The new Dutch-Flemish HLT Programme:

a concerted effort to stimulate the HLT sector.

Catia Cucchiarini and Elisabeth D'Halleweyn

Nederlandse Taalunie
(Dutch Language Union)
Lange Voorhout 19
2514 EB Den Haag
The Netherlands
ccucchiarini@taalunie.org; edhalleweyn@taalunie.org

Abstract

In 2004 a new, comprehensive Dutch-Flemish HLT programme will be launched by a number of ministries and organizations in the Netherlands and Flanders. To guarantee its Dutch-Flemish character, this large-scale programme will be carried out under the auspices of the Dutch Language Union (NTU). The aim of this new initiative, which is a continuation of the previous HLT Platform project, is to contribute to the further progress of HLT for the Dutch language. In trying to achieve this goal the project partners will make a concerted effort aimed at raising awareness of HLT results, , promoting innovation oriented strategic research in HLT, stimulating the demand of HLT products, developing HLT resources that are essential and are known to be missing, organising the management, maintenance and distribution of HLT resources.

1. Introduction

In 2004 a new, comprehensive Dutch-Flemish HLT programme will be launched by a number of partners in the Netherlands and Flanders: the Dutch Ministry of Economic Affairs, SENTER (an agency of the Dutch Ministry of Economic Affairs), the Netherlands Organisation for Scientific Research (NWO), the Dutch Ministry of Education, Culture and Science, the Flemish Institute for the Promotion of Innovation by Science and Technology (IWT), the Administration for Science and Innovation (AWI) of the Ministry of the Flemish Community, and the Flemish Fund for Scientific Research (FWO). To guarantee its Dutch-Flemish character, this programme will be carried out under the auspices of the Dutch Language Union (Nederlandse Taalunie, abbreviated NTU). The mission of the NTU, which was established by treaty between Flanders and The Netherlands in 1980, is to deal with all issues concerning the position of the Dutch language.

The aim of this large-scale programme is to contribute to the further progress of HLT for the Dutch language. In trying to achieve this goal the project partners will make a concerted effort based on three pillars: 1) raising awareness of HLT results, 2) promoting innovation oriented strategic research in HLT, developing HLT resources that are essential and are known to be missing, stimulating the demand of HLT products, 3) organising the management, maintenance and distribution of HLT resources once they are developed. In this paper we will explain how these different activities are combined in one comprehensive programme. Furthermore, we will also explain how this programme is related to previous actions in the field of HLT that were carried out in the Dutch language area (the Netherlands and Flanders) in the past five years. The overview of the activities which were undertaken may be useful for other countries that are now starting similar initiatives. It shows how co-operation can successfully be realized between different parties (language and speech technology, Flanders and the Netherlands, academia, industry and policy institutions) so as to achieve one common goal: progress in HLT.

This paper is organized as follows. In section 2 we describe the background of this new programme and explain how it is related to the previous HLT Platform project which was carried out in the Dutch language area in the past five years. In section 3 we present the new programme in more detail. Finally, in section 4 we briefly describe the current state of affairs and give some information on future activities.

.

2. Background

The new Dutch-Flemish Human Language Technologies (HLT) programme is a direct result of the Dutch-Flemish HLT Platform project, which we reported on at previous LREC conferences (2000 and 2002). The HLT Platform project envisaged four action lines:

A: Creating a 'market place', 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.

B & C: Developing a plan for the strengthening of the digital language infrastructure by defining a general BLARK (Basic LAnguage Resources Kit), identifying to what extent the BLARK exists for Dutch and which elements are missing and furthermore establishing standards and evaluation criteria for HLT resources.

D: Developing a blueprint for the management, maintenance and distribution of language resources developed with government money.

In 2003 all these action lines culminated in concrete achievements. Action line A contributed to creating transparency in the HLT field in The Netherlands and Flanders and managed to improve communication between interested partners. A co-operative framework is now available that provides a forum for discussing, exchanging and sharing experiences, best practices, information, data and tools. This success is partly due to the role of the NTU as the National Focal Point (NFP) in the Euromap project, a European Commission supported initiative dedicated to promoting greater awareness of HLT in the Information Society.

