

# Establishing the Norwegian HLT Resource Collection

**Torbjørn Breivik**  
The Language Council of Norway

Torbjorg.Breivik@sprakradet.no

## Abstract

First, this paper gives an overview of the background and the political statements that led to the final decision in the white paper presented by the Norwegian Ministry of Culture and Church in June 2008. Next, the paper presents the initiative to buy the bankrupt estate Nordisk Språkteknologi Holding AS (NST), which gave access to annotated acoustic human language resources of good quality. The third part of the paper presents the preparatory work done in order to establish the Norwegian HLT Resource Collection (Norsk språkbank) from January 1st 2009.

## 1 Introduction

Establishing a Norwegian language resource collection for use in human language technology (HLT) research and development is an idea which can be traced back to the early 1990s. The first paper presenting a complete prospect for this kind of collection was printed in 1999.

## 2 The political setting

In 1999 the Ministry for Culture and Church put two new positions into the Norwegian Language Council in order to build “a secretary for language engineering”. The government outlined their strategy for ICT in the eNorway Action Plan and pointed at the Ministry for Culture and Church to be responsible for a policy and a strategy for how the Norwegian language could meet the challenges from the rapidly developing ICT industry producing only English versions of products meant for the Norwegian market.

The Norwegian Language Council presented their action plan in 2001. Establishing a Norwegian HLT Resource Collection was one of the

efforts proposed in this plan to meet the upcoming challenges from the ICT industry.

The rightwing government coalition presented language technology as a very important subject to work on in order to keep Norwegian as the main language throughout the Norwegian society in future. In 2003 the Ministry for Culture and Church published a white paper called “Kultur fram mot 2014” where all the good arguments were presented and the conclusion was: it’s not the time yet!

The change of government in 2005 filled us with new hope as the leftwing coalition long before entering the offices had made a statement saying that they would give priority to language technology and establish a resource collection for Norwegian HLT if they got in charge.

In late 2007 and early 2008 there were signs indicating a possible change of attitude, and in June 2008 the Ministry for Culture and Church presented the project in its white paper on language policy for the future (Mål og mening). The HLT Resource Collection was the most expensive and largest effort listed in the document. The Language Council of Norway was asked to make a plan for establishing the resource collection including a detailed plan for investment and costs within the first year, a budget for the next six, organization, realistic analysis of what content there should be and where to locate the organization.

## 3 The resources from NST

### 3.1 Valuable language resources

The Norwegian-based company Nordisk Språkteknologi Holding AS (NST) intended to produce a wide range of language technology products for the Nordic market. Among other tasks they built acoustic databases, lexicons and text bases for this purpose. Unfortunately NST went bank-

rupt in 2003. The Language Council initiated negotiations with the estate in order to save the language resources for re-use by others. The trustee of the bankruptcy gave access to the databases to let us see what was in there, and two reports were published in December 2005 (one describing content and quality and one describing the legacy). Late in 2006 a consortium consisting of three universities (Oslo, Bergen and Trondheim), IBM Norway and The Language Council of Norway bought the whole estate. The aim of the initiative was to save valuable resources and hand them over to the official Norwegian HLT Resource Collection when established.

### 3.2 What is being done with the resources?

The consortium used the year 2007 to work through the estate to see if there might be valuable resources other than the acoustic bases and the lexicons. The Norwegian Research Council granted 1 MNOK for this work, which was done by Aksis at the University of Bergen. A site for the initiative has been set up, and all available resources are documented here. Demos can be found as well as license forms and contact information for possible buyers (<http://www.sprakbanken.uib.no>).

## 4 Where are we now?

The consortium's initiative has been presented to the Ministry several times. The latest presentation took place in January 2008 when the Ministry invited colleagues from three other ministries to discuss the initiative. In all our discussions with the Ministry we underlined the interdisciplinary character of the case. This would be of great value for researchers and for the development of new products by industrial companies.

In May 2008 The Language Council was asked to present a business plan for establishing the Norwegian HLT Language Collection by mid August 2008. The Ministry needed it as input in their budgeting preparations. In June the white paper "Mål og mening" was published and the decision to establish the language collection was specified there.

### 4.1 The plan for the Norwegian HLT Language Collection

The plan is based on the report from 2002 where content, organization and budget for a Language Collection was described. The basic principles

are kept: the collection will be available for commercial as well as for researchers, the contents should follow international standards or best practices, the development work should be carried out where the experts and the knowledge are, not by this new organization.

In the new plan tools are included as a vital part of the contents. A first version of a BLARK for Norwegian has been outlined but must be discussed more deeply and in detail within the subject areas at the universities and companies working in the field. All the expenses are recalculated as new tools and improved technology are available now compared to the situation in 2002. Total expenditure for the collection is calculated to approximately 100 MNOK including the administration expenses.

## References

- Action Plan for HLT Research, The Norwegian Research Council, Oslo 2000.
- Andersen, G.: Gjennomgang og evaluering av språkressurser fra NSTs konkursbo, Oslo 2005.
- Breivik, T.: Gjennomgang og evaluering av de juridiske avtalene knyttet til språkressursene i NSTs konkursbo, Oslo 2005.
- Breivik, T.: *Action Plan for Norwegian Language and ICT and the Norwegian HLT Resource Collection*, in Proceedings, IFIP 2005, Stellenbush, South Africa 2005.
- Breivik, T.: The Norwegian HLT Resource Collection, IBM Work Shop, Prague 2008.
- The KUNSTI Programme 2001-2006, The Norwegian Research Council, Oslo 2001.
- Consolidating and Increasing the Availability of Norwegian Human Language Technology resources, The Norwegian Language Council, Oslo 2002.
- eNorway Action Plan, versjon 1.0, Ministry for Trade and Industry, Oslo, 2000.
- Handlingsplan for norsk språk og IKT, The Norwegian Language Council, Oslo 2001
- Stortingsmelding nr. 48 (2002-2003) Kulturpolitikk fram mot 2014, Oslo 2003.
- Stortingsmelding nr. 35 (2007-2008) Mål og mening, Oslo 2008.
- Svendsen, T., Johansen, M., Hofland, K., Ore, C-E.: Norsk språkbank, Trondheim, 1999.