

# Representing Honorifics via Individual Constraints

Sanghoun Song

Division of Linguistics and Multilingual Studies

Nanyang Technological University

Singapore

sanghoun@ntu.edu.sg

## Abstract

Within the context of grammar engineering, modelling honorifics has been regarded as one of the components for improving machine translation and anaphora resolution. Using the HPSG and MRS framework, this paper provides a computational model of honorifics. The present study incorporates the honorific information into the meaning representation system via Individual Constraints with an eye toward semantics-based processing.

## 1 Introduction

Honorific forms express the speaker's social attitude to others and also indicate the social ranks of the participants in the discourse and the intimacy. Because honorifics are crucial for using the language in a socially correct way, they have been studied in computational linguistics as well as theories of grammar. Particularly, using the honorific information improves anaphora resolution, and helps machine translation systems provide more natural-seeming output sentences (Mima et al., 1997; Siegel, 2000; Nariyama et al., 2005).

This paper provides a way of modelling honorifics within the formalism of grammar-based language processing. Building upon Head-driven Phrase Structure Grammar (Pollard and Sag, 1994, HPSG) and Minimal Recursion Semantics (Copestake et al., 2005, MRS), the present study suggests using Individual CONstraints (henceforth, ICONS) for representing honorifics from the perspective of multilingual processing.

This paper is structured as follows: Section 2 presents some background knowledge of the current study. Section 3 proposes using Individual Constraints for modelling the honorific system. Building upon the specification, Section 4 shows how honorific expressions can be translated across

different honorific types of languages. Section 5 reports a small experiment to see if the current model contributes to semantics-based processing.

## 2 Background

### 2.1 Forms of Expressing Honorifics

A cross-linguistic survey reveals that there are three ways of expressing honorifics (Agha, 1994; Ide, 2005): (i) pronouns, (ii) inflection, and (iii) suppletives. Different languages use a different range of honorific systems, but it appears that there exists a hierarchy in the system of honorification, as presented in Table 1. Note that some languages (e.g. English) use no honorific forms.

Table 1: Honorification hierarchy

no forms	<	pronouns	<	inflection	<	suppletives
English, ...						
Chinese, German, ...						
				Javanese, Hindi, ...		
						Japanese, Korean, ...

The most widespread linguistic phenomenon regarding honorific expressions can be found in the taxonomy of personal pronouns. In many languages, personal pronouns (particularly, second pronouns) are dualized, viz. ordinary (a.k.a. informal) forms and honorific (a.k.a. formal) forms. For example, Chinese employs two second personal pronouns: 你 *nǐ* and 您 *nín*. Both sentences provided in (1) convey a meaning like “What is your name?” in English.

- (1) a. 你/您 叫 什么 名字 ?  
nǐ/nín jiào shénme míngzi ?  
2.SG be.called what name PU
- b. #你/您 贵 姓 ?  
nǐ/nín guì xìng ?  
2.SG noble last.name PU [cmn]

(1a) is a plain way to ask someone's name, in which both pronouns can be felicitously used. In contrast, (1b) is a way of asking in a courteous manner, in which the use of 你 *nǐ* is inappropriate.

That is to say, the predicate in (1b) 貴姓 *guixing* is a marked expression in terms of honorification.

Some languages employ a more complicated honorific system. In Japanese, Korean, Javanese, Hindi, and some other languages, the inflectional paradigm is conditioned by the honorific relations between dialogue participants (Siegel, 2000; Ohtake and Yamamoto, 2001; Kim and Sells, 2007). For instance, in Japanese and Korean, if the subject is in the honorific form, the predicate is preferred to be in the honorific form, as exemplified in (2). Note that 先生 *sensei* ‘teacher’ is an honorific word, and the verbal form *o+STEM+ni naru* is used to signify honor to the subject.

- (2) 先生 は 本 を お読み になりました  
 sensei wa hon o o-yomi ni nari mashi-ta  
 teacher TOP book ACC HON-read become HON-PST  
 ‘The teacher read a book.’ [jpn] (Dalrymple, 2001, p. 18)

Other elements can also be marked with respect to honorification. When non-subjects (e.g. objects and obliques) are honored, the canonical verbal form in Japanese is *o+STEM+suru*. When the speaker wants to express an honor to the hearer in Japanese, a verbal ending *masu* is used as shown in the last word of (2). On the other hand, the nominal inflectional system is also influenced by honorifics, as exemplified in (3).

- (3) a. お元気 です か  
 o-genki desu ka  
 HON-good.health COP QUES  
 ‘How are you (honored)?’  
 b. 私 は 元気 です  
 watashi wa genki desu  
 1.SG TOP good.health COP  
 ‘I am fine.’ [jpn]

In (3a), the addressee is presumed to be an honored person, and thereby the prefix *o-* canonically co-occurs with *genki* ‘good.health’. In contrast, (3b) explains the speaker’s own status, and thereby the honorific marker *o-* does not appear.

