Creation of an Open Shared Language Resource Repository in the Nordic and Baltic Countries

Andrejs Vasiļjevs[a], Markus Forsberg[f], Tatiana Gornostay[a], Dorte H. Hansen[c], Kristín M. Jóhannsdóttir[e], Krister Lindén[d], Gunn I. Lyse[b], Lene Offersgaard[c], Ville Oksanen[d], Sussi Olsen[c], Bolette S. Pedersen[c], Eiríkur Rögnvaldsson[e], Roberts Rozis[a,] Inguna Skadiņa[a], Koenraad De Smedt[b]

Tilde[a], University of Bergen[b], University of Copenhagen[c], University of Helsinki[d], University of Iceland[e], University or Gothenburg[f]

Latvia[a], Norway[b], Denmark[c], Finland[d], Iceland[e], Sweden[f]

E-mail: metanord@tilde.lv

Abstract

The META-NORD project has contributed to an open infrastructure for language resources (data and tools) under the META-NET umbrella. This paper presents the key objectives of META-NORD and reports on the results achieved in the first year of the project. META-NORD has mapped and described the national language technology landscape in the Nordic and Baltic countries in terms of language use, language technology and resources, main actors in the academy, industry, government and society; identified and collected the first batch of language resources in the Nordic and Baltic countries; documented, processed, linked, and upgraded the identified language resources to agreed standards and guidelines. The three horizontal multilingual actions in META-NORD are overviewed in this paper: linking and validating Nordic and Baltic wordnets, the harmonisation of multilingual Nordic and Baltic treebanks, and consolidating multilingual terminology resources across European countries. This paper also touches upon intellectual property rights for the sharing of language resources.

Keywords: language resources, open infrastructure, national language technology landscape, Baltic languages, Nordic languages

1. Introduction

In the past decades, numerous language resources (data and tools) have been created for all languages of the European Union, including lesser-resourced languages (e.g., the languages of the Baltic countries). However, these resources are often not readily available and used because they are not sufficiently catalogued, use different standards (if any), are not accessible or downloadable online, and are often poorly documented. As a result, language resources lack visibility and do not live up to their full potential for exploitation in research and development activities aimed at language products and services.

To address these issues, the European Commission has dedicated specific activities in its Seventh Framework Programme and Information and Communication Technologies Policy Support Programme [1]. Under the umbrella of the META-NET network[2], currently consisting of 57 research centres from 33 countries, the projects T4ME, CESAR, METANET4U, and META-NORD were initiated in 2011. These projects closely cooperate to build the technological foundations of a multilingual European information society by facilitating the creation of an open infrastructure enabling and supporting large-scale multilingual and cross-lingual services and applications.

The META-NORD project[3] (Vasiļjevs et al., 2011; Skadiņa et al., 2011) focuses on eight national languages in the Nordic and Baltic region: Danish, Estonian, Finnish, Icelandic, Latvian, Lithuanian, Norwegian, and Swedish. On January 13, 2012, the project concluded the first year of its two year life span.

In this paper we present the key objectives of META-NORD and the results achieved during the first year. We focus on the following project activities:

- mapping and describing the national language technology landscape in the Nordic and Baltic countries in terms of language use, language technology and resources, main actors (academy, industry, government and society);
- identifying and collecting the first batch of language resources in the Nordic and Baltic countries;
- documenting, processing, linking, and upgrading the identified language resources to agreed standards and guidelines.

In the second section, we describe the language technology landscape for the Nordic and Baltic languages. The third section focuses on the first batch of META-NORD language resources, their metadata, upgrade to agreed standards, and setting up an online sharing platform – the META-SHARE nodes. Further we overview the three horizontal multilingual actions in META-NORD: linking and validating Nordic and Baltic wordnets, the harmonisation of multilingual Nordic and Baltic treebanks, and consolidating multilingual terminology resources across European countries. The fifth section touches upon intellectual property rights for the sharing of language resources. Finally, we make some concluding remarks.

[1] http://ec.europa.eu/information_society/activities/ict_psp/documents/ict_psp_wp2010_final.pdf

[2] http://www.meta-net.eu

[3] http://www.meta-nord.eu

2. Language technology landscapes for the Nordic and Baltic languages

Since each of the META-NORD languages has less than 10 million speakers, the community of language resource creators and users in the Baltic and Nordic countries is small, and the viability of commercial efforts is largely dependent on public support. Even a modest increase in the availability and quality of language resources (LR) is appreciable for technology developers and end users in the Nordic and Baltic countries.

