



Dutch-Flemish Research Programme for Dutch Language and Speech Technology STEVIN (*Spraak- en Taaltechnologische Essentiële Voorzieningen In het Nederlands*)

2nd Call for Proposals:

- deadline for submitting mandatory pre-proposals for Open Call:
Tuesday April 26, 2005, 12:00 h.
- deadline for submitting full proposals for Call for Tender specific priorities:
Monday May 9, 2005, 12:00 h.

General introduction

In order to stimulate the technology sector so as to increase its innovative power, and in order to improve or at least secure the position of the Dutch language in the modern information and communication society, the *Nederlandse Taalunie* has launched the 5 year Dutch-Flemish Language and Speech Technology Programme *STEVIN*.

The *STEVIN* programme aims can be briefly summarized as follows:

- realise an appropriate digital language infrastructure for Dutch, based on the BaTaVo-priorities (see <http://taalunieversum.org/taal/technologie/docs/daelemans-strik.pdf>);
- carry out strategic research in the domains of language and speech technology, in particular in areas for which there is a large demand from specific applications and technologies;
- stimulate the creation of networks and consolidation of language and speech technology activities, educate new experts, promote demand stimulation and knowledge transfer.

More information about the objectives of the programme can be found on the *STEVIN* website <http://taalunieversum.org/stevin/>. You will also find there the electronic versions of *STEVIN*'s programme framework, the guidelines for submitting proposals and the official *STEVIN*-*subsidieregeling* of the *Nederlandse Taalunie* (the legal Dutch-Flemish basis for this Call).

Three calls have been planned within this programme:

1. a 1st open call (September 2004) - maximum budget 2 M€ - for focussed short term projects with a self-contained result (maximum duration is 2 years);
2. a 2nd open call (in the spring of 2005) for more complex projects with a longer time frame (maximum duration four years) combined with a call for tender for projects targeting some specific priorities (maximum duration 2 years) - maximum budget 4.2 M€ ; and
3. a 3rd call (in 2007) - maximum budget 2.3 M€ - the specific content of which will be determined later.

The total budget of the 2nd Call is limited to a maximum of € 4.2 million. This call consists of two parts: an **open call** and a **call for tender** targeting two specific priorities (max. € 0.8 million). Later this year the *STEVIN supporting programme* will be opened. More information can be found on p. 6 and will be available on the *STEVIN* website in the spring of 2005.

STEVIN Open Call

Practical details as to a) who can apply, b) what type and length of projects can be submitted, and c) details about the assessment procedure, etc. can be found on page 6 and 7. For the open call for proposals the same priorities apply that were presented in the official *STEVIN* programme description. These are repeated below and are considered to be equally important. Obviously, proposals targeting language and speech technology priorities that are or will be realised already with other funding are not eligible for funding in this call. It is recommended that submitters of proposals familiarize themselves with the projects approved under the first call in order to avoid overlap. Information about these projects will be available on the *STEVIN* website by the end of March.

In order to minimize the time needed for writing and evaluating the proposals in the open STEVIN call, it is compulsory to submit a preliminary proposal (pre-proposal). The deadline for the mandatory pre-proposal is Tuesday April 26, 2005.

STEVIN Priorities open call

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. More details as to the exact nature of the STEVIN priorities can be found in the official programme description in chapter 5.3.

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, inc. 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.

Assessment criteria open call

In the evaluation, ranking and selection of proposals submitted in the 2nd Call for Proposals, the following criteria will be taken into consideration. All criteria are equal in weight.

- **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 relevancy of the project for applications that are relevant for 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.

STEVIN Call for Tender

Two specific priorities have been selected for a call for tender within which full proposals for projects with a max. duration of 2 years can be submitted immediately.

These selected priorities are:

1. *A speech recognition toolkit for Dutch*
2. *A lexical resource for the semantic processing of Dutch*

For the Call for Tender a maximum budget of € 0.8 million is available.

Practical details as to a) who can apply, b) what type and length of projects can be submitted, and c) details about the assessment procedure, etc. can be found on page 6 and 7.

The deadline for submitting full proposals for call for tender specific priorities is **Monday May 9, 2005, 12:00 h.**

1. Development of a speech recognition toolkit for Dutch that meets the following requirements

One of the main priorities of the STEVIN program was to make available a speech recognition toolkit for Dutch. The toolkit must consist of software modules as well as example speech recognizers (see below) developed with this software. Both the software modules and the example recognizers developed with these modules must be properly documented. The documentation on the example recognizers can take the form of tutorials describing how the recognizers were built with the delivered modules (what was the training corpus, what scripts have been used, where did the lexicon and language model come from).