Some languages, such as Korean and Japanese, make ample use of suppletive forms of honorification. For example, Japanese has three verbs for ‘eat’: 食べる *taberu* (neutral), 召し上がる *meshiagaru* ‘(An honoree) eats’, and 頂く *itadaku* ‘(An honorer) eats’. These suppletive forms can often be used with the inflectional forms mentioned before; for example, お-召し上がりになる *o-meshiagari-ni naru* ‘(Someone highly respected) eats’ (Nariyama et al., 2005, p. 93-94).

Therefore, a single expression can involve all three types of honorification, as shown in (4).

- (4) 책을 드리셨습니다.  
 chayk-ul tuli-si-ess-supni-ta  
 book-ACC give(HON)-HON-PST-HON-DECL  
 ‘(An honoree) gave a book (to another honoree).’  
 (The hearer is also an honoree.) [kor]

The verb in (4) contains three honorific forms for the object, the subject, and the addressee. The lexeme *드리-* *tuli-* is a suppletive counterpart of *주-* *cwu-* ‘give’. This verb implies the receiver is respectable. The second one is the suffix *-si-*, which indicates the subject is an honoree. The third one is the ending suffix *-supni-*, which indicates that the speaker expresses a respect to the hearer.

There are also nominal suppletive forms. The different lexical items that denote the same referent sometimes indicate the relative degree of familiarity to the referent. For example, kinship terms in Japanese vary depending on the relationship between the speaker and the referent: When talking about the speaker’s own grandfather with others in a modest attitude, 祖父 *sofu* is normally used. When either denoting the other’s grandfather or calling the speaker’s grandfather friendly and informally, お爺さん *o-jii-san* is normally used. This contrast shows that *o-jii-san* lexically involves an honorific information, whereas *sofu* is neutral. On the other hand, because *o-jii-san* can be used to denote both the other’s grandfather and the speaker’s own grandfather, the honorific information has to be flexibly represented so as to cover the two potential relations.

In addition to the forms discussed hitherto, some particular constructions, such as passives and interrogatives, can serve to express honorification. However, the meaning is just pragmatically conveyed in this case. Such a construction is not a necessary condition but a sufficient condition for expressing honorifics. Not all passive sentences in Japanese necessarily involve an honorific relation. In contrast, if the *o+STEM+ni naru* form in Japanese is used, then the subject is presumed to be an honoree. Since the current work is exclusively concerned with honorific forms, these constructions are out of the scope of this paper.

## 2.2 Motivations

Honorifics have often been regarded as agreement phenomena just as the subject-predicate agreement in many European languages (Boeckx, 2006;

Kim et al., 2006). However, there is an opposing view to this (Choe, 2004; Bobaljik and Yatsushiro, 2006; Kim and Sells, 2007). One counterexample is provided in (5).

- (5) 김선생님이 오(시)었다  
 Kim-sensayng-nim-i o-(si)-ess-ta  
 Kim-teacher-HON-NOM come-(HON)-PST-DECL

‘Teacher Kim came.’ [kor] (Choe, 2004, p. 546)

The subject of (5) contains an honorific form *님* *nim*, but the predicate optionally takes the honorific marker *시* *-si-* though the verb with the honorific marker sounds more natural. Along this line, the current study does not constrain honorification as a way of agreement.

There are also a couple of reasons for not following honorification-as-agreement. These reasons make it necessary to model honorifics as flexibly as possible.

First, honorification is a matter of tendency rather than restriction. Notice that tendency and restriction are not on a par with each other in grammar engineering. Corpus data provide more than a few cases in which a mismatch of honorific forms happens, as exemplified in (5). Grammar engineering systems must work robustly for even less frequent items if the forms appear in naturally occurring texts and unless they critically violate the principle of human language.

Second, honorification is a matter of acceptability rather than grammaticality. Acceptability is primarily concerned with appropriateness, whereas grammaticality confirms the linguistic rules mostly provided by linguists. Thus, acceptability distinguishes not grammatical and ungrammatical sentences, but felicitous and infelicitous ones. In a similar vein, Zaenen et al. (2004) argue that animacy is mainly relevant to acceptability: For instance, the choice between the Saxon genitive and the *of*-genitive in English is sensitive to animacy, but the difference has more to do with felicity. The same goes for honorification. The choice of honorific forms leads to a difference in acceptability which forms a continuous spectrum.