Action lines B and C led to the definition of a BLARK (Basic LAnguage Resources Kit) for language technology and a BLARK for speech technology. By analyzing the availability of the various resources, priority was assigned to the development of those parts of the BLARKs that are known to be crucial and appeared to be missing. One list of priorities for speech technology and one for language technology were drawn up and were subsequently submitted to representatives from the whole HLT field (about 2,000 people). Definitive priority lists were then produced (see Daelemans and Strik, 2002, Binnenpoorte et. al 2002) and submitted to various policy institutions such as the Dutch Ministry of Economic Affairs, the Netherlands Organisation for Scientific Research (NWO), the Dutch Ministry of Education, Culture and Science, the Flemish Institute for the Promotion of Innovation by Science and Technology (IWT), the Administration for Science and Innovation (AWI) of the Ministry of the Flemish Community, and the Flemish Fund for Scientific Research (FWO). These institutions acknowledged the importance of developing the resources mentioned in the priority lists. The Dutch Ministry of Economic Affairs decided to carry out an additional study aimed at determining whether some other form of economic support, in combination with the BLARK proposal, would stimulate the HLT sector even more. This study was conducted in 2003 and its results have indeed provided interesting insights into how effective action in the HLT sector should be shaped. This will be explained in more detail in section 3.2.

In line with the structure of the HLT Platform project, the new HLT Programme is organised around three action lines: 1) Awareness, 2) Strategic research and resource development and 3) Management, maintenance and distribution of HLT resources. These three lines of activities are presented in sections 3.1, 3.2, and 3.3, respectively.

3. The new Dutch-Flemish HLT Programme

3.1. Awareness

The first pillar of the HLT Programme focuses on co-operation, information, dissemination and creating visibility. The HLT field in the Netherlands and Flanders has seen a growing tendency to cooperation. A couple of years ago surveys noted a lack of mutual visibility. A number of initiatives however, caused considerable changes. In academia, the Dutch-Flemish Spoken Dutch Corpus project brought together researchers from Flanders as well as from the Netherlands, from language technology as well as form speech technology. The HLT-Platform brought together responsible Dutch and Flemish government bodies enabling them to formulate a common agenda and to commonly launch new initiatives. A large part of the SMEs and some knowledge institutes active in the HLT field in the Netherlands, have set up the NOTaS foundation (Nederlandse Organisatie voor Taal- en Spraaktechnologie i.e. Dutch Organisation for Language and Speech Technology) in order to jointly look after their interests. Their philosophy is that although they are competitors, they are nevertheless often each other's technology or data suppliers and together they stand stronger in a new, still establishing market.

Part of the HLT Programme will concentrate on maintaining and intensifying the already existing cooperation through instruments that have proven to be successful such as the HLT website (www.taalunieversum.org/tst), newsletters, seminars and conferences.

The challenge however will be to narrow the gap between technology and the market. Only if the possible end user – in most cases through the channel suppliers – can be addressed and stimulated can the prophecies of the possibilities HLT has to offer be fulfilled and can the predicted economic potential be exploited. The government can play a crucial role in acting as an early adopter and a launching customer. An example would be to encourage public translation services to intensify the use of machine translation. To this end the Dutch Language Union actively promotes the incorporation of the results of the NL-Translex project in the systems of designated parties. This MLIS project resulted in machine translation components that handle unrestricted text and translate Dutch from and into English and French.
In 2005 awareness activities will a.o. focus on how HLT can enlarge the facilities of disabled people.

3.2. Strategic research and resource development

The survey carried out within the framework of the HLT Platform project revealed the need to develop a number of HLT resources that are known to be essential and that turned out to be missing for the Dutch language. On the basis of these findings, priorities for resource development were established (Binnenpoorte et al. 2002).
Moreover, the Dutch Ministry of Economic Affairs took the initiative to carry out an additional study which was aimed at determining the economic potential of the HLT

sector and the optimal form of government intervention to stimulate this sector.

3.2.1. Study on the economic potential of the HLT sector

In particular, this exploratory study aimed at answering the following questions:

- To what extent does the HLT sector contribute to achieving sustainable economic growth in the Netherlands and Flanders?
- Does this technology offer opportunities for the Dutch and Flemish economies?
- Hoe does the HLT innovation system work?
- How can government intervene so as to improve its functioning?
- Which form of government intervention will be supported by industry and academia?

Within the framework of this study, all important actors in the HLT sector were consulted, as had been done in the HLT survey of the Platform project. The difference in this case was that, given the economic perspective of the study, the industrial sector purposely received more attention here than it did in the previous survey.
All HLT actors were asked to answer a questionnaire that had been prepared in two forms, one for industry and one for academia. In addition, they were invited to participate in three workshops during which various topics were discussed with the aim of finding answers to the above-mentioned questions. The aim of the first workshop was to discuss the HLT innovation system in an attempt to identify possible shortcomings and potential solutions. The second workshop was devoted to discussing four scenarios that illustrated different forms of government intervention and their possible effects for the HLT sector as a whole. Finally, in the third workshop priorities for strategic research were discussed.

