

Deep Processing of Honorification Phenomena in a Typed Feature Structure Grammar

Jong-Bok Kim
School of English
Kyung Hee U.
Seoul, 130-701
jongbok@khu.ac.kr

Peter Sells
Dept. of Linguistics
Stanford U.
Stanford, CA 94305
sells@stanford.edu

Jaehyung Yang
School of Computer Eng.
Kangnam U.
Kyunggi, 449-702, Korea
jhyang@kangnam.ac.kr

Abstract

Honorific agreement is one of the main properties of languages like Korean or Japanese, playing an important role in appropriate communication. This makes the deep processing of honorific information crucial in various computational applications such as spoken language translation and generation. We argue that, contrary to the previous literature, an adequate analysis of Korean honorification involves a system that has access not only to morphosyntax but to semantics and pragmatics as well. Along these lines, we have developed a typed feature structure grammar of Korean (based on the framework of HPSG), and implemented it in the Linguistic Knowledge Builder System (LKB). The results of parsing our experimental test suites show that our grammar provides us with enriched grammatical information that can lead to the development of a robust dialogue system for the language.

1 Basic Properties of Honorific Agreement

Honorification, one of the main features of spoken language in Korean, plays a key role in proper and successful verbal communication (Chang 1996, Lee 1998, Sohn 1999). The Korean honorific system basically requires that when the subject is in the honorific form (usually with the marker *-nim*),

the predicate also be inflected with the honorific form *-(u)si* as in (1):¹

- (1) a. *sensayng-nim-i wus-usi-ess-e.*
teacher-HON-NOM laugh-HON-PST-DECL
'The teacher laughed.'
b. *#sensayng-nim-i wus-ess-e.*

This type of agreement is often assumed to be purely pragmatic, mainly because certain contexts allow disagreeing cases between the subject and the verb: the utterance of (1)b can be felicitous when the speaker does not honor the referent of the subject (marked by #). The possibility of having such disagreement has often to an assumption in the literature that using the *-nim* and *-si* form of verbs is a matter of gradience and appropriateness rather than grammaticality (cf. Chang 1996, Pollard and Sag 1994, Lee 1998).

However, one often neglected fact is that this agreement constraint must be observed when the subject is non-human as in (2) (cf. Sohn 1999):

- (2) a. *cha-ka o-(*si)-ess-e.*
cha-NOM come-HON-PST-DECL
'The car came.'
b. *kwukhoy-ka pepan-ul simuy-ha-(*si)-ess-e.*
congress bill review-HON-PST-DECL
'The congress reviewed the bill.'

If we rely only on pragmatic information, we would have difficulty understanding why, in contrast to the disagreement in (1)b, disagreement like that in (2) are rarely found in real language usages.

¹Abbreviations we use in the paper include ARG (ARGUMENT), ACC (Accusative), BAKGR (BACKGROUND), COMP (Complementizer), CTXT (CONTEXT), DECL (Declarative), HON (Honorific), IMPER (Imperative), NOM (Nominative), ORTH (ORTHOGRAPHY), PST (Past), SYN (SYNTAX), SEM (SEMANTICS), RELS (RELATIONS), and POS (part of speech).

In addition, there exist agreement-sensitive syntactic phenomena such as auxiliary verb constructions:

- (3) a. *sensayng-nim-i nolay-lul*
 teacher-HON-NOM song-ACC
pwulu-si-ci anh-(usi)-ess-e.
 sing-HON-COMP not-HON-PST-DECL
 ‘The teacher did not sing a song.’
 b. *sensayng-nim-i ton-ul mo-(*si)-e*
 teacher-NOM money-ACC save-HON-COMP
twu-si-ess-e.
 hold-HON-PST-DECL
 ‘The teacher saved money (for rainy days).’
 c. *sensayng-nim-i nolay-lul*
 teacher-HON-NOM song-ACC
*pwulu-si-na po-(*si)-e.*
 sing-HON-COMP seem-HON-DECL
 ‘The teacher seems to sing a song.’

