System for Flexibly Judging the Misuse of Honorifics in Japanese∗

Tamotsu Shirado[a], Satoko Marumoto[b], Masaki Murata[c], and Hitoshi Isahara[d]

[a]National Institute of Information and Communications Technology,
Kyoto, 619-0289 Japan
shirado@nict.go.jp

[b]The Institute of Behavioral Sciences,
Tokyo, 162-0845 Japan
smarumoto@ibs.or.jp

[c]Tottori University,
Tottori, 680-8550 Japan
murata@ike.tottori-u.ac.jp

[d]Toyohashi University of Technology,
Aichi, 441-8580 Japan
isahara@tut.ac.jp

Abstract. We propose a system for flexibly judging the misusage of honorifics in Japanese sentence. The system can point out misused words and phrases, and can also indicate how they are misused. The system uses judgment rules whose degrees of validity in modern Japanese society are quantified by psychological experiments. The system can judge sentences flexibly based on the learner's linguistic level by tuning thresholds regarding the degree of validity. The proposed system is expected to be applied in practical computer-aided education.

Keywords: Japanese, honorifics, misuse, judgment, education

1   Introduction

In Japan, linguistic honorifics play an important role in social activities, especially in conversations. There are few languages in which such linguistic honorifics are as highly developed as Japanese.

One important function of linguistic honorifics is to awaken in each person a consciousness of the social-relationship among speakers, listeners, and individuals being addressed or referred to. Here, social-relationship is assumed to be represented by a combination of {relative social position among in the people in same group} and {in-group/out-group relationship} in our study.

The correct usage of honorific expressions is indispensable to maintain appropriate social distance between individuals and assists smooth communication in Japanese society.

Recently, however, increasing misusage of honorific expressions has been noted. One origin of such misusage may be poor education regarding linguistic honorifics. So, development of a computer-aided learning system for linguistic honorifics has been needed.

Recently, Shirado (2007) proposed a system to judge the misusage of honorific expressions in Japanese. The system uses judgment rules constructed from textbooks of Japanese honorifics. However, these textbooks are not necessarily consistent with the awareness among the people regarding normative honorific expressions in modern Japanese society. Most of these textbooks

---

are based on traditional studies even though linguistic norms change over time. So, more practical system to learn linguistic honorifics is needed.

In this paper, we propose a system to flexibly judge the misusage of honorific expressions in Japanese sentences. This system uses the framework of the system proposed by Shirado, but our system is more practical than the previous one due to the following points:

1. The judgment rules are generalized.
2. The degrees of the validity of rules are quantified.
3. Judgments are performed flexibly based on the learner's linguistic level.

2 Misusage of Japanese Honorific Expressions

2.1 Types of the Japanese Honorific Expressions

Japanese honorific expressions can be divided into the following types:

Sonkeigo (in Japanese) shows respect toward a person, who is usually the subject of sentence's predicate, by elevating the person's status. _Sonkeigo_ principally includes: (1) such honorific titles as "_san_"/"_sama_," (2) respectful forms of verbs, (3) verb stems + auxiliary verbs, and (4) "_o_"/ "_go_" (prefixes) + verb stems + auxiliary verbs.

Kenjougo-I (in Japanese) shows respect toward a person, who is usually the object of the sentence's predicate, by humbling the speaker or the predicate's subject. _Kenjougo_-I principally includes: (1) humble forms of verbs, (2) verb stems + auxiliary verbs, and (3) "_o_"/"_go_" (prefixes) + verb stems + auxiliary verbs.

Kenjougo-II (or _Teichougo_) (in Japanese) shows politeness, but not necessarily respect, toward the listener by using humbling form of predicates regarding the speaker's action etc. _Kenjougo_-II includes "_mousu_"/"_mairu_"/ "_itasu_," etc.

Teineigo (in Japanese) shows politeness, but not necessarily respect, toward a person, who is usually the listener. _Teineigo_ includes an auxiliary verb at the end of a sentence, such as "_desu_"/"_masu_"/"_gozaimasu_," etc.

Table 1 show examples of morphemes corresponding to each honorific type.

Table 1: Honorific type.

