Sabine Kirchmeier-Andersen

Language technology for language institutions. What kind of technology do languages institutions use – what kind of resources can they provide?

Abstract (English)

In this paper I will examine the use of language technology (LT) in national institutions of language. After a short review of the typical tasks of language institutions and the basic aspects of language technology and its relevance for language institutions, I will discuss which types of technologies are currently put to use, and which types may be useful in solving the tasks of language institutions in the future. The empirical basis is the current language policy of the Nordic countries and a survey of language technology and language resources in the Nordic language councils in 2008. Finally, I will make some suggestions for how language technology and language institutions can profit from working together to meet the linguistic challenges that lie ahead.

Abstract (Danish)

I denne artikel undersøger jeg brugen af sprogteknologi i sproginstitutioner. Efter en kort gennemgang af de opgaver er typiske for sproginstitutioner, og af sprogteknologiens grundlæggende aspekter, diskuterer jeg hvilke slags teknologier der for øjeblikket bliver taget i anvendelse, og hvilke former for sprogteknologi der kunne være nyttige for sproginstitutioner i fremtiden. Det empiriske grundlag er den nuværende sprogpolitik i Norden og en undersøgelse af sprogteknologi og sprogresurser i de nordiske sprognævn som blev gennemført i 2008. Til sidst vil jeg fremsætte nogle forslag til hvordan sprogteknologi og sproginstitutioner kan drage nytte af at samarbejde for at blive bedre rustet til fremtidens sproglige udfordringer.

1. The tasks of language institutions

The central or national institutions of language in Europe are mainly concerned with research, documentation and policy making relating to the officially recognized standard languages within the states of the European Union. Whereas the tasks of European language institutions may vary considerably with regard to how much weight is put on language research, teaching, standardization and giving advice, Nordic Language institutions have quite a number of similar tasks. The core activities are to monitor language development that is changes in the corpus of the language mainly regarding lexis and grammar, to provide orthographic standards and to give advice on language use in private and public institutions. Most institutions also follow the development of the status of the language and provide information about the contexts in which the language is used in order to support the development of language policies. Other important activities are research, publishing and information to the public on language issues. Some institutions work on a strictly monolingual basis, whereas others, depending on the linguistic situation in their countries, provide services for multiple languages such as bilingual dictionaries or grammars. Although spoken language receives some attention, the main focus in most institutions is on the written language.

Monitoring language development means tracking the negotiation of norms in language communities and the development of neologisms. It is a huge task which traditionally has been carried out by hand – e.g. by reading through lots of carefully selected texts extracting quotes that are stored in large archives, or by recording and transcribing spoken language – the classical work of the philologist. Today, new technology is being put to use in various phases of this process: archives have become databases, quotes are extracted via a scanner or taken directly from the electronic version of the text from the internet and supplied with metadata to facilitate automatic search and further processing.

Even though language institutions have been collecting large amounts of texts and corpus linguistics and language technology play a more important role, the task of identifying a new word or phrase or a new sense of a word in most cases is still carried out manually. Today, people interact with written texts via sms, email, chat etc. much more than just 10-20 years ago, and the development of their languages has accelerated accordingly. Many language institutions would like to extend the language that is monitored to include a much larger variety of written and spoken texts, for instance from the increasing numbers of conversations in the social media, but do not have the resources to do so. The introduction of language technology might be a way to cope with this development.

2. How can language technology help?

Language technology is usually described as computer programs that work with written or spoken language as input or output, i.e. speech or text – "a kind of artificial device that is created to augment our abilities" (Sproat 2010). Various applications have been developed over time, e.g. spelling and grammar checkers, machine translation, information retrieval, computer assisted language learning, speech synthesis and speech recognition just to mention a few (cf. figure 1). Often language technology is integrated into other programs or applications such as robots, databases, user interfaces etc. There is an overlap with multimedia technology – integration with computer games, toys and teaching applications. There is also some overlap with artificial intelligence and knowledge technology, e.g. both use ontologies and taxonomies that describe relations in the world and facilitate automatic reasoning.

