Status Report: Mandarin Chinese Grammar (MCG)

Yi Zhang

LT-Lab, German Research Center for Artificial Intelligence

Computational Linguistics, Saarland University, Germany
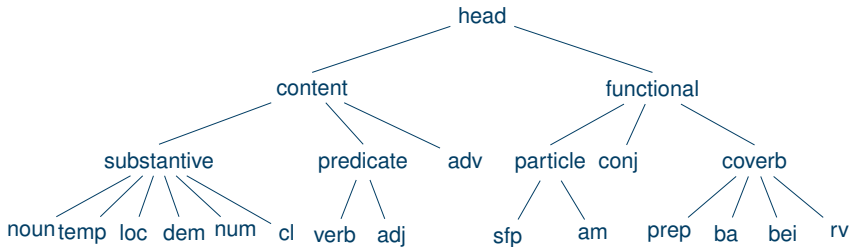
DELPH-IN

Overview

- A Chinese grammar has been under development at DFKI Saarbrücken since summer 2010
- This talk gives an ovewview of the current grammar development progress and short-term future plan
- A preview release is available for download at
  `http://www.coli.uni-saarland.de/~yzhang/mcg.php`

General Approach

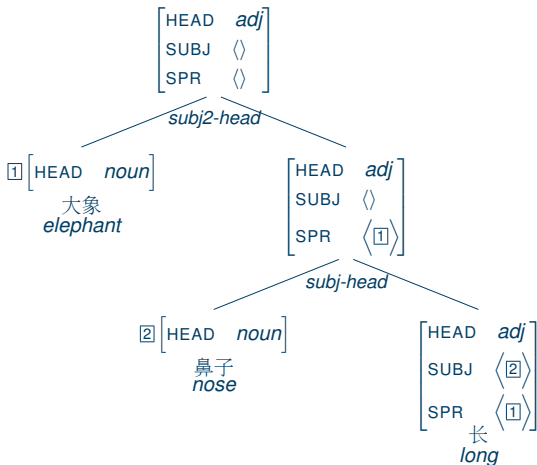- Based on Matrix customization (a few rounds of *trial and error*)
- Linguistic analyses were drawn from several grammar books
    - [Zhu, 1982]
    - [Li and Thompson, 1989]
- Analyses were reformulated in the framework of HPSG
- Other literature and implementation solutions were consulted
    - [Bender, 2000] [Gao, 2000] [Ng, 1997]
    - [Guo, 2009] [Yu et al., 2010] [Tse and Curran, 2010] [Müller and Lipenkova, 2009]

Head Types

```
                              head
                    ┌──────────┴──────────┐
                 content                functional
            ┌───────┼───────┐       ┌──────┼──────────┐
      substantive predicate adv  particle conj     coverb
     ┌──┬──┬──┬──┐   ┌──┬──┐      ┌──┬──┐        ┌──┬──┬──┐
  noun temp loc dem num  cl verb adj      sfp  am    prep ba bei rv
```

- ▶ Chinese specific head type hierarchy to accommondate for the lexical ambiguity between verbs, adjectives and nouns.

Topic constructions

$$
\begin{bmatrix} \text{HEAD} & adj \\ \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\rangle \end{bmatrix}
$$

subj2-head

$\boxed{1}\begin{bmatrix} \text{HEAD} & noun \end{bmatrix}$

大象
elephant

$$
\begin{bmatrix} \text{HEAD} & adj \\ \text{SUBJ} & \langle\rangle \\ \text{SPR} & \langle\boxed{1}\rangle \end{bmatrix}
$$

subj-head

$\boxed{2}\begin{bmatrix} \text{HEAD} & noun \end{bmatrix}$

鼻子
nose

$$
\begin{bmatrix} \text{HEAD} & adj \\ \text{SUBJ} & \langle\boxed{2}\rangle \\ \text{SPR} & \langle\boxed{1}\rangle \end{bmatrix}
$$

长
long

► Topic-comment structure is captured by SUBJ2-HEAD schema, and saturates the SPR list of the predicate.

