

META=NET

An Overview of Multilingual LT Research in Europe

Josef van Genabith

Center for Next Generation Localisation
School of Computing, Dublin City University

META=FORUM 2011 Solutions for Multilingual Europe
Budapest, Hungary, June 27/28

eu 2011.hu



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).

Multilingual LT Research in Europe

- Overview of Multilingual Language Technologies in Europe
 - Not possible to do in a fair and equitable way in 15 minutes
 - Even though we have data from META:
 - Language White Papers

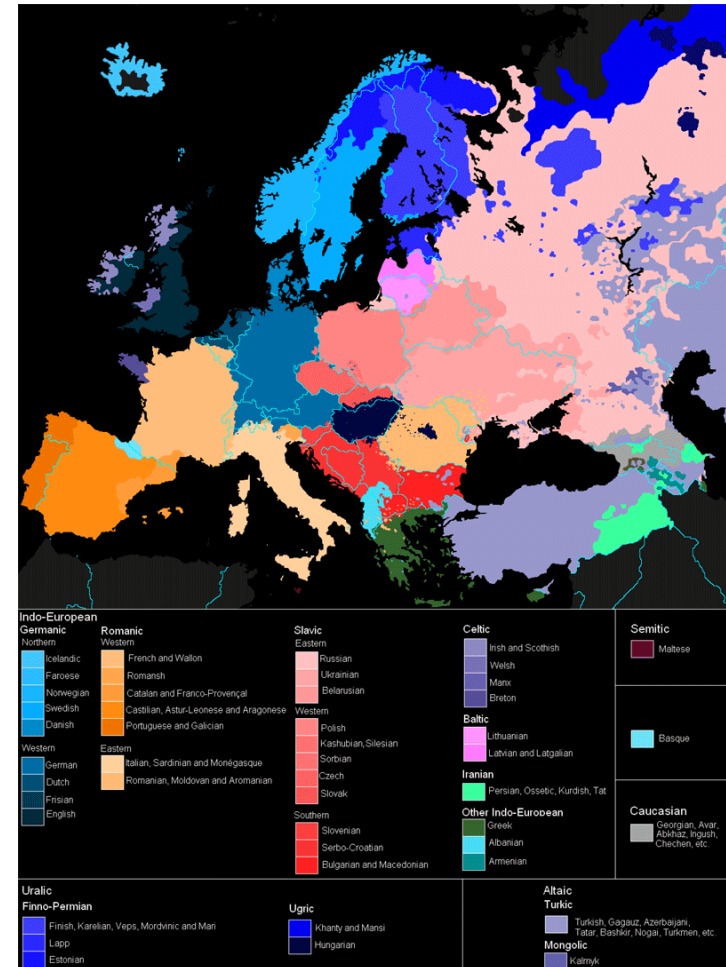
- Rather:
 - Brief overview of where we are
 - Charting map of major research challenges for where we want to go to

Multilingual LT Research in Europe

- ❑ Why language technologies in Europe?
- ❑ Challenges and Opportunities
 - Complexity
 - Fragmentation
 - Coverage, Quality, Robustness
 - New domains: professionally edited vs. user-generated content (UGC)
 - Multi(ple)-modal knowledge sources
 - Standards and interoperability
 - Affordability
 - One-size-fits-all vs. specialised approaches
 - Paradigm shift vs. incremental improvements (engineering)
 - Human/user centric language technologies
- ❑ Conclusions

Multilingual LT Research in Europe

- ❑ 23 official languages
- ❑ 60+ regional languages
- ❑ The language of Europe is Translation
- ❑ Europe leading in LT, with US and Asia

