

**Josef van Genabith - DFKI**



# **QT21: Quality Machine Translation**

**Cracking the Quality Barrier**

**presented by Hans Uszkoreit**



Funded by the 7th Framework Programme of the European Commission through the contract 296347.

# Outline of the Talk:



- European Single Digital Market
- The reality: large underexplored grey areas on the European Language Technologies map
- Difficult languages: morphologically rich, syntactically free
- Challenging resource situations: data & tools
- QT21 research focus
- Competition at the core: the QT21 workplan
- Technology Transfer at the core: the ICT-17 Technology Bridge
- Multi-disciplinarity at the core: the QT21 team
- Questions?

# European Single Digital Market:

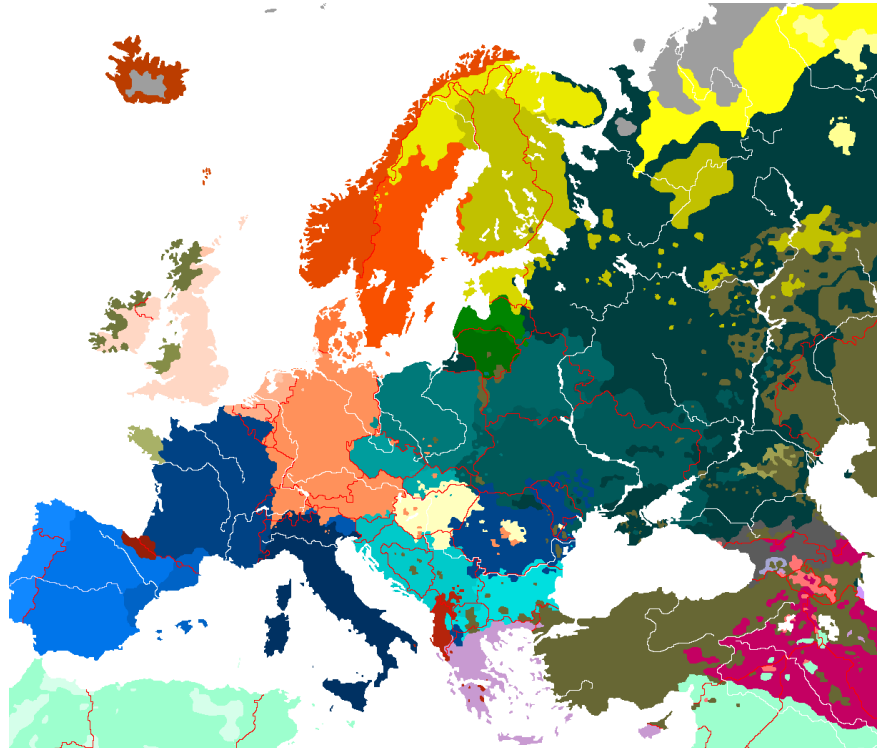


- Free of barriers
- In particular language barriers
- Free flow of ideas, commerce and people
- Stated EU goal 2020