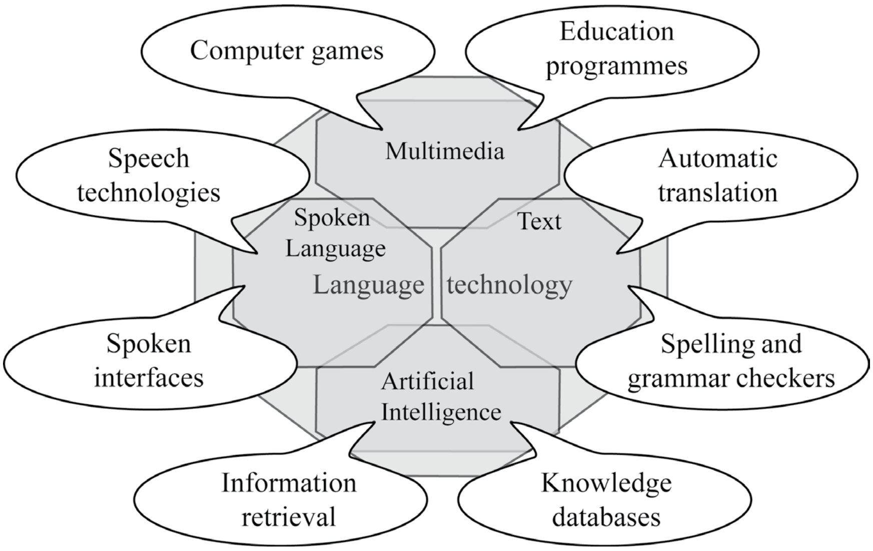


Fig. 1: Applications of language technology

Language technology today has two basic methods to perform its tasks: rule based and statistical. The rule based methods use formalized linguistic rules encoded in computational dictionaries and grammars that enable the computer to analyze or generate natural language in a number of well defined steps. If the systems are used for translation, there will also be translation rules defining the correspondences between two or more languages. Statistical systems on the other hand derive heuristic knowledge from a large collection of texts or transcribed speech which may be enriched with various types of annotations. The heuristics are then applied to new texts. Recent approaches tend to combine the two methods in order to further improve the performance of the system.

Regardless of the choice of method, LT-programs are composed of two elements: one or several software modules, such as databases, language analyzers, speech recognizers etc., and language data such as text corpora, speech data, dictionaries or formalized grammar rules (cf. figure 2).

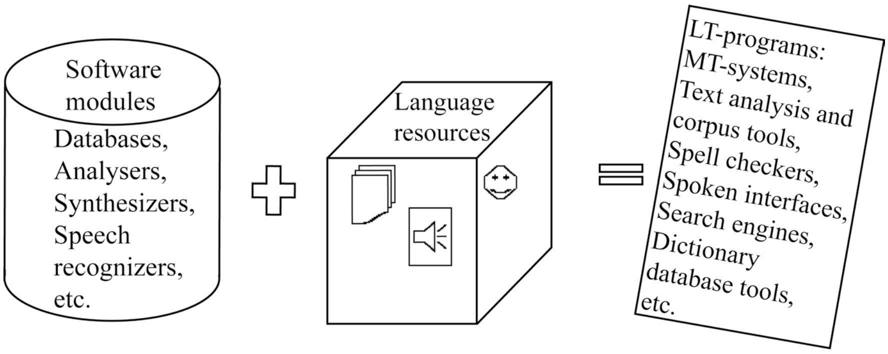
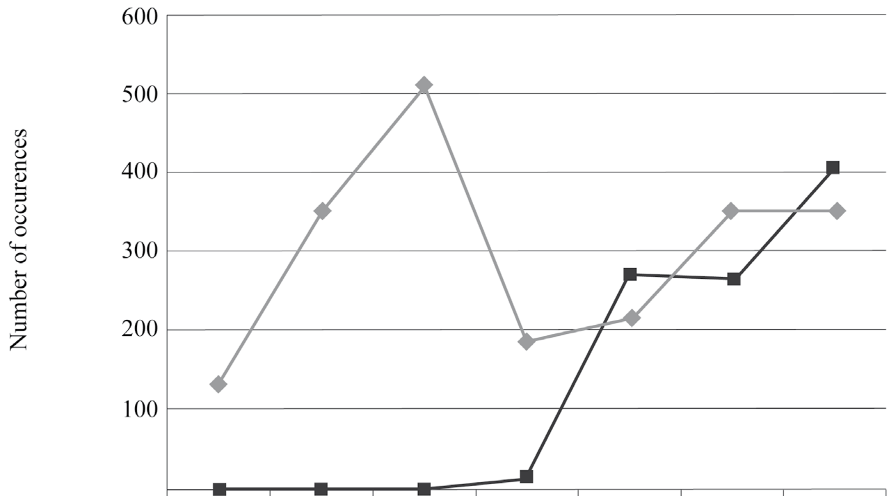

