

Linking across languages, linking across cultures

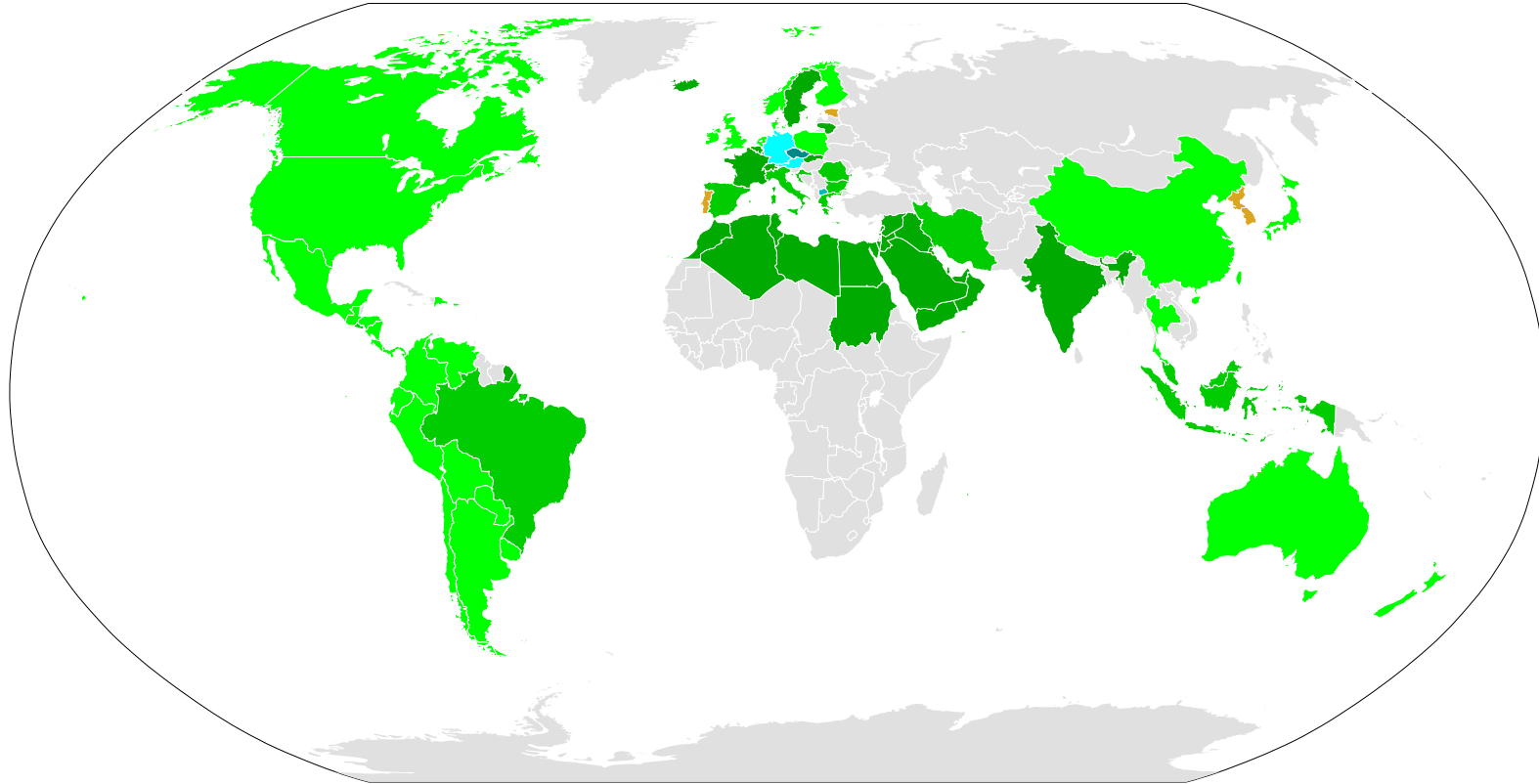
Francis Bond

Computational Linguistics Group
Linguistics and Multilingual Studies,
Nanyang Technological University

2016-01-15

Open Multilingual Wordnet Grid

- The Open Multilingual Wordnet
- The Collaborative Interlingual Index
- Princeton Wordnet



+ Romanian, Dutch, Slovak, Lithuanian, Icelandic,
Croatian, Bulgarian ↑ Polish, Arabic, Italian (new wn)

- Different wordnets built very differently
 - Big national projects
 - Small enthusiasts
 - Linguists/Computer Scientists
 - Supported/unsupported
- Linking to corpora and teaching
- New interface — release and share

- We want to add concepts not in the Princeton Wordnet (PWN)
 - Maybe they are not in English at all
 - Maybe they are in English but not (yet) in PWN
- We want these new concepts to be
 - linked to the existing network
 - shared across languages (and projects)

The solution: ILI

1. The Interlinear Index (ILI) should be a flat list of synsets.
2. The semantic and lexical relations should mean the same things for all languages.
3. Synsets should be constructed for salient and frequent lexicalized concepts in all languages.
4. Multi-word expressions (MWEs) should be included. As space is not an issue, the default should be to include common MWEs.

-
5. A formal ontology should be linked to but separate from the wordnets.
 6. The license must allow redistribution of the index
 7. Synset IDs should be persistent
 8. Each new synset **MUST** link to an existing synset in an open wordnet
 9. Each new synset should have an example
 10. Each new synset **MUST** have a definition in English
 11. Each new synset should have a name in English

12. A supertype based on the lexicographical files
13. Any project adding new synsets should first check that they do not already exist in the shared ILI

The details

- ILI as RDF — shared on github
- Each new wordnet release can suggest candidates
only projects can add new ILI entries
- Review period (1–3 months)? — accept if no
comments
- ILI's cannot be deleted — new versions link to them
- Can be linked (same as) and/or deprecated
- The ILI only has minimal information (not even POS)

- But is part of the ILL+open wordnets
- NTU/VU to host



- New POS x
non-referential words: titles, exclamatives, classifiers
- Modals
- Gap filling
dog-cart, Nusuntara, business “matter”, knock up, arrow
Merlion, sate, carrot cake
- Specialized ontologies
Geonet, Species DB, Wikipedia

➤ Linking to Grammars HPSG



References

Fellbaum, C. and Vossen, P. (2008). Challenges for a global wordnet. In J., W., Ide, N., and Fang., A., editors, *Online Proceedings of the First International Workshop on Global Interoperability for Language Resources (ICGL 2008)*, pages 75–82. City University of Hongkong.