An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age"

Georg Rehm & Hans Uszkoreit (META-NET, DFKI GmbH, Berlin, Germany), **Ido Dagan** (META-NET, Bar-Ilan University, Tel Aviv, Israel), **Vartkes Goetcherian** (META-NET, Arax Ltd., Luxembourg), **Mehmet Ugur Dogan & Coskun Mermer** (META-NET, Tübitak Bilgem, Gebze, Turkey), **Tamás Varadi** (EFNIL, META-NET, Hungarian Academy of Sciences, Budapest, Hungary), **Sabine Kirchmeier-Andersen** (EFNIL, META-NET, Danish Language Council, Copenhagen, Denmark), **Gerhard Stickel** (EFNIL, Institut für Deutsche Sprache, Mannheim, Germany), **Meirion Prys Jones** (NPLD, Network to Promote Linguistic Diversity, Cardiff, Wales), **Stefan Oeter** (Council of Europe, Committee of Experts, University of Hamburg, Hamburg, Germany), **Sigve Gramstad** (Council of Europe, Committee of Experts, Bergen, Norway)

This poster 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 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.

Introduction: META-NET

- 60 leading research centres in 34 European countries dedicated to the technological foundations of a multilingual European information society.
- Multilingual Europe Technology Alliance (META), more than 760 organisations and experts representing multiple stakeholders as of May 2014.
- The goal is monolingual, crosslingual and multilingual technology support for all European languages.

The Language White Paper Series "Europe's Languages in the Digital Age"

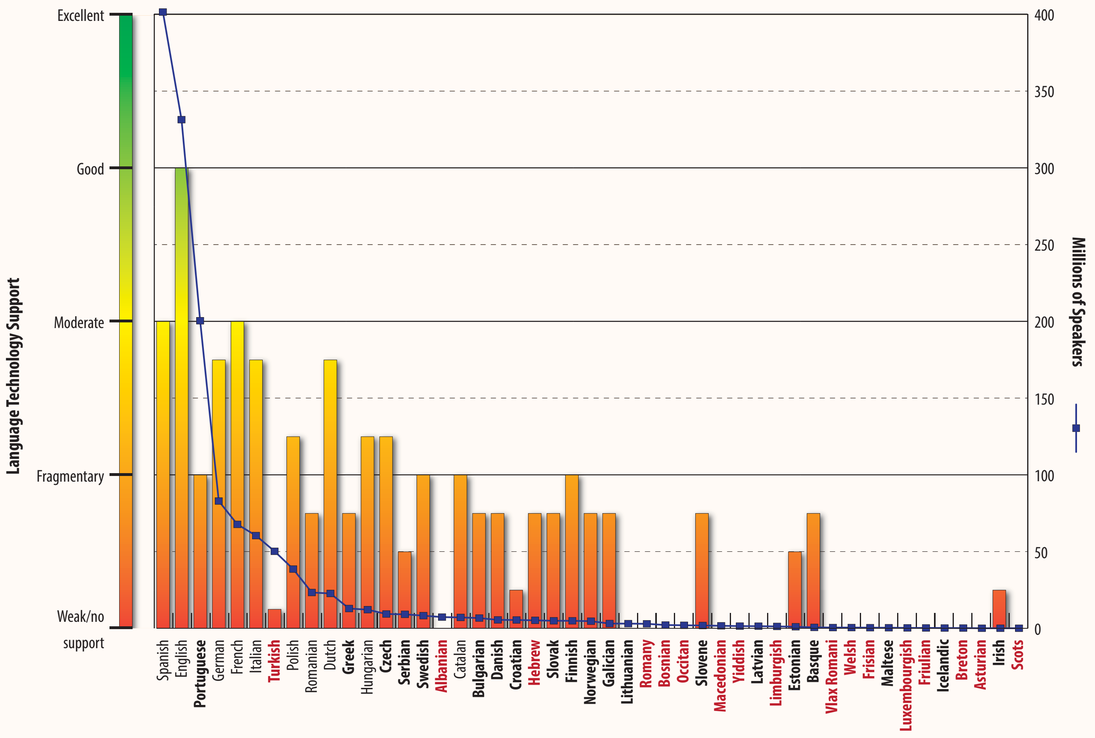
- Language White Paper Series covers 31 languages in 31 volumes, available online at http://www.meta-net.eu/whitepapers.
- Reports on the state of our languages in the digital age and the level of support through language technology.
- Press release "At least 21 European Languages in Danger of Digital Extinction," circulated on the European Day of Languages 2012 (Sept. 26).

The Set of Languages

- Original set comprised 30 languages; a few languages represented by META-NET members could not be addressed due to lack of funding.
- Multiple regional and minority languages could not be represented due to focus on official EU and national languages of the four META-NET projects' partners.
- This extended set now finally contains all languages represented by META-NET, EFNIL and NPLD as well as many of the languages monitored by the Council of Europe's Committe of Experts on Regional and Minority Languages.
- We excluded languages with less than 100,000 speakers (according to Ethnologue) and also all languages which did not originate in Europe.

