

Annotating Honorifics Denoting Social Ranking of Referents

Shigeko Nariyama^{*♥}, Hiromi Nakaiwa^{*}, Melanie Siegel^{*}

Communication Science Lab, NTT Kyoto, Japan

^{*}{shigekon, hiromi}@cslab.kecl.ntt.co.jp

[♥]The University of Melbourne, Australia shigeko@unimelb.edu.au

^{*}DFKI GmbH Saarbrücken, Germany siegel@dfki.de

Abstract

This paper proposes an annotating scheme that encodes honorifics (respectful words). Honorifics are used extensively in Japanese, reflecting the social relationship (e.g. social ranks and age) of the referents. This referential information is vital for resolving zero pronouns and improving machine translation outputs. Annotating honorifics is a complex task that involves identifying a predicate with honorifics, assigning ranks to referents of the predicate, calibrating the ranks, and connecting referents with their predicates.

1 Introduction

To varying extents, languages have ways to reflect the speaker's deference towards the addressee and people being referred to in utterances (c.f. Brown and Levinson 1987): by adopting a more polite air or tone of voice, avoiding coarse language, and modifying one's choice of specific vocabulary. This is prominent in Asian languages, Japanese and Korean in particular, which exhibit an extensive use of honorifics (respectful words).

Morphologically for example, French has a choice of the familiar *tu* and the formal *vous* (a third person plural) for the second person referent. Similarly Greek has the same choice: *esei* and *eseis* respectively. European languages commonly project one's deference by the use of different personal pronouns and titles (e.g. Mr., Dr., and Hon.).

Japanese and Korean, on the other hand, have numerous ways to say 'I' or 'you' calibrated by social position, age, gender and other factors. The projection of honorifics extends over the vocabulary of verbs, adjectives, and nouns as

well as sentence structures, to elevate a person or humble oneself. (1) and (2) below from Japanese are such examples, which use honorific verbs instead of neutral forms *kuru* 'come', *iku* 'go/accompany', and *motomeru* 'seek':

(1) いらっしゃったら、お供する。

Irasshat-tara, otomosuru.

come-when accompany

'When (an honouree) comes, (an honourer) accompanies (the honoree).'

(2) 援助を仰いだ。

Enjo-o aoida.

help-OB sought

'(A lower ranked person) turns to (a higher ranked person) for help.'

Examples (1) and (2) also reveal the notorious problem of zero pronoun resolution in Japanese, where the subject and the object of a sentence are frequently left unexpressed (Nakaiwa 2002, Nariyama 2003, inter alia). It is clear from the examples that coding the honorific relations of referents provides vital information for identifying what zero pronouns refer to; namely, to know whether or not a predicate denotes disparity of social rank between referents and to identify the rank of the referents. This is what this paper proposes to do. Siegel (2000) reported that 23.9% of Japanese zero pronouns in task-oriented dialogues can be resolved using information gleaned from honorification.

Coding of honorifics also improves machine translation outputs into Japanese in choosing the correct predicate depending on the relationship of the referents. Inappropriate use of honorifics, in particular the use of the plain form where an honorific form should be used, is rude and can be offensive.

Section 2 reviews some earlier work on this topic in NLP; Section 3 elaborates on honorifics; Section 4 formulates the ranking factors; Section 5 proposes a way to assign ranks to referents; Section 6 discusses a way to calibrate rankings of referents, as ranks are relative to the ranks of other referents in the sentence; Section 7 describes our annotation scheme, and finally conclusion in Section 8.

2 Earlier studies

The Japanese honorification system has been studied extensively in linguistics, particularly in sociolinguistics. Because of its importance and frequent use in the Japanese language, there has been some related work in NLP; within the framework of grammar formalism, GPSG by Ikeya (1983), JPSG by Gunji (1987), and more recently HPSG by Siegel (2000); work from a view point of resolving zero pronouns in dialogues by Dohsaka (1990).

Of these, the most thorough work on Japanese honorification is seen in JACY, a Japanese HPSG grammar (Siegel 2000, Siegel and Bender 2002). It extends the BACKGR (*owe – honour*) relation (Pollard and Sag 1994), which accounts only for subject honorifics, to accommodate the other types of honorification used in Japanese (see Section 3.1 for the types).

The full account of the Japanese honorification system requires syntactic and pragmatic information in many dimensions, with more input from the latter, the gathering of which is an extremely convoluted task. This paper builds on the basics from JACY and complements it in two ways to extend the JACY annotation presented in Section 7.

