

Parsing Korean Case Phenomena in a Type-Feature Structure Grammar

Jong-Bok Kim¹ and Jaehyung Yang²

¹ School of English, Kyung Hee University, Seoul, Korea 130-701

² School of Computer Engineering, Kangnam University, Kyunggi, 449-702, Korea

Abstract. For a free-word order language such as Korean, case marking remains a central topic in generative grammar analyses for several reasons. Case plays a central role in argument licensing, in the signalling of grammatical functions, and has the potential to mark properties of information structure. In addition, case marking presents a theoretical and computational test area for understanding the properties of the syntax-morphology interface of the language. This is why it is no exaggeration to say that parsing Korean starts from work on the case system of the language. This paper reports the case system of the Korean Phrase Structure Grammar (KPSG) developed as a Korean resource grammar for computational purposes and implemented in the Linguistic Knowledge Building (LKB) system. The grammar adopts the constraint-based mechanisms of feature unification and multiple inheritance type hierarchies as an extension of HPSG (Head-driven Phrase Structure Grammar) and is proved to be empirically and computationally sound and efficient.

1 Two Basic Issues

Nominal expressions in Korean can carry discourse as well as case markers that could indicate their grammatical, semantic, and discourse functions. One simple example would suffice to show the complexity of its case system:³

- (1) *sensayngnim-tul-pwuthe-ka ku chayk-ul sangca-ey-to neh-ess-ta*
teacher-PL-SRC-NOM the book-ACC box-LOC-also put-PST-DECL
'To only teachers, such a letter is also delivered.'

As noted here, the NOM and ACC assigned by the verbal predicate indicate the syntactic functions (subject and object) of the nominals they are attached to. Cases like LOC express the semantic function of the NP 'box'. The delimiter marker *-to* 'also' can be attached to the semantic case marker, LOC. A further complication arises from the co-occurrence of the grammatical case NOM with the semantic case SRC as in *sensayng-tul-pwuthe-ka*.⁴

³ The abbreviations we use here are AGT (agentivity), ARG-ST (argument-structure), COMP (complementizer), COMPS (complements), Conj (conjunctive), NOM (Nominative), ACC (accusative), DAT (dative), HON (honorific), LEX (lexical), ORTH (orthography), SRC (source), PST (Past), Pl (plural), SYN (syntax), X-Delim (X-delimiter), Z-Delim (Z-delimiter), etc.

⁴ The KPSG grammar classifies case markers into two groups: GCASE and SCASE. The GCASE includes *nom* and *acc* as a subtype of *vcase* (cases assigned by a verbal element) and *gen* as a subtype of *ncase* (case assigned by a nominal element), whereas the SCASE includes *dat*, *goal*, *src*, *inst*, etc.