3.2.2. Recommendations

On the basis of the data gathered in this study (responses to the questionnaire from industry and academia, the discussions in the workshops and interviews with representatives of companies and universities) the above-mentioned questions could be answered. It appeared that the HLT sector does indeed have potential for the economy in the Dutch language area. Furthermore, the analysis of the HLT innovation system revealed that the optimal form of support for the HLT sector should envisage three types of activities:

1) one to realise the prioritized resources and thus complete the HLT infrastructure

This corresponds to the BLARK proposal formulated in the HLT Platform project. However, it appeared that this sole action would not be enough to help the HLT sector and that some other accompanying measures would be required, as indicated below.

2) one to stimulate academic HLT research along the lines indicated by HLT companies

Aim is to strengthen innovation oriented strategic and pre-competitive fundamental research at universities and research institutes in response to industry needs via a programmed approach. In addition, transfer of knowledge and implementation of research results, stimulation of long-term collaborations and formation and expansion of networks are important targets of the programme. Stimulating strategic research (and thus increasing expertise) will enhance the profit deriving from investments in resource development and will lead to new ideas and insights and, possibly, to new start-ups.

3) one to stimulate the demand of HLT products

Since Dutch and Flemish service organizations are not eager to implement HLT products unless they are very convinced that the gains in efficiency will be at least twice the investment costs, the government should take measures to stimulate the demand of such products.

This corresponds with the awareness activities that are described in more detail above (3.1).

The view developed in this study appeared to be shared by the other financing institutions, which decided to combine their HLT subsidies in one comprehensive HLT programme which will start in 2004 under the auspices of the NTU.

3.3. Management, maintenance and distribution of HLT resources

In order to guarantee permanent use of project results, it is of utmost importance that the proper measures for management, maintenance and distribution are timely taken. Therefore the Dutch Language Union, as the owner of a number of Dutch-Flemish HLT-resources developed with public funding (a.o. the Spoken Dutch Corpus), took the initiative to set up an HLT agency. The aim is to combine the infrastructures required for different projects and thus reduce the costs for equipment, data, software, licences, experts, personnel etc.). This kind of collaboration could also create surplus value by combining resources and bringing together different kinds of expertise, as well as ensuring optimal visibility and accessibility by offering resources through a one-stop-shop supplier.
The HLT agency is discussed in more detail in the paper of J. Beeken and P. van der Kamp in this conference proceedings.

4. Current state of affairs and future perspectives

At the moment the organization of the HLT programme has been worked out. To guarantee its Dutch-Flemish character, this large-scale programme will be carried out under the auspices of the Dutch Language Union. A board consisting of the responsible Dutch and Flemish official bodies will supervise the programme. The Dutch Language Union is responsible for the activities

concerning awareness actions and the HLT agency. The activities of the HLT agency, carried out by the Institute of Dutch Lexicology, started at the beginning of 2004.

The board has installed a committee of Dutch and Flemish experts with the task to work out in more detail a five year plan for the reinforcement of the HLT infrastructure and the stimulation of strategic research along the lines indicated by HLT companies. The first tenders will follow towards the end of the summer of 2004. Projects will be effective at the beginning of 2005.

More information on this plan (procedure and content) will in due time be available on the Dutch Flemish HLT website: www. taalunieversum.org/tst.

5. References

Akkermans, J., van Berkel, B. Frowein, C., van Groos, L., Van Compernolle, D. (2004) Technologieverkenning Nederlandstalige Taal- en Spraaktechnologie, een rapport in opdracht van het Ministerie van Economische Zaken.

Binnenpoorte, D., F. De Vriend, J. Sturm, W. Daelemans, H. Strik, C. Cucchiarini (2002) A Field Survey for Establishing Priorities in the Development of HLT Resources for Dutch, Proceedings of LREC2002, Gran Canaria de Las Palmas, Spain.

Cucchiarini, C. and D'Halleweyn (2002) How to HLT-enable a Language: the Dutch-Flemish experience, http://www.hltcentral.org/htmlengine.shtml?id=996

Daelemans, W. and H. Strik (2002) Het Nederlands in taal- en spraaktechnologie: prioriteiten voor basistaalvoorzieningen; een rapport in opdracht van de Nederlandse Taalunie.

Van Sterkenburg, P. Kruyt, T., Van der Kamp, P. and Binnenpoorte, D. (2002) Blauwdruk voor onderhoud, beheer en distributie van door de overheid gefinancierde digitale materialen.