1. Ranking referents in social hierarchy
2. Calibrating the ranks

Regarding 1, honorifics tell which referent is higher in rank, so each referent must be assigned a rank to make use of honorific information. This is crucial when generating sentences to assign appropriate forms of honorific nouns and predicates in machine translation output into Japanese. In processing, ranking referents is not usually of importance when referents are overt, but it is when referents are zero pronouns. The identification of zero pronouns relies heavily on

the honorific information conveyed in the predicates.

Regarding 2, social rank is not absolute, but relative, so that the same referent may be higher or lower depending on which referent it appears with in a sentence. For example, the president of a company is socially regarded as ranked higher than the managers, who are in turn higher than clerks, but this rank is outweighed when their clients come in the sentence, in which case the president is ranked lower than their clients.

3 Honorifics

Honorifics is a term used to represent words that convey esteem or respect. Extensive studies on Japanese honorification revealed many forms of honorifics in use. The use of honorification is mandatory in many social situation.¹ Hence, every sentence can be viewed as coded for honorification if we consider the lack of an honorific marking as a sign that there is no hierarchical difference between referents.

Types of honorifics that indicate who is shown respect are described in Subsection 3.1, and forms of honorifics in Subsection 3.2.

3.1 Types of honorifics

Honorifics in modern (post-war) Japanese are generally classified into the following three categories, depending on who is shown respect (Martin 1964, Matsumoto 1997, Nariyama 2003, inter alia). The first two types are often referred to as ‘propositional (referential) honorifics’.

- i. **Subject honorifics** (called *Sonkeigo* in Japanese): to elevate or show respect towards the subject of a sentence
- ii. **Non-subject honorifics** (called Humility, or *Kenjogo*): to humble oneself by showing respect to the non-subject referent, generally the object
- iii. **Addressee honorifics** (‘polite’, *Teineigo*): to show respect towards the listener

Note that the expressions of deference are by nature made essentially with human referents, i.e. between an honouree and an honourer.

¹ However, sometimes honorification is uncoded even for respected referents, especially when the respected person is not present at the site of an utterance.

However, paying respect often extends to things and events related to the honouree in Japanese. This is often expressed with an honorific prefix *o-* or *go-* to the nouns. For example, the passage (3) is used by train conductors for ticket inspection. The use of non-subject honorific form means that the unexpressed subject (i.e. the train conductor) is showing respect towards the tickets, which belong to or have some relation to his honourees (i.e. the passengers).

(3) 切符を拝見します。

Kippu-o haikenshi-masu.

ticket-OB look[NsubH]-Polite

'(An honourer) is going to inspect (his honourees') tickets.' → 'Let me inspect your tickets, please.'

3.2 Forms of honorifics

Honorifics in Japanese take various forms that are reflected in the word forms, either in lexical choice or in inflections – verbs in Subsection 3.2.1, adjectives and nouns in Subsection 3.2.2, and also sentence structures in Subsection 3.2.3.

3.2.1 Verbs

There are five ways of expressing referent honorification in verbs, depending on the type of verb and the level of respect that is intended.² Types 3 displays the highest deference, and 3>2>1>4 in descending order. Type 5 displays a formality rather than deference towards the referent. The larger the gap in the hierarchy, the more disparity of referents in rank we expect.

3.2.1.1 Type 1: Alternation of verb forms

Verbs can be transformed into subject honorific (SubH) and non-subject honorific (NsubH) structures as follows:

SubH: *o* + verb stem + *ni naru* ('become').

NsubH: *o* + verb stem + *suru* ('do').

Accordingly, *machi* 'wait', for example, can be turned into two different forms of honorifics:

(4a) *O-machi-ni naru.* [SubH]

'(An honouree) waits (for someone/something).'

(4b) *O-machi-suru.* [NsubH]

'(An honourer) waits (for an honouree).'

The honorific prefix *o-* can be *go-*, as shown below. Basically, *o-* is used for Japanese native verbs and nouns and *go-* for Sino-Japanese (Chinese originated) words.

(5a) *Go-shichaku-ni naru.* [SubH]

'(An honouree) tries on (clothes).'

(5b) *Go-hookoku-suru* [NsubH]

'(An honourer) reports (to an honouree).'

3.2.1.2 Type 2: Suppletive forms

Different lexical items are used for some (more frequently used) verbs. For example, the following examples all mean 'ø eat':

(6a) *Taberu.* [non-honorific: neutral]

'(Someone) eats.'

(6b) *Meshiagaru.* [SubH]

'(An honouree) eats.'

(6c) *Itadaku.* [NsubH]

'(An honourer) eats.'

Table 1 shows some examples of other suppletive forms of honorification.

