

Surviving the Information Society

The development of an adequate language and speech technology infrastructure for Dutch.

by Catia Cucchiarini



The term "language resources" was originally used to refer to sets of written and spoken language data that constitute the basis for developing all sorts of language processing systems. More recently, the use of this term has been extended to include basic software tools that, together with language data, form the digital language infrastructure, necessary for conducting research and developing applications in the field of *Human Language Technologies (HLT)*. The availability of such a digital language and speech infrastructure is essential for a language to be able to survive in the information society, i.e. a society in which information and communication technologies (ICT) play a vital role. Many ICT applications rely on HLT for allowing users to address computers in natural language. Preferably, this language should be the user's mother tongue, since this is the only way to guarantee that all citizens can fully participate in the information society.

The last few years have witnessed a growing awareness of the importance of

Owing to the fragmentation of responsibilities, it was difficult in the past to conduct a coherent HLT policy.

such a digital language infrastructure, not only in the United States and in Asia, but also in Europe. This is evident from the various initiatives that have been taken at European level, such as the creation of the European Language Resources Association (ELRA), the organization of the Language Resources and Evaluation (LREC) conferences, and the various projects funded by the European Commission, e.g. SPEECHDAT, PAROLE, SIMPLE, CLASS, EAGLES, HOPE, ISLE, to name but a few. Moreover, several projects have recently been launched by the National Authorities (Ministries or their Departments) in various European countries with the specific aim of strengthening the digital language infrastructure. Projects of this kind require that a dialogue be established between the parties involved: industry, academia, and policy institutions. To establish such a dialogue is not always easy, often because the various parties have conflicting interests. Discrepancies may exist not only between industry and universities, but also between the various research

groups within industry and academia. From the contacts we have had with our European colleagues, it appears that it is just these kinds of problems that have hampered the emergence and the organization of other countries' national projects aimed at providing or improving HLT resources for their respective languages.

In this paper we report on one such initiative that was taken for the *Dutch language by the Dutch Language Union* (Nederlandse Taalunie, abbreviated NTU): the Dutch Human Language Technologies Platform. We hope that the experiences we have had in the last two years in setting up these activities may be useful to others who are now beginning with this kind of work.

The Dutch Language Union (NTU) and Human Language Technologies

The Dutch HLT Platform cannot actually be characterized as a national initiative, but rather as a supranational one, because it goes beyond national borders and concerns the Netherlands and the Flemish part of Belgium. The plan to set up a Dutch HLT platform was launched by the NTU. The NTU is an intergovernmental organization established in 1980 on the basis of the Language Union Treaty between Belgium and the Netherlands, which has the mission of dealing with all issues related to strengthening the position of the Dutch language (see www.taalunie.org). The NTU has a broad action field which encompasses teaching Dutch at home and abroad, research on the Dutch language and literature, the position of the Dutch language as a means of communication in various contexts and, more in general, the Dutch language as a tool for social interaction. The NTU enables Flanders and The Netherlands to speak with a single voice in the international arena. The Committee of Ministers,

The governments of the Netherlands and Flanders appreciate the growing importance of HLT as a specific part of information technology.

composed of the Flemish and Dutch ministers of Education and Culture, is responsible for the policy of the NTU. In

establishing its current long-term policy plan (1998–2002), the NTU has given full consideration to the rapid developments in the field of ICT that are going to have a major impact on language issues. The governments of the Netherlands and Flanders appreciate the growing importance of HLT as a specific part of information technology. Keeping up with the technological developments in this field implies major investments and the commitment of those involved, notably the knowledge infrastructure, the technology and applications developers and the policy makers at the national and European level. Co-operation among all these actors is of utmost importance, and given the size of the Dutch language area, this co-operation needs to be expanded to a cross-border Flemish-Dutch level.

The Dutch Human Language Technologies Platform

The main purpose of the Dutch HLT Platform is to contribute to the further development of an adequate language and speech technology infrastructure for Dutch. More specifically, the HLT Platform has the following objectives:

- To strengthen the position of the Dutch language in HLT developments, so that the speakers of Dutch can fully participate in the information society;
- To establish the proper conditions for a successful management and maintenance of basic HLT resources developed through governmental funding;
- To stimulate co-operation between academia and industry in the field of HLT;
- To contribute to the realization of European co-operation in HLT-relevant areas;
- To establish a network that brings together demand and supply of knowledge, products and services.

In addition to the NTU, the following Flemish and Dutch partners are involved in the HLT Platform:

- The Ministry of the Flemish Community;
- The Flemish Institute for the Promotion of Scientific-technological Research in Industry;
- The Fund for Scientific Research, Flanders;

- The Dutch Ministry of Education, Culture and Sciences;
- The Dutch Ministry of Economic Affairs;
- The Netherlands Organization for Scientific Research (NWO);
- Senter (an agency of the Dutch Ministry of Economic Affairs).

All these organizations have their own aims and responsibilities and approach HLT accordingly. Together they provide a good coverage of the various perspectives from which HLT policy can be approached.

The rationale behind the Dutch HLT platform was not to create a new structure, but rather to co-ordinate the activities of existing structures so as to favor the transfer of technology among the parties involved (see our technology transfer model in

The NTU enables Flanders and The Netherlands to speak with a single voice in the international arena.

Figure 1). The platform is a flexible framework within which the various partners adjust their respective HLT agendas to

each other's and decide whether to place new subjects on a common agenda. Initially, the Dutch HLT platform was set up for a period of five years (1999-2004).