As noted here, even though the subject is honored in each case, the honorific marker on the main predicate in (3)a is optional with the auxiliary *anh-* ‘not’; in (3)b the marker must appear only on the auxiliary verb *twu-* ‘hold’; meanwhile in (3)c the marker cannot appear on the auxiliary *po* ‘seem’. Such clear contrasts, we can hardly attribute to pragmatic factors.²

2 Honorification in a Typed Feature Structure Grammar

A closer look at the honorific phenomena of the language in the previous section suggests that an adequate theory of honorification aiming for integration into a proper communication system requires not just complex pragmatic information but also morpho-syntactic information. The basic framework of the grammar we adopt for modelling the language is the type-feature structure grammar of HPSG. HPSG seeks to model human languages as systems of constraints on typed feature structures. In particular, the grammar adopts the mechanism of a type hierarchy in which every linguistic sign is typed with appropriate constraints and hierarchically organized. This system then allows us to express cross-classifying generalizations about linguistic entities such as lexemes, stems, words, and phrases in the language (cf. Kim and Yang 2004, Kim 2004).

²In addition to this subject-verb agreement, the language employs addressee agreement marked on a verbal suffix depending on the honoring relationship between speaker and addressee. Such agreement, though implemented in our grammar, is not presented here because of space limit here.

2.1 Lexicon and Subject Agreement

Our grammar, named KPSG (Korean Phrase Structure Grammar), first assumes that a nominal with *-nim* and a verbal with *-si* bear the head feature specification [HON +]. This is supported by the contrast in the following:

- (4) a. *[[haksayng-i manna-n] sensayng-nim-i]*
 student-NOM meet-MOD teacher-HON-NOM
o-si-ess-e.
 come-HON-PST-DECL
 ‘The teacher that the student met came.’
 b. *[[sensayng-nim-i manna-si-n]*
 teacher-NOM-NOM meet-HON-MOD
*haksayng-i] o-(*si)-ess-e.*
 student-NOM come-HON-PST-DECL
 ‘The student that the teacher met came.’

As seen here, it is the honorific information on the head noun *sensayng-nim* in (4)a that agrees with that of the verb.

With this head feature information, the grammar builds the honorific nominal type (*n-hon*) from the basic lexeme (*n-lxm*) as represented in the following feature structures:³

- (5) a.
$$\left[\begin{array}{l} n\text{-lxm} \\ \text{ORTH } \langle \text{sensayng} \rangle \text{ 'teacher'} \\ \text{SYN | HEAD } \left[\begin{array}{l} \text{POS noun} \\ \text{HON boolean} \end{array} \right] \\ \text{SEM } \boxed{2} \left[\begin{array}{l} \text{INDEX } i \\ \text{RELS } \left\langle \left[\begin{array}{l} \text{PRED teacher-rel} \\ \text{INSTANCE } i \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$
- b.
$$\left[\begin{array}{l} n\text{-hon} \\ \text{ORTH } \langle \text{sensayng-nim} \rangle \text{ 'teacher-HON'} \\ \text{SYN | HEAD } \left[\begin{array}{l} \text{POS noun} \\ \text{HON +} \end{array} \right] \\ \text{SEM } \boxed{2} \end{array} \right]$$

As seen in (5)a, a nominal lexeme with no honorific marker *-nim* is underspecified for the HON feature.⁴

Meanwhile, the subject of an honorific verbal element carries the feature [HON +] in addition to the relevant pragmatic information:

³The information our grammar encodes for such lexeme entries is only the shaded part: all the other information is inherited from its supertypes defined in the grammar. For a more comprehensive system of morphology built within such a system, see Kim (2004) and Kim and Yang (2004).

⁴The boxed number here is used as a way of showing that semantic value of the lexeme, *n-lxm* is identical with that of the honorific noun *n-hon*.

- (6)
- a.
$$\left[\begin{array}{l} v\text{-}lxm \\ ORTH \boxed{1} \\ SYN | HEAD \left[\begin{array}{l} POS \text{ verb} \\ HON \text{ boolean} \end{array} \right] \\ ARG\text{-}ST \left\langle NP[INDEX \ i], \dots \right\rangle \\ SEM \boxed{2} \end{array} \right]$$
- b.
$$\left[\begin{array}{l} v\text{-}hon \\ ORTH \boxed{1} + si \\ SYN | HEAD \left[\begin{array}{l} POS \text{ verb} \\ HON + \end{array} \right] \\ ARG\text{-}ST \left\langle NP \left[\begin{array}{l} HON + \\ INDEX \ i \end{array} \right], \dots \right\rangle \\ CTXT \left[\begin{array}{l} C\text{-}INDICES | SPEAKER \ p \\ BAKGR \left\langle \left[\begin{array}{l} PRED \text{ honoring} \\ ARG1 \ p \\ ARG2 \ i \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

