NTU, NTT Site Reports

Francis Bond and Sanae Fujita[NTT]
Wang Shan, Huizhen Wang[NEU]
Fan Zhenzhen, Lê Tuấn Anh, David Moeljadi,
Michael Goodman, Luís Morgado da Costa
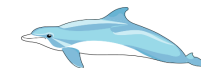**Nanyang Technological University**
[NTT]**Nippon Telegraph and Telephone Corporation**
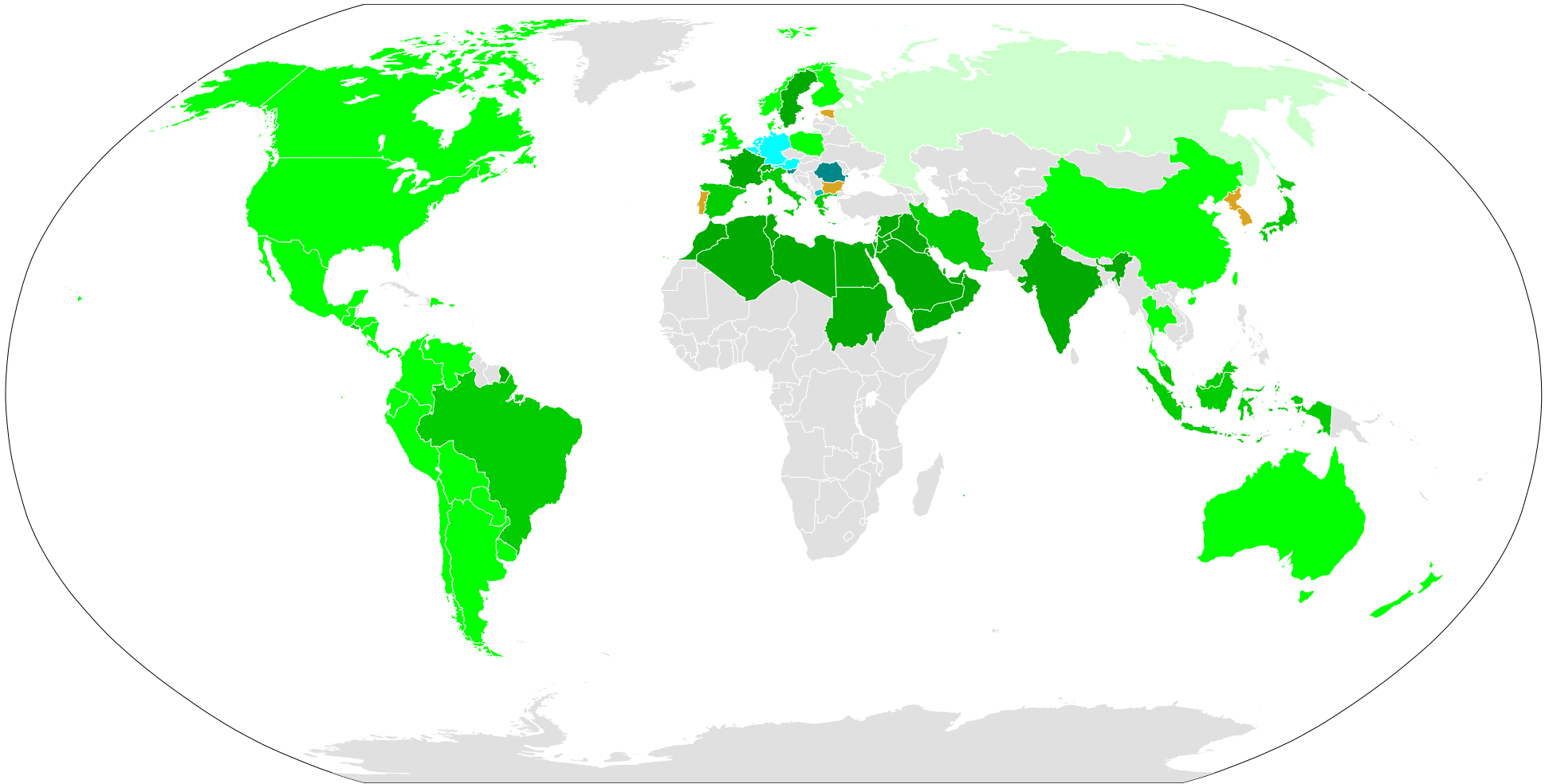[NEU]**Northeastern University, 东北大学**

<bond@ieee.org>

2014

➤ NTU

  ➤ Wordnets: Multi, Japanese, English, Chinese, Bahasa
  ➤ Grammars: new for Indonesian, Abui, Cantonese, Vietnamese
  ➤ Machine Translation (small hand built + MT)
  ➤ Several new projects

➤ NTT Report (Sanae Fujita & Takaaki Tanaka)
   Joint Research with NTU on Ontology linking, Subcat acquisition

Open Multilingual Wordnet

➤ Linking to more information (Sentiment, Time)

    ➤ Created links to the SIL Semantic domains
    ➤ Used to make dictionaries for under-resourced languages
    ➤ Working on making a wordnet for Abui (Papuan Language from Alor)

➤ Added more info: derivations, domains, . . .

