### Project consortium

1. Dr. G.J.M. van Noord (Alfa-informatica Groningen)
2. Drs. I. Schuurman (CCL Leuven)
3. Prof. dr. F. van Eynde (CCL Leuven)
4. Dr. G. Bouma (Alfa-informatica Groningen)

**STEVIN funding: 496.000 euro**

**Duration: 01/11/2006 – 31/10/2009**

### Project summary

A large corpus of written Dutch texts (1,000,000 words) is syntactically annotated (manually corrected), based on D-COI. In addition, the full D-COI corpus is syntactically annotated automatically. The project aims to extend the available syntactically annotated corpora for Dutch both in size as well as with respect to the various text genres and topical domains. In addition, various browse and search tools for syntactically annotated corpora will be further developed and made available. Their potential for applications in corpus linguistics and information extraction will be illustrated and evaluated.

**Lassy website:** <http://www.let.rug.nl/~vannoord/Lassy/>

## Missing Data Solutions (Midas)

### Project co-ordinator

Prof. dr. ir. H. Van hamme  
Katholieke Universiteit Leuven  
ESAT - PSI  
Kasteelpark Arenberg 10  
3001 Heverlee  
Telephone: + 32 16 32 18 42  
E-mail: [hugo.vanhamme@esat.kuleuven.be](mailto:hugo.vanhamme@esat.kuleuven.be)  
URL: <http://www.esat.kuleuven.be/psi/spraak/>

### Project consortium

1. Prof. dr. ir. H. Van hamme (Katholieke Universiteit Leuven)
2. Dr. ir. B. Cranen (Radboud Universiteit Nijmegen)
3. Dr. J. De Veth (Radboud Universiteit Nijmegen)
4. Ir. B. D'hoore (Nuance Communications International)

**STEVIN funding: 499.000 euro**

**Duration: 01/10/2006 – 30/09/2010**

### Project summary

Robustness to noise in automatic speech recognition is essential for the development of successful applications. Noise reduction techniques have been applied with some success in the past, but there remains a large performance gap between the best ASR implementations and human recognition, especially when the noise is non-stationary. This project tackles the noise robustness problem in ASR through missing data techniques (MDT) by addressing important open R&D issues for accuracy improvement and computational efficiency. Detectors of missing data will make minimal assumptions on the noise, while incorporating more knowledge about speech. The acoustic model in the recognizer's back-end will be refined and its evaluation will be made faster through algorithmic research. The developed algorithms will be integrated in the result of the STEVIN "call for tender - speech recognizer" (referred to as CFT-system) and made available through its distribution channels. This project contains language-independent research as well as work that is specific for Dutch, which both are of interest to the STEVIN program. It addresses three STEVIN priorities: 1) robustness of speech recognition, 2) tools and data for the development of robust speech recognition, and 3) confidence measures. How to account best for realistic environmental noise is largely language independent. However, the search for representations of speech that lead to better missing data implementations requires building new acoustic models that are language specific. In this project we will base our research on a "real-life" test suite that contains test material from the Dutch SpeechDat Car and Speecon databases.

**Midas website:** <http://www.esat.kuleuven.be/psi/spraak/projects/index.php?proj=MIDAS>



## Northern and Southern Dutch Benchmark Evaluation of Speech recognition Technology (NBest)

### Project co-ordinator

Ir. D.A. van Leeuwen  
Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek  
Technische Menskunde - Cognitieve Psychologie  
Postbus 23  
Kampweg 5  
3769 ZG De Soesterberg  
Telephone: +31 346 356 235  
E-mail: [david.vanleeuwen@tno.nl](mailto:david.vanleeuwen@tno.nl)  
URL: <http://www.tno.nl>

### Project consortium

1. Dr. D. A. van Leeuwen (TNO Coordination)
2. Dr. H. van den Heuvel (SPEX Database recording)
3. Prof. L. Boves (CLST, RU Nijmegen)
4. Dr. R. J. F. Ordeman (HMI, Twente University)
5. Prof. dr. P. Wambacq (ESAT, Leuven University)
6. Prof. dr. J.-P. Martens (ELIS, Gent University)
7. Dr. L. J. M. Rothkrantz (EWI Delft University)

**STEVIN funding: 470.000 euro**

**Duration: 01/05/2006 – 30/09/2008**

### Project summary

Over the years, standardised benchmark evaluation tests have proved indispensable for the development of several techniques in speech technology. In N-Best we will organise and execute an evaluation of large vocabulary speech recognition systems trained for Dutch (both Northern and Southern Dutch) in two evaluation conditions (Broadcast News and Conversational Telephony Speech). The goals of the project are the definition of a proper evaluation setup and a corresponding set of benchmark results. The evaluation framework can serve both as a basis for future evaluations, which can probe the progress in large vocabulary speech recognition for Dutch, and as an aid for the development of new speech recognition technologies for the Dutch language. Participants will use a common speech database, the Corpus Gesproken Nederlands (CGN), for acoustic training of their systems, as well as other common resources for language modeling and pronunciation modeling. They will co-operate through exchange of intermediate experiences, results and models of sub-technologies. The evaluation will be open to researchers outside the project, who will benefit from the common training and evaluation resources and the development experiences of the project partners. Intermediate and final exchange of experimental results and findings will be consolidated in workshops. The evaluation will be based on new speech material that will be collected and annotated for the purpose of this evaluation. All evaluation resources, materials and results will be made available via the TST-centrale.

**NBest website:** <http://speech.tm.tno.nl/n-best/>

## STEVIN can PRAAT

### Project co-ordinator

Prof. dr. P.P.G. Boersma  
Universiteit van Amsterdam  
Faculteit der Geesteswetenschappen - Fonetiek  
Herengracht 338  
1016 CG Amsterdam  
Telephone: +31 20 525 2183  
E-mail: [paul.boersma@uva.nl](mailto:paul.boersma@uva.nl)  
URL: <http://www.fon.hum.uva.nl/praat>

### Project consortium

1. Prof. dr. P. Boersma (ACLIC, University of Amsterdam)
2. Prof. dr. F. Hilgers (ACLIC, University of Amsterdam / Nederlands Kanker Instituut - Anthonie van Leeuwenhoekziekenhuis)
3. Prof. dr. V. van Heuven (University of Leiden)
4. Dr. H. van den Heuvel (SPEX: Speech Processing EXpertise centre)
5. Dr. D.J.M. Weenink (ACLIC, University of Amsterdam / SpeechMinded)

**STEVIN funding: 114.000 euro**

**Duration: 01/01/2008 – 30/09/2008**

### Project summary

Appropriate tools are indispensable for the scientist to perform his/her work. This holds true for speech science as well. The PRAAT program<sup>1</sup> is an extensive application for language, music and speech research that is used by approximately 10,000 scientists and students around the globe. Some characteristics that explain its success right from the beginning, are the wide range of features, the user-friendliness and the scriptability, i.e. the possibility to create one's own processing for a series of inputs. The other aspect that adds to the enthusiastic and widespread use is the careful support available. This encompasses user help on diverse levels online, quick response to any questions by email, immediate handling of incidents and solving of problems, and last but not least, an infrastructure for user groups. The knowledge that the PRAAT program entails, is in this means passed on to many colleagues and students. Also, users have a way to relate to one another and share their insights with regard to the possibilities the PRAAT program offers. The software is freely available for all current computer platforms like Linux, Windows and Macintosh. The manuals, FAQ and help menu are included in the package; the user group is available on the internet. Despite the multitude of features already present in the application, some important functionality is still missing. We propose to develop a number of improvements and added functionality that will then additionally and freely become available for speech scientists via the PRAAT program. This project matches the STEVIN objectives since it delivers important tools to all speech scientists who need state of the art technology to tackle the newest ideas and the largest datasets.

**STEVINcanPRAAT website:** <http://www.fon.hum.uva.nl/praat/>

Overview proposals funded in Call for proposals for applied research  
(max. budget 2,3 M€)

<i>acronym</i>	<i>coordinating institute and other academic partners</i>	<i>industrial partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed (subject)</i>	<i>planned duration</i>	<i>funding</i>
AUTONOMATA TOO	Radboud University Nijmegen – CLST (Henk van den Heuvel)  Ghent University Utrecht University	TeleAtlas Nuance	NL/VL	Speech Research  Speech Application  (ASR)	24 mnths	€ 416.000
DAISY	KU Leuven (Sien Moens)  Groningen University	Q-go R&D	VL/NL	Language Research  Language Application  (Summarization)	36 mnths	€ 457.000
DISCO	Radboud University Nijmegen – CLST (Helmer Strik)  Antwerpen University Radboud University Nijmegen – UTC	Polderland Language & Speech Technology	NL/VL	Speech research  Speech Application  (Computer assisted language learning)	36 mnths	€ 499.000
DuOMan	Universiteit van Amsterdam (Maarten de Rijke)  Groningen University Hogeschool Gent	TrendLight GridLine	NL/VL	Language Research  Language Application  (Opinion and sentiment mining)	36 mnths	€ 440.000
PaCo-MT	KU Leuven – CCL (Frank Van Eynde)  Groningen University	OneLiner Language & eBusiness Solutions BVBA	VL/NL	Language Research  Language Application  (Machine translation)	36 mnths	€ 491.000

