

Linguistic Diversity: implications for data and metadata

Boris B. Velichkovsky
Moscow State University

ECP2009, Oslo, Norway

Language issues in scientific publishing

- English – 95% of all publications listed in the SCI in 1995
- English as a language of science – does it really give everybody equal access?
- English is a common language which obviously makes scientific communication easier
- However, the dominance of English has several major drawbacks

Monolinguality: Drawbacks

- **Biasing citation practices**
- **Marginalization of peripheral scientists**
- **Imposing anglophone discursive norms through aggressive gatekeeping**
- **Dominance of research from English speaking countries**
- **In sum: *monolinguality limits access to publishing resources for peripheral scientists and leads to distortion of results***

Linguistic diversity

- **A concept from applied linguistics**
- **Various definitions all targeted at the number of languages spoken in a specific area**
- **Major trend in applied linguistics, politics and education is to support LD as a valuable asset**
- **However, LD obviously present a problem for international scientific communication**

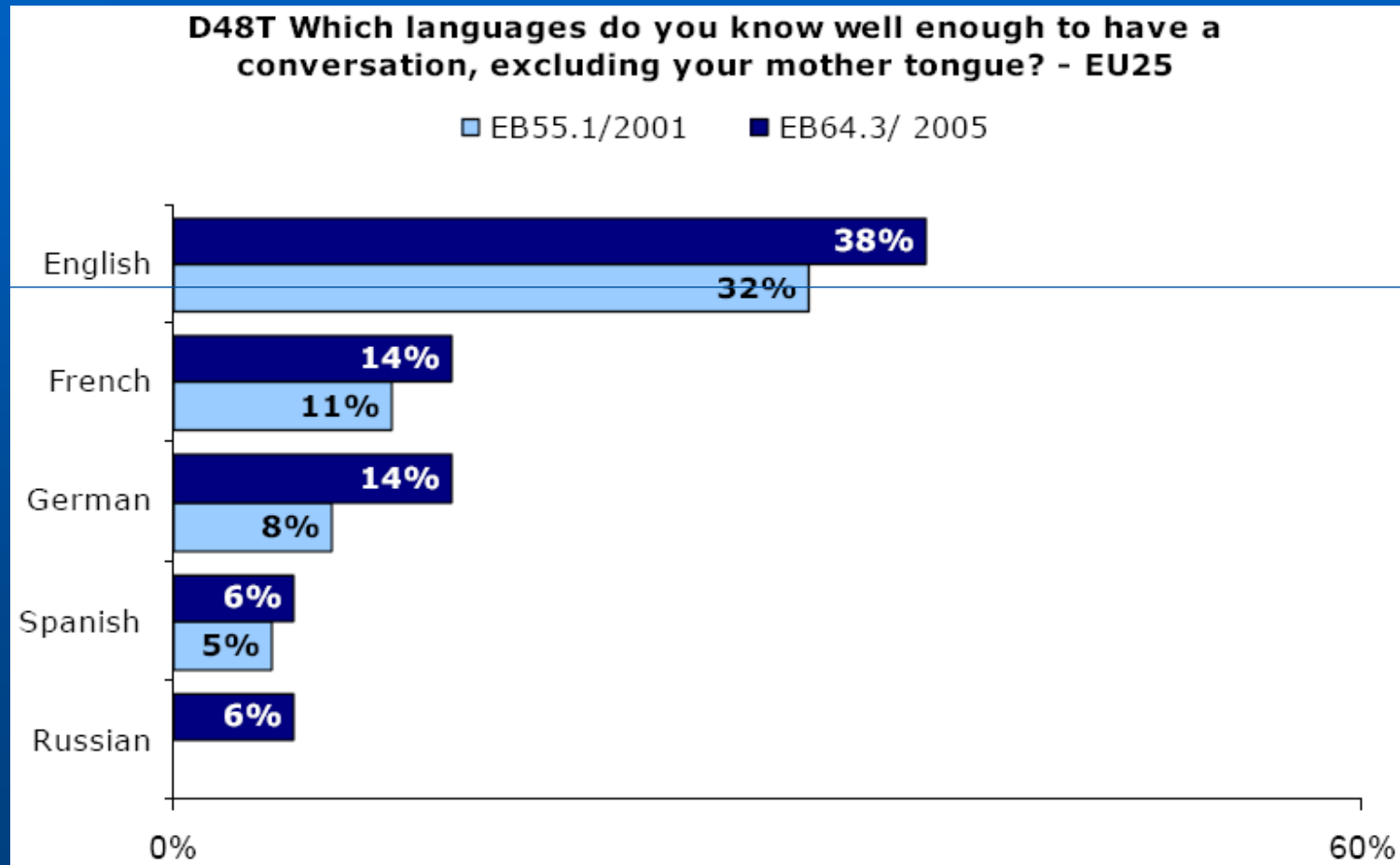
Linguistic diversity in Europe

- In total, about 80 languages are being spoken in Europe...
- Belonging to 10 language families
- Today, EU has 23 official languages