The national language technology (LT) landscape for each language addressed in the project is analysed and described in a series of white papers[4]. These reports pertain to the language service and LT industry and contain information regarding general facts about each language, its particularities, recent developments in the language and LT support, and core application areas of language and speech technology. The language white papers also present a cross-language comparison ranking the respective language within four key areas: machine translation, speech processing, text analysis, and resources (Figures 1-3). This comparison is based on expert ratings of LR for each language. Experts were asked to rate the existing tools and resources with respect to seven criteria: quantity, availability, quality, coverage, maturity, sustainability, and adaptability on a scale of 0 (no tools/resources) to 6 (well presented).

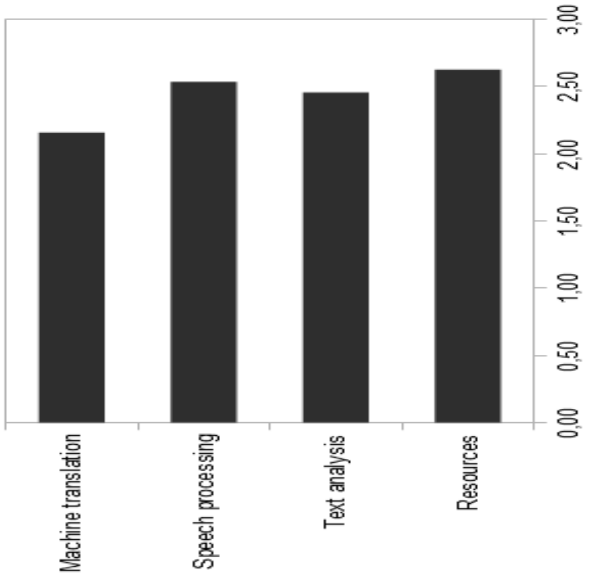

Figure 1: Average scores in four key areas across the META-NORD languages

The results indicate that only for the most basic tools and resources, such as tokenisers, PoS taggers, morphological analysers / generators, syntactic parsers, reference corpora, lexical resources and termbanks, the status is reasonably positive for all of the META-NORD languages. Furthermore, all META-NORD languages have some tools for information extraction, machine translation, and speech synthesis. There are parallel corpora, speech corpora and computational grammars for some of the META-NORD languages, though these are limited in coverage and functionality and are not always sufficiently tested and documented. When it comes to the more advanced areas (e.g., 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 areas.

An initial comparison across all 30 META-NET languages places three small languages of the Nordic and Baltic region – Icelandic, Latvian, and Lithuanian – in the bottom cluster, defined as having major gaps in all of the four key areas. The relative ranking of the remaining five META-NORD languages is slightly higher, although none of them come close to the so-called "big" languages (English, French, Spanish, and German).

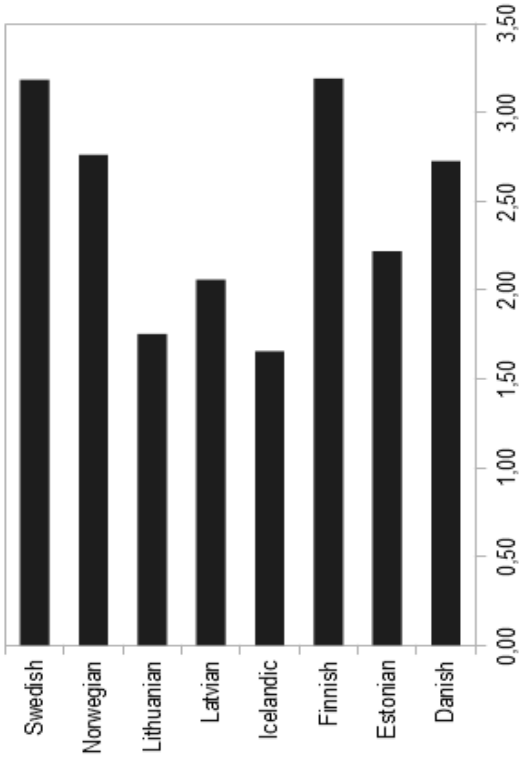

Figure 2: Average scores for each of the eight META-NORD languages

English versions of the language white papers for all of the META-NORD languages were prepared, submitted to the European Commission, and published online [5] in the beginning of June 2011. Since then, all of the white papers have gone through thorough revision and reformatting. The reports have also been translated into the respective languages.

The white papers are currently in print and will be published by Springer and distributed in the spring of 2012 – each language in a separate publication, with the native language version first, followed by the English text.