	<u>Neutral</u>	<u>SubH</u>	<u>NsubH</u>
do	<i>suru</i>	<i>nasaru</i>	<i>itasu</i>
exist/stay	<i>iru</i>	<i>irassharu/</i> <i>o-ide-ni-naru</i>	<i>oruru</i>
go	<i>iku</i>	<i>irassharu/</i> <i>o-ide-ni-naru</i>	<i>mairu/</i> <i>ukagau</i>
come	<i>kuru</i>	<i>irassharu/</i> <i>o-ide-ni-naru</i>	<i>mairu</i>
say	<i>iu</i>	<i>ossharu</i>	<i>moosu</i>
eat/drink	<i>taberu/nomu</i>	<i>meshiagaru</i>	<i>itadaku</i>

Table 1: Suppletive forms of honorification

Notice that some honorific forms are shared by very different meanings of verbs. For instance, *irassharu* can mean either 'come', 'go', or 'stay'. The nature of honorification is said to be indirect in expression. This semantic neutralization poses problems in machine translation outputs from Japanese.

² According to Wenger (1983:283-292), 70% of verbs have Subject honorific forms, while only 36% of verbs have non-subject honorific forms. He explains why not all verbs have forms of honorification, although he does not explain why there are fewer non-subject honorific forms. Honorification cannot occur, 1) unless the subject is human; this explains why there are no honorific forms for verbs such as *kooru* 'freeze' and *hoeru* 'bark'; and 2) on verbs that have negative connotations, such as *kuiarasu* 'eat greedily'.

3.2.1.3 Type 3: Combination of Types 1 & 2

This usage is restricted to some verbs, for example:

(7) *O-meshiagari-ni naru.*

'(Someone highly respected) eats.'

3.2.1.4 Type 4: Use of passive form *-rare*

The passive *-rare* is suffixed to the verb stem to display subject honorifics instead of the passive interpretation; for example:

(8) *Tabe-rare-ru.*

'(An honouree) eats.'

c.f. (6a) *Taberu.* [non-honorific: neutral]

'(Someone) eats.'

Note that there are no corresponding constructions of Types 3 and 4 for non-subject honorific forms.

3.2.1.5 Type 5: Lexical semantics

The semantics of some verbs give rise to referential restrictions, in that the subject must be higher or lower than the non-subject referent. This has been neglected in previous studies of honorification. Analogous to the example in (1), *insotsu* 引率 'take' has a restricted usage as '(a higher ranked person) leads (a group of lower ranked people).'

We used Lexeed (Bond et al. 2004) - a manually built self-contained lexicon, to extract verbs and verbal nouns with such referential restrictions. It consists of words and their definitions for the most familiar 28,000 words in Japanese, as measured by native speakers. This set is formulated to cover the most basic words, which cover 72.2% of the words in a typical Japanese newspaper. Since honorification tends to be found more in sophisticated words than in basic words, we used those extracted verbs as seeds to expand the list using the *Goi-Taikei* thesaurus (Ikehara et al. 1997).

For example, the semantic class *meirei* 命令 'command' (Class Number 1824) lists synonyms, such as *iitsukeru* 言い付ける 'tell', and *shiji* 指示 'instruct', all of which exhibit the same referential restriction: a high ranked person as the subject and a low ranked person as the object. However, this is not always the case. For instance, *kyoka* 許可 'permit' (Class Number

1735) includes as its synonyms *dooi* 同意 'agree' and *sansei* 賛成 'agree/approve' that do not exhibit the same referential restriction as *kyoka*.

We manually extracted from Lexeed 698 such verbs (397 of these are 'a higher ranked person does to a lower person' and the rest 301 are the reverse), and from *Goi-Taikei* further 429 (228, 201 respectively), 1127 in total.

3.2.2 Nouns and adjectives

Honorification is also expressed on nouns (including verbal nouns) and adjectives by the honorific prefix *o-* with variants *on-*, *go-*, and *mi-*.

Honorific prefixes have four functions:

- [1] An entity/action belongs to the honouree.
- [2] An entity/action has an implication to the honouree, even when it belongs to the speaker.³
- [3] Addressee honorifics to show formality of speech/politeness to the addressee.⁴
- [4] Conventional usage⁵

The use of the honorific particles in [1] provides important information on the type of referents. Possessors are seldom expressed and there are no definite/indefinite articles in Japanese (Bond 2005), but honorific particles can take on these functions. For example, *o-nimotsu* (honorable luggage) means 'your/his/... luggage', and *go-ryokoo* means 'your/his/... trip'. Although the exact identity of the honoree-possessors is context dependent, as the following minimal pair of sentences show, in (9a) the possessors can never be the speaker or the speaker's in-group member (see Subsection 4.2 for 'in-group'), as indicated by *. In contrast, the identity of the subject, as in (9b) without an honorific particle, is generally the speaker or his in-group member.

