

Nanyang Technological University Update

Francis **Bond**
and many, many more

Computational Linguistics Lab (凡士研)
Linguistics and Multilingual Studies,
Nanyang Technological University

[<bond@ieee.org>](mailto:bond@ieee.org)

DELPH-IN 2019-07-15



Overview

- Wordnets: Multi, Jpn, English, Chinese, Bahasa, *Myanmar, Kristang*
 - ▶ Collaborative Interlingual Index
- Grammars:
 - ▶ Indonesian (INDRA)
 - ▶ Chinese (Zhong []): Mandarin and Cantonese)
 - ▶ Japanese (JACY) — bug fixes, minor enhancements
- Tools:
 - ▶ OMW version 2.0, Graph checks
 - ▶ Enhancements to the LTDB (using doc-strings)
- Intelligent Tutoring and Feedback on Errors
- English Essay writing help
 - started treebanking student essays with the mal-grammar
 - continuing to collect new data

Open Multilingual Wordnet

- Collaborative InterLingual Index (CILI)
 - ▶ Initial testing pretty much done
 - ▶ Wow, there are a lot of interesting things you can do
 - ★ And we check for/document many of them
 - ★ Including cycles
- More languages and words
 - ▶ Sesotho, Tswana, Venda, Xhosa, Zulu, Gaelic, Abui, Moroccan Arabic, Turkish, Czech, *Cantonese*, *Open German*, ... (40+ curated languages)
 - ▶ GeoNames, English Open Wordnet
 - ▶ 10,000+ new English synsets (enwordnet)
- More links
 - ▶ Interlanguage sense level links (Harmony Project)
 - ▶ Semantic relations (dot-products, roles)
 - ▶ Better sentiment and emotive links

NTU multilingual corpus

- Small, deeply analysed corpus
 - ▶ 6,000+ sentences x 3 languages (cmn, eng, jpn); 2,000 in Indonesian
 - ★ Mainichi Newspaper (NICT translations)
 - ★ Sherlock Holmes (**DANC**, **SPEC**; **REDH**, **SCAN**, **HOUN** [eng])
used in ACL workshop on Dialogue
 - ★ Cathedral and the Bazaar (many languages)
 - ★ Spider's thread 「蜘蛛の糸」
 - ★ Singapore Tourist data (plus Korean, Viet, Indo)
 - ▶ **Hand alignment**, **WordNet tagging**, Treebanking, Sentiment, ...
- Revised the DB structure; second round of wordnet tagging
 - ▶ Much behind schedule
 - ▶ Wordnet extension is being done in parallel
- Tagged Italian, *Polish* (**SPEC**)
ready to start Bulgarian, Dutch, German, Abui (shared task)

Newish Projects

- A collaborative network for usage-based research on lesser-studied languages (JSPS: Tokyo University of Foreign Studies) — built large Indonesian corpus, morphological analyzer, learner vocab for 26 languages (with example sentences — linking to Wordnet)
- Digital Mapping the Literary Epigraph: Quantitative analysis of literary influence using network theory and thousands of epigraphs (NTU)
- *Digital Humanities Cluster* at the College of Humanities and Social Sciences
- Hosting PNC 2019 (regional DH conference at NTU in October)

Members

- Staff:
Michael Wayne Goodman (DH postdoc)
- Visiting Scholars:
Ahti Lohk (Tallinn UT) *Semantic Graphs*,
Maciej Piasecki & Arkadiusz Janz (Wrocław UT) *WSD and Sentiment*
- Students:
Luis *Technology Enhanced Language Learning*,
Hannah *Indonesian Morphological Alternations -kan and -i*
- Graduated
 - ▶ Lê Tuấn Anh 2019 (PhD: LMS, NTU) *Developing and Applying an Integrated Semantic Framework for Natural Language Understanding*
 - ▶ Zhenzhen **Fan** 2019 (PhD: LMS, NTU) *Building an HPSG Chinese Grammar (Zhong)*
 - ▶ Giulia **Bonansinga** 2019 (MA: University of Pisa) *Cross-lingual word sense disambiguation with multilingual Wordnets*

