

Nanyang Technological University Update

Francis **Bond**
and many, many more

Computational Linguistics Lab (凡土研)
Linguistics and Multilingual Studies,
Nanyang Technological University

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

DELPH-IN 2020-07-13



Overview

- Wordnets: Multi, Jpn, English, Chinese, Bahasa, Myanmar, Kristang
 - ▶ Collaborative Interlingual Index
- Grammars:
 - ▶ Chinese (Zhong [|]): Mandarin and Cantonese)
 - ▶ Japanese (JACY) — bug fixes, minor enhancements
 - ▶ Singlish (based on the ERG)
- Tools:
 - ▶ OMW version 2.0, Graph checks, better documentation selected for **Google Season of Docs**
 - ▶ LTDB (pydelphin parsing of TDL, per grammar DBs)
- 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)
 - ▶ Final testing almost finished
 - ▶ Better documentation
 - ▶ Still not released
- 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, Tokyo University of Foreign Studies Basic Vocab
 - ▶ 20,000+ new English synsets
- 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 quality control
 - ▶ Much behind schedule
 - ▶ Wordnet extension is being done in parallel
- Tagged Italian, Polish (SPEC)

Newish Projects

- *Digital Humanities Cluster* at the College of Humanities and Social Sciences
- Hosted PNC 2019 (regional DH conference at NTU in October)
- Pandemic Special
 - ▶ 4 students treebanking
 - ▶ 4 students sense tagging
 - ▶ 2 programming
 - ▶ 1 data entry and checking

- Staff:
Michael Wayne Goodman (DH postdoc)
- Students:
 - ▶ Luis *Technology Enhanced Language Learning*
 - ▶ Siew Yeng *An HPSG for Singlish*
 - ▶ Tommaso *Cross-lingual WSD*
- Graduated
 - ▶ Hannah Choi 2020 (MA: LMS, NTU) *Indonesian Morphological Alternations -**kan** and -**i***