³ For example, *o-tegami* (literally, 'honourable letter') is used when the letter is something to do with the honouree; it could be the letter that the honouree wrote, a letter sent by someone else to the honouree, or a letter written by the speaker to the honouree.

⁴ For example, *o-hana* (flowers) and *o-shokuji* (meal) are such cases where possession is not a concern.

⁵ The standard example of this type is *go-han* 'honourable-rice' meaning 'rice/meal'. Such honorific particles do not convey honorifics, but are seen as part of set phrases.

(9a) お元気である。

O-genki de iru.

Hon.-good health be stay

'(The honouree/*I/*In-group) is in good health.'

(9b) 元気である。

o-Genki de iru

good health be stay

'(*The honouree /I/ In-group) is well.'

3.2.3 Sentence structures

Honorification is manifested also in the choice of sentence structure. The causative construction can be used only when the causer is superior in social hierarchy to the causee, as shown in (10). If the causee is equal to or superior over the causer, the benefactive construction is used, conveying the same proposition with the connotation that the causee has accepted the causer's request instead of command, as in (11). Thus, the sentence structure reveals the referential disparity in rank.

(10) 私は弟に本を読ませた。

Watashi-wa ototoo-ni hon-o yom-ase-ta.

I younger brother-IO book-OB read-Caus-Past

'I made my younger brother read the book.'

(11) 私は先生に本を読んでもらった。

Watashi-wa sensei-ni hon-o yon-de morat-ta.

I teacher-IO book-OB read-and receive-Past

'(I requested my teacher to read the book for me, and) my teacher read the book for me.'

4 Ranking factors

Section 3 explained the various forms that indicate disparity of referents in rank. This section describes three factors that induce such disparity in rank: Social hierarchy, in-group and out-group distinction, and unfamiliarity of the addressee.

4.1 Social hierarchy

Social hierarchy is the core rank-inducing factor, which can be overridden by the other two factors. It refers to social ranks in such social settings as company, school, family, as well as general age/generational rank. For example, an employer is perceived as ranked higher than his employees, and a teacher is

higher than his students, and the older a person is, the higher he is ranked.

Social hierarchy functions similar to the Subject-Verb agreement in terms of person, number and gender seen in many European languages. Although Japanese has no syntactic coding of such a S-V agreement, verbs agree with the referential relation of the subject and other referents in terms of social hierarchy (the same view is held by Pollard and Sag 1994).

4.2 In-group and out-group distinction

Referents are also classified according to the in-group and out-group distinction, depending on the social relation among three parties: the speaker, the addressee, and the people being referred to.⁶ For example, in (12) an officer of a company (the speaker) talks about the president of his company (referent) to his boss (addressee). The officer is ranked lower than the president and his boss, and accordingly the subject honorific and addressee honorific ('Polite') are used. However in (13), when he reports the same proposition to people outside the company, the president is regarded as a 'in-group' member to the speaker, and therefore the description of him uses the non-subject honorific form of verb, the same as the speaker would use to describe himself. In other words, the rank assigned from social hierarchy is overridden by the in-group and out-group distinction.

(12) 社長がいらっしゃいました。

Shachoo-ga irrashai-mashi-ta.

president-SB come[SubH]-Polite-Past

'The president has arrived.'

(13) 社長が参りました。

Shachoo-ga mairi-mashi-ta.

president-SB come[NsubH]-Polite-Past

'The president has arrived.'

Thus, the dichotomy of in-group/out-group distinction is relative. This is prominently seen in the use of family terms, as shown in Table 2. When someone talks to her/his mother or about her with her/his family, 'mother' is referred to as

⁶ Generally, the type of honorific use is also determined by the three parties. To be more precise, setting and bystander also play a part in determining the type of honorifics to be used (Brown and Levinson 1987).

okaasan, using the out-group (OG) form, while when talking about her to outsiders, she is referred to as *haha*, using the in-group (IG) form.

There are three lexical types that reflect the in-group and out-group distinction.

1) the deictic prefixes:

too- hon-, hei-, setu, etc. for the in-group use, translated into English as 'my/our', and *ki-, o-, on-*, etc. for the out-group use, translated as 'your/his/her/their'. For example, *too-koo* (my/our school) versus *ki-koo* (your/their school).

2) the suffixes *-san/-sama/-dono*:

for instance, *gakusei* 'a student' is referred to as *gakusei-san* out of deference to a respected out-group person (e.g. 'your students', 'student of your school').

3) suppletive forms:

some examples are shown in Table 2.