[4] http://www.meta-net.eu/whitepapers

[5] http://www.meta-net.eu/whitepapers

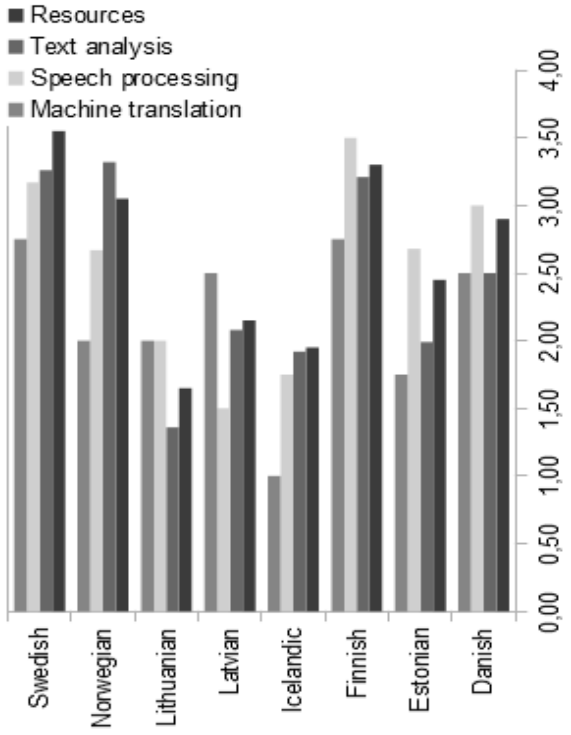

Figure 3: Evaluation results for four key areas in all eight META-NORD languages

The target group of the white papers includes politicians, government agencies, funding bodies, journalists, those who are in a position to enhance the awareness and role of LT in society. It is hoped that the Language reports will assist in making people realise the potentials of LT as an essential means towards a truly multilingual Europe.

3. First batch of META-NORD language resources: metadata description and upgrade to agreed standards

An important goal of META-NORD is to identify, upgrade and harmonise national LR within and across the META-NORD languages, in order to make them interoperable with respect to their data formats and content. Thus far, we have identified an initial pool of LR. It includes the most important LR for each language, more than 150 on the whole.

Identified resources are selected according to their quality, relevance, and usability in multilingual services. Table 1 shows the distribution of LR which were documented and made available in the first round of the project by the end of 2011, all in all 67 LR, including data and tools. These resources are available via the common META-SHARE interface[6].

A crucial measure of the project's success is the number of external LR providers participating in third-party networks and contributing to the META-SHARE infrastructure, as well as the number of LR made available by META-NORD. About half of the identified LR are owned by the consortium, while the other half is third-party owned. The work of the consortium on metadata, legal issues, and project dissemination is preparing the grounds for more third-party resources. It will hopefully result in numerous uploads of language resources to META-SHARE by the end of the project.

Resources made available in the 1st batch			
36	Lexical resources	**16**	Corpora
6	Treebanks	**3**	WordNet
5	Resources for speech	**1**	Tool
		67	**Total**

Table 1: The total number of language resources in the first batch.

3.1. Upgrade to agreed standards

Language resources to be made available by META-NORD come in many formats. The project partners put considerable effort into making LR content models as interoperable as possible. This implies adopting more strictly structured formats, e.g., Lexical Markup Format (LMF, ISO standard 24613, 2008) rather than the proprietary XML or SQL database structure for lexical resources. It also implies mapping to a set of standardised data categories, e.g., ISOcat[7].

For example, some lexical databases, such as the Danish STO, the Swedish Språkbanken's lexical resources, and the Norwegian SCARRIE lexical resources, have been upgraded to LMF.

The Danish lexicon STO consists of morphological, syntactic and, semantic levels. Currently, the morphological level (about 80 000 entries) is being upgraded by converting the original intensional morphology description to an extensional description. Unlike the extensional description, the intensional one does not explicitly list the word forms: instead the lexical entry is associated with a morphological pattern. In this process, the challenges are to express the same information and preserve as much structure as possible. The structural problems have now been resolved, and the specification of the data categories is ongoing. Nouns, adjectives, and verbs have been upgraded, and the conversion of the other part of speech is ongoing. When the morphological level has been upgraded, the next step will be to upgrade the syntactic level of STO (about 43 000 entries) to LMF.

Språkbanken's lexical resources are a steadily growing collection of freely available Swedish lexical resources for both modern and historical Swedish. The collection, at the time of writing, consists of 15 resources that are both downloadable and browsable through the open lexical infrastructure of Språkbanken[8] (Borin et al., 2012). The resources are being integrated and linked in the Swedish Framenet++ project (Borin et al., 2010) using SALDO (Borin and Forsberg, 2009) as the pivot resource – a large, freely available lexicon with morphological and semantic information. And now, through the work in META-NORD, all resources have been upgraded to LMF.