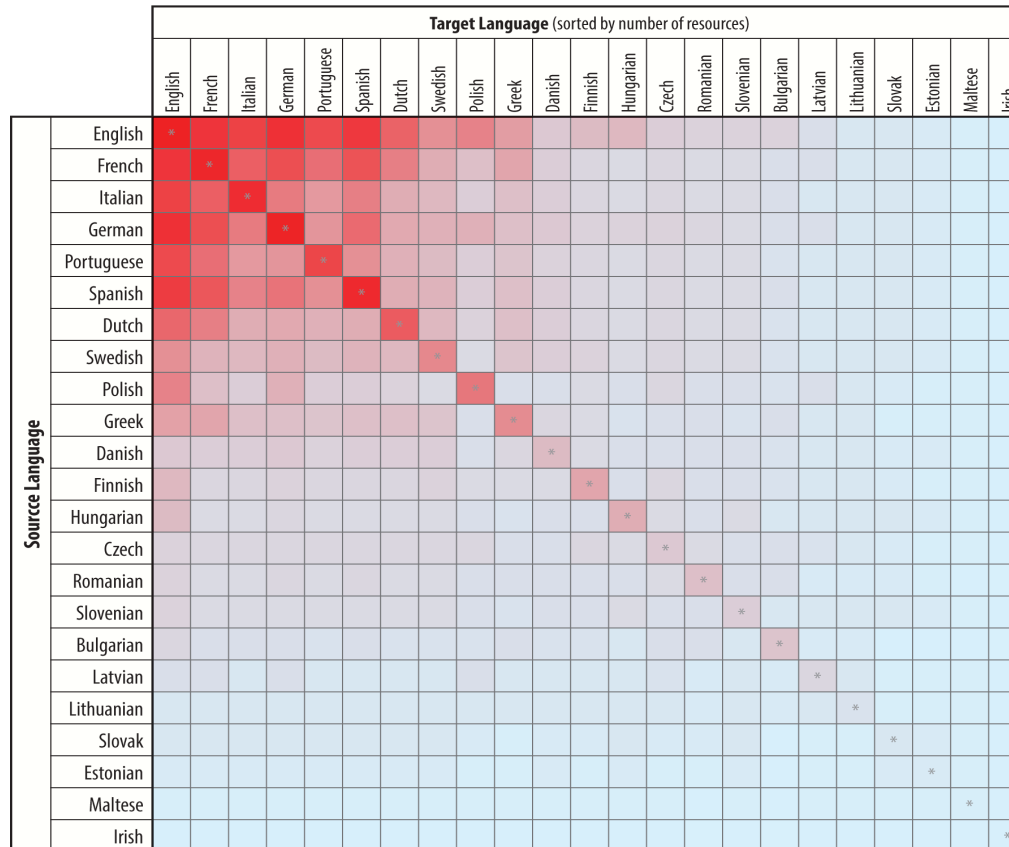
# The Reality:

- 22 official EU languages
- META-NET Language White Papers + EuroMatrixPlus
- Only 3 languages with moderate to good language technology support
- Serious obstacle to free flow of ideas, commerce and people



# The Reality: a large and under-explored area in the language technology map for European Languages

Resources for official EU languages



Source: EuroMatrix+ (2012)