### 3 Individual Constraints on Honorifics

Minimal Recursion Semantics is the formalism employed to compute semantic compositionality in the present work. In addition, the current work employs ICONS (Individual CONStraints) in order to incorporate discourse-related phenomena into semantic representation of human language

sentences. The representation method used in the present study (i.e. MRS+ICONS) has to do with not only semantic information incrementally gathered up to the parse tree, but also other components required to be accessed in the process of cross-lingual processing. MRS+ICONS enables us to model several discourse-related items within an intrasentential system (i.e. sentence-based processing). Notice that there exist several discourse-related items that can be at least partially resolved without seeing adjacent sentences. This can be conceptualized in the format of Dependency MRS (Copestake, 2009), as exemplified in (6).

- (6) a. *John likes himself.*  
*x1 x2 [x1 eq x2]*  
 b. *John likes him.*  
*x1 x2 [x1 neq x2]*

*Himself* in (6a) equals the subject *John*, while *him* in (6b) does not. The notation in the bracket in each example indicates the relationship between two individuals: **equal** and **non-equal**. That is to say, anaphora can be partially identified within an intrasentential domain via such a binary relation. There are some other phenomena that require contextual information in theory but can be partially resolved in practice in a way similar to (6), and honorification is one of them.

The current work represents honorifics as a binary relation between two individual elements. A set of honorific information is stored into a bag of constraints, and the value is only partially specified unless there is a clue to identify the honorific relation within the intrasentential context.

#### 3.1 Comparison to Previous Approaches

On the one hand, MRS+ICONS makes honorification (basically a pragmatic information) visible in semantic representation with an eye toward semantics-based language processing. In the previous HPSG-based studies, honorifics are treated as a typed feature structure under CTXT (ConTeXT). This local structure includes C-INDICES whose components are SPEAKER and ADDRESSEE (Siegel, 2000; Kim et al., 2006). In the LFG-based studies, honorification is regarded as an F-structure, given that it is one of the reliable tests to diagnose subjecthood (Dalrymple, 2001). Outside the scope of grammar-based deep processing, several studies make use of shallow processing techniques, such as POS-based pattern matching rules and regular expressions, for paraphras-

ing honorific expressions (Ohtake and Yamamoto, 2001, among others). In sum, no previous approach represents honorifics into a (near) logical form. In the semantics-based processing, all components that have a part in transfer and generation must be accessed in semantic representation.

On the other hand, the current model provides computational flexibility for handling honorification. Many previous studies on honorification employ a syntactic and/or semantic feature [HON *bool*] (Kim et al., 2006, among others). However, this feature is sometimes misleading for computational processing of honorifics for three reasons. First, there exist more than a few mismatches between honorific forms in real texts written in Korean and Japanese (i.e. no honorification-agreement (Choe, 2004)). [HON *bool*] is too restrictive to analyze rather infelicitous but acceptable honorific expressions (§2.2). For example, the two types of second personal pronouns in Chinese are interchangeable in many cases as provided in (1a), and the use of the informal pronoun 你 *nǐ* in (1b) merely results in infelicity (not ungrammaticality). The current work deals with honorifics grounded upon the premise “parsing robustly, generating strictly” (Bond et al., 2008). All potential honorific forms can be parsed robustly and flexibly, but the generation outputs are made strictly and felicitously. Second, [HON *bool*] cannot fully reflect the fact that honorifics are sometimes ambiguous and the specific meaning can be incrementally resolved up to the parse tree (Kim and Sells, 2007). For example, お爺さん *o-jii-san* ‘elderly man’ in Japanese can be used either informally or formally, and the choice between them depends on syntactic configuration. The current work makes use of a type hierarchy to constrain honorifics (see Figure 1), which manipulates the potential ambiguity and identifies the meaning throughout unification of structures. Third, the Boolean feature is too crude to place different types of constraints on subjects, objects, and addressee. For instance, the verb of (4) (in Korean) includes three HON glosses, and they have different honorific relations. MRS+ICONS represents honorifics as a binary relation amongst individuals, such as speaker, hearer, and referents.

### 3.2 Fundamentals

MRS+ICONS is structured as shown in (7). The value type is *icons* whose components are IARG1 and IARG2. Since ICONS stands for a binary re-

lation between two individuals, their value type is *individual* (a supertype of *event* and *ref-ind*).

$$(7) \left[ \begin{array}{l} \text{mrs+icons} \\ \text{HOOK} \left[ \begin{array}{l} \text{hook} \\ \text{GTOP} \quad \text{handle} \\ \text{LTOP} \quad \text{handle} \\ \text{INDEX} \quad \text{individual} \\ \text{XARG} \quad \text{individual} \\ \text{ICONS-KEY} \quad \text{icons} \\ \text{SPEAKER-KEY} \quad \text{ref-ind} \\ \text{HEARER-KEY} \quad \text{ref-ind} \end{array} \right] \\ \text{RELS} \quad \text{diff-list} \\ \text{HCONS} \quad \text{diff-list} \\ \text{ICONS} \quad \left\langle ! \dots \left[ \begin{array}{l} \text{icons} \\ \text{IARG1} \quad \text{individual} \\ \text{IARG2} \quad \text{individual} \end{array} \right] \dots ! \right\rangle \end{array} \right]$$

