

# Projections from Morphology to Syntax in the Korean Resource Grammar: Implementing Typed Feature Structures

Jong-Bok Kim<sup>1</sup> and Jaehyung Yang<sup>2</sup>

<sup>1</sup> School of English, Kyung Hee University, Seoul, Korea 130-701

<sup>2</sup> School of Computer Engineering, Kangnam University, Kyunggi, Korea, 449-702

**Abstract.** Korean has a complex inflectional system, showing agglutinative morphology and using affixation as the major mechanism for word formation. A prerequisite to the successful development of any syntactic/semantic parsers for the language thus hinges on the efficient lexicon that can syntactically expand its lexical entries and map into syntax and semantics with robust parsing performance. This paper reports the system of the Korean Resource Grammar developed as an extension of HPSG (Head-driven Phrase Structure Grammar) and the results of implementing it into the Linguistic Knowledge Building (LKB) system (cf. Copestake 2002). The paper shows that the present grammar proves to be theoretically as well as computationally efficient enough in parsing Korean sentences.

## 1 Korean Resource Grammar

The Korean Resource Grammar (KRG) is a computational grammar for Korean currently under development since October 2002 (cf. Kim and Yang 2003). Its aim is to develop an open source grammar of Korean. The grammatical framework for the KRG is the constraint-based grammar, HPSG (cf. Sag, Wasow, and Bender 2003). HPSG (Head-driven Phrase Structure Grammar) is built upon a non-derivational, constraint-based, and surface-oriented grammatical architecture. HPSG seeks to model human languages as systems of constraints on typed feature structures. In particular, the grammar adopts the mechanism of type hierarchy in which every linguistic sign is typed with appropriate constraints and hierarchically organized. The characteristic of such typed feature structure formalisms facilitates the extension of grammar in a systematic and efficient way, resulting in linguistically precise and theoretically motivated descriptions of languages including Korean. The concept of hierarchical classification is essentially assigning linguistic entities such as phrases and words to specific types, and an assignment of those types to superordinate types. Each type is declared to obey certain constraints corresponding to properties shared by all members of that type. This system then allows us to express cross-classifying generalizations about phrases and words, while accommodating the idiosyncracies of individual types on particular subtypes of phrases or words.

As the basic tool for writing, testing, and processing the Korean Resource Grammar, we adopt the LKB (Linguistic Knowledge Building) system (Copes-