Ranking by number of resources, corpora, and MT systems

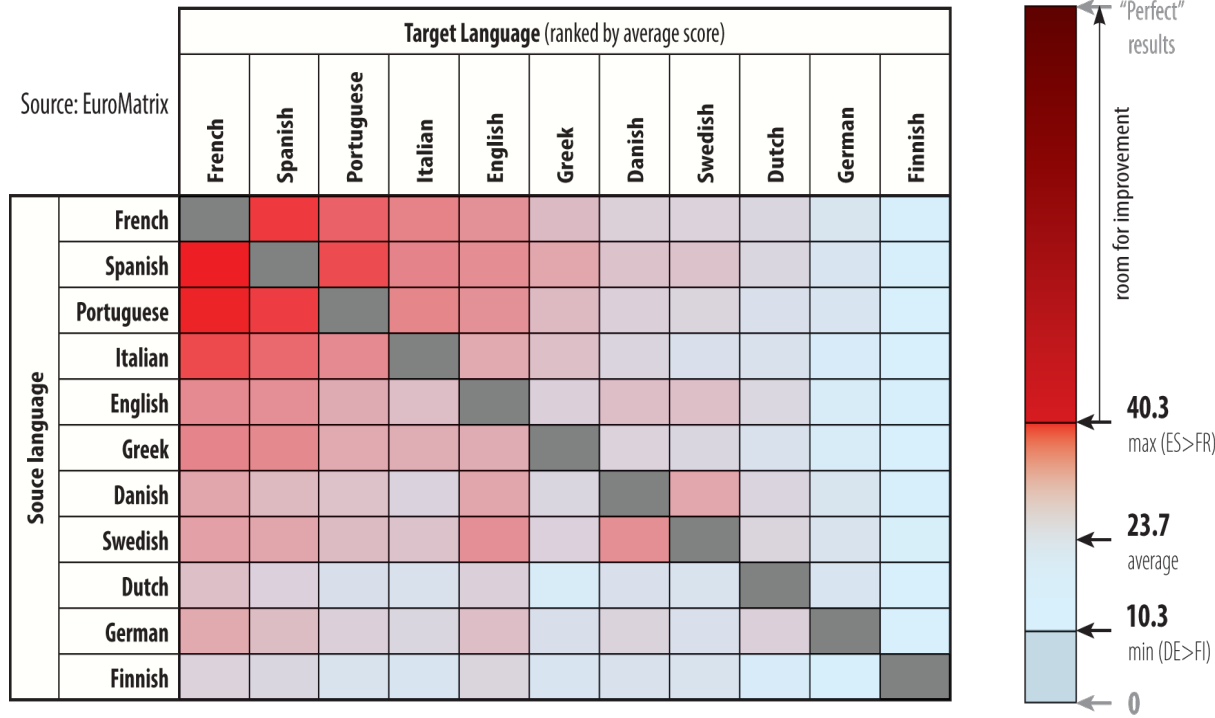
\* = monolingual resource



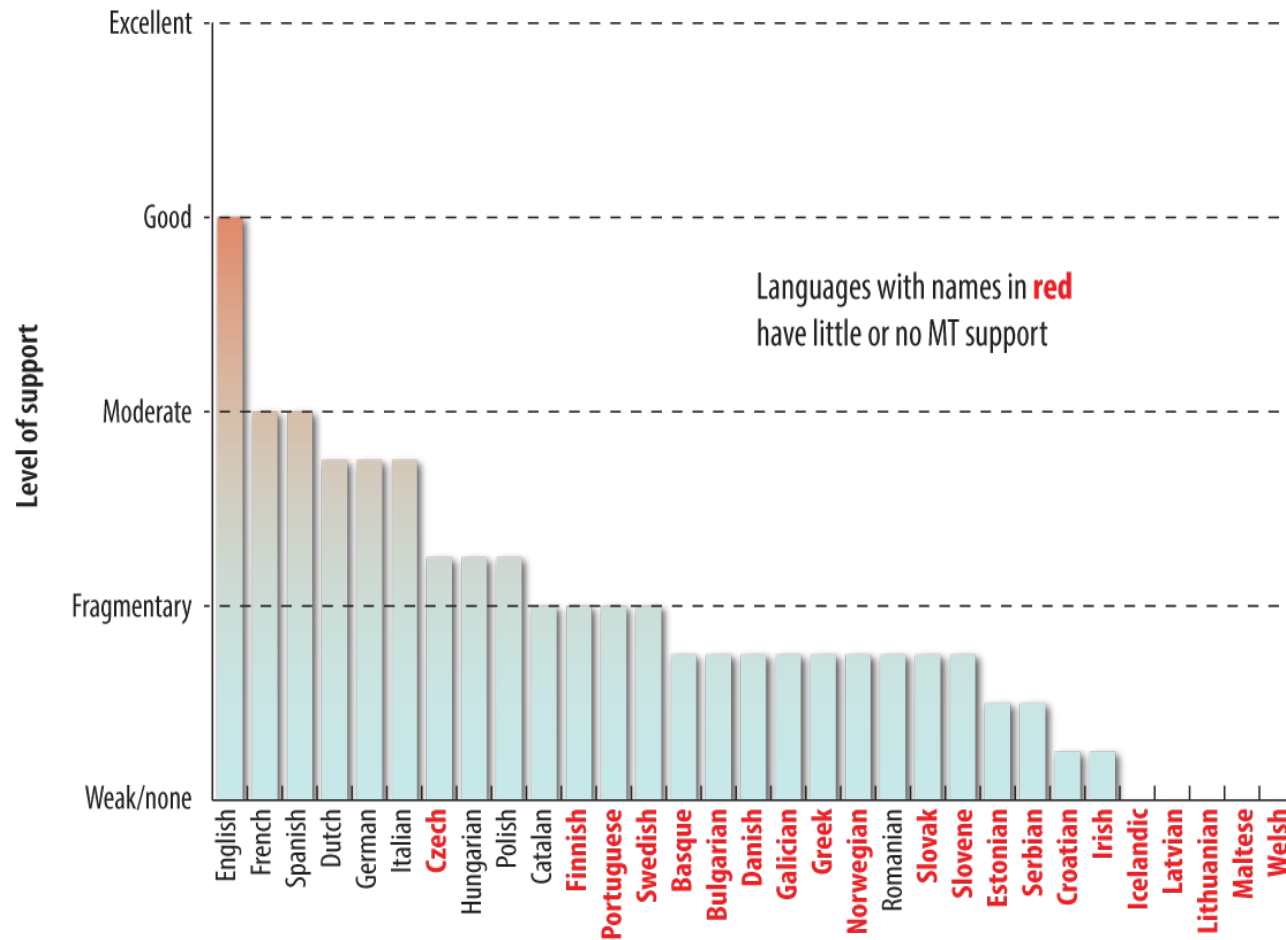
Note: Does not include Croatian, which became an EU language after this information was compiled.