The software must implement proven, state-of-the-art speech recognition technology and run at least under Windows and Linux on a high-end PC. The recognizers built with the toolkit modules should offer a good compromise between training and recognition time on the one hand and recognition accuracy on the other hand. Preferably it must be possible to trade recognition accuracy for processing time.

All software that will be developed under the contract resulting from this tender must be made available through the *TST-Centrale* in the form of source code. All software modules comprised in the toolkit must be made available in the form of libraries for Windows and Linux, and they must be designed with the aim to maximally support speech and language technology experts in their research towards more robust recognition systems and LST applications for Northern Dutch and standard Flemish. This also implies that researchers must have the opportunity to acquire the source code of the modules, under fair conditions, if they need that code for their research.

In addition to the toolkit two demonstration speech recognizers must be provided for each of the two language communities (the Netherlands and Flanders):

- A speaker independent large vocabulary continuous speech recognizer for the recognition of clean read speech. This recognizer must be provided in such a way that its acoustical, lexical and linguistic models can be substituted by alternative models created according to the conventions (formats) specified in the documentation of the tools.
- A speaker independent recognizer with user definable BNF-grammars for the recognition of telephone speech.

The proposal should demonstrate that the consortium will prove the adequacy of the software and the documentation by leaving the training of some of the demonstration recognizers to parties inside or outside the consortium whose main responsibility is not in the development and documentation of the training software, but rather in its testing and evaluation.

The IPR arrangements must comply with the accepted IPR policy of the STEVIN programme. In particular, they must define the conditions under which the software modules and the acoustical, lexical and linguistic models embedded in the example speech recognizers become available for research by future users.

The IPR arrangements should make it possible for application developers to obtain licenses to use the software modules, and preferably also some of the models supplied with the example recognizers, to build operational recognizers that can be integrated in commercial applications.

2. A lexical resource for the semantic processing of Dutch

- The lexicon should provide the information that is needed to perform tasks which require a non-trivial level of understanding, such as word sense disambiguation, content-based information retrieval, machine-aided translation and inferencing.
- The lexicon should contain semantic information on lemmata and their part-of-speech. The nature and form of the semantic information is deliberately not restricted but should be defined in relation to the purpose of exploiting this information for the above tasks.
- The lexicon should at least cover the general and more frequent words of the Dutch language and it should be large enough to evaluate the quality of the resource in NLP applications. The semantic information in the target lexicon should be domain-independent, but at the same time it should be extensible with domain-specific information. To ensure this, the consortium should select a domain and add the semantic and ontological information which is typical for that domain.
- Preferably, the semantic information should be related to other types of information about the lexical items, such as morpho-syntactic features, valence patterns, register and word-frequencies.
- Preferably, it should be connected to other semantic resources available in the NLP community, i.e. to create a multilingual resource that relates the Dutch language to other languages or to map the Dutch language to ontologies within the Semantic web community.
- The lexicon should be structured in a way which facilitates maintenance, extensibility and embedding in NLP systems.
- The resulting lexical resource should conform to international practices and standards.
- The lexicon should be publicly available in conformity with the IPR policy of the STEVIN programme.

STEVIN Supporting Programme

In order to stimulate the demand for language and speech technology, and also to ensure that the players at the technological development and application layers are represented in STEVIN, a special funding opportunity will be created in the context of the 'flankerende activiteiten' of the STEVIN supporting programme. This is intended for projects that focus on developing a specific application (rather than on research or the creation of resources) that will be actually deployed (possibly in a pilot/test phase). The resulting application should be appealing and (made) highly visible to the general public and/or decision makers in companies. Proposals for such projects should contain, apart from a concrete development plan, a promotion plan that ensures wide visibility of the application being developed and attracts attention to it and the language/speech technology used in it, and a plan for the evaluation of the actual deployment of the application. In order to avoid distortion of the market, the applications should be deployed for services by non-commercial (e.g. governmental or semi-governmental) organizations. The other general evaluation criteria also apply to proposals for such projects.

More information about the STEVIN supporting programme will be available in the spring of 2005.