## **Autonomata TOO**

### **Project co-ordinator**

Dr. H. van den Heuvel  
Radboud Universiteit Nijmegen  
Faculteit der Letteren  
Taalwetenschap  
Postbus 9103  
6500 HD Nijmegen  
Telephone: +31-24-3611686  
E-mail: [heuvel@let.ru.nl](mailto:heuvel@let.ru.nl)  
URL: [www.let.ru.nl](http://www.let.ru.nl)

### **Project consortium**

1. Dr H. van den Heuvel (CLST, Radboud University Nijmegen)
2. Prof. Dr J-P. Martens (ELIS, Ghent University)
3. Dr Ir G. Bloothoof (Utrecht institute of Linguistics (UiL-OTS), Utrecht University)
4. Ir L. Peirlinckx (TeleAtlas, Ghent)
5. ir B. D'hoore (Nuance Communications International, Merelbeke)

**STEVIN funding: 416.000 euro**

**Duration: 01/02/2008 – 31/01/2010**

### **Project summary**

The aim of this application-oriented research project is to build a demonstrator version of a Dutch/Flemish Points of Interest (POI) information providing business service, and to investigate new pronunciation modeling technologies that can help to bring the spoken name recognition component of such a service to the required level of accuracy. The demonstrator service (running on a PC) will contain a simple user interface and a restricted but realistic database of POI information. It will give a flavor of what the envisaged service can offer to the user, and it will also be used as a vehicle for testing the benefits of the newly developed speech technology in a realistic setting, involving tests with end users at strategic moments during the project.

**AUTONOMATA TOO website:** <http://lands.let.ru.nl/projects/AutonomataToo/>

## Dutch IAngeuage Investigation of Summarization technologY (DAISY)

### Project co-ordinator

Prof. dr. M.F. Moens  
Katholieke Universiteit Leuven  
Departement Computerwetenschappen  
Celestijnenlaan 200 A  
B-3001 Heverlee  
Telephone:  
E-mail: sien@cs.kuleuven.be  
URL: <http://www.cs.kuleuven.be/~sien/>

### Project consortium

1. Prof. dr. M.-F. Moens (Department of Computer Science, K.U.Leuven)
2. Dr. G.J.M. van Noord (CLCG/Computational Linguistics, RuG University of Groningen)
3. Dr. Leonoor van der Beek (Q-go Research & Development)

**STEVIN funding: 457.000 euro**

**Duration: 01/05/2008 – 30/04/2011**

### Project summary

Summarization of text is often a necessity when searching and selecting information from document repositories. However, current summarization technology is for a large part restricted to the extraction of sentences. Summarization technology for Dutch is very scarce. The aim of DAISY is to develop and evaluate essential technology for automatic summarization of Dutch informative texts. Innovative algorithms for topic salience detection, topic discrimination, rhetorical classification of content, sentence compression and text generation will be implemented. In addition, a demonstrator will be developed in collaboration with the company Q-Go.

The summarization demonstrator will be tested and evaluated in multiple ways in the QA environment of Q-go on documents in the financial and social security domains. Firstly, the system output will be compared against hand-made abstracts of the documents. Secondly, the effect of adding system-generated headline abstracts on retrieval will be measured. Finally, if suitable training and testing material can be obtained, tests will be done with automated email answering, where the summary of the email is used as input for the Q-go QA system.

**DAISY website:** <http://www.cs.kuleuven.be/~liir/projects.php?project=172>

## Development and Integration of Speech technology into COurseware for language learning (DISCO)

### Project co-ordinator

Dr. W.A.J. Strik  
Radboud Universiteit Nijmegen  
Faculteit der Letteren  
Taalwetenschap  
Postbus 9103  
6500 HD Nijmegen  
Telephone: +31-24-361 61 04  
E-mail: [strik@let.ru.nl](mailto:strik@let.ru.nl)  
URL: [www.let.ru.nl](http://www.let.ru.nl)

### Project consortium

1. Dr. H. Strik (Centre for Language and Speech Technology, Radboud University Nijmegen)
2. Prof. Dr. J. Colpaert (Linguapolis, Universiteit Antwerpen)
3. Drs. J. Bakx (Universitair Taal- en Communicatiecentrum Nijmegen)
4. Dr. I. de Mönnink (Polderland Language & Speech Technology)

**STEVIN funding: 499.000 euro**

**Duration: 01/02/2008 – 31/01/2011**

### Project summary

Language learners are known to fare best in one-on-one interactive learning situations in which they receive optimal corrective feedback. However, providing this type of tutoring by trained language instructors is time-consuming and costly, and therefore not feasible for the majority of language learners. This particularly applies to oral proficiency, where corrective feedback has to be provided immediately after the utterance has been spoken, thus making it even more difficult to provide sufficient practice in the classroom. The recent appearance of Computer Assisted Language Learning (CALL) systems that make use of Automatic Speech Recognition (ASR) and other advanced automatic techniques offers new perspectives for training oral proficiency in a second language (L2).

The present project aims to develop and test a prototype of an ASR-based CALL application for training oral proficiency for Dutch as a second language (DL2). The application optimizes learning through interaction in realistic communication situations and provides intelligent feedback on various aspects of DL2 speaking, viz. pronunciation, morphology and syntax. The communicative settings employed in Nieuwe Buren (New Neighbours, a method for DL2 training developed by Malmberg publishers) will constitute the starting point for the application.

**Disco website:** <http://lands.let.ru.nl/~strik/research/DISCO/index.html>

## Dutch Online Media Analysis (DuOMAn)

### Project co-ordinator

Prof. dr. M. de Rijke  
Universiteit van Amsterdam  
Faculteit der Natuurwetenschappen, Wiskunde en Informatica  
Instituut voor Informatica, Information & Language Processing Systems  
Telephone: +31-20-5255358  
E-mail: [mdr@science.uva.nl](mailto:mdr@science.uva.nl)  
URL: <http://staff.science.uva.nl/~mdr/>

### Project consortium

1. Prof. dr. M. de Rijke (University of Amsterdam (UvA))
2. R. Franz (TrendLight)
3. T. Spaan (GridLine)
4. Dr. G. van Noord (Rijksuniversiteit Groningen (RuG))
5. Dr. V. Hoste (Dept. Vertaalkunde, Hogeschool Gent (HoGent))

**STEVIN funding: 440.000 euro**

**Duration: 01/04/2008 – 31/03/2011**

### Project summary

When marketing campaigns or policies on sensitive or broad-ranging issues need to be defined or revised, access to the opinion of the target group is vital. An explosion in online content---both edited and user-generated---has vastly increased the range of opinions potentially available to media analysts and the general public alike, but efficient and effective access methods are needed to unlock this potential. The DuOMAn project will carry out an ambitious research agenda that will result in the development of a set of Dutch language resources and tools for identifying and aggregating sentiments in online data sources.

DuOMAn aims to transform the volumes of online information that threaten to leave media analysts information-bound into aggregates of attitudes organized by topic by employing classification, information extraction, and cross-document linking. DuOMAn will provide media analysts and members of the general public with focused access to opinionated information on people, products and topics through an online demonstrator for the general public and through integration of the tools and resources it develops into the workflow of professional media analysts. Key research contributions include sentiment-oriented lexical resources and advancement in the areas of automated sentiment analysis, parsing, and entity detection and coreference resolution. Applied research on robustness and adaptability receives central emphasis.

**DuOMAN website:** <http://staff.science.uva.nl/~mdr/Research/Projects/index.html>

## Parse and Corpus based Machine Translation (PaCo-MT)

### Project co-ordinator

Prof. dr. F. van Eynde  
Katholieke Universiteit Leuven  
Faculteit Letteren  
Centrum voor Computerlinguïstiek  
Maria Theresiastraat 21, postbus  
B-3000 Leuven  
Telephone: +32-16-325084  
E-mail: frank.vaneynde@ccl.kuleuven.be  
URL: [www.ccl.kuleuven.be](http://www.ccl.kuleuven.be)

### Project consortium

1. Prof. Dr. F. Van Eynde (Centre for Computational Linguistics (CCL), K.U.Leuven)
2. Dr. J. Tiedemann (Alfa-informatica, Rijksuniversiteit Groningen (RUG))
3. Drs. K. Desmet (OneLiner Language & eBusiness Solutions BVBA)

**STEVIN funding: 491.000 euro**

**Duration: 01/02/2008 – 31/01/2010**

### Project summary

In this project, we aim at building a hybrid machine translation system combining the positive features of corpus based and rule based systems. The primary goal is to develop an open-domain MT system for Dutch-English and Dutch-French (in both directions) integrating proper linguistic analysis and syntactic transfer into a data-driven approach. Compared to other data-driven approaches, we emphasise the improvement of translation quality and the adaptability of the system to the users requirements. This will result in a flexible MT system that is accepted by professional translators. Adaptability to users needs will be supported by a post editing interface, making the system very flexible and able to improve gradually. This novel feature increases the acceptability of the system by professional users. An evaluation of the system by human judgement and automated scores like BLEU/NIST and edit distance will be made, as well as a user test in which the translation speed will be tested.