Software modules
Databases, Analysers, Synthesizers, Speech recognizers, etc.
+
Language resources
=
LT-programs:
MT-systems,
Text analysis and corpus tools,
Spell checkers,
Spoken interfaces,
Search engines,
Dictionary database tools,
etc.

Fig. 2: Basic components of language technology

In sum, language technology is about developing various software modules that can be applied to language data in order to create a number of LT-programs or tools that can aid us in performing various tasks where the knowledge of language is involved.

Even though this description of the basic methods is very brief, it is sufficient to reveal various points of interest for language institutions: first of all a common interest in dictionaries, grammars and large collections of linguistic data, i.e. recorded speech and texts. This can be characterized as mutual interest, i.e. both groups could profit from each other if information and data could be exchanged. Secondly, language institutions could profit in various ways from the automated methods and tools developed by language technology. This is probably best illustrated by some examples.

The first example concerns tracking the distribution of the use of competing similar or synonymous words or expressions in a language, in this case the English “exit poll” vs. the corresponding Danish expression “valgstedsmåling”. Going through a large newspaper corpus to produce a graph as the one in figure 3, showing the occurrences of the two words over time, is a time consuming task which can be done faster and easier through the use of language technology and at the same time giving valuable information when it comes to discussing lexical developments and language planning. Such graphs are quite informative and are well suited to illustrate linguistic observations to a large public audience.



	2001	2002	2005	2006	2007	2008	2009
◆ exit poll	133	352	508	188	214	352	352
■ valgstedsmåling	0	0	0	13	272	265	406

Fig. 3: The frequency over time of the words “exit poll” and “valgstedsmåling” in a corpus of Danish newspapers.

The second example concerns the collections of records of answers to questions from the public about language to the Nordic language councils. Norway, Sweden and Denmark keep track of the 8,000-10,000 questions that are received at their information desks every year, as this gives valuable information about the language problems that arise on a day to day basis. Not all questions are worth keeping. Out of the approx. 500,000 questions that have been answered by the Danish Language Council since 1955 around 10,000 have be recorded in a database for future reference because they concern new linguistic phenomena that have not previously been described. Many of these answers are also made available for the public on the internet.

The three Scandinavian languages are closely related. There is frequent contact between the information services, and the possibility of searching through the databases of questions to find answers to similar problems in the other languages has long been an important point on the agenda. Since there is considerable orthographic and lexical variation between the languages, and since questions often are connected to certain words or expressions, a simple text based search is not useful. Therefore a common taxonomy of linguistic terms was developed in order to enable thematic search across the languages. Answers on, for instance, morphology of nouns in relation to definiteness can thus be retrieved for all three languages at the same time, although the answers deal with completely different word forms. Advanced language technology is used to facilitate the development of a Nordic taxonomy of linguistic terms and to enable search across languages.

The third is a recent experiment at the Danish Language Council, the development of a new tool, the wordtrawler – a computer programme that automatically scans newspaper texts for neologisms (Halskov/Jarvad 2010). Each month the system collects and processes 20 million words of text and identifies new word strings by checking them against all dictionaries and wordlists that are available at the council, including the ones that were found the month before. This results in a list of approximately 30,000 potentially new words and expressions that are filtered automatically to form a list of candidates for manual inspection. The outcome is a list of 150-250 genuine new words in the general language, i.e. 1,500-3,000 new words per year. Some of these are new compounds, some are new acronyms and others are loan words or domesticated words. Although this procedure can detect many neologisms more efficiently than the human eye, it cannot completely replace the human inspection, especially when it comes to identifying changes in the use of already existing words, but still it is a great improvement.

European language institutions vary quite a lot with regard to the use of language technology in their core activities. Generally, tools that can facilitate the creation of dictionaries and the use of text corpora are in focus. Lately, we have seen an increasing interest in more sophisticated corpus tools such as, for instance, part of speech taggers and syntactic parsers.