Morphemes	Honorific type
"A" "_san_"	honorific title of person A
"B" "_sama_"	honorific title of person B
"_osharu_"	_sonkeigo_
verb stem "_rareru_"	_sonkeigo_
"_o_" verb stem "_kudasaru_"	_sonkeigo_
"_go_" verb stem "_kudasaru_"	_sonkeigo_
verb stem "_itasu_"	_kenjougo-I_
"_o_" verb stem "_suru_"	_kenjougo-I_
"_go_" verb stem "_suru_"	_kenjougo-I_
"_mousu_"	_kenjougo-II_
"_go_" "_suru_" "_te_" "_kudasaru_"	_kenjougo-I + sonkeigo_
"_desu_" "."	_teineigo_
"_masu_" "."	_teineigo_

2.2 Categories of Misusage of Honorific Expressions

The misusage of honorific expressions can be divided into two categories:

Misusage in word form: The word form is not normative; i.e., the word form differs from the normative word forms mentioned in Section 2.1. For example, "_o_"(a pre-fix) + verb stem +

"*rareru*" (an auxiliary verb) is a misusage in word form because "*rareru*" is not allowed to be used in this word form.

Misusage in performance: The word form is normative, but is not consistent with the social-relationship among the speakers, listeners, and individuals being addressed or referred to in the sentence. For example, a sentence that does not contain *Teineigo* is usually considered a misusage in performance when the listener's social position is higher than the speaker's.

The misusage in word form can be avoided simply by preparing a sufficient list of normative and non-normative word forms. The present study concentrated on misusage in performance.

3 Framework

3.1 Restriction

In our study, each input sentence is assumed to satisfy the following conditions:

1) at most, four persons - speaker, listener, and two others (denoted as "A" and "B") being addressed or referred to - are involved in the sentence,

2) one predicate in the sentence that has one subject person and one object person

3.2 Honorific Feature

Since all sentences inputted to the system are assumed to satisfy the restrictions stated in Section 3.1, the honorific features of the sentence can be represented by the set of s, o, e, and p defined in Table 2. For example, the honorific features of sentence "*A ga kimi ni gosetsumei suru*," ("person A explains something to you") is $s=0$, $o=0$, $e=0$, $p=2$, because the subject is person A, the object is listener ("*kimi*"), there are no honorific titles for the subject and object persons, the honorific type of the auxiliary verb ("*suru*") at the end of the sentence is not polite, and the honorific type of the predicate ("*gosetsumei*") is *kenjougo*-I.

Table 2: Honorific features.

Honorific value	Explanation
$s=0$	No honorific title for subject
$s=1$	Honorific title for subject
$o=0$	No honorific title for object
$o=1$	Honorific title for object
$e=0$	Honorific type of auxiliary verb at end of sentence is not polite. / No auxiliary verb at end of sentence
$e=1$	Honorific type of auxiliary verb at end of sentence is polite
$p=0$	Honorific type of predicate is neither *sonkeigo* nor *kenjougo-I* nor *kenjougo-II*
$p=1$	Honorific type of predicate is *sonkeigo*
$p=2$	Honorific type of predicate is *kenjougo-I*
$p=3$	Honorific type of predicate is *kenjougo-I + senkeigo*
$p=4$	Honorific type of predicate is *kenjougo-II*

3.3 Judgment Rules

We constructed rules for judging the misusage of honorific expression based on the existing rule proposed by Shirado. Our rules are more general than the existing one because our rules can be applied for any number (up to two, three, or four) of persons. But the existing rules were divided into three kinds for two, three, and four persons. The experimental results using the test

sets adapted by Shirado show that our rules are identical to the existing rules regarding the judgment ability.

Table 3 shows part of the judgment rules, which define the honorific consistency between the social-relationship among persons (and specifications of the subject/object persons in some of them) and the honorific feature of sentence that is honorifically consistent with the social-relationship.

Table 3: Part of judgment rules.