**PaCo-MT website:** <http://www.ccl.kuleuven.be/~frank/Projects.html>

**Overview proposals funded in three Calls for tender for specific HLT infrastructures  
(max budget 1,6 M€)**

<i>acronym</i>	<i>coordinating institute and other academic partners</i>	<i>industrial partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed (subject)</i>	<i>planned duration</i>	<i>funding</i>
SPRAAK	KU Leuven (Patrick Wambacq)  Radboud Universiteit Nijmegen – CLST TNO Human Factors Universiteit Twente - HMI		VL/NL	Speech resources  (ASR)	26 mnths	€ 400.000
CORNETTO	Free University Amsterdam (Piek Vossen)  Universiteit van Amsterdam KU Leuven	Irion Technologies bv		Language resources  (Semantic lexicon)	24 mnths	€ 399.000
SoNaR	Radboud Universiteit Nijmegen – CLST (Nelleke Oostdijk)  Antwerpen University Hogeschool Gent Leuven University Instituut voor Nederlandse Lexicografie Groningen University Tilburg University Twente University Utrecht University Universiteit van Amsterdam SPEX, Nijmegen	Polderland Logica-CMG Dutchear Nuance IRION Van Dale Lexicografie Dutch HLT Agency	NL/VL	Language resources  speech resources  (Annotated written Dutch corpus)	36 mnths	€ 836.000

## Speech Processing, Recognition & Automatic Annotation Kit (SpraaK)

### Project co-ordinator

Prof. dr. Wambacq  
Katholieke Universiteit Leuven  
ESAT - PSI  
Kasteelpark Arenberg 10  
3001 Heverlee  
Telephone: + 32 16 32 10 57  
E-mail: [patrick.wambacq@esat.kuleuven.be](mailto:patrick.wambacq@esat.kuleuven.be)  
URL: <http://www.esat.kuleuven.be/psi/spraak>

### Project consortium

1. Prof. P. Wambacq (Katholieke Universiteit Leuven - ESAT/PSI)
2. Prof. L.W.J. Boves (Radboud Universiteit Nijmegen - Language and Speech RU)
3. Dr. Ir. D.A. van Leeuwen (TNO Human Factors (Soesterberg) TNO)
4. Dr. R. Ordeman (Universiteit Twente - Human Media Interaction UT)

**STEVIN funding: 400.000 euro**

**Duration: 01/02/2006 – 31/05/2008**

### Project summary

The availability of a speech recognition system for Dutch is mentioned as one of the essential requirements for the language and speech technology (LST) community. Indeed, researchers now are faced with the problem that no good speech recognition tool is available for their purposes or existing tools lack functionality or flexibility. This project has two primary goals that will be accomplished within a single software framework. The first goal is to develop a highly modular toolkit for research into speech recognition algorithms. It allows researchers to focus on one particular aspect of speech recognition technology without needing to worry about the details of the other components. The second goal is to provide a state-of-the-art recogniser for Dutch with a simple interface, so that it can be used by non-specialists with a minimum of programming requirements. Next to speech recognition, the resulting software will enable applications in related fields as well. Examples are linguistic and phonetic research where the software can be used to segment large speech databases or to provide high quality automatic transcriptions. We choose the existing ESAT recogniser, augmented with knowledge and code from the other partners in this project, as a starting point. This code base will be transformed to meet the specified requirements. The transformation is accomplished by improving the software interfaces to make the software package more user friendly and adapted for usage in a large user community, and by providing adequate user and developer documentation written in English, so as to make it easily accessible to the international LST community as well. Next to providing a reference speech recognition platform for the Dutch speaking community, this project also encompasses knowledge transfer between the different partners, hence strengthening the ties between the Netherlands and Flanders, and between research institutions and application developers.

**SPRAAK website:** <http://www.esat.kuleuven.be/psi/spraak/projects/index.php?proj=SPRAAK>

## Combinatorial and Relational Network as Toolkit for Dutch Language Technology (Cornetto)

### Project co-ordinator

Prof. dr. P. Vossen  
Vrije Universiteit Amsterdam  
Onderzoeksgroep Lexicologie/Terminologie  
Faculteit der Letteren  
De Boelelaan 1105, Kamer 11A-24  
Telephone: +31 20-5986466,  
E-mail: [p.vossen@let.vu.nl](mailto:p.vossen@let.vu.nl)  
URL: <http://www.let.vu.nl/organisatie/medewerkers.htm>

### Project consortium

1. Prof. Dr. W. Martin (Vrije Universiteit Amsterdam)
2. Prof. Dr. M. de Rijke (Universiteit van Amsterdam)
3. Prof. Dr. M.-F. Moens (Katholieke Universiteit Leuven)
4. Prof. Dr. P. Vossen (Irion Technologies BV)

**STEVIN funding: 399.000 euro**

**Duration: 01/04/2006 – 31/03/2008**

### Project summary

Cornetto will build a lexical semantic database for Dutch, covering 40K entries, including the most generic and central part of the language and a specialized database for the legal and finance domain. The database will go beyond the structure and content of Wordnet and FrameNet. It will contain both vertical and horizontal semantic relations and combinatorial lexical constraints such as multiword expressions, idioms and collocations on the one hand, and lexical functions and frames on the other. The concepts will be aligned with the English Wordnet so that ontologies and domain labels can be imported. The semantic layer will be validated with a formal ontology, to make it usable in Semantic Web environments. In addition, Cornetto will develop a toolkit for the acquisition of new concepts and relations and the tuning and extraction of a domain specific sub-lexicon from a compiled corpus. A sub-lexicon will be extracted for the legal and finance domain. The lexical database will be evaluated by integration in IR and QA applications and the sub-lexicon will be evaluated by a user-group of language technology companies.

**Cornetto website:** <http://www.let.vu.nl/onderzoek/projectsites/cornetto/>

## Stevin Nederlandstalig Referentiecorpus (SoNaR)

### Project co-ordinator

Dr. N. Oostdijk  
Radboud Universiteit Nijmegen  
Faculteit der Letteren  
Centre for Language and Speech Technology (CLST)  
Erasmusplein 1  
Postbus 9103  
NL-6500 HD Nijmegen  
Telephone: +31 24 361 27 65  
E-mail: [n.oostdijk@let.ru.nl](mailto:n.oostdijk@let.ru.nl)  
URL: [www.let.ru.nl](http://www.let.ru.nl)

### Project consortium

1. Dr. Nelleke Oostdijk (CLST, Radboud University Nijmegen)
2. Dr. Veronique Hoste (Dept. Vertaalkunde, Hogeschool Gent (HoGent))
3. Prof. dr. Franciska de Jong (Human Media Interaction (HMI), Twente University)
4. Dr. Martin Reynaert (Induction of Linguistic Knowledge (ILK), Tilburg University)
5. Dr. Henk van den Heuvel (CLST, Radboud University Nijmegen)
6. Dr. Paola Monachesi (UIL-OTS, Utrecht University)
7. Dr. Ineke Schuurman, (CCL, Leuven University)
8. Dr. Vincent Vandeghinste, (CCL, Leuven University)

### Members advisory Board

Beeken, INL; van den Bosch, Tilburg; Daelemans, Antwerpen; Moens & Van Eynde, Leuven; van Noord, Groningen; Vandeweghe, Gent.

### Members User Group

Bouma, Groningen; Boves, Nijmegen; Geeraerts, Leuven; van den Heuvel, Polderland; Iskra, Logica; Jongbloed, Duthear; Odijk, Nuance; de Rijke, Amsterdam; Van Veenendaal, HLT Agency; Vossen, Irion; Zuidema, Van Dale Lexicografie

**STEVIN funding: 836.000 euro**

**Duration: 01/01/2008 – 01/01/2011**

### Project summary

The project aims at the construction of a 500-million-word reference corpus of contemporary written Dutch for use in different types of linguistic (incl. lexicographic) and HLT research and the development of applications. The project will build on the results obtained in the D-COI and COREA projects which were awarded funding in the first call of proposals within the STEVIN programme. In the light of the budgetary constraints of the present call and the work conducted within other STEVIN projects (especially the LASSY project), the present project will focus on the compilation of the corpus, while the entire corpus will be (automatically) POS tagged and lemmatized by means of the D-COI tagger/lemmatizer. In addition, for a one-million-word subset of the corpus different types of semantic annotation will be provided, viz. named entity labelling, annotation of co-reference relations, semantic role labelling and annotation of spatial and temporal relations. The corpus will be made available through the Dutch HLT Agency (TST-Centrale)

**SoNaR website:** <http://lands.let.ru.nl/projects/SoNaR/>

**Overview proposals funded in three Calls for demonstration projects  
(max. budget 1,0 M€)**

<i>acronym</i>	<i>coordinating SME and other partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed</i>	<i>planned duration</i>	<i>funding</i>
Rechtsorde	C-CONTENT b.v. Polderland Language & Speech Technology b.v.	NL	Language technology demonstrator system	15 mnths	€ 90.000
Gemeente-Connect!	IRION Technologies b.v. Duthear b.v. Gemeente Gilze en Rijen	NL	Language and speech technology demonstrator system	10 mnths	€ 60.000
Spraak-gestuurde Nummerbord Retrieval Tool	Politie Utrecht Duthear b.v.	NL	Speech technology demonstrator system	10 mnths	€ 50.315
Audiokrant	Sensotec NV De Braillekrant vzw KU Leuven-SCD	VL	Speech technology demonstrator system	15 mnths	€ 92.400
Spelling- en grammatica-controle voor dyslectische gebruikers	Polderland Language & Speech Technology b.v. Technologie & Integratie bvba Die-'s-lekti-kus vzw	NL/VL	Language technology demonstrator system	12 mnths	€ 96.730
Rechtspraak-herkenning	Telecats BV Carp Technologies BV	NL	Language and speech technology demonstrator system	16 mnths	€ 97.000
Klinkende Taal	Gridline BV Utrecht University KU Leuven STIL, Tilburg Provincie Brabant Gemeente Den Haag	NL	Language technology demonstrator system	15 mnths	€ 92.200
SpelSpiek	INL ELITECH Polderland Language & Speech Technology b.v. Van Dale Lexicografie	VL/NL	Language technology demonstrator system	12 mnths	€ 72.387
Voice Assess	Telecats BV VO Consulting	NL	Speech technology demonstrator system	5 mnths	€ 45.000