[6] http://www.meta-share.eu

[7] http://www.isocat.org

[8] http://spraakbanken.gu.se

3.2 META-SHARE nodes

For all relevant types of language resources, META-NORD is preparing standardised top-level descriptions (*metadata*) based on a recommended set of metadata descriptors for documenting the META-SHARE resources[9]. This has produced consistent descriptions for each resource contributed to the shared pool.

The META-SHARE tool brings two main functionalities: it is a metadata editor, and it is a search and browse tool. The editor supports metadata authoring through, for example, providing listings of the controlled vocabularies of the metadata schema. The search and browse facilities allow for the identification and exploration of the resources described with the metadata.

The META-SHARE tool has been installed at three of the META-NORD partners' servers: Tilde [10], Swedish Språkbanken [11] and the University of Helsinki [12]. An installation is referred to as a network node, and several other nodes exist within the META-NET[13] community. The nodes remain isolated, which will be changed in the near future through metadata harvesting, i.e., automatic data sharing.

The current interface offers browsing and search based on keywords or with filtering based on language, resource type (lexical or conceptual; corpus; tool or service), and media type (text; audio; multimedia). Currently, 67 resources have been made available via this metadata repository.

It must be mentioned that even if the LR set has been collected in the Nordic and Baltic region, it includes also LR available in the region for a number of languages from outside the region, in particular in the context of multilingual resources such as EuroTermBank[14]. In fact, 30 languages are already represented in the current collection. The LR data itself is distributed and can mostly be retrieved at the place where it has been constructed. In some countries, national data repositories have been established. *Språkbanken* at the National Library of Norway and its namesake at the University of Gothenburg are cases in point. META-NORD has achieved considerable synergy with these initiatives.

4. Multilingual actions in META-NORD

META-NORD has three additional horizontal action lines: wordnets, treebanks, and terminology resources, which are each presented in this section.

4.1 Linking and validating Nordic and Baltic wordnets

The multilingual task on wordnets is concerned with the validation and pilot linking between the Nordic and Baltic wordnets. The builders of these wordnets have applied very different compilation strategies: Danish, Icelandic, and Swedish wordnets are being developed via monolingual dictionaries and corpora, and are subsequently linked to the Princeton WordNet. In contrast, Finnish and Norwegian wordnets are applying the expansion method by translating from the Princeton WordNet and Danish wordnet (DanNet). The Estonian wordnet was built as part of the EuroWordNet project (Vossen, 1999) by translating the base concepts from English as a first basis for monolingual extension. The aim of the multilingual action is to test the perspective of a multilingual linking of the Nordic and Baltic wordnets.

The linking is performed via the so-called Princeton "core wordnet", a subset of the Princeton WordNet containing the most core synsets[15] to which all the Nordic and Baltic wordnets have been linked during the first phase of META-NORD. The linked (bilingual) wordnets will be validated for their correctness, but the pilot linking will also serve as a test bed for tentatively comparing and validating the wordnets along the measures of taxonomical structure, coverage, granularity, and completeness (cf. Pedersen et al., 2012):

- ***Taxonomical structure***. Do different approaches generally lead to different taxonomical structures of the lexical networks, and can we to some extent define best practice regarding depth of structure?
- ***Coverage.*** Are frequent concepts in the target language covered well enough when compiling a wordnet via English? And when deducing it from a traditional lexical resource? Can we define a coverage "pain threshold"? These and related issues will be evaluated using corpora and existing core vocabulary lists.
- ***Granularity of the described concepts.*** Does a specific approach result in many or few sense distinctions (i.e., synonym sets) for each lemma? Is it possible to identify a technology-oriented best practice for sense granularity (i.e., something that corresponds to the main senses of traditional lexicography?)
- ***Completeness of synonym sets.*** Does a given approach bring about many or few semantic relations and/or semantic features per concept? Can a best practice set of semantic relations be established along the validated wordnets?

Current work focuses on adapting the wordnets into a common browser in order to facilitate validation. It has been decided to apply the "Andreord Browser" that is currently used for the Danish wordnet (cf. Johannsen & Pedersen, 2011). For illustration, Figure 4 and Figure 5 show the differences in the taxonomical structures of the Finnish and Danish wordnets for the concept "animal" as depicted by the "Andreord Browser".

