

# An Update and Extension of the META-NET Study “Europe’s Languages in the Digital Age”

Georg Rehm<sup>1</sup>, Hans Uszkoreit<sup>1</sup>, Ido Dagan<sup>2</sup>, Vartkes Goetcheerian<sup>3</sup>,  
Mehmet Ugur Dogan<sup>4</sup>, Coskun Mermer<sup>4</sup>, Tamás Varadi<sup>5</sup>, Sabine Kirchmeier-Andersen<sup>6</sup>,  
Gerhard Stickel<sup>7</sup>, Meirion Prys Jones<sup>8</sup>, Stefan Oeter<sup>9</sup>, Sigve Gramstad<sup>10</sup>

META-NET  
DFKI GmbH  
Berlin, Germany<sup>1</sup>

EFNIL, META-NET  
Hungarian Academy of Sciences  
Budapest, Hungary<sup>5</sup>

Council of Europe, Com. of Experts  
University of Hamburg  
Hamburg, Germany<sup>9</sup>

META-NET  
Bar-Ilan University  
Tel Aviv, Israel<sup>2</sup>

EFNIL, META-NET  
Danish Language Council  
Copenhagen, Denmark<sup>6</sup>

Council of Europe, Com. of Experts  
Bergen, Norway<sup>10</sup>

META-NET  
Arax Ltd.  
Luxembourg<sup>3</sup>

EFNIL  
Institut für Deutsche Sprache  
Mannheim, Germany<sup>7</sup>

META-NET  
Tübitak Bilgem  
Gebze, Turkey<sup>4</sup>

NPLD  
Network to Promote Ling. Diversity  
Cardiff, Wales<sup>8</sup>

## Abstract

This paper extends and updates the cross-language comparison of LT support for 30 European languages as published in the META-NET Language White Paper Series. The updated comparison confirms the original results and paints an alarming picture: it demonstrates that there are even more dramatic differences in LT support between the European languages.

**Keywords:** LR National/International Projects, Infrastructural/Policy Issues, Multilinguality, Machine Translation

## 1. Introduction and Overview

The multilingual setup of our European society imposes societal challenges on political, economic and social integration and inclusion, especially in the creation of the single digital market and unified information space targeted by the Digital Agenda (EC, 2010). Language technology is the missing piece of the puzzle, it is the key enabler and solution to boosting growth and strengthening Europe’s competitiveness.

Recognising Europe’s exceptional demand and opportunities, 60 leading research centres in 34 European countries joined forces in META-NET, a Network of Excellence dedicated to the technological foundations of a multilingual European information society. META-NET was partially supported through four projects funded by the EC: T4ME, CESAR, METANET4U and META-NORD. META-NET is forging the Multilingual Europe Technology Alliance (META) with more than 760 organisations and experts representing multiple stakeholders and signed collaboration agreements with more than 40 other projects and initiatives. META-NET’s goal is monolingual, crosslingual and multilingual technology support for all European languages (Rehm and Uszkoreit, 2013). We recommend focusing on three priority research themes connected to application scenarios that will provide European R&D with the ability to compete with other markets and achieve benefits for European society and citizens as well as opportunities for our economy and future growth.

This paper extends and updates one important result of the work carried out within the META-VISION pillar of the initiative, the cross-language comparison of LT support for 30 European languages as published in the META-NET Language White Paper Series (Rehm and Uszkoreit, 2012).

## 2. The Language White Paper Series

Answering the question on the current state of a whole R&D field is difficult and complex. For LT nobody had collected these indicators and provided comparable reports for a substantial number of European languages yet. To arrive at a first comprehensive answer, META-NET prepared the Language White Paper Series “Europe’s Languages in the Digital Age” (Rehm and Uszkoreit, 2012) that describes the current state of LT support for 30 European languages (including all 24 official EU languages). This undertaking had been in preparation with more than 200 experts since mid 2010 and was published in the summer of 2012. The study included a comparison of the support all languages receive in four areas: MT, speech, text analytics, language resources. The differences in technology support between the various languages and areas are dramatic and alarming. In the four areas, English is ahead of the other languages but even support for English is far from being perfect. While there are good quality software and resources available for a few larger languages and application areas, others, usually smaller languages, have substantial gaps. Many languages lack basic technologies for text

analytics and essential resources. Others have basic resources but semantic methods are still far away.