Rule No.	Explanation	Social-relationship among persons	Specification of subject / object	Honorific feature $s=0/1$ $o=0/1$ $e=0/1$ $p=0/1/2/3/4$
1	Rules regarding honorific title of subject (honorific feature s)	(S)(subject)	no	$s=1$
2		(subject > S) and (S, L)	no	$s=1$
3		(subject, S)(L)	no	$s=0$
4		(S => subject)	no	$s=0$
5		Any	S is subject	$s=0$
6	Rules regarding honorific title of object (honorific feature o)	(object)(S)	no	$o=1$
7		(object > S) and (S, L)	no	$o=1$
8		(object, S)(L)	no	$o=0$
9		(S => object)	no	$o=0$
10		Any	S is object	$o=0$
11	Rules regarding politeness of end of sentence (honorific feature e)	(L > S)	no	$e=1$
12		(S)(L)	no	$e=1$
13		(S => L)	no	$e=0$
14	Rules regarding type of predicate (honorific feature p)	(subject > S) and (S, L)	no	$p=1$
15		(S => L)(object) and (S => subject)	no	$p=0$ or 2
16		(S => L)(object)	S is subject	$p=0$ or 2
17		(S => subject) and (S => object) and (S => L)	no	$p=0$

In Table 3, symbols "S", "L", "subject", and "object" indicate the speaker, the listener, the subject person of the predicate, and the object person of the predicate, respectively. {"X>Y" / "X=>"Y"} means that the social position of the person corresponding to symbol X {is higher than / is higher than or equal to} that of the person corresponding to symbol Y. {"(X)(Y)"/ "(X)(Y)(Z)"} means that {X and Y / X and Y and Z} are the out-group, {"(X,Y)" / "(X,Y,Z)"} means that {X and Y / X and Y and Z} are the in-group and there is no restriction as to the relative social position between them, where X/Y/Z are S (speaker)", "L (listener)", "A (person A)", or "B (person B)" and are different.

Table 3 consists of four kinds of rule groups, each of which corresponds to each of honorific features s, o, e, p. For example, the top rule in Table 3 means that honorific feature s of the sentence must be 1 (i.e., the honorific title of the subject must appear in the sentence) when the speaker and the subject are the out-group.

4　Quantification of Degree of Validity of Judgment Rules by Psychological Experiments

4.1　Experiment

For all the honorific features (*s*, *o*, *e*, *p*) of each rule, we quantified the degree of validity by a psychological experiment using Scheffe's paired comparison method (1952).

One hundred and three peoples, all of them are over thirty years old and non-experts in Japanese linguistics, participated in the experiment.

Figure 1 shows an example of an answer sheet. The social-relationship among persons ("speaker > person A > listener"), speech intention ("person-A explains something to you"), and a pair of different types of predicates of "*setsumei suru*" (i.e., non-honorific) and "*gosetsumei suru*" (i.e., *kenjougo*-I) are described in the sheet. Each subject answered which honorific type of predicate was more honorifically appropriate to use and to what extent by marking the answer sheets with an O. If the subjects felt that both predicates were equally appropriate or inappropriate, they marked the O halfway between them.
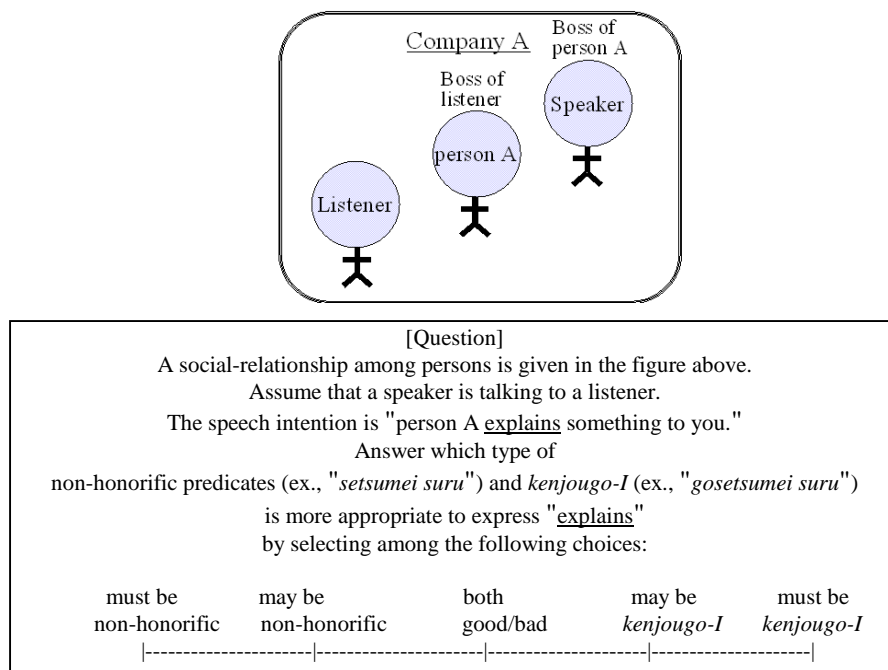
Figure 1: Example of answer sheet.

