

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 2016-06-16



- NTU
 - ▶ Wordnets: Multi, Japanese, English, Chinese, Bahasa
 - ★ the Interlingual Index
 - ▶ Grammars:
 - ★ Indonesian (INDRA)
 - ★ Chinese (Zhong []): Mandarin and Cantonese)
 - ★ Japanese (JACY)
 - ▶ Tools:
 - ★ pyDelphin, Demophon
 - ★ LingType DB
 - ★ IMI, OMWedit
 - ▶ MRS-based SMT
 - ▶ Intelligent Tutoring and Feedback on Errors

Open Multi-Lingual Wordnet

- Collaborative InterLingual Index (CILI)
 - ▶ Share new synsets between languages
 - ▶ July 15th roll out
- More lexical types
 - ▶ Exclamatives (*hi*, *tally ho!*, *checkmate*)
 - ▶ Pronouns (even more)
 - ▶ Classifiers
 - ▶ Chengyu
- More languages
 - ▶ Romanian, Open Dutch, Ita-, Slovakian, Lithuanian and Icelandic Wordnets (Ancient Greek, Gaelic, Abui queued)
 - ▶ Linked Swadesh Lists (1,200+ languages)
 - ★ Matrix starter vocab anyone?
 - ★ Also made [Singapore Sign Language](#) Swadesh entries

Open Multi-Lingual Wordnet ToDo

- Tighter Integration with Corpora
- Sense annotated and parsed definitions
- Simple definitions for core vocabulary
- Linking to ontologies (updatable)
 - ▶ Geonames (with Trento)
 - ▶ Snomed (with Amanda Hicks)
 - ▶ Wiktionary (with Patricia)
- Propagate Sentiment
- Cross-lingual sense-level linking
- More on variants

NTU multilingual corpus

- Small, deeply analysed corpus
 - ▶ 6,000+ sentences × 3 languages (cmn, eng, jpn); 2,000 in Indonesian
 - ★ Mainichi Newspaper (NICT translations)
 - ★ Sherlock Holmes (DANC, SPEC, REDH[eng])
 - ★ 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
 - ▶ Slightly behind schedule
 - ▶ Wordnet extension is being done in parallel
- Tagged Italian (SPEC)
ready to start Bulgarian, Dutch, German

Grammars and Tools

- Chinese Grammar (Zhong [||]) expanded
- Indonesian (INDRA) expanded
- [Jacy Book](#) in final copy-edit!
- Linguistic Type Database updated to work for all types (uncommitted)
- Even easier to use pydelphin, nice demo, MT, ...
- Just starting forestry again
- Very useful discussion in the VLAD
 - ▶ Any topics for tomorrow?

New Projects

- Semi-Automatic Implementation of Clinical Practice Guidelines in Singapore hospitals
- Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation Technology
- Designing Online Communication Skills Materials
Identifying common errors in Engineering undergraduate writing (EdEx)
- Detecting Meaning with Sherlock Holmes
 - ▶ Trying to make semantics more accessible
 - ▶ Would be nice to add some inference :-)
 - ▶ Also gets a range of data from students
 - paraphrases, idioms, sense annotation, sentiment
- Inter-lingual equivalence in lexical databases; NSC (Poland)



Postponed/Proposed Projects

- Shared task on annotating **SPEC**
 - ▶ Integrated Semantic Framework
 - ▶ Groningen Meaning Bank
 - ▶ Grammatical Framework
 - ▶ AMR?
 - ▶ UNL
- Illustrated Core Vocabulary
 - ▶ Redo pictures for Psych experiments, link to wordnet
 - ▶ Simpler and better definitions
- Digital Humanities
 - ▶ Epigraph network
 - ▶ Author identification

Congratulations

- Liling started as a pre-Post-Doc at Singapore University of Technology and Design (SUTD)
- New Post-Doc: [E umamaheswari Vasanthakumar](#)
 - ▶ Expert on the Universal Network Language (UNL)
- I will be on sabbatical next year!
 - ▶ UW has kindly offered to host me
 - ▶ We may have a mini summit in January
 - ▶ Main goals
 - ★ Wordnet consolidation — 4.0 release, move to NTU
 - ★ Jacy 維新 “the Jacy Restoration”
 - ★ ISF/Machine translation/QMRS