# The Reality: a large and under-explored area in the language technology map for European Languages

Quality of general-domain MT output  
(BLEU scores)



# The Reality: a large and under-explored area in the language technology map for European Languages



## “Difficult” Languages in the “grey” area of the map:

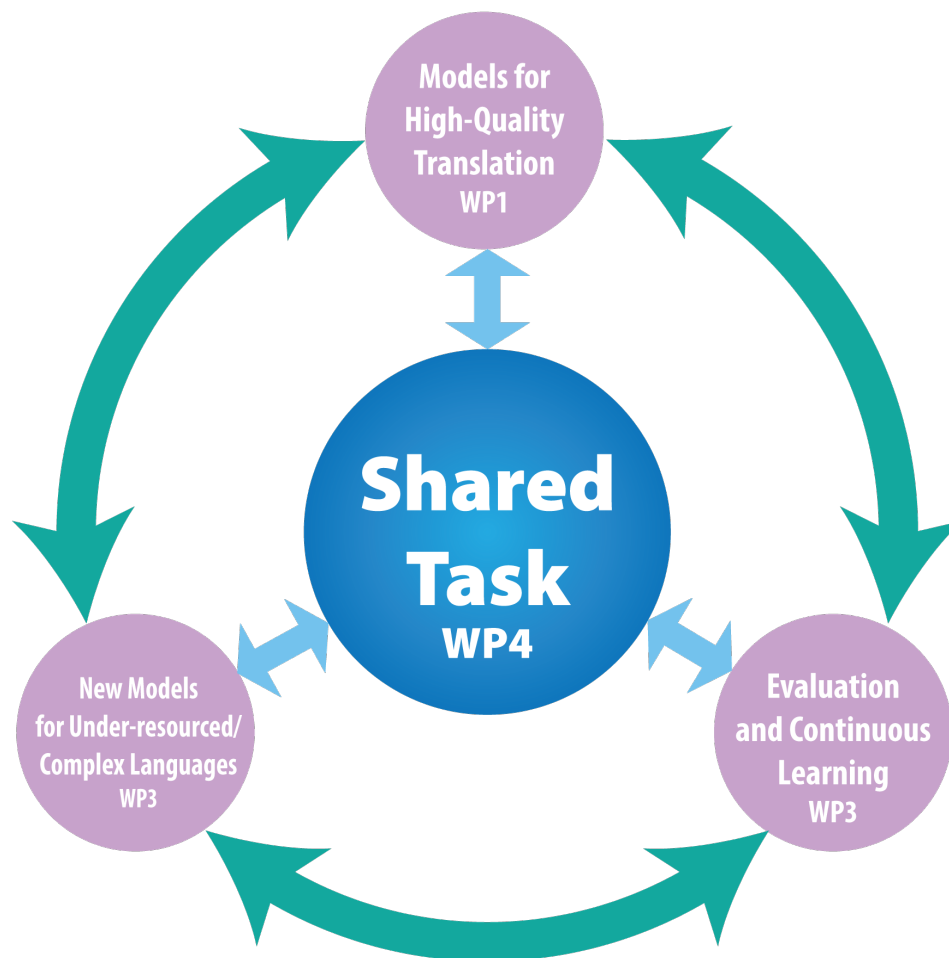


- Many languages have resisted the best of our current technologies
- Common traits:
  - Morphologically rich
  - Syntactically free
  - Lack of training and tool resources
- Often these go together ...