The basic verbal lexeme type *v-lxm* in (6)a does not carry any restriction on its subject. However, as given in (6)b, the *v-hon* type with the *-(u)si* suffix adds the information that its subject (the first element in the ARG-ST (argument structure)) is [HON +], in addition to the information that the speaker is honoring the subject referent as given in the CTXT value.

One of the key points in this system is that even though the [HON +] verb selects a [HON +] subject, the subject of a nonhonorific verb can be either in the honorific or nonhonorific form since its value is underspecified with respect to the verb. This then correctly allows disagreeing examples like (1)b where the subject is [HON +] and the verb's HON value is 'boolean':

- (7) *sensayng-nim-i wuc-ess-e*. 'The teacher laughed.'

The nonhonorific verb combines with the honorific subject with no honoring intention from the speaker since the nonhonorific verb does not bear the pragmatic constraint that the speaker honors the referent of the subject.

Yet the grammar blocks disagreeing cases like (2) where an honorific verb combines a nonhonorific subject:

- (2) a. **cha-ka o-si-ess-ta*. 'The car came.'
 b. **kwukhoy-ka ku pepan-ul simuy-ha-si-ess-e*.
 'The congress reviewed the bill.'

These are simply not parsed since the honorific verb here would combine with the [HON -] subject, violating the constraint in (6)b. A noun

like *sensayng* 'teacher' is [HON boolean], while *sensayng-nim* is [HON +], and most nouns are [HON -].

2.2 Object and Oblique Agreement

While subject honorification has a productive suffixal expression, there are some lexically suppletive forms like *poyp-e* 'see.HON-DECL' and *mosi-e* 'take.HON-DECL', which require their object to be in the honorific form:

- (8) a. **John-i Mary-lul poyp-ess-e*.
 John-NOM Mary-ACC see.HON-PST-DECL
 'John honorably saw Mary.'
 b. *John-i sensayng-nim-ul poyp-ess-e*.
 John-NOM teacher-HON-ACC
 'John honorably saw the teacher.'

Our grammar lexically specifies that these suppletive verbs require the object to be [HON +] together with the pragmatic honoring relation. The following is the lexical information that a suppletive verb like this accumulates from the inheritance hierarchy:

- (9)
$$\left[\begin{array}{l} v\text{-}lxm \\ ORTH \langle poyp\text{-} \rangle \text{ 'HON.see' } \\ SYN | HEAD \boxed{1}[HON +] \\ ARG\text{-}ST \left\langle NP[INDEX \ i], NP \left[\begin{array}{l} HON + \\ INDEX \ j \end{array} \right] \right\rangle \\ SEM \text{ see-rel} \\ CTXT \left[\begin{array}{l} BAKGR \left\langle \left[\begin{array}{l} PRED \text{ honoring} \\ ARG1 \ i \\ ARG2 \ j \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

Such lexical information can easily block examples like (8)a where the object is [HON -].

Lexically suppletive forms like *tuli-e* 'give.HON-DECL' and *yeccup-e* 'ask.HON-DECL' require their oblique argument to be in the HON form (nonhonorific forms are *cwu-e* and *mwut-e*, respectively):

- (10) a. *John-i sensayng-nim-eykey senmwul-ul*
 John-NOM teacher-HON-DAT present-ACC
tuli-ess-e.
 give.HON-PST-DECL
 'John gave the present to the teacher.'
 b. **John-i haksayng-eykey senmwul-ul tuli-ess-e*.

Just like object agreement, our grammar assigns the HON restriction on its dative argument together with the pragmatic honoring constraint:

Type III: Unlike Type II, auxiliary verbs like *po-* ‘see’ and *kath-* ‘seem’ cannot have the honorific suffix *-si* even if the subject is in the honorific form:

- (18) a. *sensayng-nim-i chayk-ul ilk-usi-na
 teacher-NOM book-ACC read-HON-COMP
 po-si-ta.
 seem-DECL
 ‘The teacher seems to read a book.’
 b. sensayng-nim-i chayk-ul ilk-usi-na po-ta.
 c. #sensayng-nim-i chayk-ul ilk-na po-ta.
 d. *sensayng-nim-i chayk-ul ilk-usi-na po-si-ta.