Conclusions

- This updated comparison confirms the original results and paints an alarming picture: there are even more dramatic differences in LT support between the European languages, i.e., the technological gap keeps widening.
- We should focus even more on fostering technology development for smaller and/or less-resourced languages and also on digital language preservation.
- Research and technology transfer between the languages along with increased collaboration across languages must receive more attention. Regional, national and international organisations as well as funding agencies should team up!
- META-NET suggests setting up a shared programme to develop resources and technologies for all European languages (cf. META-NET SRA).



Overall language technology support levels and number of speakers for 47 European languages. Languages new to the cross-language comparison are in red. Languages in bold face have no weak or no support for machine translation.

Support for Machine Translation

Excellent	Good	Moderate	Fragmentary	Weak/none
	English	French Spanish	Catalan Dutch German Hungarian Italian Polish Romanian	**Albanian** **Asturian** Basque **Bosnian** **Breton** Bulgarian Croatian Czech Danish Estonian Finnish **Frisian** **Friulian** Galician Greek **Hebrew** Icelandic Irish Latvian **Limburgish** Lithuanian **Luxembourgish** **Macedonian** Maltese Norwegian **Occitan** Portuguese **Romany** **Scots** Serbian Slovak Slovene Swedish **Turkish** **Vlax Romani** **Welsh** **Yiddish**

Support for Speech Processing

Excellent	Good	Moderate	Fragmentary	Weak/none
	English	Czech Dutch Finnish French German Italian Portuguese Spanish	Basque Bulgarian Catalan Danish Estonian Galician Greek Hungarian Irish Norwegian Polish Serbian Slovak Slovene Swedish **Turkish**	**Albanian** **Asturian** **Bosnian** **Breton** Croatian **Frisian** **Friulian** **Hebrew** Icelandic Latvian **Limburgish** Lithuanian **Luxembourgish** **Macedonian** Maltese **Occitan** Romanian **Romany** **Scots** **Vlax Romani** **Welsh** **Yiddish**

Support for Text Analytics

Excellent	Good	Moderate	Fragmentary	Weak/none
	English	Dutch French German Italian Spanish **Hebrew**	Basque Bulgarian Catalan Czech Danish Finnish Galician Greek Hungarian Norwegian Polish Portuguese Romanian Slovak Slovene Swedish	**Albanian** **Asturian** **Bosnian** **Breton** Croatian Estonian **Frisian** **Friulian** Icelandic Irish Latvian **Limburgish** Lithuanian **Luxembourgish** **Macedonian** Maltese **Occitan** **Romany** **Scots** Serbian **Turkish** **Vlax Romani** **Welsh** **Yiddish**

Support for Speech and Text Resources

Excellent	Good	Moderate	Fragmentary	Weak/none
	English	Czech Dutch French German Hungarian Italian Polish Spanish Swedish	Basque Bulgarian Catalan Croatian Danish Estonian Finnish Galician Greek **Hebrew** Norwegian Portuguese Romanian Serbian Slovak Slovene	**Albanian** **Asturian** **Bosnian** **Breton** **Frisian** **Friulian** Icelandic Irish Latvian **Limburgish** Lithuanian **Luxembourgish** **Macedonian** Maltese **Occitan** **Romany** **Scots** **Turkish** **Vlax Romani** **Welsh** **Yiddish**

Set of Languages

	No. of Speakers	White Paper?
Albanian	7,436,990	
Asturian	110,000	
Basque	657,872	Yes
Bosnian	2,216,000	
Breton	225,000	
Bulgarian	6,795,150	Yes
Catalan	7,220,420	Yes
Croatian	5,533,890	Yes
Czech	9,469,340	Yes
Danish	5,592,490	Yes
Dutch	22,984,690	Yes
English	334,800,758	Yes
Estonian	1,078,400	Yes
Finnish	4,994,490	Yes
French	68,458,600	Yes
Frisian	467,000	

	No. of Speakers	White Paper?
Friulian	300,000	
Galician	3,185,000	Yes
German	83,812,810	Yes
Greek	13,068,650	Yes
Hebrew	5,302,770	
Hungarian	12,319,330	Yes
Icelandic	243,840	Yes
Irish	106,210	Yes
Italian	61,068,677	Yes
Latvian	1,472,650	Yes
Limburgish	1,300,000	
Lithuanian	3,130,970	Yes
Luxembourgish	320,710	
Macedonian	1,710,670	
Maltese	429,000	Yes
Norwegian	4,741,780	Yes

	No. of Speakers	White Paper?
Occitan	2,048,310	
Polish	39,042,570	Yes
Portuguese	202,468,100	Yes
Romanian	23,623,890	Yes
Romany	3,017,920	
Scots	100,000	
Serbian	9,262,890	Yes
Slovak	5,007,650	Yes
Slovene	1,906,630	Yes
Spanish	405,638,110	Yes
Swedish	8,381,829	Yes
Turkish	50,733,420	
Vlax Romani	540,780	
Welsh	536,890	Yes
Yiddish	1,510,430	

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

Strategic Partners

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.

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.

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.