On the other hand, the HOOK structure, which keeps track of the features that need to be externally visible upon semantic composition, has three additional attributes, viz. ICONS-KEY, SPEAKER-KEY, and HEARER-KEY. These features function like a pointer in the compositional construction of the semantic structure. They are required to mark the constituent analyzed as the speaker or the hearer of an utterance and deliver the information up to the parse tree. In particular, first and second personal pronouns specify this value as their own index (§3.4).

CTXT (under *local*) includes C-INDICES just as Jacy does (Siegel, 2000), but the names are different as presented in the following AVM. Note that the counterpart of “speaker” must be “hearer”, and that of “addressee” must be “addressor”. The value type is *ref-ind*, because the speaker and the hearer are also referential individuals.

$$(8) \left[ \begin{array}{l} \text{local} \\ \text{CAT} \quad \text{cat} \\ \text{CONT} \quad \text{mrs} \\ \text{CTXT} \left[ \begin{array}{l} \text{ACTIVATED} \quad \text{bool} \\ \text{PRESUP} \quad \text{diff-list} \\ \text{C-INDICES} \left[ \begin{array}{l} \text{SPEAKER} \quad \text{ref-ind} \\ \text{HEARER} \quad \text{ref-ind} \end{array} \right] \end{array} \right] \end{array} \right]$$

The values of SPEAKER and HEARER remain underspecified until an utterance is established. The typed feature structure of *utterance* is presented in (9), in which SPEAKER-KEY and HEARER-KEY under CONT (i.e. *mrs*) are co-indexed with SPEAKER and HEARER under CTXT. Unless the SPEAKER-KEY and the HEARER-KEY are assigned a specific value during the construction of the parse tree, the values are still left underspecified. If the value is not specified until an utterance is built up, that means that the speaker and the hearer cannot be identified within the intrasentential domain.

$$(9) \left[ \begin{array}{l} \text{utterance} \\ \text{UTTERED} + \\ \text{CONT} \left[ \begin{array}{l} \text{SPEAKER-KEY} \boxed{1} \\ \text{HEARER-KEY} \boxed{2} \end{array} \right] \\ \text{CTXT} \left[ \begin{array}{l} \text{SPEAKER} \boxed{1} \\ \text{HEARER} \boxed{2} \end{array} \right] \\ \text{ARGS} \left\langle \text{H} \left[ \begin{array}{l} \text{sat-or-frag} \\ \text{UTTERED} - \\ \text{INDEX} \boxed{3} \end{array} \right] \right\rangle \\ \text{ICONS} \left\langle \left[ \begin{array}{l} \text{addressor} \\ \text{IARG1} \boxed{1} \\ \text{IARG2} \boxed{2} \end{array} \right], \left[ \begin{array}{l} \text{addressee} \\ \text{IARG1} \boxed{2} \\ \text{IARG2} \boxed{3} \end{array} \right] \right\rangle \end{array} \right]$$

The *utterance* rule syntactically forms a non-branching root node, whose daughter is either a saturated sentence or a fragment (*sat-or-frag*). This pseudo phrase structure rule introduces two elements into the ICONS list, as shown at the bottom of (9). They are valued as *addressor* (i.e. speaker) and *addressee* (i.e. hearer). These ICONS elements play the key role to make dialogue participants visible in semantic representation. Their IARG1s are respectively co-indexed with SPEAKER and HEARER (i.e.  $\boxed{1}$  and  $\boxed{2}$ ), and the IARG2 are commonly co-indexed with the semantic head's INDEX of the utterance (i.e.  $\boxed{3}$ ). The main reason why they have a relation to the semantic head is that it is necessary to resolve the speaker/hearer scope in quotations. For example, (10) contains two different discourse frames, viz. inner frame and outer frame.

(10) “You have been cruelly used,” said Holmes.

The two different frames may have different speakers and different hearers. For instance, the speaker in the inner frame of (10) is Holmes, while that in the outer frame is the narrator of the story. In other words, (10) includes two different utterances, and each introduces its own *addressee* and *addressor* elements into the ICONS list (i.e. four ICONS elements, in total).

### 3.3 Type Hierarchy

Going into the details, the type hierarchy of *icons* for honorification is sketched out in Figure 1.