    ➤ Created derivation links for Indonesian (hard to do automatically)

➤ Still working on the Interlingual Index

    ➤ At VU Amsterdam for the net two weeks
    ➤ Adding many words not in the Princeton Wordnet (some not in English)

➤ Christiane Fellbaum to visit in October

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 (many languages)
- ∗ Cathedral and the Bazaar (many languages)
- ∗ Singapore Tourist data (plus Korean, Viet, Indo)

   ➤ Hand alignment, WordNet tagging, Treebanking, Information Structure, . . .

➤ Revised the DB structure; second round of wordnet tagging

   ➤ Slightly behind schedule
   ➤ Wordnet extension is being done in parallel

➤ Loose cooperation with Antoni Oliver creating bitext ebooks

➤ Mandarin Chinese Grammar (Fan Zhenzhen)

   ➤ Learning more verbs from treebanks

   ➤ Adding semantics to Yi's grammar, some new phenomenon

   ➤ Applying for a grant in China with Huizhen Wang (NorthEast University) to develop it further

➤ English Resource Grammar

   ➤ *X look as though butter wouldn't melt in X's mouth*

   ➤ Around 500 possessive idioms

➤ Dan to visit in September/October

➢ Indonesian (Bahasa) Grammar (David Moeljadi: new Phd)

    ➢ Will use to parse the NTU-MC

➢ Vietnamese Grammar (Lê Tuấn Anh: new PhD) small grammar only

➢ Abui Grammar (Frantisek: Ass Prof) wants to develop with descriptive grammar

➢ Cantonese Grammar (Joanna: Lecturer) test syntactic theories

    ➢ We should use CLIMB for a shared Chinese grammar

➢ minimal MT systems built for Chinese, Indonesian, Vietnamese

    ➢ We will develop these in parallel with the grammars

    ➢ Also looking at statistical deep transfer (see Michael)

➤ *That's what you meant: A Rich Representation for Representing Meaning across Languages* $405,702, 2014–2017

   ➤ Integrating Wordnet senses and ERG/JACY/MCG/Indogram
   ➤ New RA (Luis), new PhD (Lê Tuấn Anh)

➤ *Acquisition of Physical Action Verbs by Bilingual Singaporean Preschoolers* $377,351, 2014–2017 (PI: Helena Gao)

➤ *Semi-Automatic Implementation of Clinical Practice Guidelines in Singapore hospitals* $150,000, 2014–2017 (PI: Kim JungJae)

   ➤ Match Doctor's reports to Singapore Clinical Practice Guidelines
   ➤ Experimenting with using the ERG to parse

➢ *Grammar Matrix Reloaded: Syntax and Semantics of Affectedness* $312,492, 2014–2017 (PI: František Kratochvíl)

➢ How does the degree of Affectedness affect how a participant is realised?

  * Different case
  * Argument or adjunct
  * Marking on the verb
  * Marking on the role (emphatic particle)

➢ Looking at this in multiple languages

➢ Try to include relative phenomena in the MATRIX

➢ Big issue: how to link different roles
  I hit him vs *I hit at him*

➢ New RA (Michael)

➢ Workshop this year and next year (co-located with HPSG/DELPH-IN?)

➤ Subcat acquisition for Japanese

➤ Images for lexeed/wordnet

➤ Joint research involved mainly discussions

Other Related Talks

➢ Bilingual alignment of MRS subgraphs (Michael; 15+10)

➢ Open Multilingual WordNet (Luis; 5+5)

➢ NTU Multi-lingual Corpus and cross-lingual fun (Francis; 15+10)

➢ Shallow NLP tools (lemmatizer, tokenizer, chunker) with ACE and ERG (Liling; 5+5)

➢ Discussion: Lexical Semantics, Connecting to 3rd Party Resources (Moderator: Francis)

➢ Discussion Shallow tools, Deep grammar (Moderator: Liling)

➢ Discussion: Corpus look up tools and approaches, integration with LTDB (Moderator: Francis)