Language institutions are not only potential users of language technology. As central players in the language debate of their countries, language institutions are concerned with the status of the language and with ensuring the access to high quality language technology tools such as word processors with spelling and grammar checkers, translations software, intelligent information retrieval etc. This will be further elaborated in the following sections.

3. Language technology in the Nordic countries

In the language institutions of the Nordic countries language technology has been an important concern for several years. The main language policy document of the Nordic Council of Ministers (Nordisk Ministerråd), *Declaration on a Nordic Language Policy,* which was adopted by the Nordic ministers in 2006, lays out a strategy for the relation between the global lingua franca, English, the Nordic languages essential to the state (Danish, Finnish, Icelandic, Norwegian, and Swedish), the Nordic languages which in addition to the languages of state are essential to other societies (Greenlandic, Faroese and the different varieties of Sami), languages which in some of the Nordic countries are recognized as official minority languages (Meänkieli (Tornedalian language)), the Kven language, different varieties of Romani, Yiddish, German, and the various Nordic sign languages) and, finally, the about 200 immigrant languages.

The declaration states as its main goals:

> – *that* all Nordic residents being able to read and write the language or languages that are essential to society in the area where they live
>
> – *that* all Nordic residents being able to communicate with one another, preferably in a Scandinavian language,
>
> – *that* all Nordic residents having a basic knowledge of linguistic rights in the Nordic countries and the language situation in the Nordic countries
>
> – *that* all Nordic residents having very good skills in at least one language of international importance and good skills in another foreign language
>
> – *that* all Nordic residents having a general knowledge of what language is and how it works.
>
> [...]
>
> These goals also require that all Nordic residents exhibit tolerance for variety and diversity in language, both between and within languages.
> (Declaration on a Nordic Language Policy 2006, official English version)

In order to define the relation between the various languages, the declaration operates with the term *parallel use of languages*, which in practice is used to refer to the parallelism between the Nordic languages on the one hand, and English on the other:

> The parallel use of languages refers to the concurrent use of several languages within one or more areas. None of the languages abolishes or replaces the other; they are used in parallel.
>
> Nordic residents, who internationally speaking have good English skills, have especially favorable conditions for developing skills in the parallel use of English and one or more of the languages of the Nordic countries in certain fields. A consistent policy to promote the parallel use of languages requires:
>
> – *that* it be possible to use both the languages of the Nordic countries essential to society and English as languages of science
>
> – *that* the presentation of scientific results in the languages of the Nordic countries essential to society be rewarded
>
> – *that* instruction in scientific technical language, especially in written form, be given in both English and the languages of the Nordic countries essential to society
>
> – *that* universities, colleges, and other scientific institutions can develop long-range strategies for the choice of language, the parallel use of languages, language instruction, and translation grants within their fields
>
> – *that* Nordic terminology bodies can continue to coordinate terminology in new fields
>
> – *that* business and labor-market organizations be urged to develop strategies for the parallel use of language.
>
> (Declaration on a Nordic Language Policy 2006, Goals 2.1)

As one of the means to fulfill its goals, the declaration points at the use of language technology and contains two important recommendations:

> 1. inter-Nordic dictionaries should be compiled in printed and electronic form
> 2. computer translation programs for the languages of the Nordic countries essential to society and programs for multilingual searches in Nordic databases should be developed
>
> (Declaration on a Nordic Language Policy 2006, Issue 1)

Although the declaration is not legally binding, all Nordic language councils are committed to fulfilling the goals of the declaration. The secretariat of the Council of Nordic Ministers has the task of following up with regular progress reports. In 2009 the Council of Nordic Ministers decided to establish a new body, Nordic Language Coordination, in order to intensify the follow up on the declaration and to strengthen the cooperation between the language councils and numerous other organizations working with researching and teaching Nordic languages.