	<u>IG referent</u>	<u>OG referent</u>
mother	<i>haha</i>	<i>o-kaasan</i>
father	<i>chichi</i>	<i>o-toosan</i>
wife	<i>tsuma, kanai</i>	<i>o-kusan</i>
son	<i>segare, musuko</i>	<i>go-shisoku, musuko-san</i>
daughter	<i>musume</i>	<i>o-joo-san</i>

Table 2: Referential forms by in-group and out-group

4.3 Unfamiliarity of the addressee

In apparent absence of disparity in social ranking and age, honorifics can still be used subjectively in formal settings, when communicating with unfamiliar people, particularly by female speakers.

5 Assigning referents with ranks using *Goi-Taikei* thesaurus

In order to make use of honorific information, each referent must be assigned with a rank to determine which referent in a sentence is ranked the highest. We use *Goi-Taikei* for this assignment (Ikehara et al. 1997). It has a

semantic feature tree with over 3,000 nodes of semantic classes organised with a maximum of 12 levels (see Figure 1). It includes in its semantic classes information on occupational status, generation, family composition; the sort of information needed for this assignment.

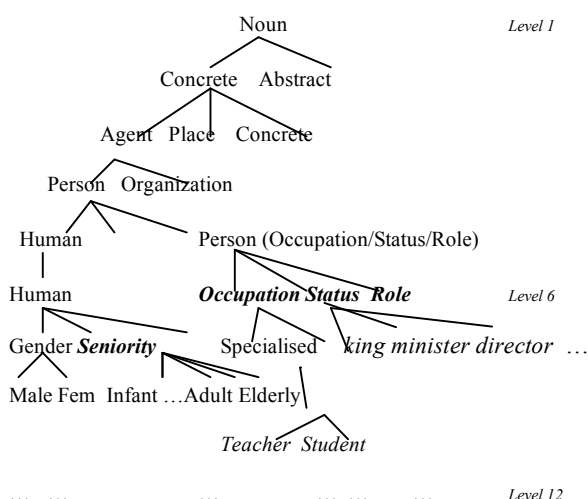


Figure 1: Excerpt from *Goi-Taikei* thesaurus

In addition, the following two tasks are required:

1) **Group** some semantic classes together from different nodes.

For the honorific use, some semantic classes that are scattered over different nodes in the tree should be grouped together. For instance, the information relevant to Social hierarchy is found not only under Occupation (status) but also under Organization, Family, and so forth.

2) **Rank** the semantic classes where relevant.

Figure 2 is a preliminary result showing ranks of referents in selected semantic classes, noted as class names followed by their semantic class numbers in *Goi-Taikei* listed in ascending order.

<p>Social hierarchy</p> <ul style="list-style-type: none"> - senior 142 > junior 143 - experienced 145 > less experienced 146 - master 139 > apprentice 140 - teacher 237 > student 238 - king/emperor 320 > aristocrats 321 - minister 322 > clerk 326 - directors 323 > deputy director 324 > executive 325 > <p>Age/Generation</p> <ul style="list-style-type: none"> - elderly 63 > adult 60 > youth 57 > boy/girl 54 > infant 51 	
<ul style="list-style-type: none"> - ancestors 84 > grandparents 81 > parents 78 > children 86 > grandchildren 89 > descendants 92 - older sibling 94 > younger sibling 97 - uncle/aunt 101 > nephew/niece 104 	

Figure 2: Strings of ranks

This list needs to be expanded. As the list is taken exhaustively from *Goi-Taikei*, these entries must be augmented with other thesauri, organisation charts, genealogical trees, and other ways as well as by hand.

6 Calibration of ranks

Ranks of referents are not absolute, but relative to the other referents in the sentence. For example, an adult referent is ranked higher than a youth, but the same referent is ranked lower when appearing with an elderly in the same sentence. Similarly, manager in a company is higher than workers with no title, but the same manager is lower than the company president. Thus, the calibration of rankings is necessary.

However, calibrating ranks while capturing relative ranks is an extremely complicated task, as any combination of referents and ranking categories can appear in a sentence as well as the fact that one referent may belong to multiple categories. For example, a measure has to be taken in case one referent is ranked in one string (e.g. ‘minister’) than the other referent (e.g. ‘clerk’) in the same sentence, but he is lower in another string (e.g. ‘less experienced’ in the profession or younger in age) (see Figure 2); or when the in-group and out-group distinction takes the precedence in the form of

honorifics, for instance a referent is senior than the other referent, but he is an in-group member to the speaker.

More complicated still, within the same class, there may exist a disparity in rank. For example, the age difference, even by one year, can determine the use of honorifics, so that honorifics is used between two referents under the same class *adult*.

Considering the above, we propose the following calibration scheme as an initial step of dealing with the complex phenomena of honorifics.