The size of the problem

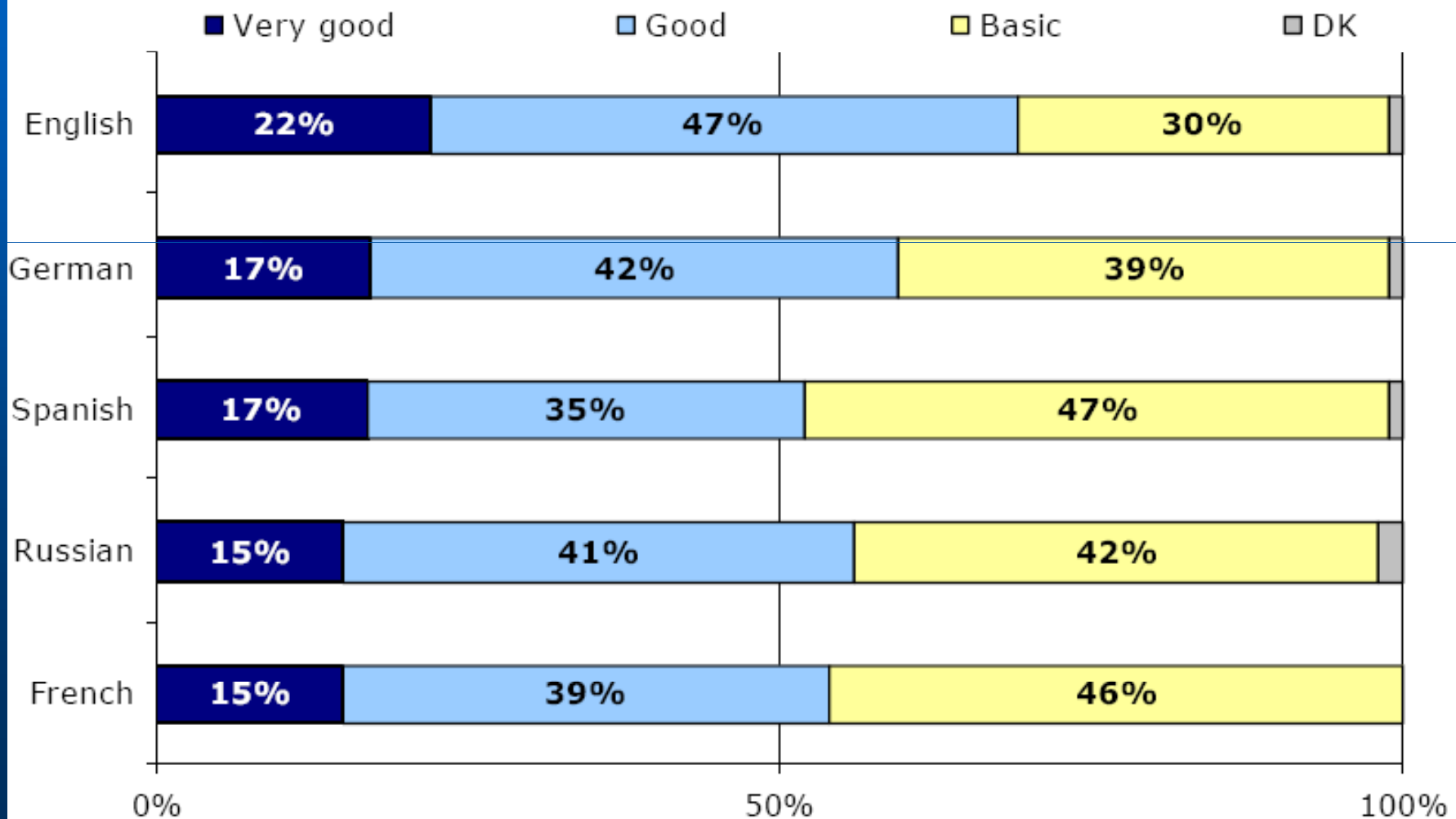
- There is a LD-problem for EPPP – it can not support all the official languages of the EU, to say nothing about the rest...
- What is the real size of the LD-problem in Europe?
- Eurobarometer 2006 Study, “Europeans and their languages” gives insights in the actual use of languages in Europe

Most used foreign languages



Proficiency level

D48f Level of the Language Spoken - % EU



Summary

- There is a well-developed multilingual culture in the EU
- Five core languages spoken in Europe can be identified: *English, German, French, Spanish, Russian*
- Significant proportion of Europeans report sufficient level of mastery for the core languages

