



Project Goal

The META-NORD project aims to establish an open linguistic infrastructure in the Baltic - Estonia, Latvia and Lithuania, and Nordic countries - Denmark, Finland, Iceland, Norway and Sweden, as part of the META-NET network.

Duration: 2011/02 – 2013/01

Main results of the first year

MAPPING AND DESCRIBING THE LANGUAGE TECHNOLOGY LANDSCAPES

in the Nordic and Baltic countries in terms of language use, language technology and resources.

- The results analysed in a series of white papers.
- The status of the most basic tools is reasonably positive in all project languages; more advanced tools are scarce or absent.
- Basic LRs available for Nordic and Baltic languages: tokenizers, PoS taggers, morphological analyzers/generators, syntactic parsers, reference corpora, and lexicons/terminologies.
- Some of the project languages have parallel corpora, speech corpora, computational grammars, though limited and not always tested/documentated.
- Lack of LRs in more advanced fields, such as sentence and text semantics, information retrieval, language generation, and multi-modal data.

IDENTIFYING, COLLECTING, UPDATING AND DISTRIBUTING OF LANGUAGE RESOURCES:

- 67 resources are documented.
- The resources are selected according to their quality, relevance, and usability in multilingual services.
- Language resources are documented, processed, linked and upgraded to agreed standards and guidelines.
- Several of the largest computational lexicons have been converted to the agreed LMF standard, e.g. SALDO (UGOT), SCARRIE (UiB), STO (UCPH).
- All Nordic and Baltic wordnets are linked to the Core Princeton WordNet.
- Several of the resources with a multilingual perspective are now actually interoperable in content.
- META-SHARE nodes operational at Tilde, Swedish Språkbanken and University of Helsinki.