STEVIN Timetable

Activity	Date
2nd Call for Proposals opening (brokerage March 2, 2005)	March 2005
Deadline mandatory Pre-proposals open call	April 26, 2005
Deadline Full proposals call for tender specific priorities	May 9, 2005
Selection pre-proposals open call and Assessment call for tender by PC	May 23, 2005
Decision on Funding by Programme Board call for tender	Summer 2005
Deadline Full-proposals open call (only if a pre-proposal was submitted)	September 1, 2005
Assessment and ranking by International Advisory Panel Open Call	October 2005
Reply applicants on Assessments IAP	October 2005
Assessment and ranking by Programme Committee Open Call	November 2005
Decision on Funding by Programme Board Open Call	November 2005
3rd Call for Proposals	2007

Practical details on the submission and evaluation procedure regarding the 2nd Call for Proposals of the STEVIN programme

Call for Proposals

The 2nd Call for Proposals for the STEVIN Programme is open from March 2, 2005. The total budget of the 2nd Call is limited to a maximum of 4.2 million Euro. This 2nd STEVIN Call consists of two parts: an open call and a call for tender targeting two specific priorities.

For the open part of this 2nd call first preliminary proposals (pre-proposals of approximately 3 A4) must be submitted. Full proposals for the open call can only be submitted by researchers who have submitted a pre-proposal. Pre-proposals must be submitted electronically in the prescribed format via <http://taalunieversum.org/stevin/> where a link can be found to the electronic proposal submission system of NWO.

For two very specific priorities a separate call for tender will be opened, the details of the tender including instructions as to who can and how to apply will be published separately on the STEVIN website. Only those proposals that specifically target these priorities will be eligible.

A 3rd Call for the STEVIN programme is envisaged for 2007.

Pre-proposals for the open call must be submitted to the NWO electronic proposal submission system. The deadline for submitting mandatory pre-proposals in the 2nd STEVIN Call is April 26, 2005, 12:00 h. Note again that full proposals cannot be submitted at this time in this open call. These should be submitted before September 1st, only by those applicants that have submitted a pre-proposal in time.

Full proposals targeting the specific priorities described in the call for tender must also be submitted to the NWO electronic proposal submission system Iris. The deadline for submitting full proposals in the call for tender for specific priorities is set for May 9, 2005, 12:00 h.

The NWO electronic proposal submission system IRIS can be accessed as of March 31 via <http://taalunieversum.org/STEVIN>.

Who can apply?

Applications can be submitted only by senior researchers of Dutch or Flemish knowledge institutes. If a research proposal is submitted by a consortium of knowledge institutes and/or companies, the main applicant and coordinator of the project must be employed by a Dutch or Flemish knowledge institute. Eligible knowledge institutes are defined in the official *STEVIN-subsidieregeling*. In principle it concerns those institutes recognized as such a) in "de Nederlandse Wet op het hoger onderwijs en wetenschappelijk onderzoek" or b) in "het Decreet van 4 april 2003 betreffende de herstructurering van het hoger onderwijs in Vlaanderen" or c) by the Board of the STEVIN programme.

A brokerage event bringing together researchers, developers and existing or potential customers has been organized on March 2, 2005 in Antwerpen to exchange existing and develop new ideas for proposals.

What can be applied for in the open call?

Proposals should specify how the proposed research or project fits in with the priorities of the STEVIN programme specified above and described more extensively in the official STEVIN programme description. Proposals can consist of self-contained focused projects of 4 years maximum. For project proposals whose budget exceeds 500 k€, matching funds from industrial partners are required.

The size of the matching funds depends on the size of the total project budget: i.e. the requested STEVIN funding plus funding provided by others either in actual funds or in kind. To determine the value in kind the following guidelines apply:

- * personal support: salary cost plus social security
- * cost price of material support
- * cost price minus depreciation of equipment

Matching must be confirmed in writing. For each Euro exceeding 500 k€, 50% must be made available from matching funds provided by industrial partners.

Format open call pre-proposals and full proposals open call

Pre-proposals must be written in English and be submitted electronically to NWO via the Iris-system (accessible via <http://taalunieversum.org/stevin/>) and not exceed three pages A4 (max. 2500 words). Full proposals (to be submitted in September 2005) must also be written in English and may not exceed 15 pages A4 (max. 8000 words). Before submitting a proposal, first read the "Richtlijnen indienen vooraanmeldingen tweede STEVIN Oproep" which can be downloaded from <http://taalunieversum.org/stevin/> and contains specific directions as to the format in which pre-proposals and full proposals must be submitted.