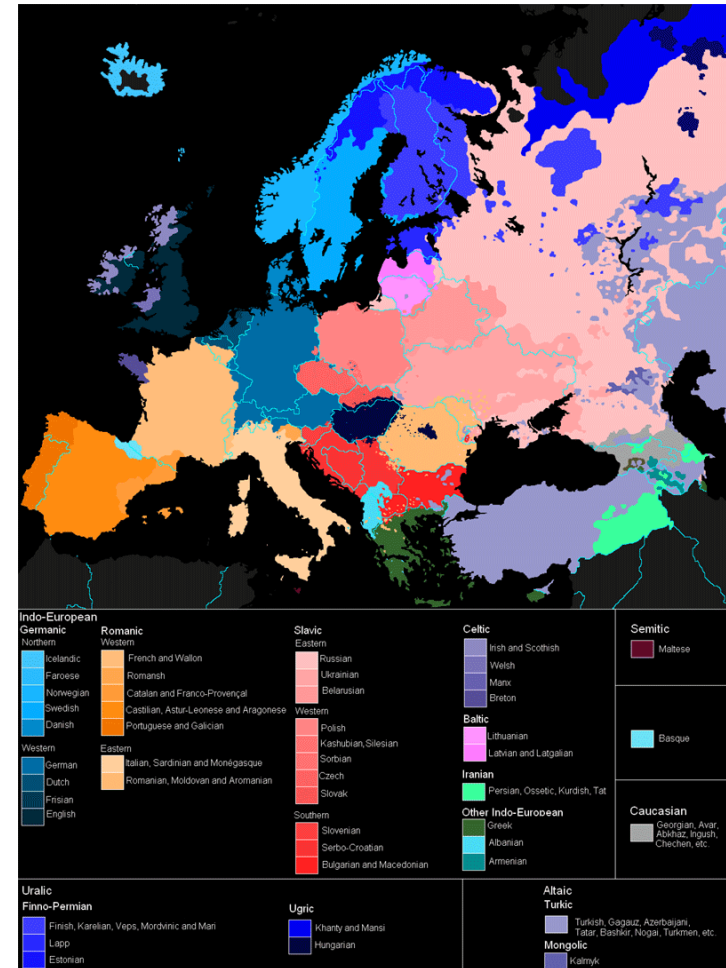


Multilingual LT Research in Europe

- ❑ Europe 500M citizens
- ❑ EU GDP 16+ trillion \$
- ❑ EU huge internal common market
- ❑ EU external trade

- ❑ Languages:
 - Commerce and trade
 - Cultural heritage and identity
 - Social stability and development

- ❑ Languages and communication are crucial



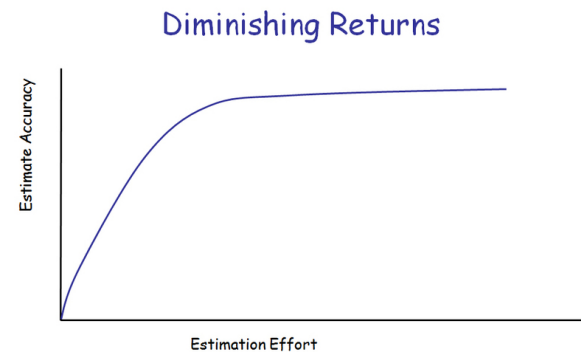
Multilingual LT Research in Europe

- ❑ Human languages are extraordinarily complex
- ❑ Many challenges to basic and applied research and applications

- ❑ Much progress has been made
- ❑ Overall picture remains patchy

- ❑ Coverage
 - LT resources for many “big” languages
 - Massive gaps for others

- ❑ Quality and robustness
 - Leaves much to be desired
 - Even for big languages



Multilingual LT Research in Europe

- ❑ New domains: professionally edited vs. user-generated content UGC
- ❑ Traditionally LTs designed for professionally edited content
 - Newspaper text, manuals, ...
 - Clean data
 - (Exception: speech and dialog processing)
- ❑ Move from WEB user as consumer to producer (WEB 2.0)
 - Prosumer
 - Social networking sites, blogs, fora, etc.
 - Noisy data
 - Huge commercial value
 - Customer care, sentiment analysis, social dyna intelligence



and other

Multilingual LT Research in Europe

- ❑ Research fragmented
- ❑ Standards and interoperability
- ❑ Maturity of technology: nuts and bolts
- ❑ Plug & play architectures
- ❑ LT app store
- ❑ Required for commercial exploitation of LTs
- ❑ Supports incremental progress



Multilingual LT Research in Europe

- ❑ Affordability of LTs
- ❑ Maturity of technology: nuts and bolts
- ❑ Plug and play architectures
- ❑ Expertise
- ❑ Data

