

Parsing Korean Honorification Phenomena in a Typed Feature Structure Grammar

Jong-Bok Kim¹, Peter Sells², and Jaehyung Yang³

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

² Dept. of Linguistics, Stanford University, USA

³ School of Computer Engineering, Kangnam University, Kyunggi, 446-702, Korea

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 morpho-syntax 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 (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

1.1 Subject Agreement

Honorification, one of the main features of spoken language in Korean, plays a key role in proper and successful verbal communication ([?, ?, ?]).⁴ 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):⁵

- (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.

⁴ We thank three anonymous reviewers for their helpful comments and suggestions. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2005-042-A00056).

⁵ Abbreviations 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).

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 (1b) can be felicitous when the speaker does not honor the referent of the subject (marked by #). The possibility of having such disagreement has often led to an assumption in the literature that the cooccurrence of *-nim* on the subject and *-si* on the verb is a matter of gradience and appropriateness rather than grammaticality (cf. [?, ?, ?]).

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

- (2) a. cha-ka o-(*si)-ess-e.
 car-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.’

In both examples, the nonhuman subject does not allow the presence of the honorific marker *-si* in the verb. If we rely only on pragmatic information, we would have difficulty understanding why, in contrast to the disagreement in (1b), disagreement like that in (2) are rarely found in real language usages.

In addition, there exist agreement-sensitive syntactic phenomena such as auxiliary verb constructions. Consider examples with the negative auxiliary verb *anh-* ‘not’:

- (3) a. sensayng-nim-i nolay-lul pwulu-si-ci anh-(usi)-ess-e.
 teacher-HON-NOM song-ACC sing-HON-COMP not-(HON)-PST-DECL
 ‘The teacher did not sing a song.’
 b. sensayngnim-i ton-ul mo-(*si)-e twu-si-ess-e.
 teacher-NOM money-ACC save-HON-COMP held-(*HON)-PST-DECL
 ‘The teacher saved money (for rainy days).’
 c. sensayng-nim-i nolay-lul pwulu-si-na po-(*si)-e.
 teacher-HON-NOM song-ACC 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 (3a) is optional with the auxiliary verb *anh-* ‘not’; in (3b) the marker must appear only on the auxiliary verb *twu-* ‘hold’; meanwhile in (3c) the marker cannot appear on the auxiliary *po* ‘seem’. Such clear contrasts, we can hardly attribute to pragmatic factors.

1.2 Addressee Agreement

Matters become more complicated when we consider the agreement triggered by different types of verbal endings. Korean has at least two different endings depending on the honoring relationship between speaker and addressee (cf. [?]):

- (4) a. haksayng-i o-ass-e/o-ass-eyo.
 student-NOM come-PST-plain.DECL/come-PST-resp.DECL
 ‘The student came.’

- b. *sensayng-nim-i* *o-si-ess-e/o-si-ess-eyo*.
 teacher-HON-NOM come-HON-PST-plain.DECL
 ‘The teacher came.’

As noted here the verbal endings *-e* and *-eyo* are different with respect to addressee agreement. The ‘respectful declarative (resp.DECL)’ ending *-eyo* is used when the social status of the addressee is higher than that of the speaker. The data implies that not only the speaker but also the addressee plays a role in proper communication strategies with respect to the honorification system.

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 syntactic and semantic information. The basic framework of the grammar we adopt for modelling the language is the typed feature structure grammar of Head-Driven Phrase Structure Grammar. This framework, 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. [?, ?, ?]).

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:

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

As seen here, it is the honorific information on the head noun *sensayng-nim* in (5a) 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:⁶

⁶ The 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 [?, ?].

$$(6) \left[\begin{array}{l} n-lxm \\ \text{ORTH } \boxed{1} \langle \text{sensayng} \rangle \text{ 'teacher'} \\ \text{SYN | HEAD } \left[\begin{array}{l} \text{POS } noun \\ \text{HON } boolean \end{array} \right] \\ \text{SEM } \boxed{2} \left[\begin{array}{l} \text{INDEX } i \\ \text{RELS } \left\langle \left[\begin{array}{l} \text{PRED } \text{teacher-rel} \\ \text{INSTANCE } i \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

$$\text{a.}$$

$$\left[\begin{array}{l} n-hon \\ \text{ORTH } \boxed{1} + \langle \text{nim} \rangle \text{ 'teacher + HON'} \\ \text{SYN | HEAD } \left[\begin{array}{l} \text{POS } noun \\ \text{HON } + \end{array} \right] \\ \text{SEM } \boxed{2} \end{array} \right]$$

$$\text{b.}$$

As seen in (6a), a nominal lexeme with no honorific marker *-nim* is underspecified for the HON feature.⁷

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

$$(7) \left[\begin{array}{l} v-lxm \\ \text{ORTH } \boxed{1} \\ \text{SYN | HEAD } \left[\begin{array}{l} \text{POS } verb \\ \text{HON } boolean \end{array} \right] \\ \text{ARG-ST } \left\langle \text{NP} \left[\begin{array}{l} \text{INDEX } i \\ \dots \end{array} \right] \right\rangle \\ \text{SEM } \boxed{2} \end{array} \right]$$

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

The basic verbal lexeme type *v-lxm* in (7) does not carry any restriction on its subject. However, 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

⁷ The 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*.