[1] **create referential links**, example modules of which are suggested in Figure 3. Each string of ranks in Figure 2 constitutes a module, which is connected to another module. Figure 3 shows that a referent ‘JOHN’ is a student as well as a child of his parents that is depicted. JOHN belongs to other modules of strings; he may be an elder brother at home, and may be a senior student at school, each of which is a member of a module and is connected to other modules. Connections between two modules may be more than one, for example, ‘grandparent’ may be a teacher of ‘teacher’ of JOHN.

It is necessary to identify as many modules as identifiable and to link them in order to accurately determine the ranks of each referent for a sentence.

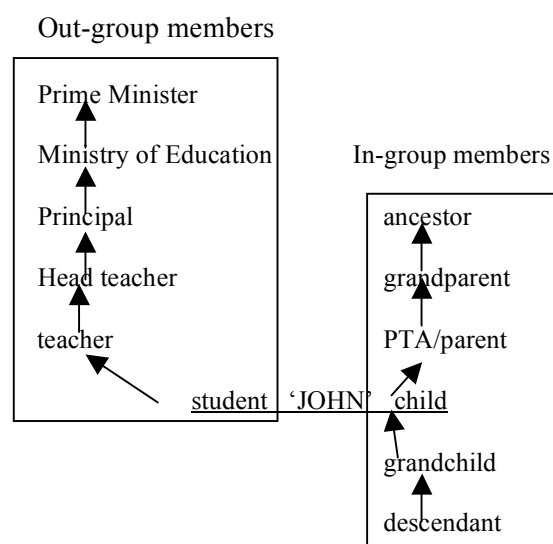


Figure 3: Two modules of referential links

[2] a diagram for calibrating ranks

Figure 4 is proposed to capture the mechanisms of honorifics that determine the ranks of referents for a sentence.

The referential links are the first (core) rank determining factor. When one referent belongs to multiple strings, for instance, a string from Social hierarchy and another from Age, then the former takes the higher rank, which is noted as ‘Social > Age’. The case where two referents belong to the same class but still appear with honorifics is due to the subtle difference in rank, noted as ‘The same class’.

These ranks assigned by the referential links can be overridden by ‘in/out group precedence’, which is determined by the type of modules, as shown in Figure 3.

The use of honorifics in absence of disparity in social rank is interpreted as lack of familiarity of the addressee.

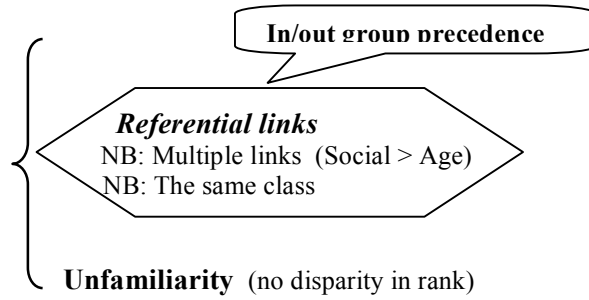


Figure 4: Diagram of calibrating ranks

7 Annotation

Our annotation method is an extension of the framework of JACY, a Japanese HPSG grammar (Siegel 2000), as discussed in Section 2. Subsection 7.1 describes the JACY annotation and Subsection 7.2 is the extension we made from this research.

7.1 JACY annotation

The JACY annotation scheme for honorification can be seen in Figure 5 with examples on the bottom of the tree. It annotates honorification concerning referential nouns (honorific entity), predicative honorifics (subject honorifics) that are triggered by honorific entities, and predicates of the addressee honorifics. The notion of polarity is used to denote the three types of

value; a polarity value “+” means a subject honorific form, “-” denotes a non-subject honorific form, and “bool” is indeterminate. It is capable of accounting for the basic types of honorification, as being expressed by verb forms, suppletive forms, passive, nouns and adjectives.

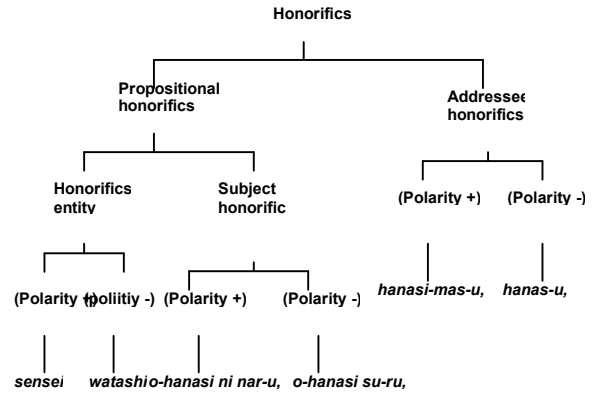


Figure 5: JACY annotation for honorifics (with examples)

7.2 Extended JACY annotation

Based on our findings, we extend the JACY annotation Figure 5 to Figure 6 by adding two relations in the honorification, Social ranking and In-group relation.