[9] http://www.meta-net.eu/meta-share/metadata-schema/
[10] http://metanode.tilde.com
[11] http://spraakdata.gu.se/ws/metashare/
[12] http://metashare.csc.fi
[13] http://www.meta-share.org/
[14] www.eurotermbank.com
[15] http://wordnet.princeton.edu/wordnet/download/standoff/

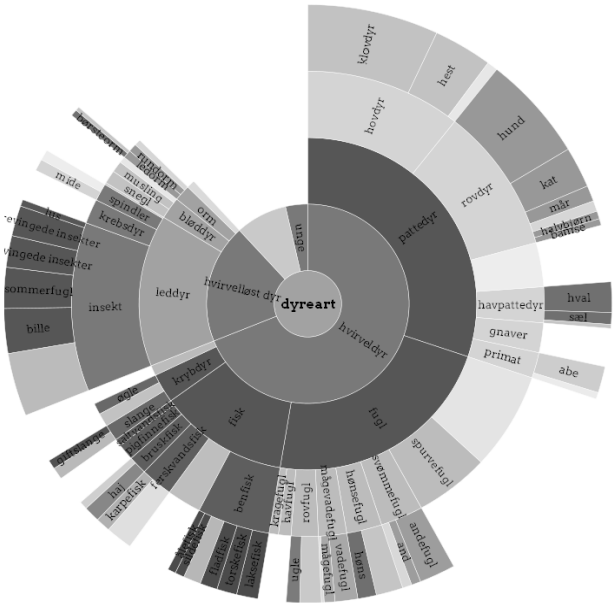

Figure 4: Taxonomical structure of subconcepts of animal in the Danish wordnet

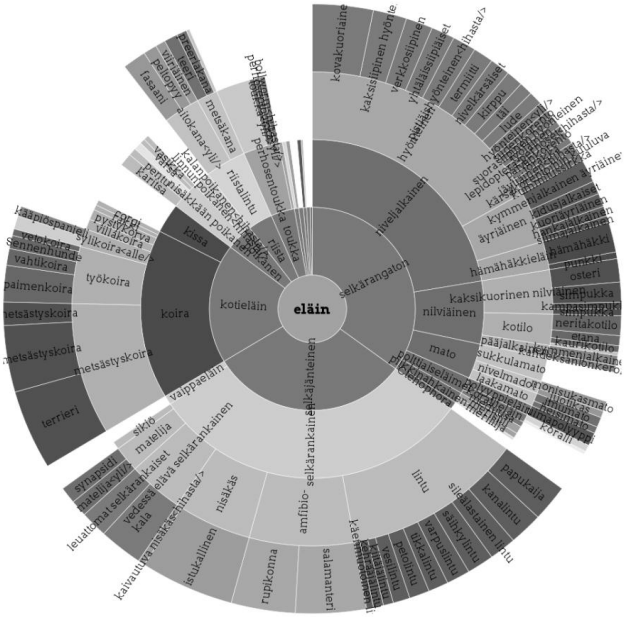

Figure 5: Taxonomical structure of subconcepts of animal in the Finnish wordnet

4.2 Multilingual Nordic and Baltic treebanks

The horizontal task on multilingual treebanks aims to harmonise treebanks by making treebanks for various languages accessible through a uniform web interface, and by linking treebanks across languages. In cooperation with INESS [16], a growing collection of treebanks is made available for browsing, search, visualisation and download. Indexing and filtering are provided for all treebanks with a search engine that is a reimplementation of TigerSearch with extensions and improvements of the search query syntax (Meurer, 2012).

Currently, INESS provides access to treebanks in the following languages in the region covered by META-NORD: Norwegian Bokmål (9), Icelandic (2), Northern Sami (2), Danish (1), Estonian (1), Swedish (1), in addition to treebanks in other languages. The annotation types cover LFG, dependency, and constituency annotation.

This action has also delivered a pilot parallel treebank, aligned at the sentence level across a limited number of languages. The text material for this treebank has been taken from the Norwegian novel *Sofies verden* [Sophie's world] (Gaarder, 1991). This work was chosen because it is linguistically rich and it has been translated professionally into many languages. Based on previous coordinating work by the Text Laboratory in Oslo, the Norwegian META-NORD partner has cleared rights for research use of the initial chapters in annotated form with some limitations on copying. The resulting *Sofie* parallel treebank is linked at the sentence level and is currently available for the 15 language pairs: Danish ⇄ Icelandic, Danish ⇄ Norwegian (LFG), Danish ⇄ Swedish, Danish ⇄ Estonian, Danish ⇄ German, German ⇄ Icelandic, German ⇄ Norwegian (LFG), German ⇄ Swedish, German ⇄ Estonian, Estonian ⇄ Icelandic, Estonian ⇄ Norwegian (LFG), Estonian ⇄ Swedish, Icelandic ⇄ Norwegian, Icelandic ⇄ Norwegian (LFG), Icelandic ⇄ Swedish.

A Finnish version will be added at the end of March 2012. Access to this material, including online services and download, is provided in cooperation with INESS (Figure 6).