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’:

- (1) a. *sensayng-nim-i wus-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 (6b). A noun like *sensayng* ‘teacher’ is [HON boolean], while *sensayng-nim* is [HON +], and canonical lexeme nouns are [HON –].⁸

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/po-ass-e*.
 John-NOM Mary-ACC *see.HON-PST-DECL/see-PST-DECL
 ‘John honorably saw Mary.’
 b. John-i sensayng-nim-ul *poyp-ess-e*.
 John-NOM teacher-HON-ACC see.HON-PST-DECL
 ‘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)
$$\left[\begin{array}{l} v\text{-}lxm \\ \text{ORTH } \langle \text{poyp-} \rangle \text{ ‘HON.see’} \\ \text{SYN | HEAD } \square[\text{HON +}] \\ \text{ARG-ST } \left\langle \text{NP}[\text{INDEX } i], \text{NP} \left[\begin{array}{c} \text{HON +} \\ \text{INDEX } j \end{array} \right] \right\rangle \\ \text{SEM } \textit{see-rel} \\ \text{CTXT } \left[\text{BAKGR} \left\langle \left[\begin{array}{c} \text{PRED } \textit{honoring} \\ \text{ARG1 } i \\ \text{ARG2 } j \end{array} \right] \right\rangle \right] \end{array} \right]$$

⁸ Nouns such as *taythonglyeng* (‘president’) are inherently [HON +] without the honorific marker *nim*.

Such lexical information can easily block examples like (8a) where the object is [HON −].⁹

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 (non-honorific forms are *cwu-e* and *mwul-e*, respectively). This is why the nonhonored oblique argument *haksayng-eykey* ‘student-DAT’ in (10b) is not acceptable here:

- (10) a. John-i sensayng-nim-eykey senmwul-ul tuli-ess-e.
 John-NOM teacher-HON-DAT present-ACC 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:

$$(11) \left[\begin{array}{l} v\text{-}lxm \\ \text{SYN} | \text{HEAD} | \text{HON} + \\ \text{ARG-ST} \left\langle \left[\text{INDEX } i, [], \left[\begin{array}{c} \text{HON} + \\ \text{INDEX } k \end{array} \right] \right\rangle \\ \text{CTXT} | \text{BAKGR} \left\langle \left[\begin{array}{c} \text{PRED } \textit{honoring} \\ \text{ARG1 } i \\ \text{ARG2 } k \end{array} \right] \right\rangle \end{array} \right]$$

Once again the grammar rules out examples like (10b) in which the dative argument *haksayng-eykey* ‘student-DAT’ is nonhonorific. However, nothing blocks the grammar from generating examples like (12) where the dative argument *sensayng-nim-eykey* ‘teacher-HON-DAT’ is [HON +] even if the verb *cwu-* ‘give’ is in the nonhonorific (unspecified) form:

- (12) John-i sensayng-nim-eykey senmwul-ul cwu-ess-e. (≈(10)a)

2.3 Multiple Honorification

Given this system, we can easily predict that it is possible to have multiple honorific examples in which subject agreement cooccurs with object agreement:

⁹ Notice here that unlike the case with subject agreement, here the pragmatic background information involves the honoring relationship between the subject and the object. This implies that if there is a situation where the speaker honors the object, a given example can be felicitous. In fact, we find a corpus example like the following from our test suites:

- (i) *pro* chayk-ul mosi-ko ...
 pro book-ACC attend.to.HON-COMP
 ‘He/she attends to the books ...’

We leave the issue of dealing with such examples for future research.

- (13) ape-nim-i sensayng-nim-ul poyp-(usi)-ess-e.
 father-HON-NOM teacher-HON-ACC HON.see-(HON)-PST-DECL
 ‘The father saw the teacher.’

The honorific suffix *-si* on the verb here requires the subject to be [HON +] whereas the suppletive verb stem asks its object to be [HON +]. In such examples, the honorific marker in the verb can be optional or the verb can even be replaced by the nonsuppletive form *po-* ‘seem’. However, the grammar does not generate cases like the following:

- (14) a. *John-i sensayng-nim-ul poyp-usi-ess-e.
 John-NOM teacher-HON-ACC HON.see-HON-PST-DECL
 ‘John saw the teacher.’
 b. *ape-nim-i John-lul poyp-ess-e.
 father-HON-NOM John-ACC HON.see-HON-PST-DECL
 ‘The father saw the teacher.’