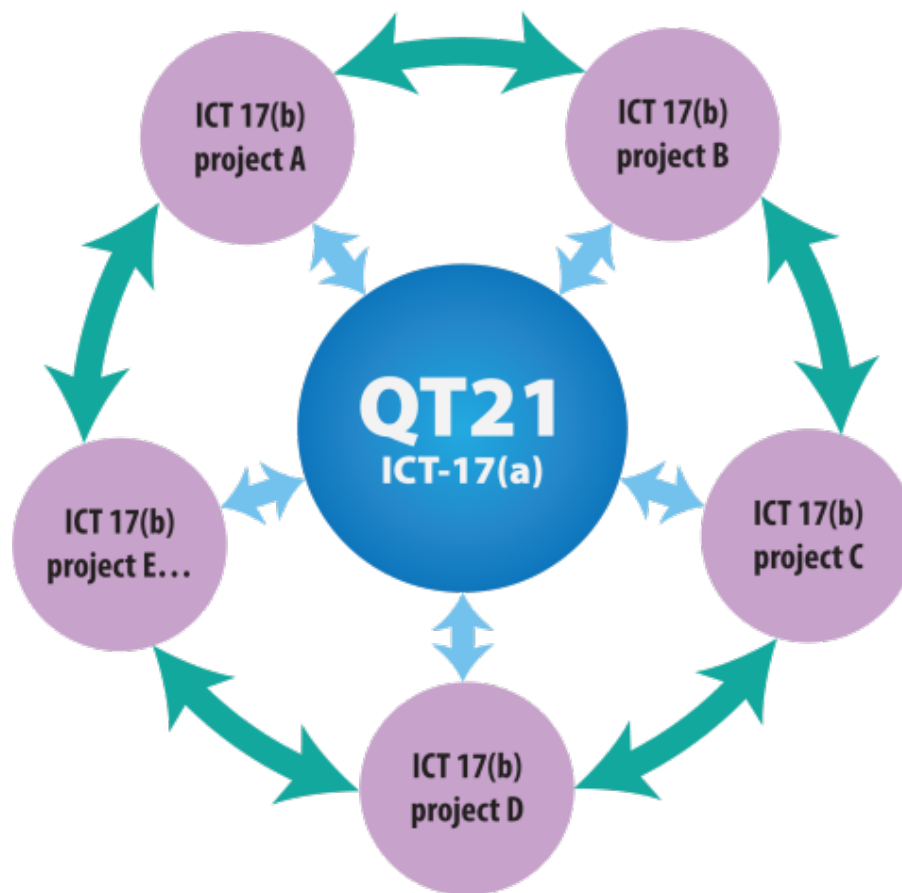


- 
- Challenging linguistic phenomena
  - Challenging resource scenarios
  - Focus of QT21 research:
    - Substantially improved statistical and machine-learning based models for challenging languages and resource scenarios
    - Improved evaluation
    - Continuous learning from mistakes
    - Guided by systematic analysis of quality barriers, informed by human translators
    - Learning and decoding efficient
    - Reliance on data (annotated or not) minimised

# QT21 Workplan: Competition at the Core



# ICT-17(a) – ICT-17(b) Technology Bridge



- DFKI
- CNRS (LIMSI)
- RWTH
- KIT
- University of Edinburgh
- FBK
- CUNI (Prague)
- Dublin City University
- University of Amsterdam
- University of Sheffield
- HKUST (Hong Kong)

- Tilde
- text&form
- GALA
- FIT
- TAUS
- DGT (associated, non-costed)

## QT21: Quality Machine Translation for the 21<sup>st</sup> Century

### Questions?

Many thanks to Arle Lommel for the graphics!