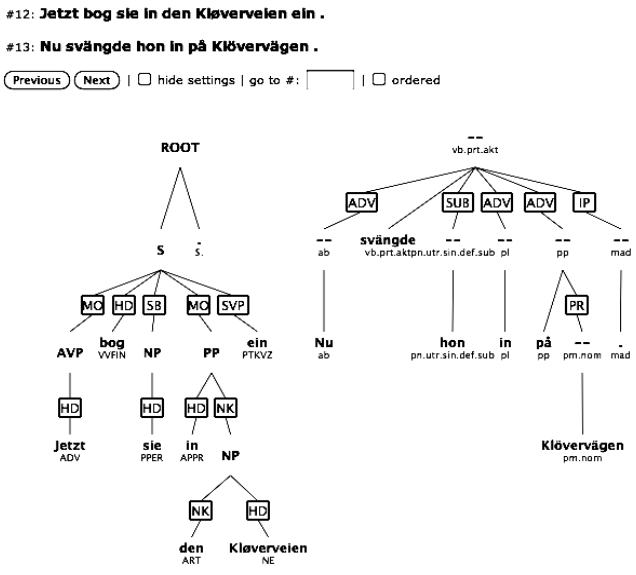

Figure 6: Syntactically analysed German-Swedish sentences in the parallel Sofie treebank

[16] Infrastructure for the Exploration of Syntax and Semantics, http://iness.uib.no

Furthermore, this action aims at extending and syntactically annotating a small part of the JRC-Acquis Multilingual Parallel Corpus (Steinberger et al., 2006), which contains texts in all EU languages. Whereas Norwegian is not an official language of the EU, a partial translation to Norwegian with over 5000 texts has been obtained from the Norwegian Ministry of Foreign Affairs for inclusion in this parallel corpus.

In a subsequent phase, phrase alignment will be attempted between at least Danish and Norwegian, using technology from the XPAR project (Dyvik et al., 2009).

4.3 Consolidating multilingual terminology

META-NORD also addresses a growing demand for the consolidation of distributed terminology resources across languages and domains by extending the open linguistic infrastructure with multilingual terminology resources. META-NORD partners (Tilde, the Institute of Lithuanian Language, the University of Tartu, and the University of Copenhagen) have already established the solid terminology consolidation platform EuroTermBank (Vasiljevs et al., 2008) which provides a single access point to more than 2 million terms in 27 languages (Figure 7). However, terminology coverage for some languages in EuroTermBank (Latvian, Lithuanian, Polish, and others) surpasses that of other languages, especially Western and Nordic languages, whose terminology resources have not been as comprehensively integrated into the EuroTermBank platform.

The goal of the terminology task in META-NORD is twofold:

- to increase the number of publicly available terminology resources by identifying and consolidating more resources;
- to integrate EuroTermBank with META-SHARE as a terminology storage and access node in the META-SHARE infrastructure.

The broadening of terminology resources coverage is done by META-NORD partners identifying and addressing holders of terminology databanks, be it the National terminology databanks, university databases, national terminology portals, or alike in each respective country. Common understanding must be reached to facilitate the sharing of terminology resources through cross-linking and the federation approach as well as elaborating the mechanism for consolidated multilingual representation of monolingual and bilingual terminology entries.