First, the agreement in Type I simply follows from the general assumption that this kind of auxiliary verbs acts like a raising verb whose subject is identical with that of the main verb:

- (19) a. $\left[\begin{array}{l} \text{aux-v} \\ \text{ORTH } \langle \text{anh-a} \rangle \text{ 'not-DECL'} \\ \text{SYN} \mid \text{HEAD} \mid \text{AUX } + \\ \text{ARG-ST} \left\langle \left[\begin{array}{l} \text{1} \\ \text{2} \end{array} \right], \left[\begin{array}{l} \text{LEX } + \\ \text{ARG-ST} \langle \text{1}, \dots \rangle \end{array} \right] \right\rangle \\ \text{SEM } \textit{not-rel} \end{array} \right]$
 b. $\left[\begin{array}{l} \text{aux-hon-v} \\ \text{ORTH } \langle \text{anh-usi-e} \rangle \text{ 'not-HON-DECL'} \\ \text{SYN} \mid \text{HEAD} \left[\begin{array}{l} \text{AUX } + \\ \text{HON } + \end{array} \right] \\ \text{ARG-ST} \left\langle \left[\begin{array}{l} \text{1} \\ \text{HON } + \end{array} \right], \left[\begin{array}{l} \text{2} \end{array} \right] \right\rangle \\ \text{SEM } \textit{not-rel} \end{array} \right]$

The negative auxiliary verb with or without the *-(u)si* suffix selects as its arguments a subject and a lexical complement whose subject is identical with the auxiliary’s subject. This means when either one of the verbs requires an HON subject, then the combination of the main verb as a complex predicate will also require an HON subject.⁵

The absence of the HON on the main verb for the Type II AUX is due to the language’s morphological constraints. Such an auxiliary verb forms a verbal complex together with a main verb that bears the COMP suffix *-a/e*: this suffix morphologically requires its verb stem to have no honorific *-(u)si* (cf. Kim and Yang 2004). This morphological constraint can be attested by the fact that suppletive honorific form with no

⁵This treatment assumes that the auxiliary verb combines with the preceding (main or auxiliary) verb and forms a complex predicate. See Kim and Yang (2004) for this line of treatment.

productively-formed *-si* marking can occur in the Type II construction:⁶

- (20) a. sensayng-nim-i sakwa-lul tusi-e
 teacher-NOM apple-ACC HON.eat-COMP
 po-si-ess-e.
 try-HON-PST-DECL
 ‘The teacher tried to eat the apple.’
 b. sensayng-nim-i chayk-ul ilk-(*usi)-e
 teacher-NOM book-ACC read-HON-COMP
 po-si-ess-e.
 try-HON-PST-DECL
 ‘The teacher tried to read the book.’

Within the grammar we developed where each specific verb stem has its own type constraint, the stem value of the COMP suffix *-a/e* must be a verb lexeme with no suffix *-si*.

As for the Type III AUX, the grammar needs to rely on semantics: AUX verbs like *po-* ‘see’ and *kath-* ‘seem’ select propositions as their semantic argument:

- (21) $\left[\begin{array}{l} \langle \text{po-} \rangle \text{ 'see'} \\ \text{SYN} \mid \text{HEAD} \left[\begin{array}{l} \text{AUX } + \\ \text{HON } - \end{array} \right] \\ \text{ARG-ST} \langle \text{S}[\text{INDEX } s2] \rangle \\ \text{SEM} \left[\begin{array}{l} \text{INDEX } s1 \\ \text{RELS} \left\langle \left[\begin{array}{l} \text{PRED } \textit{seem-rel} \\ \text{ARG0 } s1 \\ \text{ARG1 } s2 \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$

The honoring relation applies not to a proposition but to a human individual: it is such a semantic property that places a restriction on the HON value of the auxiliary verb.