The original study was limited to 30 languages (most of them official and several regional languages). These were, in essence, the languages represented by the membership of META-NET at the time of preparing the study. Since then, META-NET has grown and added members in countries such as Israel and Turkey. When we presented pre-prints of the series at LREC 2012 in Istanbul (also elsewhere), volunteers approached us and explained their interest to prepare white papers on additional languages. The first new white paper, reporting on Welsh, has recently been published (Evas, 2014).

The series is available at <http://www.meta-net.eu>. Here, we also present the press release “At least 21 European Languages in Danger of Digital Extinction”, circulated on the European Day of Languages 2012 (Sept. 26). It generated more than 600 mentions internationally (newspapers, blogs, radio and television interviews etc.). This shows that Europe is very passionate and concerned about its languages and that it is also very interested in the idea of establishing a solid LT base for overcoming language barriers.

In 2010, META-NET initiated a collaboration with the European Federation of National Institutions for Language (EFNIL) and started presenting its goals at the annual EFNIL conferences. Along the same lines, META-NET approached the Network to Promote Linguistic Diversity (NPLD) and, in 2013, the Council of Europe’s Committee of Experts that is responsible for the Charter on Regional and Minority Languages. Representatives of the three organisations were invited to a panel discussion at META-FORUM 2013 (Berlin, Germany, September 19/20) where it was agreed to intensify the collaboration between all organisations.

### 3. Language Communities

In addition to the update of the cross-language comparison, this paper extends the co-authorship and support of the META-NET study by three organisations representing the language communities.

#### 3.1. EFNIL

Formed in 2003, the European Federation of National Institutions for Language has institutional members from 30 countries whose role includes monitoring the official language(s) of their country, advising on language use or developing language policy. It provides a forum for these institutions to exchange information about their work and to gather and publish information about language use and policy within the EU. EFNIL encourages the study of the official EU languages and a coordinated approach towards mother-tongue and foreign-language learning, as a means of promoting linguistic and cultural diversity within the EU.

There is an increasing awareness among EFNIL members of the relevance and importance of LT on several counts. First, as a vital component and indeed a requirement for the sustainability of their respective national languages in the digital age. Second, as a research and productivity tool that has increasing impact on their daily work. Third, EFNIL members, many representing the central academic institutions for their language, can contribute to the technology support for their language through the invaluable language resources they develop. As a modest homegrown effort, EFNIL is running a pilot project (EFNILEX) aimed at developing LT support for the production of bilingual dictionaries between language pairs which are considered by mainstream publishing houses as commercially unviable.

#### 3.2. NPLD

The Network to Promote Linguistic Diversity is a pan-European network which works with constitutional, regional and smaller state languages. It has 35 members, 10 of these being either member state or regional governments and the others major NGOs who have a role or are interested in language planning and management. NPLD was established in 2007 and has already asserted itself as the main voice of those linguistic communities that are not the official languages of the EU. NPLD’s formation is a reflection of the growing interest in lesser used languages in Europe. Many governments from across the continent have established departments charged with the specific task of revitalizing and promoting the use of these languages. Many of these governments are represented within NPLD.

NPLD has two main goals. The first is to take advantage of the growth in knowledge and expertise which is now available in the area of language regeneration by ensuring that it is shared. This is done mainly through meetings and seminars, and is in the process of being further developed through the expansion of a digital library on language planning for its members. The second goal concerns the issue of policy development at a European level. Although much is said by the European Institutions about the importance of linguistic diversity, very few policy initiatives are undertaken and less funding is provided to support European linguistic diversity. We aim to highlight this deficiency and to promote the need for more support for all indigenous languages of Europe to ensure that our rich landscape of languages, many of them highly endangered, survive into the future.

ICT and social media will play a vital role in the future survival of most, if not all of the languages of Europe. Working together on a European stage to develop technical resources in areas such as translation and voice recognition will be vital if we are to avoid the digital extinction of many of our languages.

### 3.3. Council of Europe Committee of Experts on the Language Charter

The European Charter for Regional or Minority languages is a treaty of the Council of Europe with the purpose to protect and promote the regional and minority languages used in Europe. The two main political goals are the preservation of Europe's cultural heritage and diversity, and the promotion of democracy. The historic cultural and linguistic diversity in Europe is an integral part of European identity, and policies that acknowledge and promote this diversity also facilitate intercultural exchange and the participation in democratic processes. 33 European states have signed the treaty, and 25 states of those have ratified. The Languages Charter is applied to more than 190 regional or minority languages (or language situations), with around 40 million users. Most of these languages are small, less than 50,000 users. Only a handful are spoken by more than a million.

