

# A Head-driven Phrase Structure Grammar for Chinese

Zhong [ ]

**Fan** Zhenzhen (范 真真)  
Francis **Bond**, Sanghoun **Song** and many more

Linguistics and Multilingual Studies,  
Nanyang Technological University

<bond@ieee.org>

2016-06-16



A meta-grammar for Chinese ([wiki.delph-in.net/moin/ZhongTop](http://wiki.delph-in.net/moin/ZhongTop))

- model various varieties of Chinese in a single hierarchy
  - ▶ share some elements
    - ★ basic word order
  - ▶ some elements are separate
    - ★ lexemes
    - ★ some grammar rules
- **zhong.tdl** is inherited by all
  - ▶ **yue.tdl** is the Cantonese grammar
  - ▶ **cmn.tdl** is inherited by the Mandarin grammars
    - ★ **zhs.tdl** is the Mandarin grammar with simplified characters
    - ★ **zht.tdl** is the Mandarin grammar with traditional characters

# Overview

- Data: Penn Chinese Treebank, HKCanCor, Academica Sinica Corpus, NTU-MC (<http://compling.hss.ntu.edu.sg/ntumc/>)
- Tools: LKB, ACE, ITSDB
- Preprocessors: Stanford Tools; ZPAR (segmentizer, POS tagger) default lexical entries for unknown words
- Lexical Acquisition: about 25,000 items in Mandarin Chinese extracted from corpora
- Phenomena Coverage: major constructions in Chinese
- Corpus Coverage: very, very low

# What's new?

- Increased the coverage of Zhong (mainly zhs and yue)
  - ▶ Relative Clauses
  - ▶ Basic clause structure
  - ▶ Pumping rules (NP-PP, ...)
  - ▶ Better reduplication
  - ▶ Many small fixes
- Full-Forest Treebank of the MRS test suite
- DEMOPHIN with pre-processing
- ZZ will switch to full-time next semester and come to UW  
Focused development for the home stretch!

# Where do we go from here?

- Keep modeling major Chinese language phenomena
  - ▶ Complement structures and the role of DE (得)
  - ▶ Serial Verb Constructions
  - ▶ Subordinate clauses
- Improve the unknown word handling of Zhong to make it more robust in parsing unseen data
- Investigate how different syntactic structures of verbs in Chinese sentences influence the affectedness of participants
- Test the Zhong grammar with phenomena-based test suites and coverage-based test suites to ensure sufficient coverage gain.
- Build a prototype Chinese to English machine translation system with Zhong and the ERG (an English HPSG resource grammar), and evaluate its performance on test corpora
- Treebank parts of the NTU-MC Chinese corpus and train a statistical ranking model

# Thanks

- The initial version of Zhong was based on ManGO (by Justin Yang and Dan Flickinger), which used the Grammar MATRIX (Bender et al., 2002) to generate the initial grammar.
- We have benefited greatly from discussions within the NTU computational linguistics group, we especially thank Min Huan, Bo Chen, Huizhen Wang, Joanna Sio, Luis Morgado Da Costa, Michael Goodman, Shan Wang, Wenjie Wang, as well the wider DELPH-IN community, in particular Yi Zhang and Rui Wang
- This research is partly supported by the Singapore MOE ARF Tier 2 grants *That's what you meant: A Rich Representation for Manipulating Meaning* (MOE2013-T2-1-084) and *Grammar Matrix Reloaded: Syntax and Semantics of Affectedness* (MOE2013-T2-2-016)