DE (的)-constructions

- DE-constructions are divided into two major types
  - Associative DE express an "associative" (possessive) relation between the noun phrase before and after DE
  - Nominalizing DE creates a nominal phrase from the S, S/NP or VP in front of DE
    - Subject-gapping DE phrase
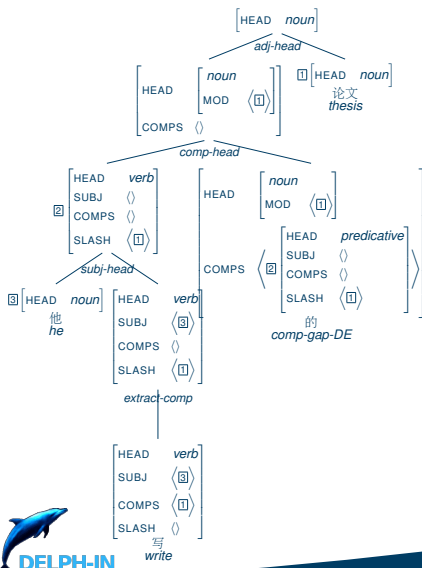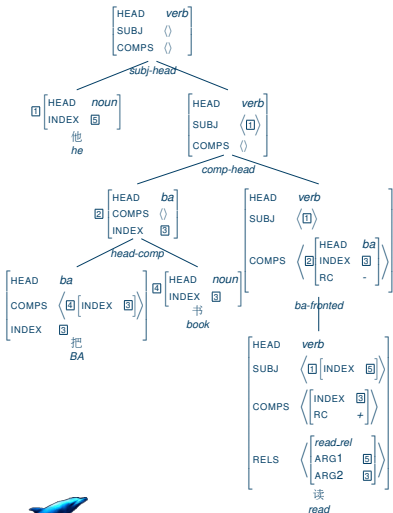   - Complement-gaping DE phrase
    - Non-gapping DE phrase

DE (的)-constructions

Associative DE

DE (的)-constructions

Subject-gapping nominalizing DE

DE (的)-constructions

Complement-gapping nominalizing DE

DE (的)-constructions

Non-gapping nominalizing DE

$$\begin{bmatrix} \text{HEAD} & noun \end{bmatrix}$$

adj-head

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} noun \\ \text{MOD} & \langle \boxed{1} \rangle \end{bmatrix} \\ \text{COMPS} & \langle \rangle \end{bmatrix}$$

$$\boxed{1}\begin{bmatrix} \text{HEAD} & noun \end{bmatrix}$$

时间
time

comp-head

$$\boxed{2}\begin{bmatrix} \text{HEAD} & verb \\ \text{SUBJ} & \langle \boxed{3} \rangle \\ \text{COMPS} & \langle \rangle \\ \text{SLASH} & \langle \rangle \end{bmatrix}$$

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} noun \\ \text{MOD} & \langle \boxed{1} \rangle \end{bmatrix} \\ \text{COMPS} & \langle \boxed{2}\begin{bmatrix} \text{HEAD} & predicative \\ \text{SLASH} & \langle \rangle \end{bmatrix} \rangle \end{bmatrix}$$

的
no-gap-DE

head-comp

$$\begin{bmatrix} \text{HEAD} & verb \\ \text{SUBJ} & \langle \boxed{3} \rangle \\ \text{COMPS} & \langle \boxed{4} \rangle \\ \text{SLASH} & \langle \rangle \end{bmatrix}$$

$$\boxed{4}\begin{bmatrix} \text{HEAD} & noun \end{bmatrix}$$

俄
Russia

访
visit

BA (把)-constructions

HEAD *verb*
SUBJ ⟨⟩
COMPS ⟨⟩
subj-head

① HEAD *noun*
INDEX ⑤
他
he

HEAD *verb*
SUBJ ⟨①⟩
COMPS ⟨⟩
comp-head

② HEAD *ba*
COMPS ⟨⟩
INDEX ③
head-comp

HEAD *ba*
COMPS ⟨④[INDEX ③]⟩
INDEX ③
把
BA

④ HEAD *noun*
INDEX ③
书
book

HEAD *verb*
SUBJ ⟨①⟩
COMPS ⟨②[HEAD *ba* / INDEX ③ / RC −]⟩
ba-fronted

HEAD *verb*
SUBJ ⟨①[INDEX ⑤]⟩
COMPS ⟨[INDEX ③ / RC +]⟩
RELS ⟨[*read_rel* / ARG1 ⑤ / ARG2 ③]⟩
读
read

- BA is treated as a special category *coverb*
- Rule BA-FRONTED moves the last right complement (direct-object) to the left of the verb

DELPH-IN

BEI (被)-constructions
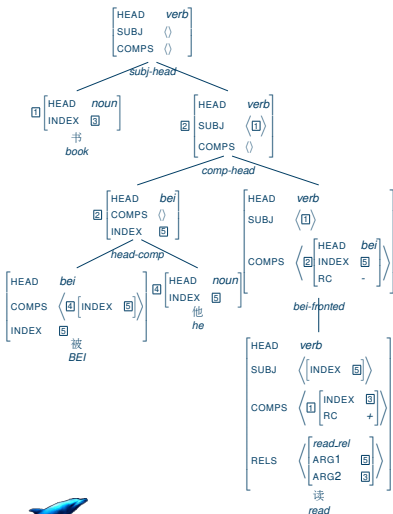
- Similar to BA, BEI is also treated as *coverb*
- Unary rule BEI-FRONTED switches the indices of the syntactic subject and the direct object, moves the complement to the preverbal position