<i>acronym</i>	<i>coordinating SME and other partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed</i>	<i>planned duration</i>	<i>funding</i>
Alfabetisering Anderstaligen Plan (AAP)	Polderland Language & Speech Technology b.v. BEMO-materiaalontwikkeling Uitgeverij Boom Radboud University Nijmegen	NL	Language technology demonstrator system	15 mnths	€ 56.000
Easy Info	Irion Technologie BV Carp Technologies MD Info	NL	Language technology demonstrator system	6 mnths	€ 20.000
Hulp bij Auditieve Training na Cochleaire Implantatie (HATCI)	Advanced Binics NV ONICI KU Leuven	VL	Speech technology demonstrator system	8 mnths	€ 48.482
Nederlands-talige Ondertiteling (Neon)	Telecats bv, Vlaamse Radio en Televisie, Ned. Publi. Omroep, K.U.Leuven - ESAT/PSI, Universiteit Gent - ELIS, Universiteit Antwerpen - CNTS	NL/VL	Language and speech technology demonstrator system	12 mnths	€ 86.000
Sprekende zelfcorrigerende woordvoorspeller voor dyslectische gebruikers (WooDy)	Sensotec NV Lexima bv	VL	Language technology demonstrator system	15 mnths	€ 90.000

**Overview proposals funded in 1st Call for educational projects  
(max. budget € 27.500)**

<i>acronym</i>	<i>coordinating institute and other partners</i>	<i>VL/NL consortium nationality</i>	<i>STEVIN priorities addressed (subject)</i>	<i>planned duration</i>	<i>funding</i>
Vooronderzoek TST in het voortgezet onderwijs	Stichting Studio Taalwetenschap	NL	raising awareness for HLT in secondary schools	9 mnd	€ 15.000
TST op Kennislink	Utrecht University Kennislink	NL	raising awareness for HLT in secondary schools	12 mnths	€ 27.500

# ນຸ | STEVIN

## Rechtsorde

**STEVIN funding: 90.000 euro**

### Consortium partners

1. C-CONTENT b.v., contact person: Marcel Mooren, [m.mooren@c-content.nl](mailto:m.mooren@c-content.nl)
2. Polderland Language & Speech Technology b.v., contact person: Wilko Apperloo

**Duration: 01/01/2006 – 01/04/2007 (15 months)**

### Project summary

De Nederlandse overheid is er de laatste jaren meer en meer toe overgegaan om elektronische informatie op het gebied van wet- en regelgeving (W&R) publiek toegankelijk te maken. Helaas wordt deze informatie verspreid over vele (niet gestandaardiseerde) websites van de overheid gepubliceerd. Dit maakt het haast onmogelijk voor een professionele gebruiker om de gezochte informatie snel boven water te krijgen. Er is daarom grote behoefte aan één centrale ingang waar alle openbare W&R informatie volledig en snel doorzocht kan worden. C-CONTENT is begin 2005 in dit "gat" gestapt en heeft een systeem "Rechtsorde.nl" gebouwd dat dagelijks, (geautomatiseerd) alle wet- en regelgeving informatie vergaart van verschillende vrij toegankelijke overheidssites en deze informatie vervolgens middels één portaal, [www.rechtsorde.nl](http://www.rechtsorde.nl), doorzoekbaar maakt. Rechtsorde.nl is gericht op de professionele eindgebruiker en bevat o.a. wetten, jurisprudentie, CAO's, ministeriele regelingen, officiële publicaties, verordeningen van lokale overheden etc. In dit demonstratieproject zal de zoekfunctionaliteit van Rechtsorde.nl uitgebreid worden met tal van taalondersteunende gereedschappen van Polderland. Het doel is dat de gezochte documenten gebruiksvriendelijker en efficiënter gevonden kunnen worden en dat de gebruiker middels suggesties meer geholpen wordt bij het vinden van de juiste documenten.

**Rechtsorde website and demonstration (in Dutch):** <http://www.rechtsorde.nl/>

## GemeenteConnect!

**STEVIN funding: 60.000 euro**

### Consortium partners

1. Irion Technologies BV, contact person: Joop van Gent: [gent@irion.nl](mailto:gent@irion.nl)
2. Dutchear BV, contact person: Victor Huisman
3. Gemeente Gilze en Rijen, contact person: Frank Meulendijks

**Duration: 01/01/2006 – 01/11/2006 (10 months)**

### Project summary

Gemeentes in Nederland werken aan een overbrugging van de kloof tussen overheid en burger. Zij kampen echter alle met een groot probleem: de hoeveelheid vragen die telefonisch of in direct baliecontact op ze afkomen is dermate groot dat de vraag vaak de capaciteit overstijgt. Het project GemeenteConnect! wordt opgezet om aan te tonen dat een slimme combinatie van spraak- en taaltechnologie dit probleem voor een fors deel kan oplossen: de meest voorkomende telefonische burgervragen aan gemeentes moeten ermee kunnen worden afgehandeld. Irion en Dutchear, beide spin-offs van TNO gevestigd in Delft, hebben een systeem ontwikkeld, waarmee via de telefoon interactief en op natuurlijke wijze informatie kan worden opgevraagd uit grote databases, zonder dat de gebruikers steeds met menu-toetsen worden geconfronteerd. De voordelen van het systeem voor een gemeente zijn onder andere:

- (a) geen wachttijden voor de burger;
- (b) geen menu-toetsen;
- (c) het systeem heeft verstand van alle onderwerpen, dus er hoeft niet te worden doorverbonden;
- (d) het systeem is zelflerend, op basis van gesprekken met burgers, en kan dus steeds beter antwoord geven;
- (e) het systeem kan omgaan met emoties;
- (f) het systeem kan ook als digitaal loket op de website worden geplaatst, waardoor een "chat"-functie ontstaat.

Een belangrijk onderdeel van het project betreft PR-werkzaamheden om deze specifieke en succesvolle combinatie van taal- en spraaktechnologie voor gemeentes landelijke bekendheid te geven in zowel Nederland als Vlaanderen.

**Gemeenteconnect website and demonstration (in Dutch):** <http://www.gemeenteconnect.nl/>



## **Spraakgestuurde Nummerbord Retrieval Tool**

**STEVIN funding: 50.315 euro**

### **Consortium partners**

1. Politie Utrecht, contactpersonen: Janneke Huijssoon, René Anker
2. Dutchear BV, contact person: Els Nachtegaal, E.Nachtegal@dutchear.nl

**Duration: 15/03/2006 – 30/09/2006 (10 months)**

### **Project summary**

Dutchear ontwerpt in samenwerking met de Politie Utrecht de Nummerbord Retrieval Tool. De Nummerbord Retrieval Tool zorgt ervoor dat agenten van Politie Utrecht altijd op een snelle, gemakkelijke en veilige manier voertuiginformatie kunnen krijgen. Momenteel belt een agent met zijn GSM naar de meldkamer of naar de infodesk, wanneer hij een kentekenplaat wil natrekken. De snelheid waarmee hij geholpen wordt is geheel afhankelijk van de beschikbaarheid van medewerkers op de meldkamer of bij de infodesk. De lijnen zijn echter regelmatig bezet waardoor de wachttijd voor de agent oploopt. De huidige situatie is daarom onwenselijk. Hoe sneller een agent over de relevante informatie beschikt, hoe veiliger de situatie voor hem en de maatschappij is. In de tijd dat de agent moet wachten op de informatie blijft mogelijk de onverzekerde auto doorrijden, of laat de agent een bestuurder van een gestolen auto wegrijden. Agenten kunnen lopend, op de mountainbike, in de auto en op de motor bellen met de Nummerbord Retrieval Tool (NRT). De agent spreekt het kenteken in en krijgt informatie (naam eigenaar, APK, verzekering, gestolen) over het betreffende voertuig teruggekoppeld via een Text-To-Speech engine (sprekende computer). Naast de terugkoppeling van de informatie door de telefoon ontvangt de agent bovendien een SMS met de aan hem voorgelezen informatie.