There are three main regional or minority language (RML) situations: 1. A RML in one country is a majority language in another country (as German, Ukrainian and Hungarian); 2. A RML is a minority language in more than one country (as Basque, Romani and Sami); 3. A RML is only found in one country (as Galician, Sorbian and Welsh). The content provisions are found in two parts of the Charter. Part II sets out that the state party shall base its policies, legislation and practise on certain objectives and principles. They cover the acknowledgement of the RML as an integral part of the state's cultural wealth, securing the language area, the use of the RML in public and private life, education, also regarding non-speakers, the elimination of unjustified discrimination, raising awareness and tolerance among the majority population. Part III contains concrete undertakings a state may apply to specific languages in the areas where the languages are in traditional use. Topics covered in Part III are education, judicial authorities, administrative authorities and public services, the media, cultural activities and facilities, and economic and social life. A Committee of Experts (Comex) monitors how the states comply with their obligations under the Charter. The monitoring is primarily based on three-yearly, national reports, visits to the country and information from NGOs.

LT may serve as a vehicle for the protection and promotion also of RML. At present, LT is primarily used in relation to national and large regional languages, partly due to the investment required. However, from the perspective of the Language Charter: To preserve the historical cultural and linguistic diversity of Europe and to facilitate an active participation of all European citizens in our democratic processes, it is also important for the smaller languages in Europe to make use of LT. The challenge to all of us, governments, research, the

industry and RML users, is therefore to identify which tools are the most important ones. The development of tools that will serve the needs of these languages, and to make them available in practice, both from an economic and user-friendly perspective, is the task ahead of us.

## 4. The Set of Languages

The original set covered by the META-NET White Paper Series comprised 30 languages (see table 1). Back then, several of the languages represented by research centres that are members in META-NET could not be addressed because due to a lack of funding for those members (e. g., Hebrew, Luxembourgish). Multiple regional and minority languages could not be taken into account because META-NET's focus were the official EU languages and the official national languages of all partners of the four funded projects.

The extended set of languages addressed in this paper now finally contains *all* official languages represented by META-NET and also by EFNIL. It also contains all regional and minority languages represented by NPLD and many of the languages monitored by Council of Europe's Committee of Experts on Regional and Minority Languages. About 40 of the languages that fall under the mandate of the Committee of Experts were excluded to keep this extension and update of the cross-language comparison manageable. We excluded languages which were not listed in (Ethnologue, 2013), which had less than 100,000 speakers (according to Ethnologue) and also all languages which did not originate in Europe.

## 5. Cross-Language Comparison

As already reported in the White Paper Series (Rehm and Uszkoreit, 2012), the current state of LT support varies considerably from one language community to another. In the following, we briefly recapitulate how the original cross-language comparison was prepared. In order to compare the situation between languages, we selected two sample application areas (machine translation, speech), one underlying technology (text analytics), and the area of basic language resources. Languages were categorised using a five-point scale: 1. Excellent support; 2. Good support; 3. Moderate support; 4. Fragmentary support; 5. Weak or no support. For the original 30 languages, LT support was measured according to the following criteria:

**MT:** Quality of existing MT technologies, number of language pairs covered, coverage of linguistic phenomena and domains, quality and size of existing parallel corpora, amount and variety of available applications.

**Speech:** Quality of existing speech recognition technologies, quality of existing speech synthesis technologies, coverage of domains, number and size of existing speech corpora, amount and variety of available speech-based applications.