(Tree diagram of typed feature structures)

```
            ┌ HEAD   verb ┐
            │ SUBJ   ⟨⟩   │
            │ COMPS  ⟨⟩   │
            └             ┘
               subj-head
      ┌─────────────┴─────────────┐
  1 ┌ HEAD  noun ┐      2 ┌ HEAD   verb  ┐
    │ INDEX  3   │        │ SUBJ   ⟨1⟩   │
    └            ┘        │ COMPS  ⟨⟩    │
      书                 └               ┘
     book                   comp-head
                  ┌────────────┴────────────┐
          2 ┌ HEAD  bei ┐        ┌ HEAD  verb ┐
            │ COMPS ⟨⟩  │        │ SUBJ  ⟨1⟩  │
            │ INDEX 5   │        │            ┌ HEAD  bei ┐ │
            └           ┘        │ COMPS ⟨2 │ INDEX 5   │⟩│
              head-comp          │            │ RC    -   │ │
    ┌──────────┴──────────┐      └            └           ┘ ┘
┌ HEAD  bei ┐   4 ┌ HEAD  noun ┐       bei-fronted
│ COMPS ⟨4 INDEX 5⟩│ INDEX  5  │      ┌ HEAD  verb ┐
│ INDEX 5   │      └           ┘      │ SUBJ ⟨[INDEX 5]⟩ │
└           ┘         他              │        ┌ INDEX 3 ┐ │
   被                 he              │ COMPS ⟨1│ RC    + │⟩│
   BEI                               │        └          ┘ │
                                     │        ┌ read_rel ┐ │
                                     │ RELS ⟨│ ARG1  5   │⟩│
                                     │        │ ARG2  3   │ │
                                     └        └           ┘ ┘
                                          读
                                          read
```

DELPH-IN

Other Phenomena

- Resultative verb compound
- Aspect marking (着, 了, 过)
- Sentence final particle (question, 了)
- Numeral-classifier & demonstrative
- Basic serial verb constructions
- …

Preview Release 20110615

`http://www.coli.uni-saarland.de/~yzhang/mcg.php`

- ▶ Developers: Yi Zhang, Rui Wang, Yu Chen
- ▶ About 200 newly added types
- ▶ Over 500 manually compiled lexical entries
- ▶ A (Chinese) phenomena oriented test-suite with over 700 positive & negative items (only 129 included in the release)
- ▶ Released under the LGPL-LR (excluding the test suite)
- ▶ A reference publication available soon

DELPH-IN

Future Plan

- Validation of MRS
- Corpus-based lexical acquisition
- Coverage boosting
- Treebanking
- CFG approximation
- Collaborations with other sites

DELPH-IN

References I

Bender, E. M. (2000).
The syntax of mandarin ba: Reconsidering the verbal analysis.
Journal of East Asian Linguistics, 9(2):105–145.

Gao, Q. (2000).
Argument Structure, HPSG, and Chinese Grammar.
PhD thesis, The Ohio State University.

Guo, Y. (2009).
Treebank-Based Acquisition of Chinese LFG Resources for Parsing and Generation.
PhD thesis, School of Computing, Dublin City University.

Li, C. N. and Thompson, S. A. (1989).
Mandarin Chinese: A functional reference grammar.
University of California Press, London, England.

References II

Müller, S. and Lipenkova, J. (2009).
Serial verb constructions in chinese: A hpsg account.
In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, pages 234–254, Germany.

Ng, S. K. (1997).
A double-specifier account of chinese nps using head-driven phrase structure grammar.
Master's thesis, Department of Linguistics. University of Edinburgh.

Tse, D. and Curran, J. R. (2010).
Chinese ccgbank: extracting ccg derivations from the penn chinese treebank.
In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1083–1091, Beijing, China.

DELPH-IN

Yu, K., Yusuke, M., Wang, X., Matsuzaki, T., and Tsujii, J. (2010).
Semi-automatically developing chinese hpsg grammar from the penn chinese treebank for deep parsing.
In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1417–1425, Beijing, China.

Zhu, D. (1982).
语法讲义.
商务印书馆.

DELPH-IN