As for Social ranking, Subsection 3.2.1.5 introduced those verbs with referential restrictions, such as *insotsu* 引率 ‘take’ has a restricted usage as ‘(a higher ranked person) leads (a group of lower ranked people).’ These lexical items are added to honorific information in JACY, as being part of the lexical type hierarchy. In addition, the use of causative that imposes the interpretation ‘a high ranked person acts on the lower’ is accounted for under Social ranking (see Subsection 3.2.3).

We notate the relation deriving from social ranking as *social_ranking-rel*. It has two arguments, which show the semantic indices of the verbal arguments, the first (or left) argument being ranked higher. The relation is triggered by the lexical types and the causative usage. Example 10, ‘I made my younger brother read the book’, is annotated with *social_ranking-rel* (*watashi, ototo*), while example 11 ‘my teacher read the book for me’ is annotated with *social_ranking-rel* (*sensei, watashi*).

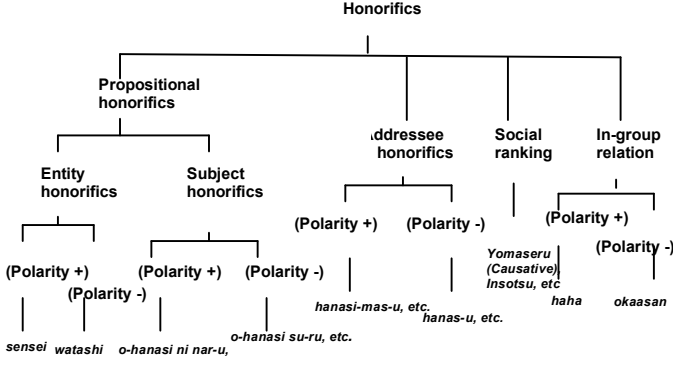


Figure 6: Extended JACY annotation for honorifics (with examples)

The distinction between in-group and out-group makes it necessary to add a further relation, called *in_group_rel*. It has two arguments, relating the speaker with the predicate's subject. As in the other honorific relations, it gets a POLARITY feature, showing an in-group relation with [POLARITY +] and an out-group relation with [POLARITY -], and “bool” for indeterminate. The nominal expressions that trigger in-group relations (such as *okaasan* and *haha* in Table 2) add this relation to the CONTEXT.

For a predicate, such as Example 13, with subject honorific information [POLARITY -] and a subject with honorific entity information [POLARITY +], an *in_group_rel* is added to relate the speaker and the subject, annotated as *in_group_rel* (*speaker, shachoo*).

To better understand the interaction of Social ranking and In-group relation, we refer to examples 12 and 13. In processing, predicative honorifics is identified not by the referential nouns, but by the predicates. So, if the predicate is *minus_shon* (- SubH) and the subject is *plus_ohon* (+ entity honorifics), i.e. (13), then there is an in-group relation. On the other hand, with an out-group relation as in (12), the predicate is *plus_shon* (+ SubH) and the subject is *plus_ohon* (+ entity honorifics).

- (12) 社長がいらっしゃいました。
Shachoo-ga irrashai-mashi-ta.
 president-SB come[SubH]-Polite-Past
 ‘The president has arrived.’

- (13) 社長が参りました。
Shachoo-ga mairi-mashi-ta.
 president-SB come[NsubH]-Polite-Past
 ‘The president has arrived.’

(14) is an example that combines different types of honorific information. Its CONTEXT annotation is described in Figure 7.⁷ The usage of the noun *haha* triggers an *in_group_rel* (*speaker, haha*) with [POLARITY +], while the usage of the noun *okaasan* will trigger an *in_group_rel* (*speaker, okaasan*) with [POLARITY -]. The extraction of social ranking information from *Goi-Taikei* shown in Figure 2 makes use of this relation *social_ranking_rel* (*arg1, arg2*) between the entities in the sentence, for example 14 *social_ranking_rel* (*sensei, haha*).

- (14) 母は先生に電話をしてもらいました。
Haha-wa senseo-ni denwa-o
 Mother-Top teacher-Dat call-Acc
site-morai-mashi-ta.
 do-receive-Polite-past
 ‘My mother got the teacher to call.’