Even if the Netherlands and Flanders cooperate in funding the development of basic language resources, the investments

The availability of such a digital language and speech infrastructure is essential for a language to be able to survive in the information society.

for the different partners involved remain substantial. This absolutely requires that efforts be cumulative and not duplicated, that insight be provided into the resources that are needed for a language in general and for Dutch in particular and that a plan be drawn up for the development of the resources that are totally lacking or insufficiently available for Dutch. Furthermore, attention should be paid to such matters as evaluation of resources and project results, standardization, maintenance, distribution etc. In other words, it is necessary to create the preconditions to maximize the out-

come of efforts in the field of HLT. To this end, an *Action plan for Dutch in language and speech technology* has been defined, which is funded jointly by the different partners in the HLT platform. The activities described in this action plan are organized in four action lines as illustrated in the text box.

Soon after the HLT Platform was set up it was decided that survey (action line B) and evaluation (action line C) be carried out simultaneously because the actual availability of a product is not determined merely by its existence, but depends heavily on the quality of the product itself.

Concluding Remarks

After only two years of activity the HLT Platform has already produced important results. Various seminars and workshops have been organized within action line A, which were attended by people from industry, academia and policy institutions. The survey and evaluation in action lines B and C have now been completed and the results were discussed in public during a supranational seminar held in The Hague on 15 November. The blueprint to be defined in action line D will be ready by the end of 2001 (for further information, visit www.taalumieversum.org/tst).

It should be noted that in the two years that the Dutch HLT Platform has been active much effort has been spent in set-

Action plan for Dutch in language and speech technology

Action line A: construction of 'broking and linking' function

The main goals of this action line are to encourage co-operation between the parties involved (industry, academia and policy institutions), to raise awareness and give publicity to the results of HLT research so as to stimulate market take-up of these results.

Action line B: plan to strengthen digital language infrastructure

The aims of action line B are to define what the so-called BLARK (Basic Language Resources Kit) for Dutch should contain and to carry out a survey to determine what is needed to complete this BLARK and what costs are associated with the development of the material needed. These efforts should result in a priority list with cost estimates which can serve as a policy guideline.

Action line C: working out standards and evaluation criteria

This action line is aimed at drawing up a set of standards and criteria for the evaluation of the basic materials contained in the BLARK and for the assessment of project results.

Action line D: management, maintenance and distribution plan

The purpose of this action line is to define a blueprint for management (including intellectual property rights), maintenance, and distribution of HLT resources.

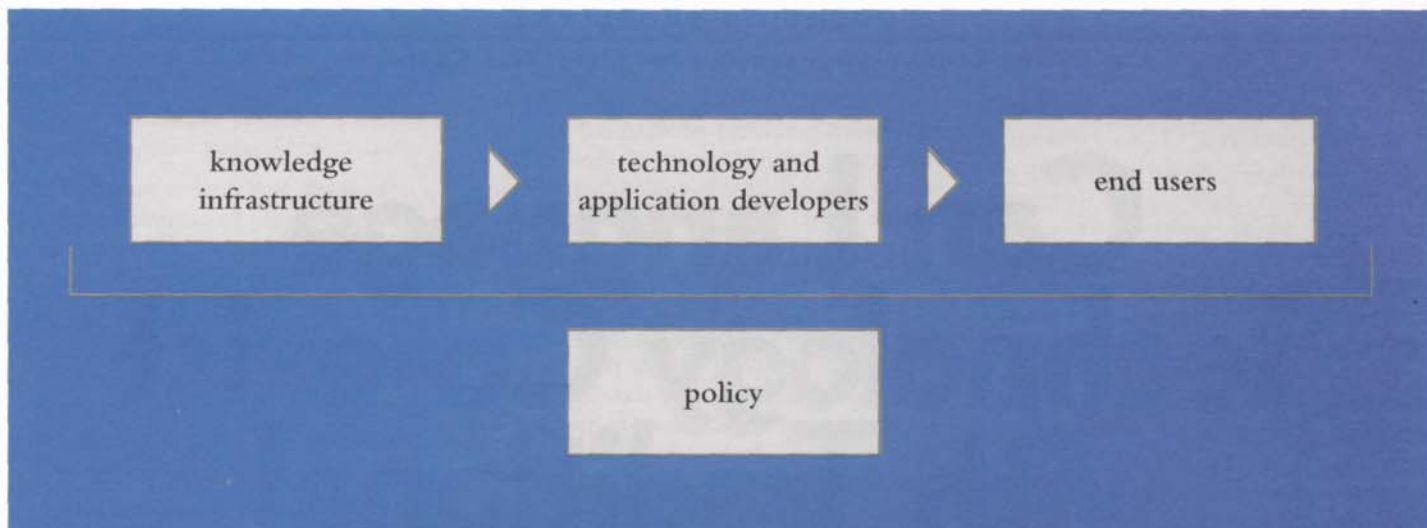


Figure 1 Technology transfer model



ting up the whole platform structure, i.e. in finding the representatives of the appropriate responsible bodies and expertise centers. Owing to the fragmentation of responsibilities, it was difficult in the past to conduct a coherent HLT policy. We hope that the HLT platform will contribute to creating more transparency in this respect. Up to now our experiences have been positive across the board and we can only hope that this trend will continue, since there is still much work to be done.

After graduating in translation in Triest (Italy), Catia Cucchiariini obtained a PhD in Phonetic Sciences from KUN (The Netherlands). She has done research on phonetics, automatic speech recognition (ASR) and Computer Assisted Language Learning (CALL). At present she is employed as a researcher in ASR and CALL at KUN and as a project manager for Speech and Language Technology at NTU.



Catia Cucchiariini