Since 2005 the Network of the Nordic language councils has entrusted a special language technology group, ASTIN,[1] with the task of developing LT solutions for language institutions, stimulating research and development of language technology for the Nor-

[1] ASTIN: Arbejdsgruppen for sprogrøgt og sprogteknologi (Nordic task force for language and language technology) was initiated in 2005 by the Network of the Nordic Language Councils and currently consists of: Torbjørg Breivik (Language Council in Norway), Rickard Domeij (Language Council in Sweden), Per Langgård (Language Council of Greenland), Sjur Nøstlebo Moshagen (Sametinget/Sami Council), Jakob Halskov (Language Council in Denmark).

dic languages, ensuring that important language technology is adapted to the Nordic languages, and monitoring the consequences of the use of language technology for the development of each language. ASTIN has a special focus on facilitating the dialogue between language institutions, LT-researchers, LT-developers and policy makers through conferences, workshop and publications.

In 2008 ASTIN made a survey of the kind of data and tools that are already in use in the Nordic language councils. The survey asked for information on which software modules and language data were available, which LT-tools were most frequently used, and which tools the institutions would like to use in the near future.

All language institutions were asked to indicate whether they made use of software in the following categories:

Software

- Parsers (e.g. automatic morphological analysis, sentence analysis)
- Tools for information extraction
- Tools for automatic mark-up (e.g.part-of-speech, syntax, topics)
- Concordance programs (e.g. frequency counts and keyword in context)
- Tools for annotation (e.g. manual annotation of language data)
- Transcription tools (e.g. transcription of spoken data)
- Statistical tools
- Automatic analysis of document structure
- Spell checkers
- Grammar checkers
- Tools for automatic extraction of new words
- Databases with linguistic questions and answers
- Dictionary databases
- Tools to develop language teaching programs
- Other

Furthermore, institutions were asked to indicate whether they were using or had access to the following types of resources:

Language resources

- Digital lexica
    - Monolingual
    - Bilingual
    - Multilingual
- Corpora
    - Standards
    - Speech corpora
    - Text corpora
    - Multilingual corpora
    - Parallel multilingual corpora

- Terminology databases
- Thesauri/word nets
- Ontologies
- Digitalized records of questions and answers about languages
- Digital word collections
- Digital language training suites
- Other

In all cases, institutions were asked to indicate whether they had developed the indicated software or resources by themselves and/or were holding the copyright, or whether software and resources were bought or licensed from other public institutions or private vendors.

The answers to the questionnaires, which were given by Denmark, Iceland, Norway and Sweden, showed that the following software was most widely used: tools for information extraction, word databases, databases for linguistic questions and answers, spell checkers, statistics programs and concordance programs. Regarding language resources, the Nordic language institutions could report frequent use of the following: monolingual lexica, digitalized questions and answers about language problems, text corpora and terminology databases.

59% of the software was commercial, 41% produced by the language institutions themselves or freeware. 18% of the language resources were commercial products whereas 82% were produced by the language institutions or freeware.

The language councils were also asked what kind of tools and resources they envisage using in the future. The list below only shows the top priorities:

Software

- Automatic analysis of document structure
- Programs for automatic extraction of new words and new word senses
- Tools for automatic/semi-automatic mark-up/tagging
- Tools for language training
- Specialized analyzers and parsers
- Transcription tools for spoken text

Language resources

- Bilingual lexica
- Wordnets and ontologies
- Thesauruses
- Multilingual corpora
- Parallel multilingual corpora
- Common standards for corpora

From the example of the Nordic language councils we can conclude that:

- language institutions can make use of language technology tools
- language institutions expect to be using more language technology in the future
- language institutions have been and still are collecting large repositories of language resources
- language institutions are getting prepared to maneuver in a multilingual context

4. How can language technology developers and language institutions work together?

It has emerged from the previous sections that there are several areas of mutual interest between language technology and languages institutions. Here we shall focus only on four important areas: 1. development of language technology software for language institutions, 2. sharing language resources, and 3. cooperating in policy making for a language technology infrastructure, 4. cooperating on research to improve the adaptation of language technology products to language change.

Better language technology software can help language institutions to do their work more efficiently with text analysis tools to monitor language use, with databases and workbenches to facilitate the development of mono- and multilingual dictionaries, with language teaching programmes, with translation tools for all language pairs including minority languages, etc.

Language institutions can help to improve language technology applications by developing and sharing language resources such as text collections, dictionaries and other linguistic information in a common infrastructure such as CLARIN (www.clarin.eu), METANET (www.meta-net.eu). In this context it is absolutely imperative that the obstacles that exist due to the current legislation on intellectual property rights can be overcome – not in the sense that intellectual property rights should no longer be granted, but in the sense that an exception should be made for projects that aim at using the text as a collection of words to produce a linguistic tool in which the original text can no longer be reproduced and thus not be misused.