3 Testing the Feasibility of the Analysis

In testing the performance and feasibility of the grammar, we implemented our grammar in the LKB (Linguistic Knowledge Building) system (cf. Copestake 2002). The test suites we used consist of the SERI Test Suites ’97 (Sung and Jang 1997), the Sejong Corpus, and sentences from the literature on honorification. The SERI Test Suites (Sung and Jang 1997), designed to evaluate the performance of Korean syntactic parsers,

⁶The verb in Korean cannot be an independent word without inflectional suffixes. The suffixes cannot be attached arbitrarily to a stem or word, but need to observe a regular fixed order. Reflecting this, the verbal morphology has traditionally been assumed to be templatic:

- (i) V-base + (Passive/Causative) + (HON) + (TENSE) + MOOD

consists of total 472 sentences (292 test sentences representing the core phenomena of the language and 180 sentences representing different types of predicate). Meanwhile, the Sejong Corpus has 179,082 sentences with about 2 million words. We randomly selected 200 simple sentences (the average number of words in each sentence is about 5) from the corpus. These sentences are classified according to their honorification types (agreement target \times predicate) and the ratio of parsed sentences:⁷

(22)	(target) (predicate)	# of Sentences	# Parsed Sentences
	nonHON (tgt) \times nonHON (pred)	514 (76.4%)	455 (88.5%)
	HON (tgt) \times HON (pred)	64 (9.5%)	58 (90%)
	HON (tgt) \times nonHON (pred)	90 (13.3%)	82 (91%)
	nonHON (tgt) \times HON (pred)	4 (0.05%)	0 (0%)
	Total	672	595 (88.5%)

In addition to these sentences, we selected 100 sentences (including the ones given in the paper) from the literature on Korean honorification: 51 sentences with *-si* marked verbs, 31 with auxiliary verb constructions, and 18 with suppletive verb forms. We obtained similar results: the grammar parsed a total of 96 sentences.

Among the total of 691 parsed sentences, we checked the meaning representations (minimal recursion semantics: MRS) and the pragmatic representations of 100 randomly selected sentences, and could see that the representations contain the correct information that the grammar is designed for. We believe that the enriched deep processing of grammatical honorific information that the grammar successfully composed in the parsing process can well function for the proper understanding of natural data.

4 Conclusion

Honorification, one of the most salient features of the language, involves various grammatical levels

⁷The four nonHON \times HON sentences are cases where the nominals are not in the honorific form. One way to accept such examples is to remove the [HON +] restriction on the object of such verbs while keeping the pragmatic honoring relationship between the subject and object.

of information: morphology, syntax, semantics, and pragmatics. It is thus necessary for a parser to have not only shallow but also deep processing of the honorific information, so that we can check that a given sentence is felicitous. Such deep processing is a prerequisite to the success of dialogue processing, zero pronominal/anaphoric resolution, and so forth.

The grammatical architecture we adopt is a typed feature structure grammar, couched upon HPSG, that allows us to handle morpho-syntactic, semantic, and also pragmatic information. The implementation of this grammar in the LKB system proves that a type-feature structure grammar can provide us with a proper deep processing mechanism for Korean honorification that opens doors for promising applications in such areas as machine translation and dialogue systems.

References

- Chang, Suk-Jin. 1996. *Korean*. Amsterdam: John Benjamins.
- Copetake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.
- Kim, Jong-Bok and Jaehyung Yang. 2004. Projections from Morphology to Syntax in the Korean Resource Grammar: Implementing Typed Feature Structures. In *Lecture Notes in Computer Science* Vol. 2945: 13–24. Springer-Verlag.
- Kim, Jong-Bok. 2004. *Korean Phrase Structure Grammar (In Korean)*. Seoul: Hankwuk Publishing.
- Lee, Dong-Young. 1998. *Information-based processing of Korean dialogue with reference to English*. Seoul: Thaeahak Publishing.
- Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Sells, Peter. 1998. Structural Relationships within complex predicates. In B.-S. Park and J. Yoon (eds.), *The 11th International Conference on Korean Linguistics*. Seoul: Hankwuk Publishing, 115–147.
- Sohn, Ho-Min. 1999. *The Korean Language*. Cambridge: Cambridge University Press.
- Sung, Won-Kyung and Myung-Gil Jang. 1997. SERI Test Suites '95. In *Proceedings of the Conference on Hangeul and Korean Language Information Processing*.