The META-NET Language White Paper Series: Overview and Key Results

Bolette Sandford Pedersen

University of Copenhagen, Denmark
bspedersen@hum.ku.dk

META FORUM 2012

A Strategy for Multilingual Europe
Brussels, Belgium, June 20/21, 2012

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

Language White Papers

- Key communication instruments to address decision makers and journalists.
- Cover all EU languages (30 volumes).
- Report on the state of a language (general, social, strategic and technological aspects) and the level of support through language technology.
- Inform target group about societal and technological problems and challenges as well as economic opportunities.
- Printed documents published by Springer; will be distributed by META-NET.
- PDF versions available for free.

30 Languages Covered

META-NET

- Basque
- Bulgarian*
- Catalan
- Czech*
- Danish*
- Dutch*
- English*
- Estonian*
- Finnish*
- French*
- Galician
- German*
- Greek*
- Hungarian*
- Icelandic
- Irish*
- Italian*
- Latvian*
- Lithuanian*
- Maltese*
- Norwegian
- Polish*
- Portuguese*
- Romanian*
- Serbian
- Slovak*
- Slovene*
- Spanish*
- Swedish*
- Croatian

* = Official EU language

META NET

White Paper Series Séria bielych kníh

THE SLOVAK LANGUAGE IN THE DIGITAL AGE

SLOVENSKÝ JAZYK V DIGITÁLNOM VEKU

Mária Šimková
Radovan Garabík
Katarína Gajdošová
Michal Laclavík
Slavomír Ondrejovič
Jozef Juhár
Ján Genči
Karol Furdík
Helena Ivoríková
Jozef Ivanecký

Springer

OBSAH CONTENTS

SLOVENSKÝ JAZYK V DIGITÁLNOM VEKU

THE SLOVAK LANGUAGE IN THE DIGITAL AGE

A few Numbers ...

- >160 national experts contributed as authors or co-authors (ca. 5 per language on average).
- >50 additional experts have contributed data and information.
- >8,000 copies will be printed and distributed by META-NET.

Cross-Lingual Ranking

- Journalists and politicians need simple and clear messages.
- In four application areas, each language is assigned to one of five clusters, ranging from *excellent LT support* to *weak/no support*:
  1. Machine Translation
  2. Speech Processing
  3. Text Analysis
  4. Resources
- Results finalised at a meeting in Berlin with representatives of all 30 languages (October 21/22, 2011).

MT (top) & Speech Processing (bottom)

META-NET

excellent	good	moderate	fragmentary	weak or no support
	English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish

excellent	good	moderate	fragmentary	weak or 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	Croatian, Icelandic, Latvian, Lithuanian, Maltese, Romanian

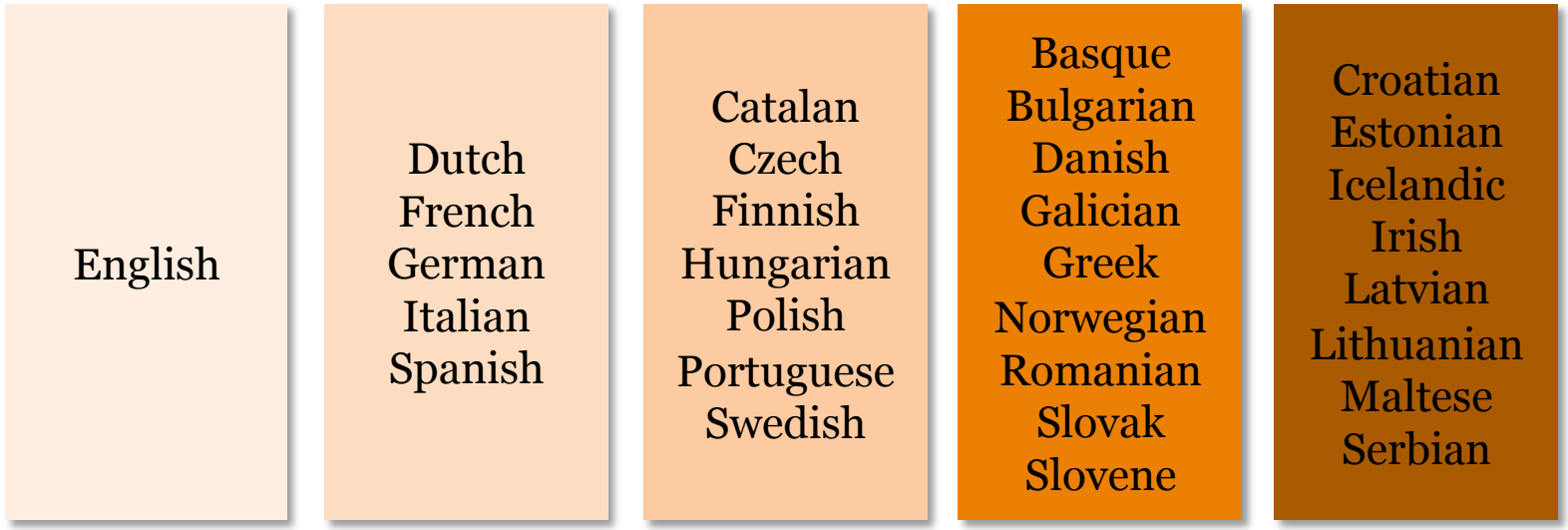
Text Analysis (top) & Resources (bottom)