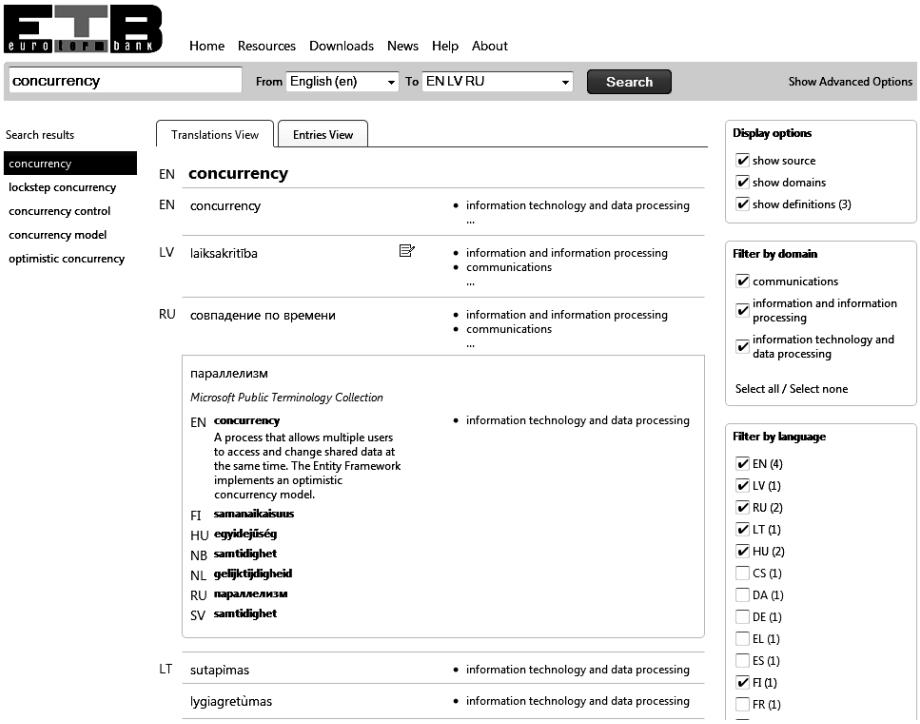

Figure 7: Consolidated representation of terminology entries from different bilingual and multilingual resources

Many terminology resources created by multinational companies for translation and localization of their products are still proprietary and even confidential: various industry players do not disclose their terminology to public users. At the same time, the idea of sharing is slowly becoming a common practice (Vasiljevs et al., 2010), and the sharing of terminology resources is also being encouraged in META-NORD. For terminology resources which cannot be granted for full sharing via download, EuroTermBank will provide an access to individual terms through search and lookup facilities.

The integration of terminology resources into META-SHARE will be gradual. First, the metadata of each terminology resource will be added to META-SHARE. In

parallel, the integration and data exchange between EuroTermBank and META-SHARE will be implemented, and the new terminology resources will be integrated and interlinked with EuroTermBank. Eventually, META-SHARE should be able to perform live mining of metadata from EuroTermBank, and terminology search and lookup from META-SHARE should be possible by transfer to EuroTermBank. The sharing of terminological data is based on the TBX standard (TBX, ISO standard 30042, 2008).

It is anticipated that the integration of terminology resources into the META-SHARE infrastructure through EuroTermBank can be further extended to other European countries by other projects of the META-NET network, respectively, CESAR, METANET4U, and T4ME.

5. Intellectual Property Rights (IPR)

For promoting the use of open data and following the Creative Commons and Open Data Commons principles, META-NORD applies the most appropriate license schemes from the set of templates provided by META-NET. Model licenses are checked by the consortium with respect to regulations and practices at national levels, taking into account possible different regimes due to the ownership, type, and/or pre-existing arrangements with the owners of the original content from which the resource was derived. Resources resulting from the project are being cleared, i.e., made compliant with the legal principles and provisions established by META-NET, as completed and amended by the consortium and accepted by respective right holders.

Of special interest, from an IPR point of view, are the collective works, i.e., parallel treebanks and multilingual core wordnets. To avoid cumbersome joint international ownership, LR are distributed as separate works for each language. To create a multilingual resource, individual resources have a standard format that can be loaded into a database and joined using commonly agreed interfaces upon interlingual indexes, and in this case, the only thing that needs to be checked is that individual works have compatible licenses.

One of the most difficult legal problems that very large data sets have is the number of rights holders. The transaction cost for reaching every single rights holder is prohibitive. However, this situation is quite similar to another field, i.e., the re-use of radio and TV programs. The solution there has been to use collective licensing, which has worked relatively well. Collective licensing means that a Copyright Society, which administers and promotes the rights of all (or a substantial portion of) copyright holders, licenses a collection of works on behalf of all the rights holders.

This approach should also be possible for scientific databases, i.e., collections of works for scientific use. There are of course several problems, and the most optimal solution would be to redefine the scope of copyright legislation to exclude at least non-competing scientific use of content, i.e., a copyright exception for scientific use. However, content owners are likely to fight such changes because they see them as a threat to their future business models. Therefore, collective licensing is more likely to get wide support politically.

If the Collective Licensing approach is chosen, several questions have to be solved:

- What kind of content is included? Does the license cover only texts? Or are audio and visual works are also included? What about software and databases? The broader the license is, the more useful the data set is, but it also makes negotiations harder.
- What kind of organisations are in charge of the system? In Finland, there are three major organisations for different aspects of collective licensing: Teosto for music copyrights, Gramex for related rights, and Kopiosto for literature, photo-copying, and broadcasting rights. From a customer-service point of view, it might be preferable to have one organisation granting all of the rights for research. In Finland, that would then most likely be Kopiosto, which already has a wide range of education-related agreements. (From a negotiation point of view, it could, of course, be even better to have several organisations granting all of the rights.)
- What is the right price for the license and who can get one? What about research in commercial organisations, and what about individual researchers who are not part of any academic research organisation? For the supply side, the questions are: how is the use measured and how is the collected money shared between the rights holders in the value chain? For music, there are several different models for deciding on pricing. The negotiations are typically difficult due to the inherently monopolistic nature of the system and the problem of public rent seeking. The best solution might be the one used in Canada, in which a board of financial experts is used for estimating a reasonable price[17].
- What kind of uses would the license cover? It is quite common that at least part of the research has direct commercial outcomes. Should there be different pricing categories for basic research and commercial research?
- Should the system be national or should the license cover the whole EU? The European Commission is currently pushing collective rights management systems to open their licenses for EU-wide licensing and also plans to create only one open data license for European content. Therefore, this is most likely the only viable approach for a new system.