For all honorific features (*s*, *o*, *e*, *p*), all the pairs of the stimuli corresponding to each value of each honorific feature were presented one at a time to each subject. For example, all pairs of non-honorifics (*p*=0) / *sonkeigo* (*p*=1) / *kenjougo*-I (*p*=2) / *kenjougo*-I + *sonkeigo* (*p*=3) / *kenjougo*-II (*p*=4) were presented in the experiments regarding the honorific feature *p*.

All kinds of social-relationships, each of which represents a typical social-relationship described in each rule, are presented in the answer sheet. For example, we assumed that social-relationship "speaker > person A > listener" is the typical social-relationship described in rule No. 17 of Table 3.

4.2 Experimental Results

Table 4 shows part of the values of the degrees of validity obtained by the experiment. Common rule numbers are used between Tables 4 and 3. Each value of the degree of validity is shown in parentheses followed by each value of the honorific feature, where values with/without bold face mean that they are/are not described in Table 3. The larger the value, the more the degree of validity. Due to the quantification method, the value of the validity theoretically ranges from -2.0 to 2.0 for $s=0$, $s=1$, $o=0$, $o=1$, $e=0$, and $e=1$ and from -1.6 to 1.6 for $p=0$, $p=1$, $p=2$, $p=3$, and $p=4$.

Table 4: Part of values of degree of validity.

Rule No.	Honorific feature (value of degree of validity)				
1	$s=0$ (-1.592)	**$s=1$** (1.592)			
2	$s=0$ (-1.495)	**$s=1$** (1.495)			
3	**$s=0$** (0.883)	$s=1$ (-0.883)			
4	**$s=0$** (0.611)	$s=1$ (-0.611)			
5	**$s=0$** (0.922)	$s=1$ (-0.922)			
6	$o=0$ (-1.311)	**$o=1$** (1.311)			
7	$o=0$ (-1.417)	**$o=1$** (1.417)			
8	**$o=0$** (0.932)	$o=1$ (-0.932)			
9	**$o=0$** (0.340)	$o=1$ (-0.340)			
10	**$o=0$** (0.650)	$o=1$ (-0.650)			
11	$e=0$ (-1.282)	**$e=1$** (1.282)			
12	$e=0$ (-0.320)	**$e=1$** (0.320)			
13	**$e=0$** (-0.204)	$e=1$ (0.204)			
14	$p=0$(-0.111)	**$p=1$** (0.231)	$p=2$ (-0.295)	$p=3$ (0.214)	$p=4$ (-0.039)
15	**$p=0$** (0.633)	$p=1$ (-0.351)	**$p=2$** (0.039)	$p=3$ (-0.435)	$p=4$ (0.115)
16	**$p=0$** (0.664)	$p=1$ (-0.400)	**$p=2$** (0.039)	$p=3$ (-0.404)	$p=4$ (0.101)
17	**$p=0$** (0.920)	$p=1$ (-0.322)	$p=2$ (-0.216)	$p=3$ (-0.328)	$p=4$ (-0.054)

5 Algorithm

Figure 2 shows an example of the process flow of our system. In this example, the social-relationship among persons "(S>A>L)" and sentence "*A ga kimi ni gosetsumei suru*" ("person A explains something to you"), are given as input.

The honorific judgment is executed by the flowing steps.

Step 1 Obtain morphemes by analyzing the sentence using "ChaSen," the Japanese morphological analysis program.

Step 2 Identify subject / object of the predicate ("*gosetsumei*") using our parsing program (subject / object are identified as "A" (i.e., person-A) / "*kimi*"(i.e., listener (L)) for this example).

Step 3 Determine the honorific features of the sentence using the results of Steps 1 and 2 ($s/o/e/p$ are determined to 0/0/0/2 for this example).

Step 4 Select all rules whose description regarding the social-relationship and the subject / object coincide with those obtained in Steps 1 and 2 (rules No. 4 (for s), No.9 (for o), No13 (for e), and No.17(for p) are selected for this example).