**For privacy reasons no live demonstration is available, a movie in which the system is re-enacted is being made at the moment**

## Audiokrانت

**STEVIN funding: 92.400 euro**

### Consortium partners

1. Sensotec NV, contact person: Frank Allereersch, [frank.allemeersch@sensotec.be](mailto:frank.allemeersch@sensotec.be)
2. De Braillekrانت vzw, contact person: Katty Kloeck
3. Katholieke Universiteit Leuven-SCD, contact person: Jan Engelen

**Duration: 15/02/2007 – 14/05/2008 (15 months)**

### Project summary

Voor personen met een leeshandicap is de toegankelijkheid tot kranteninformatie allesbehalve evident. Er bestaan dan ook al sinds enkele jaren speciale voorzieningen om deze toegankelijkheid te bewerkstelligen. In Vlaanderen zijn dat de initiatieven Braillekrانت en DiGiKrانت (gecoördineerd door De Braillekrانت vzw), waarbij respectievelijk een extractie van de krانت in Braille en een volledige krانت in digitale vorm wordt aangeboden. Voor het lezen van de krانت in digitale vorm dient men te beschikken over een pc uitgerust met vergrotingssoftware, synthetische spraakoutput en/of een braille leeslijn. De beperking tot lezers met kennis van braille of die kunnen beschikken over pc met extra uitrusting gecombineerd met een voldoende basiskennis in pc gebruik heeft als gevolg dat het gedeelte van de doelgroep dat de krانت kan lezen toch nog vrij beperkt blijft. Anderzijds is er sinds 2004 voor wat betreft de gesproken boeken voor personen met een leeshandicap zowel in Vlaanderen als in Nederland de overstap gemaakt van verspreiding op cassette naar verspreiding op data-CD. Voor de verstrekking op CD maakt men gebruik van de internationale DAISY standaard, waarmee zowel audio als tekst op eenzelfde drager kan geplaatst worden. Voor het beluisteren van de Daisy cd's bestaan er specifieke voorleesapparaten en ongeveer iedere regelmatige gebruiker van gesproken boeken in Vlaanderen en Nederland beschikt ondertussen over zo'n (draagbaar) voorleesapparaat. Het gaat hierbij om een paar tienduizend dergelijke apparaten. Binnen het AudioKrانت project zullen we dagelijks een versie van de krانت produceren die conform is met de Daisy standaard en kan voorgelezen worden met die voorleesapparaten. Vanwege het tijdskritische karakter van de productie van een krانت, is het uitgesloten dat we, zoals voor de productie van gesproken boeken, gaan gebruik maken van voorlezers. Naar onze overtuiging kan de aanwending van spraaktechnologie (synthetische spraak) en hoogtechnologische taaltechnologie (voor de optimalisatie ervan) hier echter de oplossing brengen.

**Audiokrانت website and demonstration (in Dutch):** [http://www.braillekrانت.be/nieuws\\_detail.php?nr=13](http://www.braillekrانت.be/nieuws_detail.php?nr=13)

## Spelling- en grammaticaconrole voor dyslectische gebruikers

STEVIN funding: 96.730 euro

### Consortium partners

1. Polderland Language & Speech Technology bv, contact person: Inge de Mönnink, inge@polderland.nl
2. Technologie & Integratie b.v.b.a., contact person: Jo Cremelie
3. Die-'s-lekti-kus vzw, contact person: Dirk Callebaut

**Duration: 15/02/2007 – 14/02/2008 (12 months)**

### Project summary

Het resultaat van dit project is een spellingcontrole en een grammaticaconrole aangepast voor dyslectische gebruikers. De standaard spelling- en grammaticaconrole in Microsoft® Office worden in dit project zodanig aangepast dat ze beter aansluiten bij de typische fouten die dyslectische gebruikers maken (bijvoorbeeld 'eemoscho nele' i.p.v. 'emotionele' en 'brugste' i.p.v. 'beruchtste'). Bovendien wordt aan de spelling- en grammaticaconrole de mogelijkheid toegevoegd om suggesties voorgelezen te krijgen door een spraaksyntheseprogramma. Omdat dyslectische gebruikers behalve spellingproblemen ook leesproblemen hebben ondersteunt de combinatie van een aangepaste spellingcontrole en een spraaksyntheseprogramma de dyslectische gebruiker maximaal in hun schrijfproces. Als laatste wordt ook de interface van de spellingcontrole aangepast op dyslectische gebruikers. Het project richt zich op dyslectische kinderen. Hierdoor kan het product al in de onderwijssituatie optimaal worden ingezet en zal het aantal kinderen dat door hun taalbeperking in het onderwijs buiten de boot valt verder beperkt kunnen worden. Aangezien de spelling- en grammaticaconrole ingebed zitten in Office en dus onder andere te gebruiken zijn in Word en Outlook is het eindresultaat ook zeer nuttig te gebruiken door volwassen dyslectische en door anderen meet een taalbeperking zoals niet-moedertaalsprekers van het Nederlands, slechtzienden en kinderen met leerproblemen.

**Spelling- en grammaticaconrole voor dyslectische gebruiker website:** [www.polderland.nl](http://www.polderland.nl)

## Rechtspraakherkenning

STEVIN funding: 97.000 euro

### Consortium partners

1. Telecats BV, contact person: W. Luimes, [w.luimes@telecats.nl](mailto:w.luimes@telecats.nl)
2. Carp Technologies BV, contact person: D. Lie

**Duration: 15/02/2007 – 14/06/2008 (16 months)**

### Project summary

Rechtbanken in Nederland zien zich in toenemende mate verplicht om de geluidsopnamen in de rechtszaal volledig uit te schrijven. Met behulp van bestaande taal- en spraaktechnologie is het mogelijk hulpmiddelen te ontwikkelen die de tijd die gemoeid is met het uitschrijven van gesproken geluidsopnamen, aanzienlijk kan verkorten. Bovendien kan vervolgens op relatief eenvoudige wijze de eenmaal uitgeschreven tekst doorzoekbaar worden gemaakt zodat gevonden passages dmv een muisklik ook beluisterbaar worden. Bijkomend voordeel van het inzetten van deze technologie is dat daarmee een goede basis wordt gelegd voor additionele toepassingen en innovaties, zoals bijvoorbeeld het (semi-) automatisch samenvatten van conversaties. Centraal in dit voorstel is dat technologie hier moet worden ingezet als hulpmiddel en niet als substitutie. Dat houdt in dat het werk nog steeds door (dezelfde) mensen wordt gedaan, maar dat door het inzetten van hulpmiddelen de benodigde tijd en dus werkdruk sterk verlaagd wordt.

**For privacy reasons no live demonstration will be made available, a movie in which the system is re-enacted is being made at the moment**

## Klinkende Taal

**STEVIN funding: 92.200 euro**

### Consortium partners

1. GridLine BV, contact person: Oele Koornwinder, [oele@gridline.nl](mailto:oele@gridline.nl)
2. Faculteit der Letteren van de Universiteit Utrecht - UiL OTS, contact person: H. Pander Maat
3. Faculteit Letteren van de Katholieke Universiteit Leuven - Centrum voor Computerlinguïstiek, contact person: Frank Van Eynde
4. Stichting Toepassing Inductieve Leertechnieken, contact person: Antal Van den Bosch
5. Provincie Brabant, contact person: H. Maaskant
6. Gemeente Den Haag - Dienst Voorlichting en Ext. Betrekkingen, contact person: H. De Kievith

**Duration: 15/02/2007 – 14/05/2008 (15 months)**

### Project summary

Van de Nederlandse overheid wordt in toenemende mate verwacht dat zij klare taal spreekt. Overheidsinstellingen produceren veel publieksgerichte teksten, in brochures en brieven en op websites. De leesbaarheid van de publieksgerichte communicatie kan worden verbeterd door de teksten van ambtelijk jargon te ontdoen. Het demonstratieproject speelt in op deze opgave door een dynamische jargon-bewaker op de markt te brengen. Het betreft een op maat aangeboden toepassing, die overheidsinstellingen in staat stelt hun teksten begrijpelijker te maken, namelijk door de opsporing en vervanging van termen die de doelgroep als jargon zal ervaren. Deze dynamische Jargonbewaker onderscheidt zich van bestaande woordkeuzetools doordat hij zich automatisch aanpast aan het kennisdomein van de organisatie en de doelgroep, alsmede aan de veranderingen die hierin optreden. De tool wordt aangeboden in een laagdrempelige vorm die aansluit op de bestaande werkwijze van de gebruiker. Het project richt zich speciaal op jargon-bewaking in publieksteksten van de lagere overheid, te weten provincies en gemeenten. Om deze lagere overheden te overtuigen van het nut van de applicatie zal een Jargonbewaker-op-maat worden gebouwd voor twee proefgebruikers, te weten de provincie Brabant en de gemeente Den Haag. De effectiviteit van deze demonstrators wordt aangetoond door middel van een leesexperiment met proefpersonen. Het project voorziet tot slot in een grootscheeps marketing-offensief, waarbij overheidsinstellingen en communicatie-adviesbureaus via presentaties en workshops kennis zullen maken met de doeltreffendheid van automatische jargon-opsporing.

