

Korean Resource Grammar: Status Quo and Future Directions

Jong-Bok Kim & Sanghoun Song
Kyung Hee Univ. & UW
jongbok@khu.ac.kr & sanghoun@uw.edu

DELPH-IN 2010 Summit
Paris, 07-02-2010

- has been developed since 2003 with the aim of building an open source grammar of Korean.
- main developers: Jong-Bok Kim and Jaehyung Yang
- focusing on linguistically precise and theoretically motivated descriptions of Korean
- covered major constructions in Korean such as case marking, word ordering

Weak points of the KRG 1

- no generation modular
- not covering naturally occurring data
- low parsing efficiency

- for the last two years (2009 – 2010) with the help of Sanghoun Song and Francis Bond
- started using shared grammar libraries (the Grammar Matrix) and data-driven expansion (using the Korean portions of multilingual texts)
- added the generation modular

Main improvements

- fully customizing the KRG to the Grammar Matrix system
- exploiting naturally occurring texts as the generalization corpus.
- improving/revising the grammar for the generation modular

Revisions and Improvements

- Modifying the Modular Structure: streamlining .tdl files
- Expanding the test suites: 6,180 grammatical sentences, 118 ungrammatical sentences
- Simplifying the inflectional system
- Adding the sentence style feature

Using more natural-occurring data

- *Sejong Bilingual Corpora*
- *Basic Travel Expression Corpus* (BTEC)

Table: Comparison between KRG1 and KRG2

	KRG1	KRG2
# of default types	121	159
# of lexical types	289	593
# of phrasal types	58	106
# of inflectional rules	86	244
# of syntactic rules	36	96
# of lexicon	2,297	39,190

Parsing and Generation Coverage

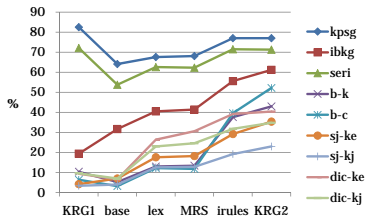


Figure: Parsing Coverage

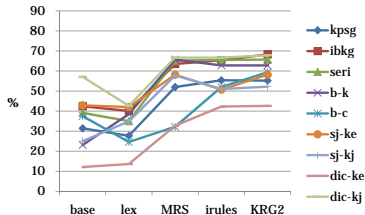


Figure: Generation Coverage

Parsing and Generation Performance

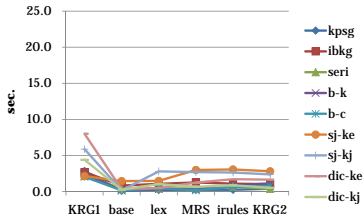


Figure: Parsing Performance

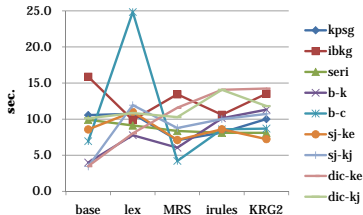


Figure: Generation Performance

Grammar Released and Site Update

- homepage: <http://krg.khu.ac.kr>
- demo: <http://krg.khu.ac.kr:8106/logon>
- The 8th Workshop on Asian Language Resources (in conjunction with COLING 2010)
 - Development of the Korean Resource Grammar: Towards Grammar Customization
 - (Sanghoun Song, Jong-Bok Kim, Francis Bond and Jaehyung Yang)

- Cleft constructions
- Comparative constructions
- Serial verb constructions

- Korean HPSG/MRG Treebanks
 - building the International Corpus of Korean (role-modeling the International International Corpus of English, Great Britain)
 - a collection of the KRG-parsed sentences
- MT baseline: Korean \leftrightarrow English/Japanese
- a collaborative work with Emily Bender, Francis Bond, and others.