Language	Speakers	White Paper
1. <b>Albanian</b>	7,436,990	
2. <b>Asturian</b>	110,000	
3. <b>Basque</b>	657,872	(Hernández et al., 2012)
4. <b>Bosnian</b>	2,216,000	
5. <b>Breton</b>	225,000	
6. <b>Bulgarian</b>	6,795,150	(Blagoeva et al., 2012)
7. <b>Catalan</b>	7,220,420	(Moreno et al., 2012)
8. <b>Croatian</b>	5,533,890	(Tadić et al., 2012)
9. <b>Czech</b>	9,469,340	(Bojar et al., 2012)
10. <b>Danish</b>	5,592,490	(Pedersen et al., 2012)
11. <b>Dutch</b>	22,984,690	(Odijk, 2012)
12. <b>English</b>	334,800,758	(Ananiadou et al., 2012)
13. <b>Estonian</b>	1,078,400	(Liin et al., 2012)
14. <b>Finnish</b>	4,994,490	(Koskenniemi et al., 2012)
15. <b>French</b>	68,458,600	(Mariani et al., 2012)
16. <b>Frisian</b>	467,000	
17. <b>Friulian</b>	300,000	
18. <b>Galician</b>	3,185,000	(García-Mateo and Arza, 2012)
19. <b>German</b>	83,812,810	(Burchardt et al., 2012)
20. <b>Greek</b>	13,068,650	(Gavrilidou et al., 2012)
21. <b>Hebrew</b>	5,302,770	
22. <b>Hungarian</b>	12,319,330	(Simon et al., 2012)
23. <b>Icelandic</b>	243,840	(Rögnvaldsson et al., 2012)
24. <b>Irish</b>	106,210	(Judge et al., 2012)
25. <b>Italian</b>	61,068,677	(Calzolari et al., 2012)
26. <b>Latvian</b>	1,472,650	(Skadiņa et al., 2012)
27. <b>Limburgish</b>	1,300,000	
28. <b>Lithuanian</b>	3,130,970	(Vaišnien and Zabarskaitė, 2012)
29. <b>Luxembourgish</b>	320,710	
30. <b>Macedonian</b>	1,710,670	
31. <b>Maltese</b>	429,000	(Rosner and Joachimsen, 2012)
32. <b>Norwegian</b>	4,741,780	(Smedt et al., 2012a; Smedt et al., 2012b)
33. <b>Occitan</b>	2,048,310	
34. <b>Polish</b>	39,042,570	(Milkowski, 2012)
35. <b>Portuguese</b>	202,468,100	(Branco et al., 2012)
36. <b>Romanian</b>	23,623,890	(Trandabăţ et al., 2012)
37. <b>Romany</b>	3,017,920	
38. <b>Scots</b>	100,000	
39. <b>Serbian</b>	9,262,890	(Vitas et al., 2012)
40. <b>Slovak</b>	5,007,650	(Šimková et al., 2012)
41. <b>Slovene</b>	1,906,630	(Krek, 2012)
42. <b>Spanish</b>	405,638,110	(Melero et al., 2012)
43. <b>Swedish</b>	8,381,829	(Borin et al., 2012)
44. <b>Turkish</b>	50,733,420	
45. <b>Vlax Romani</b>	540,780	
46. <b>Welsh</b>	536,890	(Evas, 2014)
47. <b>Yiddish</b>	1,510,430	

Table 1: Languages included in the updated cross-language comparison (new languages in bold, number of world-wide speakers according to Ethnologue)

**Text Analytics:** Quality and coverage of existing text analytics technologies (morphology, syntax, semantics), coverage of linguistic phenomena and domains, amount and variety of available applications, quality and size of existing (annotated) text corpora, quality and coverage of existing lexical resources (e. g., WordNet) and grammars.

**Resources:** Quality and size of existing text corpora, speech corpora and parallel corpora, quality and coverage of existing lexical resources and grammars.

Figures 1, 2, 3 and 4 show that there are massive differences between the 47 languages surveyed. The four updated comparisons can be considered a solid first draft that the authors of this contribution agree upon. The updated tables have been circulated and discussed by the

organisations and communities involved in this article in order to arrive at a coherent result that all organisations and language communities are in agreement with.

## 6. Conclusions

In the original series of white papers, we provided the very first high-level comparison of LT support, taking into account 30 European languages. Even though more fine-grained analyses are needed, the first draft of the extended and updated comparison presented in this paper confirms the original results and paints an alarming picture: in its extended form, the comparison demonstrates that there are even more dramatic differences in LT support between the European languages, i. e., the technological gap keeps widening. While there are good-quality software and resources available for a few languages and application areas only, other (usually smaller) languages have substantial gaps. Many languages lack basic technologies for text analytics and essential resources. Others have a few basic tools and resources, but there is little chance of implementing semantic methods in the near future.