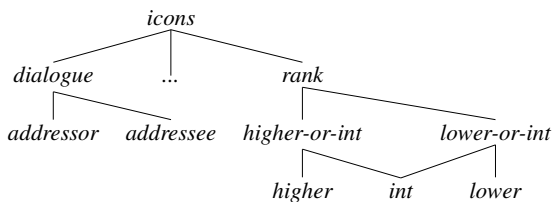


Figure 1: Type hierarchy of *icons*

Regarding honorification, *icons* includes two immediate subtypes: namely, *dialogue* and *rank*. The former branches out into *addressor* and *addressee*, and the latter includes two levels of subtypes. *Higher-or-int* indicates that one individual is socially higher than the other or intimate to the other. Recall that お爺さん *o-jii-san* in Japanese can be canonically used when the referent is higher than the speaker (formal) or intimate to the speaker (less formal). The word itself has the [ICONS-KEY *higher-or-int*] feature, which can be further constrained by the value that the predicate assigns to the word. Honorification is normally relevant to which is “higher” than which, but the linguistic forms can sometimes be altered when talking to someone in the lower position. For instance, Korean employs six levels of imperative inflections conditioned by the relationship between the speaker and the hearer. *Lower-or-int* and *lower* work for this case. Finally, note that *int* inherits from both *higher-or-int* and *lower-or-int*.

### 3.4 Specifications

First, pronouns are specified with respect to the speaker and the hearer, as shown in (11).

$$(11) \left[ \begin{array}{l} \text{pr-1sg} \\ \text{STEM} \langle \text{“我”} \rangle \\ \text{HOOK} \left[ \begin{array}{l} \text{INDEX} \boxed{1} \\ \text{S-KEY} \boxed{2} \end{array} \right] \\ \text{ICONS} \langle ! ! \rangle \end{array} \right] \left[ \begin{array}{l} \text{pr-2sg-hon} \\ \text{STEM} \langle \text{“您”} \rangle \\ \text{HOOK} \left[ \begin{array}{l} \text{INDEX} \boxed{1} \\ \text{S-KEY} \boxed{2} \\ \text{H-KEY} \boxed{3} \end{array} \right] \\ \text{ICONS} \left\langle ! \left[ \begin{array}{l} \text{higher} \\ \text{IARG1} \boxed{1} \\ \text{IARG2} \boxed{2} \end{array} \right] ! \right\rangle \end{array} \right]$$

The first personal pronoun has a co-index between its own INDEX and SPEAKER-KEY, and the ICONS list is empty because it does not contribute to honorification by itself. Likewise, the second personal pronouns link their INDEX to HEARER-KEY. If the pronoun is honorific, one ICONS element is introduced. Otherwise (e.g. 你 *ni*), the ICONS list is empty. The ICONS element of the right AVM indicates that the hearer  $\boxed{1}$  is higher than the speaker  $\boxed{2}$ .

Second, several inflectional rules introduce an ICONS element as exemplified in (12) for the subject-honorific form and the addressee-honorific form in Japanese. The left AVM's ICONS element represents that the subject  $\boxed{1}$  is higher than the speaker  $\boxed{3}$ . Likewise, the right AVM's ICONS element specifies the relation between the hearer  $\boxed{2}$  and the speaker  $\boxed{1}$ .

$$(12) \left[ \begin{array}{l} \text{ninaru} \\ \text{STEM} \langle \text{“に”, “なる”} \rangle \\ \text{SUBJ} \langle \left[ \begin{array}{l} \text{INDEX} \square \\ \text{I-KEY} \square \end{array} \right] \rangle \\ \text{S-KEY} \square \\ \text{ICONS} \langle \left[ \begin{array}{l} \text{!} \square \left[ \begin{array}{l} \text{higher} \\ \text{IARG1} \square \end{array} \right] \square \text{!} \\ \text{IARG2} \square \end{array} \right] \rangle \end{array} \right] \left[ \begin{array}{l} \text{masu} \\ \text{STEM} \langle \text{“ます”} \rangle \\ \text{S-KEY} \square \\ \text{H-KEY} \square \\ \text{ICONS} \langle \left[ \begin{array}{l} \text{higher} \\ \text{IARG1} \square \\ \text{IARG2} \square \end{array} \right] \rangle \end{array} \right]$$

Third, the suppletive forms themselves do not introduce an ICONS element, but the ICONS-KEY is specified in order to place a partial constraint on polarity. This pointer value functions similarly to [HON *bool*], but operates more flexibly (§3.1). They are instantiated in (13). Note that お休み *oyasumi* is a suppletive counterpart of 寝る *neru* ‘sleep’ in Japanese.