Language technology is also important for the status and the use of a language in every domain of a society (Crystal 2000). Language institutions have an obligation to ensure that language technology products are developed and continuously kept up to date. Cooperation in the area of policy development between language institutions and language technology could thus include

- convincing political decision makers to support the development of language technology,
- convincing political decision makers to make available language data for language technology,
- convincing political decision makers that it is important that there exist experts in language technology for their language.

Last but not least, a joint research effort should be made in order to improve the adaptation of language technology to language change. Computer programs are generally of static nature; once a dictionary or a grammar has been loaded or the system has been trained on available language data, there is hardly any adaptation to changes in the grammar and vocabulary of a given language. The figures from the word-trawler described in section 2 above showed that in general language between 2000 and 3000 words enter the language each year. We have not tried to estimate the growth in the vocabulary of a language if specialized domains were taken into account as well.

National institutions of language are experts on language change and would be able to provide substantial contributions in collaborative research projects with researchers on language technology.

5. Language technology for lesser used languages

Widely used languages such as English, Chinese, Spanish, French, German, Russian etc. constitute attractive markets for the big players of the IT-industry, and language technology based tools for these languages are made readily available and constantly being improved. With the development of social communication platforms and more interactive communication tools, the field for language technology has broadened immensely. In the coming years we will see more:

- Multilingual knowledge sharing using encyclopedia, knowledge bases and terminology databases.
- Automatic or semi-automatic translation on the web.
- Automatic interpretation through combinations of translation systems with speech recognition and speech synthesis.
- Quick access to knowledge through multilingual information retrieval combined with automatic summarization and translation.
- Linguistic services such as mono- and multilingual dictionaries and translation services on mobile platforms.
- Better support for the disabled such as language controlled ambient computing.
- More and more programs with spoken interfaces
- More intelligently personalized web services and social communication platforms.

However, very little of all this will be available for the less widely spoken languages unless political decision makers contribute substantially through public funded research and development.

It is generally acknowledged that the costs to produce high quality language technology for a given language are the same regardless of the number of speakers, and thus the smaller the number of speakers and potential customers, the less the return on investment. For less widely used languages, the lack of high quality language technology tools and advanced language resources is a disadvantage as the use of the language is no longer supported in all domains. Users of the language will eventually be inclined to use a more widely spoken language in his or her communication or search for information because the tools are better and make it easier and faster to reach a given goal.

Political decision makers who wish to maintain the status of, for instance, the state language and/or preserve and develop the linguistic diversity of the country, should be alert about this development and in due course develop research and development programs that support language technology for the languages of their countries.

Otherwise, if technology does not adapt to people and their language, people will adapt themselves and their language to technology. One striking example: in spite of the remarkable progress that has been made in many areas of IT, a fundamental problem such as the support of different character systems for different languages still has not found a satisfactory solution, and for Danish there are still IT-products, especially on the web, that cannot cope with the 3 national characters *æ*, *ø*, *å*. Danes are still forced to use ae, oe, aa for instance in email-addresses, not to mention the fact that basic functions such as alphabetization do not work and that in some applications the Danish characters are simply ignored or replaced by arbitrary symbols making information retrieval a rather arduous task. With explicit reference to the view that it was difficult to market Århus, one of the major Danish cities, in an international digital setting, the city in 2010 changed its name to Aarhus countering the latest major orthographic reform for Danish which took place more than 60 years ago.

Projects such as Euromatrix have already documented the biased situation for the less widely used languages. The matrix shows the available software and language resources for machine translation products for each country and for different language pairs. For English there is almost 8 times as much material (1,320) than there is available for Danish (181). For a language like Estonian the ratio 80 to 1.