(14a) is ruled out since the HON form *-(u)si* requires the subject to be [HON +] whereas (14b) is ruled out since the suppletive form *poyp-* selects a [HON +] object. We also can see that oblique agreement can occur together with subject agreement:

- (15) a. eme-nim-i sensayng-nim-eykey senmwul-ul tuli-si-ess-e.
 mother-HON-NOM teacher-HON-DAT present-ACC give.HON-PST-DECL
 ‘Mother gave the teacher a present.’
 b. #eme-nim-i sensayng-nim-eykey senmwul-ul tuli-ess-e.
 c. #eme-nim-i sensayng-nim-eykey senmwul-ul cwu-(si)-ess-e.
 d. *John-i sensayng-nim-eykey senmwul-ul tuli-si-ess-e.
 e. *eme-nim-i John-eykey senmwul-ul tuli-si-ess-e.

Since the nonhonorific verb places no restriction on the subject, the grammar allows the disagreement in (15b) and (15c). However, (15d) and (15e) cannot be generated: the former violates subject agreement and the latter violates object agreement.

2.4 Agreement in Auxiliary Constructions

The present honorification system in the KPSG can offer us a streamlined way of explaining the agreement in auxiliary verb constructions we noted in section 1.1. Basically there are three types of auxiliaries with respect to agreement (see [?]):

Type I: In the construction with auxiliary verbs like *anh-* ‘not’, when the subject is in the honorific form, the honorific suffix *-si* can optionally appear either on the preceding main verb or on the auxiliary verb or on both:

- (16) a. sensayng-nim-i o-si-ci anh-usi-ess-e.
 teacher-NOM come-HON-COMP not.HON-PST-DECL
 ‘The teacher did not come.’

- b. *sensayng-nim-i o-si-ci anh-ass-e.*
- c. *sensayng-nim-i o-ci anh-usi-ess-e.*
- d. *#sensayng-nim-i o-ci anh-ass-e.*

Type II: When the head auxiliary verb is one like *po-* ‘try’, *twu-* ‘hold’, and *ci-* ‘become’, subject honorification occurs only on the auxiliary verb. That is, the preceding main verb with the specific COMP suffix form *-a/e* cannot have the honorific suffix *-si*:

- (17) a. **sensayng-nim-i John-ul cap-usi-e twu-si-ess-e.*
 teacher-HON-NOM John-ACC catch-HON-COMP do-HON-PST-DECL
 ‘The teacher hold John for future.’
- b. *sensayng-nim-i John-ul cap-a twu-si-ess-e.*
- c. **sensayng-nim-i John-ul cap-usi-e twu-ess-e.*
- d. *sensayng-nim-i John-ul cap-a twu-ess-e.*

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 po-si-ta.*
 teacher-HON-NOM book-ACC read-HON-COMP seem-HON-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-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:¹⁰

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

¹⁰ The semantic relation *not-rel* represents the predicate relation induced by the negative auxiliary verb *anh-*.

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.¹¹

The 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:

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

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. [?]). As can be seen from the above template, verb suffixes, attaching to the preceding verb stem or word, mark honorific, tense, and mood functions. COMP suffixes are classified depending on which slot they can occur here: for example the COMP suffix *a/e* occupies the HON slot:

- (21) a. *sensayng-nim-i sakwa-lul tusi-e po-si-ess-e.*
 teacher-NOM apple-ACC HON.eat-COMP try-HON-PST-DECL
 'The teacher tried to eat the apple.'
- b. *sensayng-nim-i chayk-ul ilk-(*usi)-e po-si-ess-e.*
 teacher-NOM book-ACC read-(*HON)-COMP 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-* 'seem' and *kath-* 'seem' select an event (*e1* or *e2*) as their semantic argument:

(22)
$$\left[\begin{array}{l} \langle \text{po-} \rangle \text{ 'see' } \\ \text{SYN | HEAD } \left[\begin{array}{l} \text{AUX } + \\ \text{HON } - \end{array} \right] \\ \text{ARG-ST } \langle \text{S[INDEX } \boxed{e2} \rangle \\ \text{SEM } \left[\begin{array}{l} \text{INDEX } \boxed{e1} \\ \text{RELS } \left\langle \left[\begin{array}{l} \text{PRED } \textit{seem-rel} \\ \text{ARG0 } \boxed{e1} \\ \text{ARG1 } \boxed{e2} \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

¹¹ This treatment assumes that the auxiliary verb combines with the preceding (main or auxiliary) verb and forms a complex predicate. See [?] for this line of treatment.