$$(13) \text{ a. } \left[ \begin{array}{l} \text{sofu} \\ \text{STEM} \langle \text{“祖父”} \rangle \\ \text{ICONS} \langle \text{!} \text{!} \rangle \end{array} \right] \left[ \begin{array}{l} \text{neru} \\ \text{STEM} \langle \text{“寝る”} \rangle \\ \text{ICONS} \langle \text{!} \text{!} \rangle \end{array} \right]$$

$$\text{ b. } \left[ \begin{array}{l} \text{ojiisan} \\ \text{STEM} \langle \text{“お爺さん”} \rangle \\ \text{I-KEY} \text{ higher-or-int} \\ \text{ICONS} \langle \text{!} \text{!} \rangle \end{array} \right] \left[ \begin{array}{l} \text{oyasumi} \\ \text{STEM} \langle \text{“お休み”} \rangle \\ \text{I-KEY} \text{ higher} \\ \text{ICONS} \langle \text{!} \text{!} \rangle \end{array} \right]$$

While the neutral forms provided in (13a) are underspecified, the honorific forms in (13b) place a constraint on ICONS-KEY. Notably, お爺さん *ojiisan* ‘elderly man’ in Japanese assigns *higher-or-int* to the ICNOS-KEY covering the ambiguity.

### 3.5 Sample Representation

The example sentence is illustrated in (14). Note that the nominative marker *が* *ga* and the two verbal ending forms are semantically empty.

$$(14) \begin{array}{ccccccc} \text{お爺さん} & \text{が} & \text{お休み} & \text{になり} & \text{ます} & & \\ \text{o-jiisan} & \text{ga} & \text{o-yasumi} & \text{ni nari} & \text{masu} & & \\ \text{HON-elderly.man} & \text{NOM} & \text{HON-sleep} & \text{become} & \text{HON} & & \\ \text{‘The elderly man is sleeping.’ (to an honoree) [jpn]} & & & & & & \end{array}$$

Therefore, only underlined elements are left in the semantic representation as shown in (15a). In addition, there are two invisible elements as provided in (15b), such as the speaker and the hearer. Recall that MRS+ICONS includes these invisible referential individuals into the semantics. These four individuals have four relations as presented in (15c), and they are added into the ICONS list.

$$(15) \text{ a. } \begin{array}{ll} \text{ojiisan} & \text{oyasumi} \\ x1 & e2 \end{array}$$

$$\text{ b. speaker: } x3, \text{ hearer: } x4$$

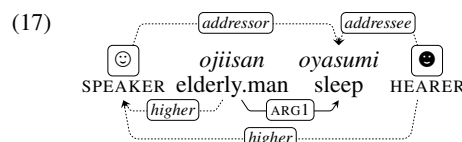
$$\text{ c. } \begin{array}{l} [x3 \text{ addressor } e2] \text{ (utterance)} \\ [x4 \text{ addressee } e2] \text{ (utterance)} \end{array}$$

$$\begin{array}{l} [x4 \text{ higher } x3] \text{ (ます } \textit{masu}) \\ [x1 \text{ higher } x3] \text{ (になる } \textit{ni naru}) \end{array}$$

Each relation given in (15c) is in the format as  $[\alpha \mathbf{X} \beta]$ , which is read as “ $\alpha$  has an  $\mathbf{X}$  relation to  $\beta$ ”. For instance,  $[x1 \text{ higher } x3]$  means that  $x1$  (i.e. the subject) is higher than  $x3$  (i.e. the speaker). The first two relations are introduced when the utterance is built up (see (9)). The last two relations came from the verbal ending forms (see (12)). The MRS representation for (14) is provided in (16), in which IARG1 and IARG2 respectively correspond to  $\alpha$  and  $\beta$  in the  $[\alpha \mathbf{X} \beta]$  format.

$$(16) \left[ \begin{array}{l} \text{INDEX} \square \\ \text{RELS} \left\langle \left[ \begin{array}{l} \text{udef}_q\_rel \\ \text{LBL} \square \\ \text{ARG0} \square \\ \text{RSTR} \square \\ \text{BODY} \square \end{array} \right], \left[ \begin{array}{l} \text{ojiisan}_n\_1\_rel \\ \text{LBL} \square \\ \text{ARG0} \square \end{array} \right], \left[ \begin{array}{l} \text{oyasumi}_s\_2\_rel \\ \text{LBL} \square \\ \text{ARG0} \square \\ \text{ARG1} \square \end{array} \right] \right\rangle \\ \text{HCONS} \left\langle \left[ \begin{array}{l} \text{qeq} \\ \text{HARG} \square \\ \text{LARG} \square \end{array} \right] \right\rangle \\ \text{ICONS} \left\langle \left[ \begin{array}{l} \text{addressor} \\ \text{IARG1} \square \\ \text{IARG2} \square \end{array} \right], \left[ \begin{array}{l} \text{addressee} \\ \text{IARG1} \square \\ \text{IARG2} \square \end{array} \right], \left[ \begin{array}{l} \text{higher} \\ \text{IARG1} \square \\ \text{IARG2} \square \end{array} \right], \left[ \begin{array}{l} \text{higher} \\ \text{IARG1} \square \\ \text{IARG2} \square \end{array} \right] \right\rangle \end{array} \right]$$

The traditional representation (16) can be converted into a dependency graph for ease of exposition. In (17), ☺ and ☹ tentatively stand for the dialogue participants.