**Klinkende Taal website and demonstration (in Dutch):** <http://www.klinkendetaal.nl/>

## SpelSpiek

**STEVIN funding: 72.387 euro**

### **Consortium partners**

1. Instituut voor Nederlandse Lexicologie, dependance Vlaanderen, contact person: Katrien Van pellicom, [pellicom@inl.nl](mailto:pellicom@inl.nl)
2. Elitech, contact person: J. Brouwers
3. Polderland Language & Speech Technology bv, contact person: Inge de Mönnink
4. Van Dale Lexicografie bv, contact person: Johan Zuidema

**Duration: 15/02/2007 – 14/02/2008 (12 months)**

### **Project summary**

Op 1 augustus is de nieuwe spelling ingegaan. De spellingregels en meest recente bijstellingen aan die regels zijn lang niet bij iedereen bekend. Vooral jongeren zijn vaak niet op de hoogte van de spellingregels, maar ook de professionele taalgebruiker heeft wel eens zijn twijfels over de manier waarop je een bepaald woord moet schrijven. Er bestaan al verschillende kanalen via welke je de spelling van woorden kunt opzoeken, of de officiële regels van de spelling van de Nederlandse taal kunt bestuderen. De Taalunie heeft een website waar je de woorden uit de Woordenlijst van de Nederlandse Taal kunt opzoeken, en waar je de regels kunt lezen. Het Groene Boekje bestaat bovendien zowel in boekvorm als op cd-rom, en er is bovendien een elektronische versie van het Groene Boekje gratis online beschikbaar. Dynamische communicatiemiddelen als MSN en sms zijn erg populair, vooral onder jongeren. Het hierboven beschreven project maakt het mogelijk om deze communicatiemiddelen te gebruiken als spellinghulp, door het inzetten van een chatbot. Dat is een robot waarmee je via MSN kunt chatten. In dit geval is het een spellingchatbot: je kunt er bijvoorbeeld aan vragen: "Hoe spel je bjoetiekees?" De chatbot geeft dan direct het juiste antwoord terug. Op die manier heb je een snelle feedback over de juiste spelling van een woord. Zowel achter de computer als onderweg, want dezelfde service stellen we ook via sms beschikbaar. Daarnaast is de service ook gewoon via de webbrowser te bereiken. Drie moderne, populaire communicatiemiddelen dus. Bovendien wordt de bot door de tijd heen slimmer: woorden die de bot niet kent (of foutieve spellingen daarvan), worden bekeken door een spellingdeskundige, waarna die informatie wordt toegevoegd aan de bot. Op die manier wordt hij dus steeds beter in het corrigeren van woorden.

**Klinkende Taal website and demonstration (in Dutch):** <http://www.spelspiek.nl/>

## Voice Assess

**STEVIN funding: 45.000 euro**

### Consortium partners

1. Telecats BV, contact person: W. Luimes, [w.luimes@telecats.nl](mailto:w.luimes@telecats.nl)
2. VO Consulting, contact person: Geert van Ouwerkerk

**Duration: 15/02/2007 – 14/07/2007 (5 months)**

### Project summary

Bedrijven besteden erg veel tijd en geld aan het selecteren van geschikte kandidaten voor het werken in call centers omdat slechts 10% van degene die zich aanmelden daadwerkelijk geschikt blijkt te zijn. Een goede automatische voorselectie geeft bedrijven de mogelijkheid om meer tijd en aandacht te besteden aan de geschiktheid van de geselecteerde kandidaten. Om dit te kunnen doen wordt een applicatie gemaakt die geheel automatisch een (min-of-meer voorgebakken) conversatie met de kandidaten aangaat. Spraakherkenning wordt gebruikt om te meten of bepaalde essentiële woorden wel of niet gezegd zijn. De dialoog verloopt op basis van de gegeven antwoorden omdat een vraag nogmaals (op een andere wijze) wordt gesteld wanneer één of meerdere sleutelwoorden ontbreken. De kandidaten die door het systeem gebeld worden, moeten eerst een reeds bestaande web-applicatie met goed gevolg doorlopen hebben. Deze web-applicatie die een gedegen uitleg geeft over het werken in het call center, is er op gericht de kandidaten te testen op hun kennis van de verschillende telefoniesystemen die ze gaan gebruiken. Als de kandidaten de web-applicatie met goed gevolg doorlopen hebben, kunnen ze het telefoonnummer invullen waarop ze bereikbaar zijn. De hier voorgestelde applicatie gaat ze dan op dat nummer bellen en begint dan de dialoog. Op deze gecombineerde manier (web en telefonie) kunnen veel kandidaten snel en tegen geringe kosten beoordeeld worden op hun mogelijke geschiktheid om als call center medewerker aan de slag te gaan. De applicatie is dus bedoeld voor de voorselectie om het kaf van het koren te scheiden. De eigenlijke selectie gebeurt daarna op de "ouderwetse" manier.

**Voice Assess website and demonstration (in Dutch):** <http://www.webassess.nl/>

## **Alfabetisering Anderstaligen Plan (AAP)**

**STEVIN funding: 56.000 euro**

### **Consortium partners**

1. Polderland Language & Speech technology bv, contact person: Peter Beinema, peter@polderland.nl
2. BEMO-materiaalontwikkeling, contact person: Ad Bakker
3. Uitgeverij Boom, contact person: Geert van der Meulen
4. Radboud Universiteit Nijmegen, contact person: I. Van de Craats

**Duration: 01/04/2008 – 30/06/2009 (15 months)**

### **Project summary**

Dit project implementeert een demonstrator die bestaande spraaktechnologie toepast in het kader van alfabetisering. Hierbij is onmiddellijke feedback essentieel. De methode AAP (alfabetisering anderstaligen plan) wordt hiervoor gevolgd. De technologie zal kunnen geïntegreerd worden in toepassingen van derden.

## Easy Info

**STEVIN funding: 20.000 euro**

### Consortium partners

1. Irion Technologies bv, contact person: Joop van Gent, [gent@irion.nl](mailto:gent@irion.nl)
2. Carp Technologies, contact person: Danny Lie
3. MD Info contact person: Bert Ponsen

**Duration: 15/02/2008 – 14/08/2008 (6 months)**

### Project summary

In het kader van de nieuwsvoorziening is er een tendens naar dienstverlening zoals “news brokers” of knipseldiensten. Klanten van deze dienstverlening kunnen een profiel opgeven in de vorm van trefwoorden. Dat profiel wordt dan gebruikt om een selectie te maken uit de actuele nieuwsberichten. Het aanmaken van profielen op basis van trefwoorden vereist veel handwerk en de “matching” blijft laag. Automatische methoden daarentegen falen dikwijls omdat er gebruik gemaakt wordt van eenvoudige zoektechnologie of statistische methodes. Dit project zal een betere “matching” verwezenlijken. Als demo koppelt men een classificatiesysteem en een samenvattingsgenerator aan het standaardplatform van een aanbieder van gepersonaliseerde informatie. Met behulp van een testgroep worden er evaluaties uitgevoerd om de kwaliteit van het systeem te testen.

## Hulp bij Auditieve Training na Cochleaire Implantatie (HATCI)

**STEVIN funding: 48.482 euro**

### **Consortium partners**

1. Advanced Bionics NV, contact person: Filiep Vanpoucke, [filiepv@abionics.fr](mailto:filiepv@abionics.fr)
2. ONICI contact person: Leo De Raeve
3. K.U.Leuven - ESAT/PSI, contact person: Hugo Van hamme

**Duration: 01/04/2008 – 30/11/2008 (8 months)**

### **Project summary**

Tijdens dit project wordt een applicatie gebouwd die m.b.v. een automatische spraakbeoordeling een therapeut ondersteunt bij het toepassen van de "speech tracking" als hoortherapie en -evaluatie bij revalidatie na cochleaire implantatie. Na cochleaire implementatie dient de patiënt te leren spreken en horen met zijn nieuwe implantaat. De doelgroep zijn vooral patiënten die reeds tot een goede articulatie komen, maar voor wie de hoornauwkeurigheid, het taalgevoel en de grammaticaverwerving verder gestimuleerd moeten worden. De demonstrator zal vooraf opgenomen teksten aan de patiënt aanbieden en hij/zij moet de tekst herhalen. De correctheid van deze herhaling wordt beoordeeld d.m.v. automatische spraakherkenning revalidatiestap.

## Nederlandstalige Ondertiteling

**STEVIN funding: 86.000 euro**

### **Consortium partners**

1. Telecats bv, contact person: Michel Boedeltje, [m.boedeltje@telecats.nl](mailto:m.boedeltje@telecats.nl)
2. Vlaamse Radio en Televisie, contact person: Bernard Dewulf
3. Nederlandse Publieke Omroep, contact person: Jurgen Lentz
4. K.U.Leuven - ESAT/PSI, contact person: Patrick Wambacq
5. Universiteit Gent - ELIS, contact person: Jean-Pierre Martens
6. Universiteit Antwerpen - CNTS, contact person: Walter Daelemans

**Duration: 01/04/2008 – 31/03/2009 (12 months)**

### **Project summary (in Dutch)**

In dit project zal een geavanceerde en minder arbeidsintensieve spraakherkenningstoepassing geïmplementeerd worden voor ondertiteling van televisieprogramma's, met name gerealiseerd door het gecondenseerd aligneren van bestaande teksten of scripts met gesproken audio. Dit zal leiden tot een (semi-)automatische ondertiteling in het Nederlands. Dit gebeurt m.b.v. een spraakherkenningssysteem, waardoor automatisch rechtsstreekse transcriptie van de audiostroom (het resultaat van de spraakherkenning) altijd in de achtergrond aanwezig is om op terug te vallen.

### **Project summary (in English)**

This project will implement a less labour intensive application of speech recognition for television subtitling, in particular by using condensed alignment of existing texts or scripts with the speech audio. This should lead to (semi-)automatic subtitling in Dutch. The use of speech recognition provides an automatic direct transcription of the speech in the background, to replace the text or script as a fall back.