Back in September 2012, the original results were disseminated using a press release with the headline “At least 21 European languages in danger of digital extinction” (Rehm et al., 2014). The updated and extended comparison demonstrates, drastically, that the real number of digitally endangered languages is, in fact, significantly larger; also see (Soria and Mariani, 2013). Overcoming language borders through multilingual language technologies is one of our key goals. The comparison shows that, in our long term plans, we should focus even more on fostering technology development for smaller and/or less-resourced languages and also on language preservation through digital means. Research and technology transfer between the languages along with increased collaboration across languages must receive more attention.

One key problem in this regard is the following: the number of speakers of a certain language seems to correlate with the amount and quality of technologies available for that language. For companies there is simply no sustainable business case which is why they refrain from investing in the development of sophisticated language technologies for a language that is only spoken by a small or very small number of speakers. This is why regional, national and international organisations as well as funding agencies should team up in order to address this issue. META-NET suggests setting up and actively supporting a shared programme to develop at least basic resources and technologies for all European languages (Rehm and Uszkoreit, 2013).

Our results show that such a large-scale effort is needed to reach the ambitious goal of providing support for *all* European languages, for example, through high-quality

machine translation. The long term goal of META-NET is to enable the creation of high-quality LT for all languages. This depends on all stakeholders right across politics, research, business, and society uniting their efforts. The resulting technology will help transform barriers into bridges between Europe's languages and pave the way for political and economic unity through cultural diversity.

## Acknowledgments

META-NET was co-funded by FP7 and ICT-PSP of the European Commission through the contracts T4ME (grant agreement no.: 249 119), CESAR (no.: 271 022), METANET4U (no.: 270 893) and META-NORD (no.: 270 899). The work presented in this article would not have been possible without the dedication and commitment of the 60 member organisations of the META-NET network of excellence and the more than 200 authors of and contributors to the META-NET Language White Paper Series.

## 7. References

- Ananiadou, S., McNaught, J., and Thompson, P. (2012). *The English Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Blagoeva, D., Koeva, S., and Murdarov, V. (2012). *Българският език в дигиталната епоха – The Bulgarian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Bojar, O., Cinková, S., Hajič, J., Hladká, B., Kuboň, V., Mírovský, J., Panevová, J., Peterek, N., Spoustová, J., and Žabokrtský, Z. (2012). *Čeština v digitálním věku – The Czech Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Borin, L., Brandt, M.D., Edlund, J., Lindh, J., and Parkvall, M. (2012). *Svenska språket i den digitala tidsåldern – The Swedish Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Branco, A., Mendes, A., Pereira, S., Henriques, P., Pellegrini, T., Meinedo, H., Trancoso, I., Quaresma, P., de Lima, V.L. Strube, and Bacelar, F. (2012). *A língua portuguesa na era digital – The Portuguese Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Burchardt, A., Egg, M., Eichler, K., Krenn, B., Kreutel, J., Leßmöllmann, A., Rehm, G., Stede, M., Uszkoreit, H., and Volk, M. (2012). *Die Deutsche Sprache im digitalen Zeitalter – German in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Calzolari, N., Magnini, B., Soria, C., and Speranza, M. (2012). *La Lingua Italiana nell'Era Digitale – The Italian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- EC. (2010). A Digital Agenda for Europe. European Commission. [http://ec.europa.eu/information\\_society/digital-agenda/publications/](http://ec.europa.eu/information_society/digital-agenda/publications/).
- Ethnologue. (2013). Ethnologue – Languages of the World. <http://www.ethnologue.com>.
- Evas, J. (2014). *Y Gymraeg yn yr Oes Ddigidol – The Welsh Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- García-Mateo, C. and Arza, M. (2012). *O idioma galego na era dixital – The Galician Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Gavrilidou, M., Koutsombogera, M., Patrikakos, A., and Piperidis, S. (2012). *H Ελληνική Γλώσσα στην Ψηφιακή Εποχή – The Greek Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Hernáez, I., Navas, E., Odriozola, I., Sarasola, K., de Ilaraza, A. Diaz, Leturia, I., de Lezana, A. Diaz, Oihartzabal, B., and Salaberria, J. (2012). *Euskara Aro Digitalean – Basque in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Judge, J., Chasaide, A. Ní, Dhubhda, R. Ní, Scannell, K.P., and Dhonnchadha, E. Uí. (2012). *An Ghaeilge sa Ré Dhigiteach – The Irish Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Koskenniemi, K., Lindén, K., Carlson, L., Vainio, M., Arppe, A., Lennes, M., Westerlund, H., Hyvärinen, M., Bartis, I., Nuolijärvi, P., and Piehl, A. (2012). *Suomen kieli digitaalisella aikakaudella – The Finnish Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Krek, S. (2012). *Slovenski jezik v digitalni dobi – The Slovene Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Liin, K., Muischnek, K., Müürisep, K., and Vider, K. (2012). *Eesti keel digiajastul – The Estonian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Mariani, J., Paroubek, P., Francopoulo, G., Max, A., Yvon, F., and Zweigenbaum, P. (2012). *La langue française à l'Ère du numérique – The French Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Melero, M., Badia, T., and Moreno, A. (2012). *La lengua española en la era digital – The Spanish Language in the*

- Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Miłkowski, M. (2012). *Język polski w erze cyfrowej – The Polish Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Moreno, A., Bel, N., Revilla, E., Garcia, E., and Vallverdú, S. (2012). *La llengua catalana a l'era digital – The Catalan Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Odiijk, J. (2012). *Het Nederlands in het Digitale Tijdperk – The Dutch Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Pedersen, B. Sandford, Wedekind, J., Bøhm-Andersen, S., Henriksen, P. Juel, Hoffensetz-Andersen, S., Kirchmeier-Andersen, S., Kjærum, J.O., Larsen, L. Bie, Maegaard, B., Nimb, S., Rasmussen, J.-E., Revsbech, P., and Thomsen, H. Erdman. (2012). *Det danske sprog i den digitale tidsalder – The Danish Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Rehm, G. and Uszkoreit, H., editors. (2012). *META-NET White Paper Series: Europe's Languages in the Digital Age*. Springer, Heidelberg, New York, Dordrecht, London. 31 volumes on 30 European languages. <http://www.meta-net.eu/whitepapers>.
- Rehm, G. and Uszkoreit, H., editors. (2013). *The META-NET Strategic Research Agenda for Multilingual Europe 2020*. Springer, Heidelberg, New York, Dordrecht, London. <http://www.meta-net.eu/sra>.
- Rehm, G., Uszkoreit, H., Ananiadou, S., Bel, N., Bielevičienė, A., Borin, L., Branco, A., Budin, G., Calzolari, N., Daelemans, W., Garabík, R., Grobelnik, M., García-Mateo, C., van Genabith, J., Hajič, J., Hernáez, I., Judge, J., Koeva, S., Krek, S., Krstev, C., Lindén, K., Magnini, B., Mariani, J., McNaught, J., Melero, M., Monachini, M., Moreno, A., Odiijk, J., Ogródniczuk, M., Pęzik, P., Piperidis, S., Przepiórkowski, A., Rögnvaldsson, E., Rosner, M., Pedersen, B. Sandford, Skadiņa, I., Smedt, K. De, Tadić, M., Thompson, P., Tufiş, D., Váradi, T., Vasiljevs, A., Vider, K., and Zabarskaite, J. (2014). The Strategic Impact of META-NET on the Regional, National and International Level. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014)*, Reykjavik, Iceland, May.
- Rosner, M. and Joachimsen, J. (2012). *Il-Lingwa Maltija Fl-Era Digitali – The Maltese Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Rögnvaldsson, E., Jóhannsdóttir, K.M., Helgadóttir, S., and Steingrímsson, S. (2012). *Íslensk tunga á stafrænni öld – The Icelandic Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Simon, E., Lendvai, P., Németh, G., Olaszy, G., and Vicsi, K. (2012). *A magyar nyelv a digitális korban – The Hungarian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Skadiņa, I., Veisbergs, A., Vasiljevs, A., Gornostaja, T., Keiša, I., and Rudzīte, A. (2012). *Latviešu valoda digitālajā laikmetā – The Latvian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Smedt, K. De, Lyse, G. Inger, Gjesdal, A. Müller, and Losnegaard, G.S. (2012a). *Norsk i den digitale tidsalderen (bokmålsversjon) – The Norwegian Language in the Digital Age (Bokmål Version)*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Smedt, K. De, Lyse, G. Inger, Gjesdal, A. Müller, and Losnegaard, G.S. (2012b). *Norsk i den digitale tidsalderen (nynorskversjon) – The Norwegian Language in the Digital Age (Nynorsk Version)*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Soria, C. and Mariani, J. (2013). Searching LTs for Minority Languages. In *Proceedings of TALN-RECITAL 2013*, pages 235–247.
- Tadić, M., Brozović-Rončević, D., and Kapetanović, A. (2012). *Hrvatski Jezik u Digitalnom Dobu – The Croatian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Trandabăt, D., Irimia, E., Mititelu, V. Barbu, Cristea, D., and Tufiş, D. (2012). *Limba română în era digitală – The Romanian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Vaišnien, D. and Zabarskaitė, J. (2012). *Lietuvių kalba skaitmeniniame amžiuje – The Lithuanian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Vitas, D., Popović, L., Krstev, C., Obradović, I., Pavlović-Lažetić, G., and Stanojević, M. (2012). *Српски језик у дигиталном добу – The Serbian Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.
- Šimková, M., Garabík, R., Gajdošová, K., Laclavík, M., Ondrejovič, S., Juhár, J., Genčí, J., Furdík, K., Ivoríková, H., and Ivanecký, J. (2012). *Slovenský jazyk v digitálnom veku – The Slovak Language in the Digital Age*. META-NET White Paper Series, Rehm, G. and Uszkoreit, H. (eds.). Springer, Heidelberg, New York, Dordrecht, London.