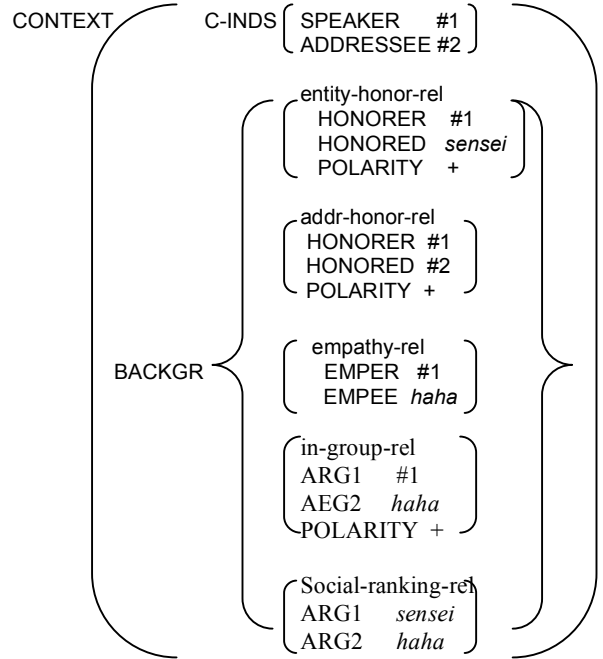


Figure 7: Context annotation of complex honorification

⁷ The values of HONORED or ARGx are actually pointers to the indices of the entities, which are written here as the orthographic realization for readability.

8 Conclusion

This paper has proposed a scheme to realise the complex linguistic phenomena of the Japanese honorifics in tangible forms for auto-processing. Ranking referents is an extremely complex task that requires a combined understanding of syntax, semantics and pragmatics in many dimensions.

In future work, the referential links and their calibration need to be expanded to make an annotation more meaningful. This will be an incremental process and takes a substantial amount of work, perhaps comparable to that required in creating a thesaurus or knowledge base.

The annotated data will be a valuable resource for research on zero pronoun resolution and Machine Translation of generating Japanese sentences. As the Korean honorification system is quite similar to the Japanese, it will be feasible to make use of the approach also for Korean. Furthermore, a part of the approach can be extended as well for Chinese, since Japanese makes use of the Chinese characters.

Acknowledgement

We are grateful to Kyonghee Paik for her invaluable advice.

References

- F. Bond. 2005. *Translating the untranslatable: A solution to the problem of generating English determiners*. CSLI Publications. California
- F. Bond et al. 2004. The Hinoki Treebank: A Treebank for Text Understanding, In Proc. of the First *IJCNLP, Lecture Notes in Computer Science*. Springer Verlag
- P. Brown and S.C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge University Press
- K. Dohsaka. 1990. Identifying the referents of zero-pronouns in Japanese dialogues. In Proc. Of the 9th European Conference on Artificial Intelligence. 240-245
- T. Gunji. 1987. *Japanese phrase structure grammar*. Dordrecht: Reidel
- S.I. Harada. 1976. Honorifics. In Shibatani (ed). *Japanese Generative Grammar*. Syntax and Semantics Series Vol.5. New York: Academic Press. 499-561
- M. Hoelter. 1999. *Lexical-semantic information in Head-Driven Phrase Structure grammar and natural language processing*. Lincom Theoretical Linguistics
- S. Ikehara et al. (eds). 1997. *Japanese Lexicon*. Iwanami Publishing
- A. Ikeya. 1983. Japanese honorific systems. In Proc. of the 3rd Korean-Japanese Joint Workshop. Seoul
- S. Martin. 1964. Speech levels in Japanese and Korean. In Hymes. *Language in culture and society*. Harper and Row. 407-415
- Y. Matsumoto. 1997. The rise and fall of Japanese nonsubject honorifics. *Journal of Pragmatics*. 28:719-740
- H. Nakaiwa. 2002. *Studies on zero pronoun resolution for the Japanese-to-English Machine translation* (in Japanese).
- S. Nariyama. 2003. *Ellipsis and Reference-tracking in Japanese*. SLCS 66, Amsterdam: John Benjamins
- S. Nariyama et al. 2005. Extracting Representative Arguments from Dictionaries for Resolving Zero Pronouns. In Proc. of the MT Summit X, Phuket
- C. Pollard and I.A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press
- M. Siegel. 2000. Japanese Honorification in an HPSG Framework. In Proc. of the 14th Pacific Asia Conference on Language, Information and Computation. 289-300. Waseda University, Tokyo
- M. Siegel and E M. Bender. 2002. Efficient deep processing of Japanese. In Proc. of the 3rd Workshop on Asian Language Resources and International Standardization at *COLING*. Taipei
- J. Wenger. 1983. Variation and change in Japanese honorific forms. In Miyagawa and Kitagawa (eds). *Studies in Japanese language use*. vol.16. Edmonton: Linguistic Research Inc. 283-292