## **Spreekende zelfcorrigerende woordvoorspeller voor dyslectische gebruikers (WooDy)**

**STEVIN funding: 90.000 euro**

### **Consortium partners**

1. Sensotec NV, contact person: Frank Allemeersch, [frank.allemeersch@sensotec.be](mailto:frank.allemeersch@sensotec.be)
2. Lexima bv, contact person: Ria Janssen

**Duration: 15/02/2008 – 14/05/2009 (15 months)**

### **Project summary**

Dit project bouwt een spreekende zelfcorrigerende woordvoorspeller voor dyslectische gebruikers d.m.v. van een combinatie van zelfcorrectie en woordvoorspelling. De kern bestaat uit de ontwikkeling van een basisset van woordenlijsten waaruit voorspelling wordt afgeleid, en van algoritmes ter bepaling van welke woorden aangereikt zullen worden rekening houdend met persoon-specifieke beperkingen. Dit alles wordt geïmplementeerd en gedemonstreerd met een prototype spreekende woordvoorspeller. Doelgroepen zijn individuele gebruikers met lees- en taalbeperkingen, en omkaderende dienstverlening.

## TST-pagina's voor Kennislink

**STEVIN funding: 27.500 euro**

### Consortium partners

1. Landelijke Onderzoekschool Taalkunde (LOT), contact person: Mw. drs. M.M. Jansen, redacteur taalwetenschappen, [Mathilde.Jansen@kennislink.nl](mailto:Mathilde.Jansen@kennislink.nl)
2. Kennislink (Stichting Nationaal Centrum voor Wetenschap en Technologie): contact person R. Smalenburg, manager & projectleider, [ronald@ijburglaan.nl](mailto:ronald@ijburglaan.nl)

**Duration: 01/03/2008 – 28/02/2009 (12 months)**

### Project summary

Het project 'Taal- en Spraaktechnologie op Kennislink' bestaat uit twee componenten:

1. Het populariseren van beschikbaar materiaal uit het vakgebied. Uit eerdere gesprekken tussen A. van Hessen (Notas) en Kennislinkredacteur M. Jansen is gebleken dat er veel materiaal voorhanden is binnen de Taal- en Spraaktechnologie dat geschikt is om toegankelijk gemaakt te worden voor een breed publiek. Als voorbeeld kunnen de artikelen uit het tijdschrift over Toegepaste Taal- en Spraaktechnologie, Dixit, genoemd worden. Het onderzoek binnen de TST leent zich erg goed voor popularisering, omdat veel van de onderwerpen een groot publiek aanspreken. De redacteur TST zal zich daarom voornamelijk bezig houden met het beschikbaar maken binnen Kennislink van reeds voorhanden materiaal. Daarnaast zal een netwerk van correspondenten binnen de Taal- en Spraaktechnologie worden opgezet. Van de correspondenten wordt gevraagd artikelen aan te dragen over het eigen onderzoek, dat door de redacteur TST zal worden geredigeerd. Naast het schrijven en redigeren van artikelen, houdt de redacteur zich bezig met het samenstellen van themagestuurde dossiers. De dossiers vormen een aparte categorie op Kennislink, en vormen een introductie op een bepaald thema. De dossiers worden vooral gebruikt door scholieren voor thema- en profielwerkstukken.

2. Een verkennend onderzoek naar de aansluiting van TST op een brede doelgroep. De doelgroep van Kennislink bestaat uit een gevarieerd publiek, van scholieren tot beleidsmakers. Om de aansluiting van TST op deze markt te verkennen, schrijft de redacteur TST-nieuwsberichten over recente ontwikkelingen binnen het vakgebied. Binnen het jaar waarin de redacteur TST is aangesteld, zullen regelmatig evaluatiemomenten worden ingebouwd, waarin bepaald moet worden of de artikelen aanslaan bij de verschillende doelgroepen van Kennislink. Hieruit zal worden afgeleid op welke wijze Taal- en Spraaktechnologie het beste een blijvend onderdeel kan gaan uitmaken van de vakpagina Taalwetenschappen (mogelijk in combinatie met de vakpagina Techniek).

**Kennislink website:** <http://www.kennislink.nl/web/show>

## STEVIN IPR and Standards policy

IPR policy is an integral part of the STEVIN programme. One of the major aims of this programme is to make the basic digital language infrastructure for Dutch – and above all the results of this programme – available in a non-discriminatory way to all stakeholders. It must be considered a major challenge to formulate and implement an IPR-policy for all new language resources created with public funds acceptable to all parties involved. In the STEVIN programme the situation is more complicated as not only newly created language resources are involved but also those that have been implemented in the past either with or without national or European funding for which IPR has not been satisfactorily settled.

The basic principle of STEVIN IPR policy is that all basic resources – both new and existing ones – should be actively maintained by the *TST-Centre* (Dutch Language and Speech Technology archive) of the Dutch Language Union. This involves both making available the language resources and protecting their IPR. The *TST-Centre* has started to define the rules and regulations to be followed by the STEVIN programme. These rules and regulations will be based on experiences gained within the context of the Dutch-Flemish Corpus of Spoken Dutch (CGN) that was recently finished and for which close cooperation was established with ELRA and LDC. Both ELRA and LDC have developed IPR-standards which allow the development of resources on the basis of existing resources and are widely accepted by all parties involved, i.e. government, research institutes and industry.

To keep IPR within the STEVIN programme as simple and transparent as possible, before the project actually starts STEVIN project partners must contractually lay down in which way project results will be made available for all stakeholders. These contracts will be based on rules and regulations that are currently being developed and formulated. If a project builds on existing resources for which industrial IPR has been established, it must be contractually stated that the existing resources will be made available against reasonable conditions comparable to the way this has been arranged for pre-existing knowledge in IPR contracts for 6th Framework projects (cf. best practice guide ([www.cordis.lu/fp6/find-doc.htm#ipr](http://www.cordis.lu/fp6/find-doc.htm#ipr))).

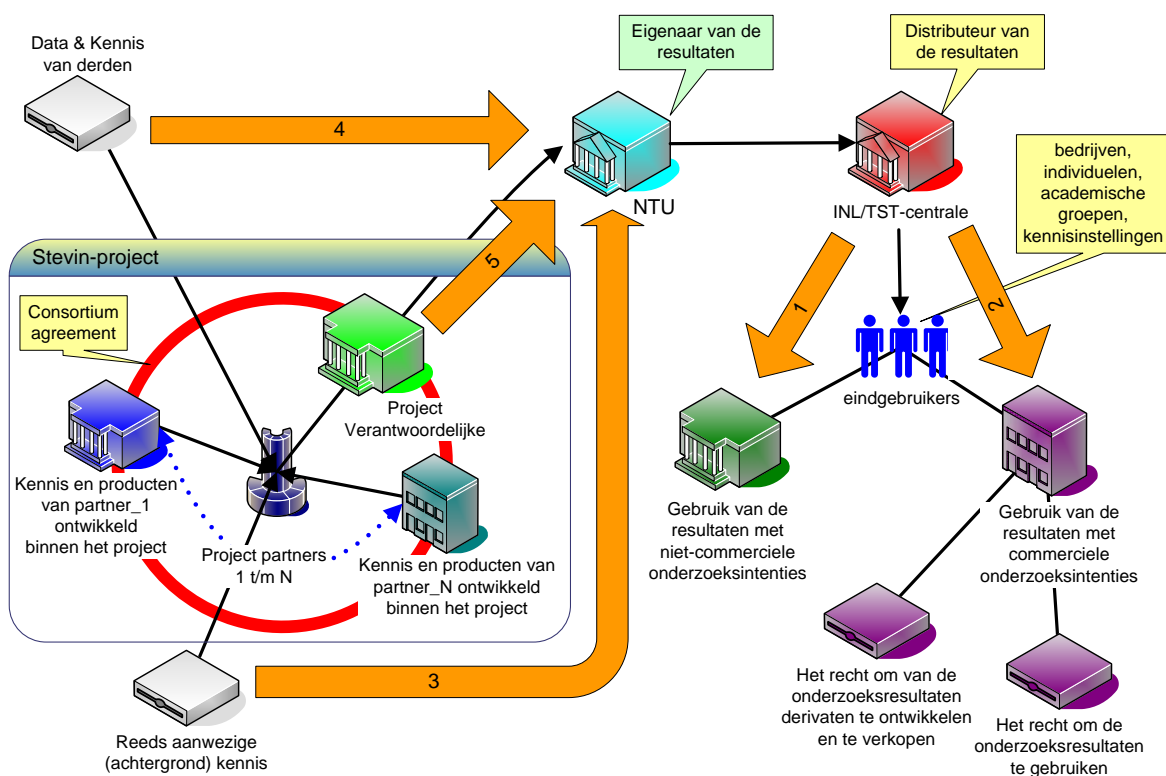
The reusability of some of the language resources developed in the past has been hindered by the use of idiosyncratic formats and data structures. Fortunately, the HLT community has been very active in developing and promoting standards. These were partly developed within European collaborative programmes such as EAGLES and ISLE. Other important institutes concerned with normalisation are ISO/T37, W3C and LISA. For the Dutch HLT-industry, with its relatively small market, it is especially important that international standards are realised and supported by the industry. The Programme Committee demands that projects apply existing standards and cooperate in developing new standards.