The solid line in (17) means that the relation is specified in the RELS list. The dotted line stands for the ICONS element. The relational value is labelled on the arrow, and the direction of the arrow indicates which individual is co-indexed with which IARG. For instance, the arrow from *ojiisan* to ☺ means the same as  $[x1 \text{ higher } x3]$  presented in (15c) and the last ICONS element of (16).

## 4 Translating Honorifics

With respect to translating honorific expressions across languages, there are different types of translation strategies. Notice that paraphrasing is regarded as a specific type of translation (i.e. monolingual translation) in the current study, given that it is also carried out via the same procedure consisting of parsing, (transfer), and generation.

First, if both the source language and the target language have a complex honorific system (e.g. Japanese→Japanese), all ICONS elements gathered in the parsing stage persist in the transfer and generation stage. The four sentences in Japanese provided in (18) convey a meaning like “Did you/someone sleep?” in English, but the preference in the choice hinges on the social relation. The felicity condition is presented in Table 2.

Table 2: Choice of (18a-d)

	higher subject	plain subject
higher hearer	(18a)	(18b)
plain hearer	(18c)	(18d)

- (18) a. お休み になりましたか？  
 o-yasumi ni nari mashi ta ka ?  
 HON-sleep become HON PST QUES PU
- b. 寝ましたか？  
 ne mashi ta ka ?  
 sleep HON PST QUES PU
- c. お休み になった？  
 o-yasumi ni nat ta ?  
 HON-sleep become PST PU
- d. 寝た？  
 ne ta ?  
 sleep PST PU [jpn]

The monolingual translation (i.e. paraphrasing) is carried out as follows: First, if a suppletive form is used (e.g. *oyasumi*), the corresponding form in the generation output should be the same because the suppletive form is more informative and specific than the neutral form. If suppletives are converted into neutrals, loss of information happens. Notice that the suppletive forms normally have different PRED names as shown earlier in (15-17). For this reason, (18a-b) and (18c-d) are not interchangeable. Second, if there is an element in the ICONS list of the input MRS, the element persists in the output MRS in order not to lose a piece of information. In other words, a completely underspecified output for each ICONS element is not allowed in generation. For instance, both (18a) and (18c) use *oyasumi* (suppletive), but (18a) cannot be paraphrased into (18c) because (18a) has one more honorific relation in the ICONS list (i.e. *ni naru*). The same goes for (18b) and (18d): Since *masu* in (18b) makes the sentence more informative than (18d), (18b) cannot be paraphrased into (18d). Third, the opposite direction is acceptable. Even if a constituent introduces no ICONS element, the output can include an honorific constituent. For instance, (18c) can be paraphrased into (18a), and (18d) can be paraphrased into (18b). Translating from a less infor-

mative form to a more informative form is plausible because there is no discarded information.

Second, if the source language has rich honorifics, and the target language places an honorific constraint on only pronouns (e.g. Japanese→Chinese), the ICONS elements are selectively transferred: The element not linked to pronouns are filtered out. In the opposite direction, the underspecified ICONS element in the input MRS can be resolved in the output as discussed above. For instance, the subject and the hearer in (19) is the honorific second pronoun 您 *nín* in Chinese. (19) cannot be translated into (18c-d) in which *masu* does not show up (see (12)). In contrast, all sentences given in (18) can be translated into (19) without loss of information. Recall that the current model analyzes the neutral form as underspecified (not [HON -]).

- (19) 您 睡 了 吗 ？  
 nín shuì le ma ?  
 2.SG(HON) sleep PERF QUES PU [cmn]

Third, if the source language employs rich honorification and the target language has no honorific form (e.g. English), the transfer system turns off ICONS. For example, (18a-d) are commonly translated into “Did you sleep?” in English. The other direction (e.g. English→Japanese) raises no problem because all underspecified elements are restored on the target language’s side.

## 5 Experiment

In order to verify whether the current model works for semantics-based processing, one experiment was conducted with ACE (<http://sweaglesw.org/linguistics/ace>). The HPSG used for this experiment is the Jacy (Siegel and Bender, 2002). The basic analysis for honorific expressions in Japanese discussed in this paper was implemented. The testset was the first 4,500 sentences in the Tanaka corpus (Tanaka, 2001). Using these resources, paraphrasing in Japanese (i.e. monolingual translation) was carried out with the 5-best option for parsing and 512MB memory capacity. After paraphrasing was completed, the two results (i.e. without or with ICONS) were compared, as provided in Table 3. The comparison was made with respect to (A) the average output numbers, (B) the number of the items with end-to-end-success, and (C) the number of the items with exact-match-output out of (B).

