



META-NORD

**Baltic and Nordic Branch of the European Open Linguistic
Infrastructure**

Project no. 270899

**Deliverable D2.1
Language Report for Each Language
Covered in the Project**

Version No. 1.0

06/06/2011

Document Information

Deliverable number:	D2.1
Deliverable title:	Language Report for each language covered in the project
Due date of deliverable:	31.05.2011
Actual submission date of deliverable:	06.06.2011
Main Author(s):	Tilde, UCPH, UT, UIB, UHEL, HI, LKI, UGOT
Participants:	Tilde, UCPH, UT, UIB, UHEL, HI, LKI, UGOT
Internal reviewer:	Tilde, HI
Work package:	WP2
Work package title:	Analysis and selection of language resources
Work package leader:	UT
Dissemination Level:	PU
Version:	V1.0
Keywords:	Language reports, language in the European information society, language community, language in education, language policy, language technology support, language technology programs, META-NET

History of Versions

Version	Date	Status	Name of the Author (Partner)	Contributions	Description/ Approval Level
V0.1	07.03.2011	Draft template	DFKI	Draft template for all reports	Template approved
V0.2	25.03.2011	Partners drafts	All META-NORD partners	Draft template evaluated and modified	Template approved
V0.7	14.04.2011	Stable draft	All META-NORD partners	Stable draft	Stable draft
V0.8	09.05.2011	Pre-final draft	All META-NORD partners	Pre-final version of each partners Language report	Pre-final drafts evaluated
V0.9	31.05.2011	META-NET final contribution	DFKI and All META-NORD partners	Final document modifications	Ready for delivery to EC
V1.0	06.06.2011	Delivered to EC	Tilde	Delivered to EC	Delivered to EC

EXECUTIVE SUMMARY

The META-NORD partners have compiled overviews of the language service and language technology industry for all the languages targeted by the project. These languages include the main official languages spoken in the Nordic and Baltic geographical area: Danish, Estonian, Finnish, Icelandic, Latvian, Lithuanian, Norwegian (Nynorsk and Bokmål) and Swedish. The state of affairs in language technology in the Nordic and Baltic countries is described in the eight reports. Each report describes the situation of a language community and the position of the language service and language technology industry for that particular language. The reports are written as a series of separate publications for each language, but they are closely coordinated in structure. For each language, an analysis of the language community has been conducted and the role of the language in the respective country/language community is described. The language technology research community and the language service and language technology industry are identified. The importance of language technology products and services in the language community are assessed. Legal provisions related to language resources and tools, which may differ from country to country, are outlined.

The first section of the reports, which is common to all the reports and written by experts from the DFKI, is intended to explain the opportunities and challenges for language technology in the modern information society. The next section is different for each language and written by experts in the particular language. It contains subsections on general facts of the language (number of speakers, official status, dialects, etc.), particularities of the language, recent developments in the language, language cultivation, language in education, international aspects, as well as the role of the language on the Internet. The reports further contain an important section on language technology support for the language in question. It contains subsections on the core application areas of language and speech technology, such as language checking, web search, speech interaction, machine translation, etc., and describes the situation in the language with respect to the application areas. Furthermore, there are language particular subsections on language technology in education and language technology programs in the country in question. The common parts of this section are written by the DFKI experts but the language particular parts are written by experts on each language.

The reports also contain detailed tables of Language Technology tools and resources for each language, including numerical estimates of a large number of technological aspects. The results indicate that only with respect to the most basic tools and resources such as tokenizers, PoS taggers, morphological analyzers/generators, syntactic parsers, reference corpora, and lexicons/terminologies, the status is reasonably positive for all META-NORD languages. Furthermore, all the languages seem to have some tools for information extraction, machine translation and speech recognition and synthesis, as well as resources such as parallel corpora, speech corpora, and grammar, although these tools and resources are rather simple and have limited functionality for some of the languages. When it comes to more advanced fields such as sentence and text semantics, information retrieval, language generation, and multimodal data, it appears that one or more of the languages lack tools and resources for these fields. For the most advanced tools and resources such as discourse processing, dialogue management, semantics and discourse corpora, and ontological resources, most of the languages either have nothing of the kind or their tools and resources have a quite limited scope.

The means for all the languages together indicate that quantity and availability may be a greater concern than quality; this need is the very *raison d'être* of the META-NORD project.

META-NORD Language Reports

Eight Language Reports are added to this document as separate .pdf files:

- Language Report for the Latvian language prepared by Tilde – File: D2.1 Language Report for each language covered in the project_Latvia.pdf
- Language Report for the Danish language prepared by UCPH – File: D2.1 Language Report for each language covered in the project_Denmark.pdf
- Language Report for the Estonian language prepared by UT – File: D2.1 Language Report for each language covered in the project_Estonia.pdf
- Language Report for the Norwegian language prepared by UIB – File: D2.1 Language Report for each language covered in the project_Norway.pdf
- Language Report for the Finnish language prepared by UHEL – File: D2.1 Language Report for each language covered in the project_Finland.pdf
- Language Report for the Icelandic language prepared by HI – File: D2.1 Language Report for each language covered in the project_Iceland.pdf
- Language Report for the Lithuanian language prepared by LKI – File: D2.1 Language Report for each language covered in the project_Lithuania.pdf
- Language Report for the Swedish language prepared by UGOT – File: D2.1 Language Report for each language covered in the project_Sweden.pdf