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	French Spanish	Catalan Dutch German Hungarian Italian Polish Romanian	<b>Albanian</b> <b>Asturian</b> Basque <b>Bosnian</b> <b>Breton</b> Bulgarian Croatian Czech Danish Estonian Finnish <b>Frisian</b> <b>Friulian</b> Galician Greek <b>Hebrew</b> Icelandic Irish Latvian <b>Limburgish</b> Lithuanian <b>Luxembourgish</b> <b>Macedonian</b> Maltese Norwegian <b>Occitan</b> Portuguese <b>Romany</b> <b>Scots</b> Serbian Slovak Slovene Swedish <b>Turkish</b> <b>Vlax Romani</b> Welsh <b>Yiddish</b>

Figure 1: Machine translation – state of language technology support for 47 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Czech Dutch Finnish French German Italian Portuguese Spanish	Basque Bulgarian Catalan Danish Estonian Galician Greek Hungarian Irish Norwegian Polish Serbian Slovak Slovene Swedish	<b>Albanian</b> <b>Asturian</b> <b>Bosnian</b> <b>Breton</b> Croatian <b>Frisian</b> <b>Friulian</b> <b>Hebrew</b> Icelandic Latvian <b>Limburgish</b> Lithuanian <b>Luxembourgish</b> <b>Macedonian</b> Maltese <b>Occitan</b> Romanian <b>Romany</b> <b>Scots</b> <b>Turkish</b> <b>Vlax Romani</b> Welsh <b>Yiddish</b>

Figure 2: Speech processing – state of language technology support for 47 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Dutch French German Italian Spanish	Basque Bulgarian Catalan Czech Danish Finnish Galician Greek <b>Hebrew</b> Hungarian Norwegian Polish Portuguese Romanian Slovak Slovene Swedish	<b>Albanian</b> <b>Asturian</b> <b>Bosnian</b> <b>Breton</b> Croatian Estonian <b>Frisian</b> <b>Friulian</b> Icelandic Irish Latvian <b>Limburgish</b> Lithuanian <b>Luxembourgish</b> <b>Macedonian</b> Maltese <b>Occitan</b> <b>Romany</b> <b>Scots</b> Serbian <b>Turkish</b> <b>Vlax Romani</b> Welsh <b>Yiddish</b>

Figure 3: Text analytics – state of language technology support for 47 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Czech Dutch French German Hungarian Italian Polish Spanish Swedish	Basque Bulgarian Catalan Croatian Danish Estonian Finnish Galician Greek <b>Hebrew</b> Norwegian Portuguese Romanian Serbian Slovak Slovene	<b>Albanian</b> <b>Asturian</b> <b>Bosnian</b> <b>Breton</b> <b>Frisian</b> <b>Friulian</b> Icelandic Irish Latvian <b>Limburgish</b> Lithuanian <b>Luxembourgish</b> <b>Macedonian</b> Maltese <b>Occitan</b> <b>Romany</b> <b>Scots</b> <b>Turkish</b> <b>Vlax Romani</b> Welsh <b>Yiddish</b>

Figure 4: Speech and text resources – state of language technology support for 47 European languages