Table 3: Evaluation

(A, #)		(B, %)		(C, %)	
plain	ICONS	plain	ICONS	plain	ICONS
132.67	280.98	50.38	45.40	59.64	67.89

Table 3 presents that the current model aids in producing more precise outputs, as indicated in (C): The translation accuracy grows by 8.25%. The number of outputs (A) also grows because all potential forms of expressing honorifics are generated without loss of information: All ambiguous interpretations are generated as long as the information is provided in the semantic representation. The end-to-end-success rate (B) decreases, but it is mainly due to the memory limitation, not the model itself: If the size of generated outputs exceeds the given value of memory limitation (512MB in the current experiment), all the outputs are ignored in comparison. If a bigger value is chosen, this rate also increases though it takes much longer time to yield the outputs.

## Acknowledgments

I am very grateful to Francis Bond, Jae-Woong Choe, Yasunari Harada, Jong-Bok Kim, Michael Wayne Goodman, David Moeljadi, and Zhenzhen Fan for their help and comments. This research was supported in part by the MOE Tier 2 grant *That's what you meant: a Rich Representation for Manipulation of Meaning* (MOE ARC41/13).

## References

- Asif Agha. 1994. Honorification. *Annual Review of Anthropology*, pages 277–302.
- Jonathan David Bobaljik and Kazuko Yatsushiro. 2006. Problems with Honorification-as-Agreement in Japanese: A Reply to Boeckx & Niinuma. *Natural Language & Linguistic Theory*, 24(2):355–384.
- Cedric Boeckx. 2006. Honorification as Agreement. *Natural Language & Linguistic Theory*, 24(2):385–398.
- Francis Bond, Eric Nichols, Darren Scott Appling, and Michael Paul. 2008. Improving Statistical Machine Translation by Paraphrasing the Training Data. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 150–157, Hawaii.
- Jae-Woong Choe. 2004. Obligatory Honorification and the Honorific Feature. *Studies in Generative Grammar*, 14(4):545–559.
- Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation*, 3(4):281–332.
- Ann Copestake. 2009. Slacker Semantics: Why Superficiality, Dependency and Avoidance of Commitment can be the Right Way to Go. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 1–9, Athens.
- Mary Dalrymple. 2001. *Lexical Functional Grammar*. Academic Press, New York.
- Sachiko Ide. 2005. How and Why Honorifics can Signify Dignity and Elegance. In Robin Tolmach Lakoff and Sachiko Ide, editors, *Broadening the Horizon of Linguistic Politeness*, pages 45–64. John Benjamins Publishing.
- Jong-Bok Kim and Peter Sells. 2007. Korean Honorification: a Kind of Expressive Meaning. *Journal of East Asian Linguistics*, 16(4):303–336.
- Jong-Bok Kim, Peter Sells, and Jaehyung Yang. 2006. Parsing Korean Honorification Phenomena in a Typed Feature Structure Grammar. In Luc Lamontagne and Mario Marchand, editors, *Advances in Artificial Intelligence*, pages 254–265. Springer.
- Hideki Mima, Osamu Furuse, and Hitoshi Iida. 1997. A Situation-based Approach to Spoken Dialog Translation Between Different Social Roles. In *Seventh International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-97*, pages 176–183, Santa-Fe.
- Shigeko Nariyama, Hiromi Nakaiwa, and Melanie Siegel. 2005. Annotating Honorifics Denoting Social Ranking of Referents. In *Proceedings of the 6th International Workshop on Linguistically Interpreted Corpora (LINC-05)*, pages 91–100, Jeju.
- Kiyonori Ohtake and Kazuhide Yamamoto. 2001. Paraphrasing Honorifics. In *Workshop Proceedings of Automatic Paraphrasing: Theories and Applications*, pages 13–20, Jeju.
- Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago, IL.
- Melanie Siegel and Emily M. Bender. 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization*, pages 1–8, Taipei.
- Melanie Siegel. 2000. Japanese Honorification in an HPSG Framework. In *Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation*, page 289–300, Tokyo.
- Yasuhito Tanaka. 2001. Compilation of a Multilingual Corpus. In *Proceedings of the PACLING*, pages 265–268, Kyushu.
- Annie Zaenen, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M Catherine O’Connor, and Tom Wasow. 2004. Animacy Encoding in English: why and how. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pages 118–125, Stroudsburg.