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.

2.5 Addressee Agreement

As noted in Section 1.1, Korean mood marking may also indicate an honoring relationship between the addressee and the speaker. Our grammar, in which the inflected verbal element is built from a basic verbal lexeme within a type hierarchy system (cf. [?]), systematically allows addition of this honoring relationship in the lexical information:

- (23)
- a.
$$\left[\begin{array}{l} v\text{-plain-decl} \\ \text{ORTH } \langle \text{o-si-ess-e} \rangle \text{ 'come-HON-PST-plain.DECL'} \\ \text{SYN } \boxed{1} | \text{HEAD } | \text{HON } + \\ \text{ARG-ST } \left\langle \text{NP} \begin{bmatrix} \text{HON } + \\ \text{INDEX } i \end{bmatrix} \right\rangle \\ \text{SEM } \boxed{3} \\ \text{CTXT} \left[\begin{array}{l} \text{C-INDICES } \begin{bmatrix} \text{SPEAKER } p \\ \text{ADDRESSEE } q \end{bmatrix} \\ \text{BAKGR} \left\langle \begin{array}{l} \begin{bmatrix} \text{PRED } \textit{honoring} \\ \text{ARG1 } p \\ \text{ARG2 } i \end{bmatrix} \\ \begin{bmatrix} \text{PRED } \textit{soc-higher} \\ \text{ARG1 } p \\ \text{ARG2 } q \end{bmatrix} \end{array} \right\rangle \end{array} \right] \end{array} \right]$$
- b.
$$\left[\begin{array}{l} v\text{-resp-decl} \\ \text{ORTH } \langle \text{o-si-ess-eyo} \rangle \text{ 'come-HON-PST-resp.DECL'} \\ \text{SYN } \boxed{1} | \text{HEAD } | \text{HON } + \\ \text{ARG-ST } \left\langle \text{NP} \begin{bmatrix} \text{HON } + \\ \text{INDEX } i \end{bmatrix} \right\rangle \\ \text{SEM } \boxed{3} \\ \text{CTXT} \left[\begin{array}{l} \text{C-INDICES } \begin{bmatrix} \text{SPEAKER } p \\ \text{ADDRESSEE } q \end{bmatrix} \\ \text{BAKGR} \left\langle \begin{array}{l} \begin{bmatrix} \text{PRED } \textit{honoring} \\ \text{ARG1 } p \\ \text{ARG2 } i \end{bmatrix} \\ \begin{bmatrix} \text{PRED } \textit{soc-higher} \\ \text{ARG1 } q \\ \text{ARG2 } p \end{bmatrix} \end{array} \right\rangle \end{array} \right] \end{array} \right]$$

The plain declarative ending adds the information that the speaker is higher than the addressee whereas the respective one the opposite relation. The treatment of address agreement follows the same vein as subject/object agreement.

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 Builder) (cf. [?]). The test suites we used consist of the SERI Test Suites '97 ([?]), the Sejong Corpus, and sentences from the literature on honorification. The SERI Test Suites ([?]), designed to evaluate the performance of Korean syntactic parsers, consists of a total of 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.¹²

(24)

(target) \times (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.¹³ 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.

¹² The 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.

¹³ The failed sentences are due to the unwritten parts of our grammar. For example, the current version of our grammar does not cover postposing, floating quantifiers, gapping, and so forth.

4 Conclusion

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

1. Chang, S.J.: Korean. John Benjamins, Amsterdam (1996)
2. Lee, D.Y.: Information-based processing of Korean dialogue with reference to English. Thaeak Publishing, Seoul (1998)
3. Sohn, H.M.: The Korean Language. Cambridge University Press, Cambridge (1999)
4. Pollard, C., Sag, I.A.: Head-Driven Phrase Structure Grammar. University of Chicago Press, Chicago (1994)
5. Lee, I., Ramsey, R.: The Korean Language. State University of New York Press (2000)
6. Kim, J.B., Yang, J.: Projections from morphology to syntax in the Korean resource grammar: implementing typed feature structures. In: Lecture Notes in Computer Science. Volume 2945. Springer-Verlag (2004) 13–24
7. Kim, J.B.: Korean Phrase Structure Grammar. Hankwuk Publishing, Seoul (2004) In Korean.
8. Sells, P.: Structural relationships within complex predicates. In Park, B.S., Yoon, J., eds.: The 11th International Conference on Korean Linguistics, Seoul, Hankwuk Publishing (1998) 115–147
9. Copestake, A.: Implementing Typed Feature Structure Grammars. CSLI Publications, Stanford (2002)
10. Sung, W.K., Jang, M.G.: Seri test suites '95. In: Proceedings of the Conference on Hangeul and Korean Language Information Processing. (1997)