Evaluation criteria

- **A multilingual support strategy must be implemented in EPPP to solve the LD-problem. Three questions have to be answered:**
 - **Who is the translator: Professionals / Authors**
 - **What is the scope of translation: Text + metadata / Metadata only**
 - **What are the target languages: EU official languages / Subset of core languages / Selected languages**

Multilinguality in OA: Current practices

- **Directory of Open Access Journals (www.doaj.org)**
- **106 journals in the subject field of psychology**
- **Multilinguality support**
 - 19,8% - 2 languages
 - 6,6% - 3 languages
 - 3,8% - up to 5 languages
- **Multilinguality is valuable but hard to implement**

Implementation of multilinguality support

- **Many “bilingual” journals even not having submission instructions in the second language**
- **Texts submitted mostly in one language**
 - Language-editing if not a native speaker (costs!)
 - Copy-editing by native speakers
 - Translation on behalf of the author
 - Translation services
 - Translation partly on behalf of the journal in the case of acceptance
- **Abstracts submitted mostly in two languages**
 - Translated by the author
 - Original and a core “language” (mostly English)
 - Sometimes two core language

Implementation of multilinguality support

- **Keywords submitted mostly in two languages**
 - Translated by the author
 - Original and a core “language” (mostly English)
 - Sometimes two core languages
 - Predefined keyword-list (APA-based)
- **Titles submitted mostly in two languages**
 - As with other metadata

Summary

- **Only metadata are multilingual**
 - Titles
 - Abstracts
 - Keywords
- **Multilinguality support is off-loaded on the author**
- **Current practices clearly reveal the simple *impossibility* of implementing full-fledged multilinguality support within current OA business-models**

Multilinguality support in EPPP:

Outline

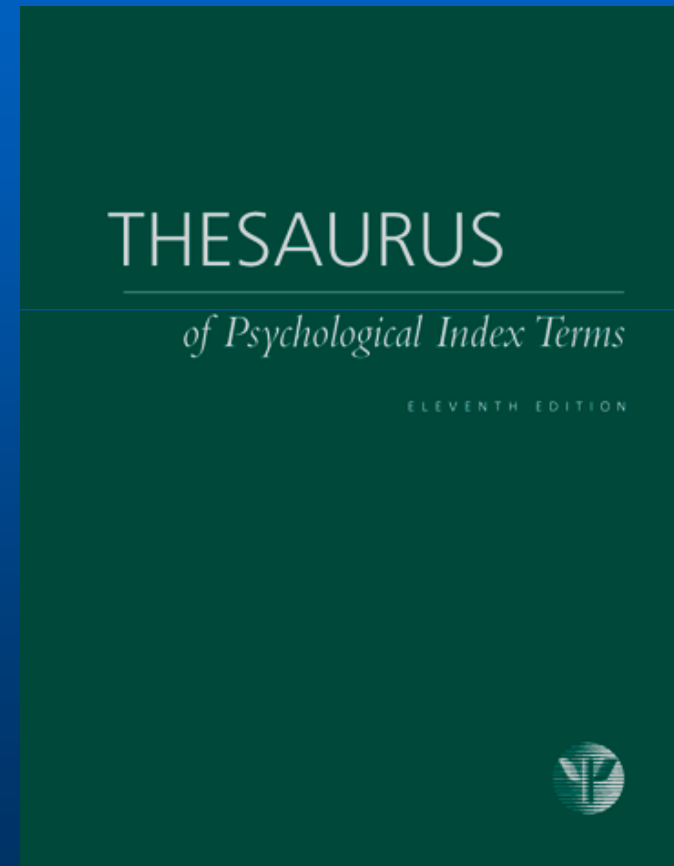
- Core languages must be supported
- Only metadata are to be translated
- Titles and abstracts can only reasonably be translated by the authors. However, the number of supported languages will reduce to two
- For keywords, there is an option of using *multilingual thesauri*

Thesaurus

- A thesaurus contains a list of terms together with the relationships between them
- A thesaurus can be hierarchically organized depicting the generic relationships between the terms
- A multilingual thesaurus contains the mappings between identical terms in various languages

Psychological thesauri

- **APA Thesaurus of Psychological Index Terms, 11th edition**



Multilingual thesauri

- **Allows for controlled selection of keywords in the native language**
- **The document is made accessible through searches in all supported languages**
- **Technology already used for cross-language retrieval in computer sciences**

Multilingual thesauri for Europe

- **Inter-active terminology for Europe, 24 languages, 1,4 million entries**
- **ETB Thesaurus (education and learning), 14 languages, 1155 terms**
- **European thesaurus on health promotion, 11 languages**
- **There is still no European thesaurus in psychology ...**

Thank you!