Step 5 Enable/disable rules whose degrees of validity are greater/less than thresholds, where the thresholds are set separately for each *s*, *o*, *e*, *p* (rules No. 4, No.9, and No.17 are enabled, rule No.13 is disabled for this example).

Step 6 If some misusages are detected by the rules enabled in Step 5, output all the misusages, otherwise, output "input sentence is correct" (only one misusage for Rule No.17 was detected for this example).
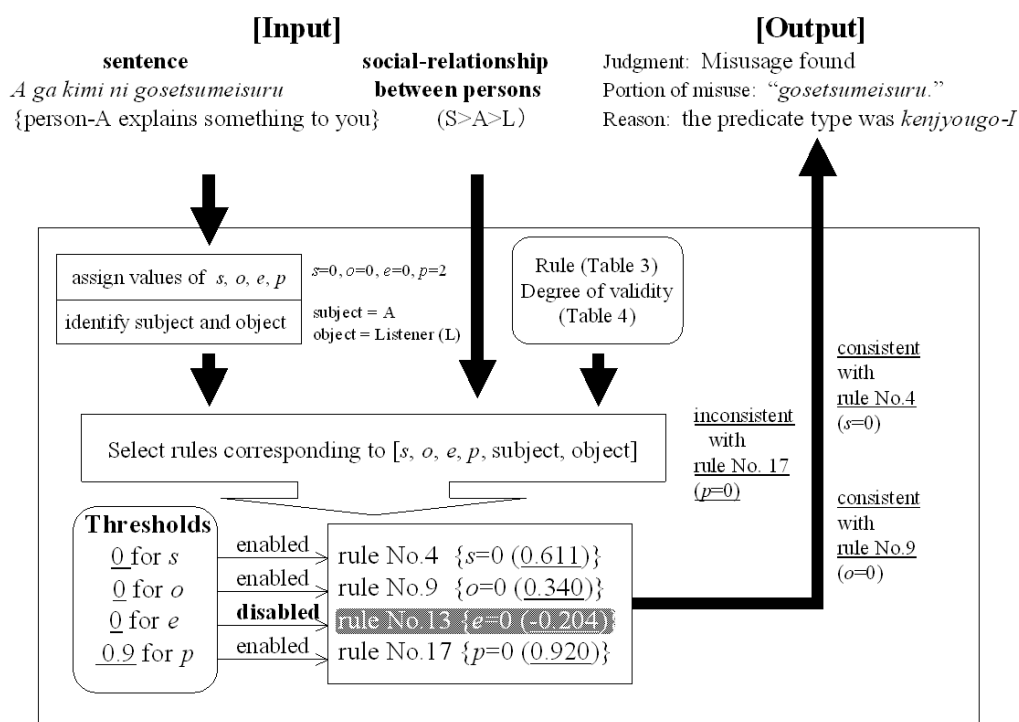
[Input]   **[Output]**

sentence   social-relationship   Judgment: Misusage found

A ga kimi ni gosetsumeisuru   **between persons**   Portion of misuse: *"gosetsumeisuru."*

{person-A explains something to you}   (S>A>L)   Reason: the predicate type was *kenjyougo-I*

assign values of *s*, *o*, *e*, *p* $s=0, o=0, e=0, p=2$

identify subject and object   subject = A
object = Listener (L)

Rule (Table 3)
Degree of validity
(Table 4)

Select rules corresponding to [*s*, *o*, *e*, *p*, subject, object]

inconsistent
with
rule No. 17
(*p*=0)

consistent
with
rule No.4
(*s*=0)

Thresholds
0 for *s*   enabled   rule No.4 {*s*=0 (0.611)}
0 for *o*   enabled   rule No.9 {*o*=0 (0.340)}
0 for *e*   **disabled**   rule No.13 {*e*=0 (0.204)}
0.9 for *p*   enabled   rule No.17 {*p*=0 (0.920)}

consistent
with
rule No.9
(*o*=0)

Figure 2: Process flow.