	eng	fra	deu	spa	ita	por	nld	swe	ell	pol	dan	ces	fin	rom	hun	bul	slv	lav	lit	slk	est	mlt	gle
eng	1320	109	111	107	100	84	50	30	24	35	13	11	15	10	16	10	10	7	5	5	4	3	2
fra	109	847	79	66	52	41	36	19	22	14	11	9	9	8	8	7	8	7	5	5	4	3	2
deu	111	79	720	42	38	26	20	18	14	18	12	10	10	9	10	7	8	7	5	5	4	3	2
spa	105	65	40	650	35	29	19	17	14	11	12	9	9	8	8	7	8	5	5	5	4	3	2
ita	100	52	38	36	599	25	19	16	14	11	11	9	9	8	8	7	8	5	5	5	4	3	2
por	85	41	25	29	25	497	18	15	13	11	11	9	9	8	8	6	8	5	5	5	4	3	2
nld	49	36	20	19	19	18	376	16	14	10	11	9	9	8	8	6	8	5	5	5	4	3	2
swe	30	17	18	17	16	15	16	272	13	8	12	9	10	8	8	6	8	5	5	5	4	3	2
ell	23	22	14	14	14	13	14	13	267	7	9	7	8	7	6	7	6	5	5	3	3	3	2
pol	35	14	18	11	11	11	10	8	7	250	7	9	8	7	7	6	7	7	5	4	3	3	2
dan	13	11	13	12	11	10	11	12	9	7	181	7	8	7	7	6	7	5	3	4	4	3	2
ces	10	9	9	9	9	9	9	9	7	9	7	168	9	8	8	7	7	6	5	4	3	3	2
fin	16	9	10	9	9	9	9	10	8	8	8	9	157	7	7	6	7	5	5	4	3	3	2
rom	10	8	9	9	8	8	8	8	7	7	7	8	7	144	7	7	7	4	4	4	3	2	1
hun	15	8	9	8	8	8	8	8	6	7	7	8	7	7	129	5	8	5	5	5	4	3	2
bul	9	7	7	7	7	6	6	6	7	6	6	7	6	7	5	151	5	4	4	2	2	2	1
slv	10	8	8	8	8	8	8	8	6	7	7	7	7	7	8	5	112	5	5	5	4	3	2
lav	7	7	7	5	5	5	5	5	5	7	5	6	5	4	5	4	5	93	5	3	3	3	2
lit	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	4	5	5	75	3	3	3	2
slk	5	5	5	5	5	5	5	5	3	4	4	4	4	4	5	2	5	3	3	5	4	3	2
est	4	4	4	4	4	4	4	4	3	3	4	3	3	3	4	2	4	3	3	4	4	3	2
mlt	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	3	3	3	3	3	2
gle	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2

Fig. 4: Euromatrix for MT – systems and corpora

6. Conclusions

Language technology and language institutions are becoming more closely linked and for good reasons: both can gain a lot from working together by sharing software and language resources. Language technology is important for keeping languages alive, relevant and useful in all domains of society in our digital and global age. For their own sake and for the sake of their languages, language institutions should continuously promote the development of all aspects of language technology, and engage in sharing their unique knowledge about language and language change.

To give an impression of the task at hand we can take a look at a recent experiment at the Danish Language Council, the development of the a new tool, the wordtrawler – a computer programme that automatically scans newspaper texts for neologisms (Halskov/Jarvad 2010). Each month the systems collects and processes 20 million words of text and identifies new word strings by checking them against all dictionaries and wordlists that are available at the council, including the ones that were found the month before. This results in a list of approximately 30,000 potentially new words and expressions that are filtered automatically to form a list of candidates for manual inspection. The outcome is a list of 150-250 genuine new words in the general language, i.e. 1,500-3,000 new words per year. Some of these are new compounds, some new acronyms, others are loan words or domesticated words. Although this procedure more efficiently than the human eye can detect many neologism, it cannot completely replace the human inspection, especially when it comes to identify changes in the use of already existing words.

7. References

Crystal, D. (2000): *Language death*. Cambridge et al.: Cambridge University Press.

Declaration on a Nordic Language Policy (2007). Copenhagen: Nordic Council of Ministers.

Halskov, J./Jarvad, P. (2010): Automated extraction of neologisms for lexicography. In: Granger, S./Paquot, M. (eds.): *eLexicography in the 21st Century: new challenges, new applications. Proceedings of eLex 2009. Cahiers du CENTAL*. Louvain: Presses universitaires de Louvain.

Nordic Council: Internet: www.norden.org/en/nordic-council.

Sproat, R. (2010): *Language, technology, and society*. Oxford: Oxford University Press.