Even with all of these open questions, the collective licensing approach is something that can offer a relatively easy way forward. The rights holders know the system and experiences are quite good, e.g., with photo-copying. Therefore, this could be the middle ground, which is not perfect for everyone, but is still acceptable.

[17] http://www.editionsyvonblais.com/description.asp?DocID=8469

6. Conclusion

META-NORD lays the ground for fruitful cooperation in identifying, enhancing, and sharing of LR created in the Nordic and Baltic countries. The language white papers have shown that these countries still have a long way to go to implement the vision of the region as a leader in LT. Currently, 67 initial resources have been included into the META-SHARE repository, and their number will be increasing in the second year of the project.

7. Acknowledgements

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

8. References

Borin, L. and Forsberg, M. (2009). All in the family: A comparison of SALDO and WordNet. In *Proceedings of the NODALIDA 2009 Workshop on WordNets and other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*, Odense.

Borin, L., Dannélls, D., Forsberg, M., Toporowska, M., Kokkinakis, G. and Kokkinakis, D. (2010). The past meets the present in Swedish FrameNet++. In *Proceedings of the 14th EURALEX International Congress*.

Borin, L., Forsberg, M., Olsson, L. and Uppström, J. (2012). The open lexical infrastructure of Språkbanken. To appear in *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul: ELRA.

Dyvik, H., Meurer, P., Rosen, V. and De Smedt, K. (2009). Linguistically motivated parallel parsebanks. In Marco Passarotti, Adam Przepiórkowski, Sabine Raynaud and Frank Van Eynde (Eds.), *Proceedings of the 8th International Workshop on Treebanks and Linguistic Theories*, Milan, Italy, ED-UCat, pp.71-82.

Gaarder, J. (1991). *Sofies verden: roman om filosofiens historie*. Aschehoug, Oslo, Norway.

ISO 24613 (2008). Language resource management – Lexical Markup Framework (LMF).

ISO 30042 (2008). Systems to manage terminology, knowledge and content – TermBase eXchange (TBX).

Johannsen, A. and Pedersen, B. S. (2011). *"Andre ord" – a wordnet browser for the Danish wordnet, DanNet* . In *Proceedings from 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)*, Riga, Latvia. Nothern Association for Language Technology, Vol. 11, pp. 295-298, University of Tartu.

Meurer, P. (2012, submitted). INESS-Search: A search system for LFG (and other) treebanks. In *Proceedings of the LFG 2012 Conference*.

Pedersen, B.S., Borin, L., Forsberg, M., Lindén, K., Orav, H. and Rögnvalssson, E. (2012). Linking and Validating Nordic and Baltic Wordnets – A Multilingual Action in META-NORD. In *Proceedings of 6th International Global Wordnet Conference*, Matsue, Japan, pp. 254-260.

Skadiņa, I., Vasiljevs, A., Borin, L., De Smedt, K., Linden, K. and Rognvaldsson, E. (2011). META-NORD: Towards Sharing of Language Resources in Nordic and Baltic Countries. In *Proceedings of Workshop on Language Resources, Technology and Services in the Sharing Paradigm*, Chiang Mai, Thailand, pp. 107-114.

Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D., Varga, D. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, pp. 2142-2147.

Vasiļjevs, A., Pedersen, B. S., De Smedt, K., Borin, L., Skadiņa, I. (2011). META-NORD: Baltic and Nordic Branch of the European Open Linguistic Infrastructure. In *Proceedings of the NODALIDA 2011 workshop on Visibility and Availability of LT Resources*, Riga, Latvia, pp. 18-22.

Vasiljevs, A., Rirdance, S. and Gornostay T. (2010). Reaching the User: Targeted Delivery of Federated Content in Multilingual Term Bank. In *Proceedings of the Terminology and Knowledge Engineering Conference "Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges" (TKE 2010)*, Dublin City University, Ireland.

Vasiljevs, A., Rirdance, S. and Liedskalnins. A. (2008). EuroTermBank: Towards Greater Interoperability of Dispersed Multilingual Terminology Data. In *Proceedings of the 1st International Conference on Global Interoperability for Language Resources (ICGL 2008)*, Hong Kong.

Vossen, P. (ed.). (1999). *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic Publishers, The Netherlands.