- ❑ Not just for the big players: MNO
- ❑ SMEs, NGOs and individuals (e.g. translators)



Multilingual LT Research in Europe

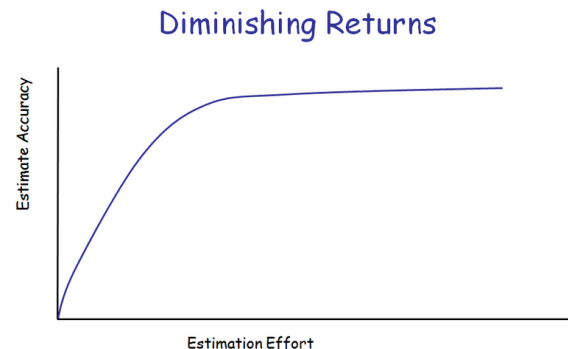
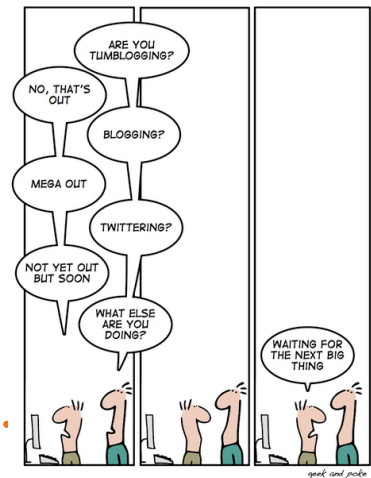
- ❑ One-size-fits-all vs. specialised approaches
- ❑ One-size-fits-all: e.g. PB-SMT
 - Simplicity
 - Scale
 - Affordable
 - Quality?
- ❑ Specialised approaches: for example special treatments/resources of free word-order languages or morphologically rich languages
 - Better quality
 - Costly?
 - Scalability?



Un-/semi-supervised machine learning the rescue?

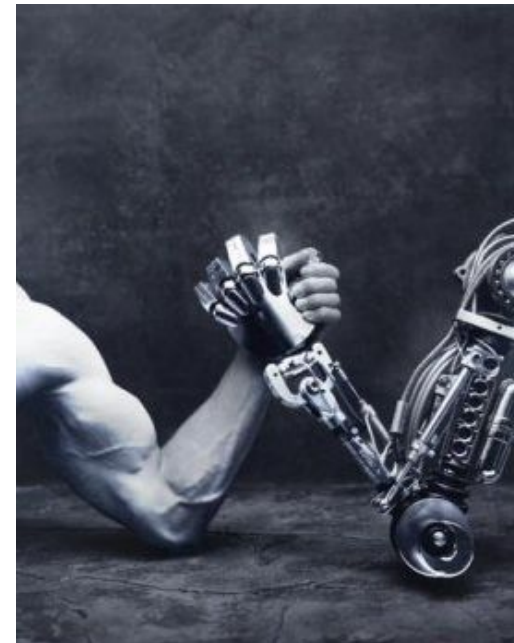
Multilingual LT Research in Europe

- ❑ Paradigm shift vs. incremental improvements (engineering)
- ❑ Paradigm shift : **the next big thing ...**
 - Cannot augur the future
 - Most predictions tend to be wrong
 - Paradigm shifts tend to take time
- ❑ Incremental improvements/engineering: **the hard slog ...**
 - Struggle with point of diminishing returns



Multilingual LT Research in Europe

- Human/user centric language technologies
 - Translators are using machine translation
 - Interactive MT rather than post-editing MT output
 - Human in control
 - Disconnect between labs and users
 - LTs and HCI and software engineering



Multilingual LT Research in Europe



- ❑ Outstanding LT research, development and commercial expertise in Europe
- ❑ Landscape is fragmented
- ❑ Need:
 - Cross-European partnerships
 - Industry-academia-user partnerships
 - Shared research agenda



Q/A

META  NET

Thank you.

office@meta-net.eu

<http://www.meta-net.eu>

<http://www.facebook.com/META.Alliance>

<http://www.meta-net.eu>