6   Tuning of Thresholds Based on Learner's Linguistic Level

As mentioned above, only rules whose degrees of validity exceed the thresholds of *s*, *o*, *e*, *p*, are enabled for judgment. The degree of validity is closely correlated with the importance to be learned in modern Japanese society. Our system can perform flexible judgments based on the learner's linguistic level by tuning these thresholds. For example, when the thresholds are set to appropriately large values, the system urges learners to learn only rules whose importance is large.

As a primitive study to investigate the appropriate tuning point of these thresholds, we introduce comprehensive index "*T*," which is related to these thresholds. Concretely, we defined the thresholds of *s*/*o*/*e* as -2.0 + (2.0+2.0)/10*(10-*T*) and the thresholds of *p* as -1.6 + (1.6 + 1.6) / 10 * (10-*T*), where they are normalized based on the ranges of the degrees of validity for *s*, *o*, *e*, *p* mentioned in Section 4.2. The thresholds of *s*/*o*/*e* decrease from 2.0 to -2.0 and the threshold of *p* decreases from 1.6 to -1.6 when *T* increases from 0 to 1.

Two kinds of test sets (each test set is a pair of a social-relationship and a speech sentence) were inputted to our system. One is called "low," which includes elementary level honorific

misusages prepared by us. The other is called "middle," which was collected by non-experts of Japanese linguistics.

Figure 3 shows the number of sentences (percentage of all input included in each test set) judged as "misusage" versus T.

Figure 3 suggests that the system can urge elementary level learners to learn rules whose importance is in the upper half of all rules when T is set to around 2. It also suggests that the system can urge middle level learners to learn rules whose importance is in the upper half of the rules that haven't been learned yet (i.e., about 30% of rules in our study) when T is set to around 4, assuming that they have already learned rules whose importance is in the upper 70% of all rules.
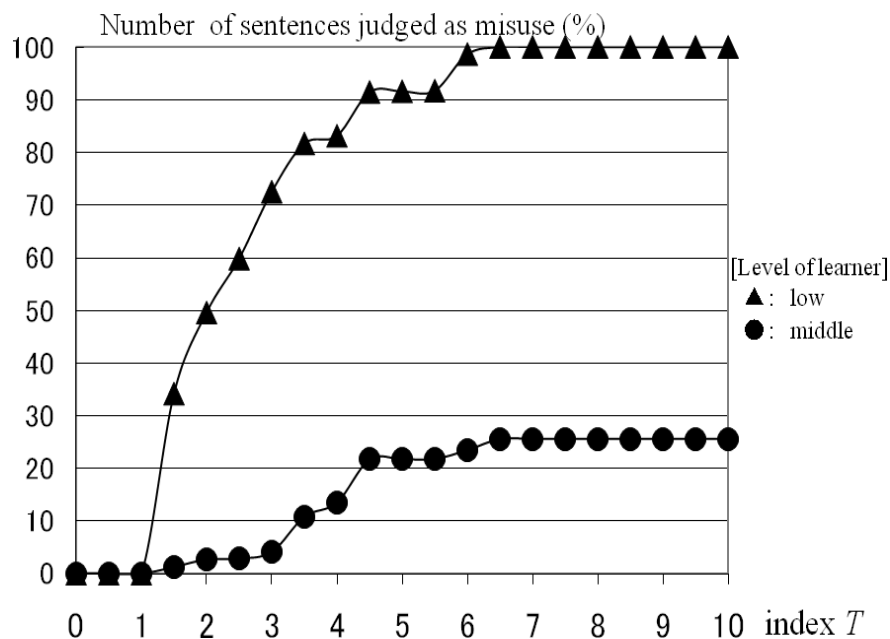

Figure 3: Dependency on T

7 Conclusion

We proposed a system that flexibly judges the misusage of honorifics in Japanese sentence. The system uses judgment rules whose degrees of validity in modern Japanese society were quantified by psychological experiments. The system can judge sentences flexibly based on the learner's linguistic level by tuning thresholds regarding the degree of validity. The proposed system is expected to be applied in practical computer-aided education.

References

Scheffe H. 1952. An analysis of variance for paired comparisons, *Journal of American Statistical Association*, 47, 381-400.

Shirado T., Marumoto S., Murata M., Uchimoto K., Kawahara D., and Isahara H. 2007. A System to Point Out Misuse of Honorific Japanese Expressions(in Japanese), *IEICE Transactions*, J90-A(11): 1-13.