

Status of ZHONG [|], a Chinese HPSG Shared-Grammar

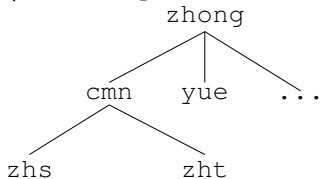
Computational Linguistics Lab
School of Humanities, Nanyang Technological University, Singapore
Zhenzhen Fan, Francis Bond, and many more

2017 Aug 7th



ZHONG [] (中)

- Chinese - a family of various languages including Mandarin Chinese, Cantonese, Min, etc.
- ZHONG [] is a shared HPSG grammar for Chinese (Fan et al., 2015) (wiki.delph-in.net/moin/ZhongTop)



- With a large portion of shared grammatical components
- Differ in certain syntax rules, lexicon, and orthography (simplified and traditional characters)

Collective efforts on developing Zhong

- Main contributors: Dan Flickinger, Justin Yang Chunlei, Fan Zhenzhen, Sanghoun Song, Francis Bond
- NTU Computational Linguistics group: Joanna Sio, Luis Morgado da Costa, Wenjie Wang, . . .
- DELPH-IN community, particularly Emily Bender, Michael Goodman, Yi Zhang, Rui Wang, . . .
- And many more who have given us great suggestions in discussions.

Current Status of Zhong

- 2840 types in multiple-inheritance monotonic hierarchy
- 116 leaf lexical types
- 38258 lexemes
- 58 syntactic rules
- 4 morphological rules

Grammar Enhancement

- What have been enhanced in the grammar:
 - ▶ DE constructions: relativizing DE, possessive DE, associative DE, nominalizing DE, Shi-DE(cleft) structures
 - ▶ Interrogatives
 - ▶ Reimplementation of reduplications using regex pattern matching and lexical rules, without using irreg.s.tab
 - ▶ Particles, fragments, interjections, coordination, vocatives, topic-comment, NP structures, argument structure, adpositions,
- Grammar for syntax error detection: MAL rules for several common grammatical errors made by second language learners
- A new test-suite: cmnedu
 - ▶ 798 sentences from Chinese textbooks for second language learners
 - ▶ Current coverage: 93.48%
 - ▶ Treebanking in progress

Where do we go from here?

- Continue with grammar enhancement: serial verb constructions, subordinate clauses, etc
- Expand the MAL-extension of the grammar and test on learners' errors corpus
- Long sentences causing out-of-memory issue
- Train a statistical ranking model from CMNEDU treebank
- System: Wordnet incorporation, transfer rules, machine translation,

Thanks!