META-NET

excellent	good	moderate	fragmentary	weak or no support
	English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catalan, Czech, Danish, Finnish, Galician, Greek, Hungarian, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish	Croatian, Estonian, Icelandic, Irish, Latvian, Lithuanian, Maltese, Serbian

excellent	good	moderate	fragmentary	weak/no support
	English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, Serbian, Slovak, Slovene	Icelandic, Irish, Latvian, Lithuanian, Maltese

Europe's Languages and LT



Key Results

- When it comes to Language Technology support, there are massive differences between Europe's languages and technology areas.
- Language Technology support for English is ahead of any other language – but far from being perfect.
- For 16 of 30 languages, Language Technology support is only *fragmentary*, *very weak* or *non-existent*!
- We now have a list of gaps and needs for all 30 languages which need to be addressed in the years to come.

Recent Developments

- Today we launch a new version of the Language White Paper website: http://www.meta-net.eu/whitepapers.
- Ca. 20 of the 30 Language White Papers available online in their final versions (PDF).
- First batch of printed copies (published by Springer) will be available soon.
- New volumes upcoming (such as, “European Sign Languages in the Digital Age”).
- Dissemination of printed volumes to politicians and journalists to start in July.

META-NET White Paper Series

Aims and Scope

META-NET, a Network of Excellence consisting of 57 research centres from 33 countries, is dedicated to building the technological foundations of a multilingual European information society.

META-NET is forging META, the Multilingual Europe Technology Alliance. The benefits offered by Language Technology differ from language to language. So do the actions that need to be taken within META-NET, depending on the factors such as the complexity of the respective language, the size of its community, and the existence of active research centres in this area.

The META-NET Language White Paper series "Languages in the European Information Society" reports on the state of each European language with respect to Language Technology and explains the most urgent risks and chances. The series will cover all official European languages and several other languages spoken in geographical Europe. While there have been a number of valuable and comprehensive scientific studies on certain aspects of languages and technology, there exists no generally understandable compendium that takes a stand by presenting the main findings and challenges for each language. The META-NET white paper series will fill this gap.

Languages (31)

Basque, Bulgarian, Catalan, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, Galician, German, Greek, Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian (bokmål), Norwegian (nynorsk), Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Spanish, Swedish.

Top Quotes

- Andrius Kubilius (Prime Minister of the Republic of Lithuania): *"Having preserved a close link with the old Indo-European parent languages, the Lithuanian language today satisfies the needs of the modern society perfectly well. However, active users of the Lithuanian language only amount to several million. Conserving it for future generations is a responsibility of the whole of the European Union. How we proceed with developing information technology will pretty much determine the future of the Lithuanian language."*
- Dr. Danilo Türk (President of the Republic of Slovenia): *"It is imperative that language technologies for Slovene are developed systematically if we want Slovene to flourish also in the future digital world."*
- Valdis Dombrovskis (Prime Minister of Latvia): *"Diversity of cultures, traditions and languages is one of the most important treasures of Europe and it is our duty to preserve this heritage for generations to come. For such small languages like Latvian keeping up with the ever increasing pace of time and technological development is crucial. The only way to ensure future existence of our language is to provide its users with equal opportunities as the users of larger languages enjoy. Therefore being on the forefront of modern technologies is our opportunity."*

People

See the authors and contributors (97) of all issues.

See the team behind the META-NET White Paper Series.

Thank you.

office@meta-net.eu

http://www.meta-net.eu
http://www.facebook.com/META.Alliance