SHARED LANGUAGE RESOURCES

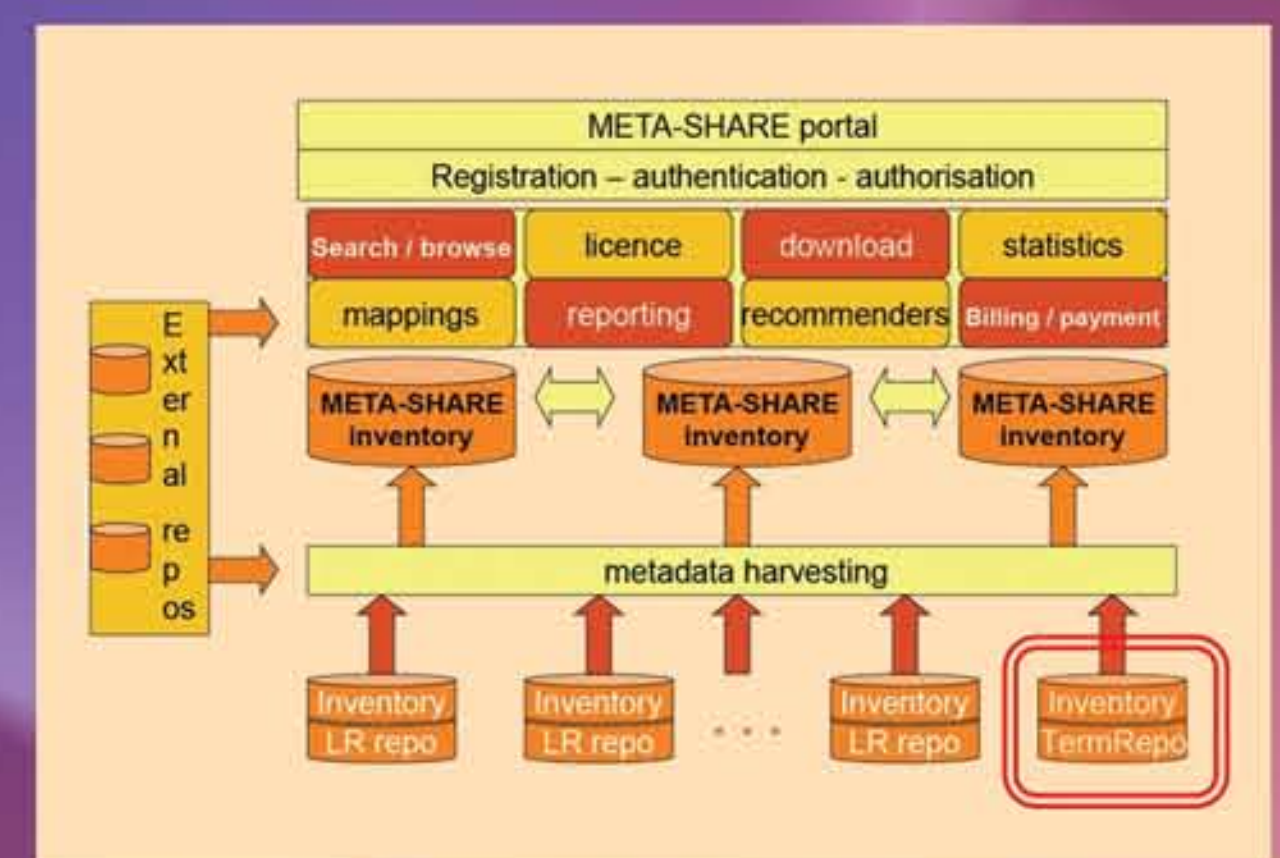
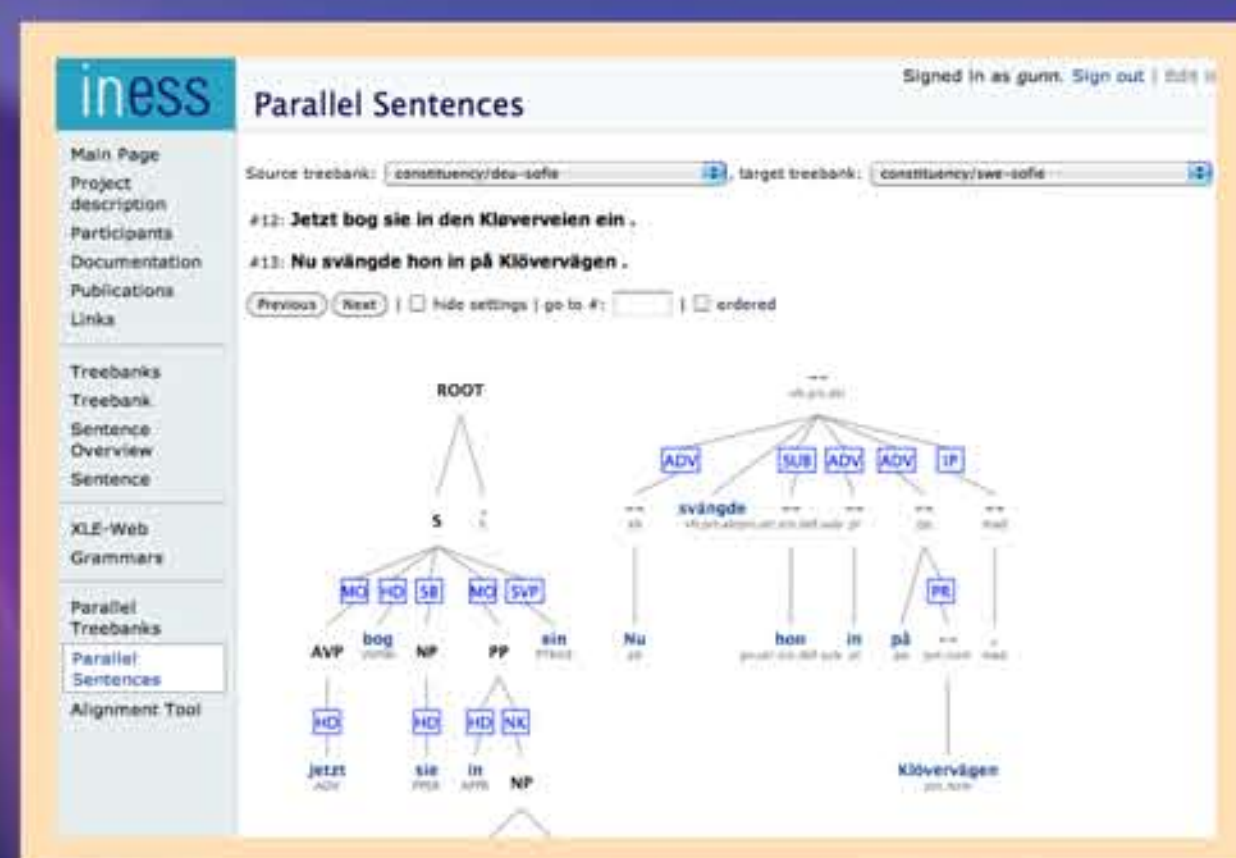
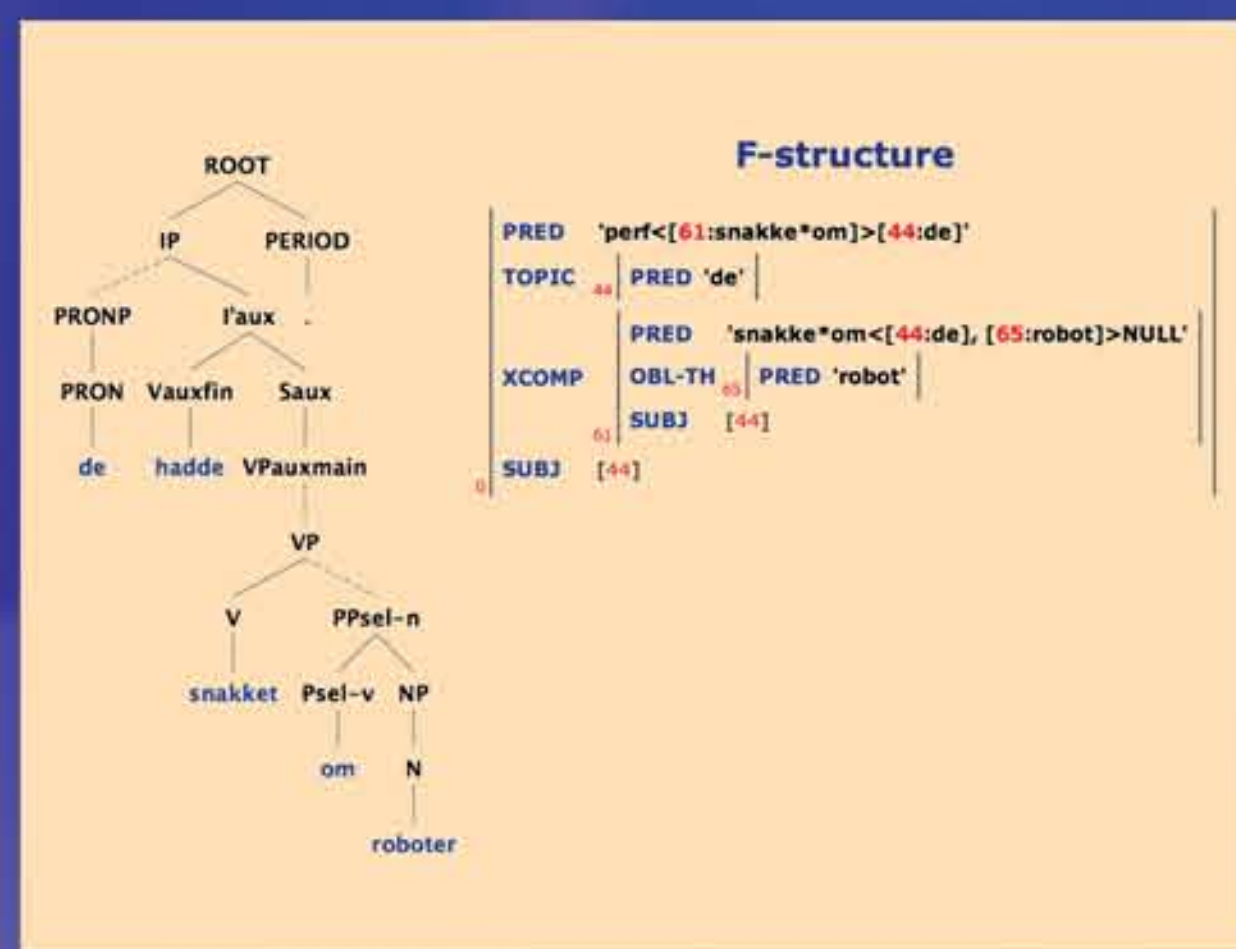
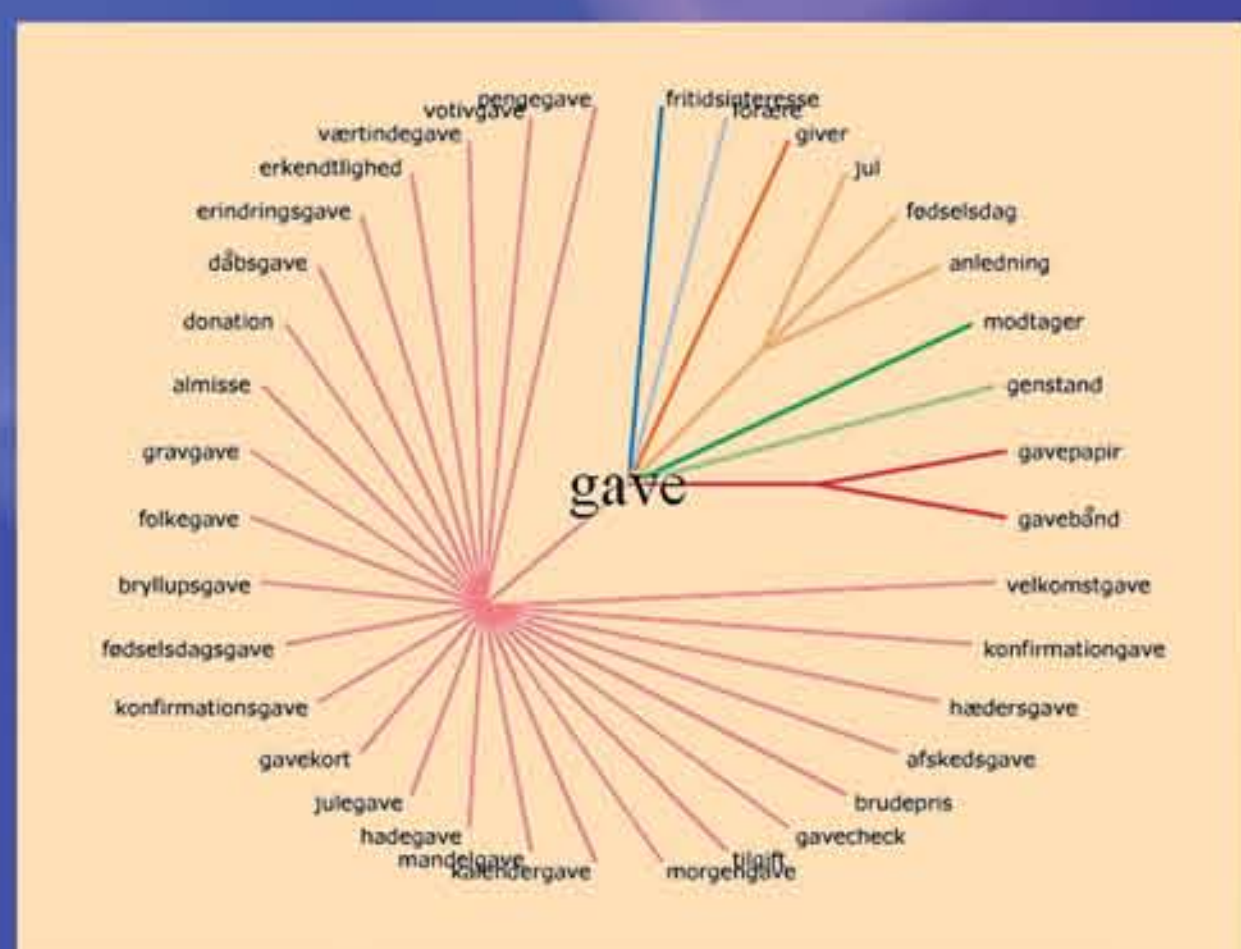
Resources in First Batch	
36	Lexical resources
16	Corpora
6	Treebanks
5	Resources for speech
3	WordNet
1	Tool
67	Total

The total number of the types of resources in the 1st batch

WORDNETS. Validation and pilot linking between the Baltic and Nordic wordnets. The aim is to test the perspective of multilingual linking of the Baltic and Nordic wordnets and, through such pilot linking, tentatively compare and validate the wordnets along the measures of taxonomical structure, coverage, granularity, and completeness.

TREEBANKS. Harmonising treebanks by making them accessible through a uniform web interface and by linking across languages.

TERMINOLOGY. Consolidation of distributed terminology resources across languages and domains by extending the open linguistic infrastructure with multilingual terminology resources.



Join the META community. Share your resources.

PROJECT PARTNERS



CONTACT INFORMATION

ANDREJS VASIĻJEVS
andrejs@tilde.com
Tilde, Vienības gatve 75a, Rīga
LV1004, Latvia



The META-NORD project has received funding from the European Commission through the ICT PSP Programme, grant agreement no 270899.