## IPR and use and re-use of STEVIN results

To enable the use and re-use of STEVIN results, a particular IPR-arrangement has been set up. The materials (software, data etc.) must be handed over to the Dutch Language Union and will be made available to third parties through the Dutch HLT Agency ('TST Centrale' [www.tst.inl.nl](http://www.tst.inl.nl)). The Dutch HLT Agency is responsible for IPR issues, and for the management, maintenance and distribution of HLTD-materials. In addition, the Dutch HLT Agency provides HLT-related information, advice and training to third parties.

### Schematic overview HLT actors (in Dutch)

The scheme below had been produced by the STEVIN IPR working group. This committee is led by the Dutch Language Union (NTU) and consists of academic and industrial HLT experts on IPR, legal experts and representatives from the Dutch HLT Agency (TST Centrale). They advise the STEVIN PC and HLT Board in order to co-ordinate and optimize STEVIN IPR practices.



In deze figuur is getracht informatiestromen (zwarte lijnen) en de overeenkomsten die moeten worden afgesloten (oranje pijlen) naast elkaar te leggen. Binnen een Stevin-project (gesymboliseerd door het zuiltje) werken academische (en eventueel niet-academische) partners samen. Er kan door alle partijen achtergrondkennis ingebracht worden die soms ook aan de NTU gelicentieerd moet worden in het geval de achtergrondkennis onderdeel uitmaakt van de projectresultaten. Daarnaast kan een Stevin-project data en kennis van "derden", niet betrokken partijen, gebruiken (bv krantenarchieven of archieven met audio-visueel materiaal), die eveneens aan de NTU gelicentieerd moet worden in het geval deze data en kennis onderdeel uitmaken van de projectresultaten. De NTU sluit daartoe licentieovereenkomsten af met de projectpartners (pijl 3) en met de derden (pijl 4). De rechten op de binnen het Stevin-project verworven kennis en data moeten aan de NTU overgedragen worden: dat gebeurt in pijl 5.

De TST-centrale geeft namens de NTU de resultaten van de Stevin-projecten in licentie aan eindgebruikers. De eindgebruikers kunnen de resultaten (kennis, data of derivaten) voor niet-commercieel onderzoek gebruiken (pijl 1) of er commerciële bedoelingen mee hebben (pijl 2). In dat laatste geval kunnen ze de resultaten van de projecten gebruiken voor eigen onderzoek en/of om er zelf derivaten mee te maken en/of om de al gemaakte derivaten van de projecten commercieel uit te nutten.

IPR flyer prepared by IPR working group to help convince data provider to make their data available for HLT R&D (in Dutch)

Nederlandse Taalunie Verzoek om data

**U wilt efficiënt in uw archieven kunnen zoeken...**  
Naar alle waarschijnlijkheid beschikt u, als content provider, over grote archieven met digitale tekst, beeld of geluid. U hebt zo uw eigen methoden om efficiënt te zoeken in die data. Grote kans dat u daarbij gebruik maakt van TST: Taal- en Spraaktechnologie. Zoekt u met behulp van trefwoorden? Gebruikt u tools om lange documenten samen te vatten? Of gebruikt u geavanceerde zoektechnologie om (gesproken) documenten terug te vinden? In al deze gevallen komt de onderliggende techniek voort uit de TST.

**... en TST helpt u daarbij.**  
Zoals dat bij iedere technologische ontwikkeling gaat, werkt ook in TST het mechanisme van voortschrijdend inzicht. Er wordt doorlopend onderzoek gedaan naar nieuwe en nog betere toepassingen. Ideeën zijn er genoeg en aan kennis ontbreekt het niet. Nederland en Vlaanderen lopen wat TST betreft voorop in Europa. Maar van één ding is er nooit genoeg: grote hoeveelheden spraak en tekst.

**TST heeft veel data nodig...**  
Voor het ontwikkelen van betere classificatietechnieken voor gesproken en geschreven teksten, betere spraakherkenners, betere automatische samenvattingen en betere zoekstrategieën is heel veel taalmateriaal nodig. De nieuwste generatie taal- en spraaksoftware leert van realistische voorbeelden: hoe schrijft een journalist voor de krant? Hoe spreekt een sportverslaggever bij Studio Sport of bij Sporza? Hoe meer voorbeeldmateriaal, hoe beter de resultaten. TST-resultaten drijven op statistische analyse en de grondstof daarvoor is veel, meer, en nog meer taal.  
Teksten en spraakopnamen zijn doorgaans goed beschermd door auteursrechten. TST-onderzoekers maken alleen gebruik van rechtenvrij taalmateriaal, en soms van collecties die de rechthebbenden voor onderzoeksdoelinden beschikbaar hebben gesteld.

**... helpt u TST daarbij?**  
U kunt TST een impuls geven door taalcollecties te doneren. De Nederlandse Taalunie heeft een bronnen centrum, de TST-Centrale, ingesteld, die deze collecties zal beheren. U krijgt de garantie dat het materiaal alléén gebruikt zal worden voor het (verder) ontwikkelen en testen van TST. Teksten en geluidsopnamen mogen na gebruik voor statistische analyses niet verder verspreid worden door TST-onderzoekers. En ... doordat u de mogelijkheid krijgt de ontwikkelingen vanaf het eerste moment te volgen kunt u straks nog sneller de weg vinden in uw digitale archieven.

**De Nederlandse Taalunie** is een internationale verdragsorganisatie waarin Nederland, België en Suriname samenwerken op het gebied van de Nederlandse taal en letteren en het onderwijs in en van het Nederlands.

**STEVIN** is een meerjarig onderzoeks- en stimuleringsprogramma met als doel het stimuleren van de taal- en spraaktechnologische sector in Vlaanderen en Nederland zodat de positie van het Nederlands in de informatie- en communicatiewereld wordt versterkt. STEVIN wordt gefinancierd en begeleid door de Nederlandse en Vlaamse overheden en uitgevoerd onder auspiciën van de Nederlandse Taalunie. Nederlandse partners zijn de Ministeries van Economische Zaken en Onderwijs, Cultuur en Wetenschappen, de Nederlandse Organisatie voor Wetenschappelijk Onderzoek en SententNovem; Vlaamse partners zijn het Departement Economie, Wetenschap en Innovatie, het Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen en het Fonds Wetenschappelijk Onderzoek Vlaanderen.

De **TST-centrale** (Centrale voor Taal en Spraaktechnologie) is verantwoordelijk voor het verwerven, beheren, onderhouden en verspreiden van diverse spraak- en taalmaterialen die voor het Nederlands taalgebied met overheidsmiddelen zijn ontwikkeld.

**Voor meer informatie over STEVIN:**  
I <http://www.taalunieversum.org/stevin>  
E [stevin@taalunie.org](mailto:stevin@taalunie.org)

**Voor meer informatie over de TST-centrale:**  
I <http://www.tst.in.nl>  
E [tst@in.nl](mailto:tst@in.nl)

## List of HLT activities organised or supported by the STEVIN programme

<i>Date</i>	<i>Location</i>	<i>HLT activity (* = organised by STEVIN)</i>
<b>2004</b>		
September 15	Tilburg	STEVIN Brokerage and kick off STEVIN programme (160 participants)*
<b>2005</b>		
March 2	Antwerpen	STEVIN Brokerage - HLT and the ICT market (163 participants)*
November 22	Eindhoven	Taal in Bedrijf (290 participants)*
December 16	Amsterdam	CLIN 2005, the 16th meeting of computational linguists in the Netherlands Symp. on Speech Technology for Clinical and Educational Applications (STCEA 2005)
<b>2006</b>		
March 13,14	Delft	DIR 2006, 6th Dutch-Belgian Information Retrieval Workshop (TNO)
June 20	Utrecht	NoTaS speed dating session
September 10	Antwerpen	STEVIN programme meeting*
December 15	Nijmegen	HLT in the care sector in St. Maartens kliniek
<b>2007</b>		
January 12	Leuven	CLIN 2006, the 17th meeting of computational linguists in the Netherlands
May 16	Amsterdam	Machine Learning for NLP 2007
June 11-22	Leuven	2007 LOT Summerschool
August 27-31	Antwerpen	Interspeech 2007
September 21	Hoeven	STEVIN programme meeting*
November 23	Antwerpen	TST Archive meeting: de gebruiker centraal

## List of publications about the STEVIN programme

- English [folder](#) about the Dutch Language Union and HLT for Dutch
- a bilingual [Nederlands-Engelse folder](#) about the STEVIN-programme

### Publications in Dutch

- "[Computer plaatst vragen in de juiste context](#)", SenterNovem Innovatiekrant SenterNovem 26 april 2006, p.8
- [SenterNovem Monitor 2006 \(4\): 7-9](#)
- "[De computer begint steeds meer mee te praten](#)" door dr. Peter-Arno Coppen in Taalschrift 20/10/06
- Het Dixiteindejaarsnummer 2006 bevat een uitgebreide thematische STEVIN-sectie. DIXIT wordt uitgegeven door en is te verkrijgen via de [Stichting NoTaS](#).
- "[Innovatief spraakherkenningssysteem als back-up voor justitie](#)", SenterNovem Innovatiekrant 24 april 2007, p.12
- "Klinkende Taal voor Ambtenaren", SenterNovem Innovatiekrant 4 december 2007, p.17
- "Digi-revolutie in de rechtbank" in [De Twentsche Courant Tubantia 11 mei 2007 \(pdf-bestand\)](#)
- Bea Ross, "[Computer luistert beleefd en geeft netjes antwoord](#)", NWO Hypothese 2007, p. 18-20