Format full proposals specific priorities submitted in the call for tender

Full proposals must be written in English and be submitted electronically to NWO via the Iris-system (accessible via <http://taalunieversum.org/stevin/>). Before submitting a proposal, first read the "Richtlijnen indienen call for tender tweede STEVIN Oproep" which can be downloaded from <http://taalunieversum.org/stevin/> which contains specific directions as to the format in which proposals for the specific priorities set for the call for tender must be submitted.

Eligible costs

- Personnel costs (100%) directly related to the project to which 25% will be added for overhead costs.
- A bench fee of 11.000 € per FTE for computers, material costs and conference visits. This amount will be raised by 50% if the project consortium contains at least one Dutch and one Flemish partner.
- Costs related to realising databases.

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

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

All eligible full proposals submitted in the open call will be presented to a panel of international experts in language and speech technology. For this call the same experts will be 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 will be asked to serve in this panel. The composition of the International Assessment Panel can be found on the STEVIN website. This international panel will evaluate and rank the eligible applications based on the assessment criteria mentioned earlier and will formulate a set of recommendations for the STEVIN's Programme Committee.

The panel's assessment will be 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 will assess the applications and determine the order of priority of the eligible proposals. On the basis of the Programme Committee's advice, the Board of the STEVIN programme will finally determine which projects will be funded.

Evaluation procedure full proposals specific priorities submitted in the call for tender

The assessment and ranking of full proposals targeting the specific priorities will be carried out by the Programme Committee. The proposals and the PC assessment will subsequently be 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 will finally determine which projects will be funded.

STEVIN subsidieregeling

More information as to which legal rules apply for this specific Dutch-Flemish granting scheme are laid down in the *Subsidieregeling van de Nederlandse Taalunie tot subsidieverstrekking in het kader van Nederlandse taal- en spraaktechnologie "STEVIN" (STEVIN-subsidieregeling)*, which can be found on the STEVIN website.

STEVIN Organisation

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

- o de Nederlandse Taalunie;
- o de administratie Wetenschap en Innovatie (AWI) van het Ministerie van de Vlaamse Gemeenschap (MVG);
- 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).

Programme Board

The Board of the STEVIN programme will be responsible for the final funding decision. 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*.

In the Board representatives of all funding parties will be represented which may be complemented with some senior experts from the field.

Programme Committee

- o prof. dr. J. Odijk (Scansoft/Universiteit Utrecht) - chair;
- o prof. dr. J.P. Martens (Universiteit Gent);
- o prof. dr. F. Van Eynde (Katholieke Universiteit Leuven);
- o prof. dr. W. Daelemans (Universiteit Antwerpen/Universiteit van Tilburg);
- o mrs. D. Kenyon-Jackson (Polderland Language & Speech Technology BV);
- o dr. P. Vossen (Irion Technologies);
- o dr. A. van Hessen (Telecats BV/Universiteit Twente);
- o prof. dr. L. Boves (Katholieke Universiteit Nijmegen);
- o dr. W. Ceusters, Universität des Saarlandes, European Centre for Ontological Research (ECOR);
- o mrs. dr. J. Beeken (INL/TST-centrale);
- o dr. K. Van Bruwaene (VRT);
- o vacancy.

Programme Bureau

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

For further information, please contact the Programme Bureau or the *Nederlandse Taalunie*:

<i>NWO-Geesteswetenschappen</i>	<i>SenterNovem</i>	<i>Nederlandse Taalunie</i>
Alice Dijkstra	Hans Kruithof	Erwin Dewallef
dijkstra@nwo.nl	jak@senternovem.nl	edewallef@taalunie.org
+31 (0)70 34 40 736	+31 (0)70 37 35 363	+ 31 (0)70 34 69 548
<i>Visiting address</i>	<i>Visiting address</i>	<i>Visiting address</i>
Laan van Nieuw Oost Indië 300, Den Haag	Juliana van Stolberglaan 3, Den Haag	Lange Voorhout 19, Den Haag
<i>Postal address</i>	<i>Postal address</i>	<i>Postal address</i>
Postbus 93425	Postbus 93144	Postbus 10595
NL 2509 AK Den Haag	NL 2509 AC Den Haag	NL 2501 HN Den Haag

STEVIN website: <http://taalunieversum.org/stevin/>.