A Paraphrasing System for Transforming Regular Expressions into Honorifics

Dongli Han, Shuntaro Kamochi, Xin Song, Naoki Akegawa, Tomomasa Hori
Department of Computer Science and System Analysis,
College of Humanities and Sciences, Nihon University
han@cssa.chs.nihon-u.ac.jp

Abstract

Honorifics in Japanese plays an incredibly important role in all walks of social life. The demand to transform regular expressions in Japanese into honorifics automatically has increased rapidly especially in business situations. This paper reviews existing studies and proposes a system to fill this demand with more practicable functions. The experiment shows the effectiveness of our strategy.

1 Introduction

The Japanese language is a kind of highly specific language in establishing hierarchical relations among people, or paying respects to people comparing with other languages. The honorifics in Japanese include different levels of respectful, humble, and polite speeches which are frequently used in various social or business situations. The mechanism of honorifics in Japanese is so complicated that recent young generations in Japan could hardly master it or use it properly.

This situation has encouraged the study dealing with honorifics in Japanese including automatic paraphrasing. For instance, Noguchi et al. generate all kinds of honorific forms for single verbs automatically (Noguchi et al. 2007). In their study, verbs are considered exclusively, and hence no contextual information has been employed.

In another study, Tazoe et al. have proposed a computer model to translate regular expressions into respectful speeches (Tazoe et al. 2005). They determine the type and level of honorifics for a verb in a sentence based on the subject of the sentence and the listener level, the situation level, and the topic level retrieved from the entire article. Comparing with the study of Noguchi et al., this one is more practical. However, there exist some problems in this work. No strategy seems to have been prepared in case multiple verbs with different agents appear in a same sentence. Another problem is the omission of subjects in Japanese sentences. This will obstruct the determination of the honorific form of the verb. Worst of all, the method proposed in this work seems to be remaining as a computer model without being implemented at all.

This paper describes a practical system developed to transform regular expressions in Japanese into honorific forms automatically. Specifically, we manage to retrieve the hierarchical relationships among characters in a sentence so as to determine different honorific forms for multiple verbs with different agents in the same sentence. Another major difference from previous studies is that we employ a series of strategies to cope with the problem of subject-omission. Here in our study, we mainly concentrate our attention on the situation of composing business e-mails.

We first describe the framework of our system in Section 2, and then some main modules in the following sections. Finally we discuss the experiment and conclude this paper in Section 7.

2 Framework of Our System

Our system contains four main parts: Information-retrieval Unit, Subject-complement Unit, Honorific-Form-Determination Unit, and Paraphrasing Unit. We illustrate the whole framework in Figure 1.

Coling 2010: Demonstration Volume, pages 1–4,
Beijing, August 2010

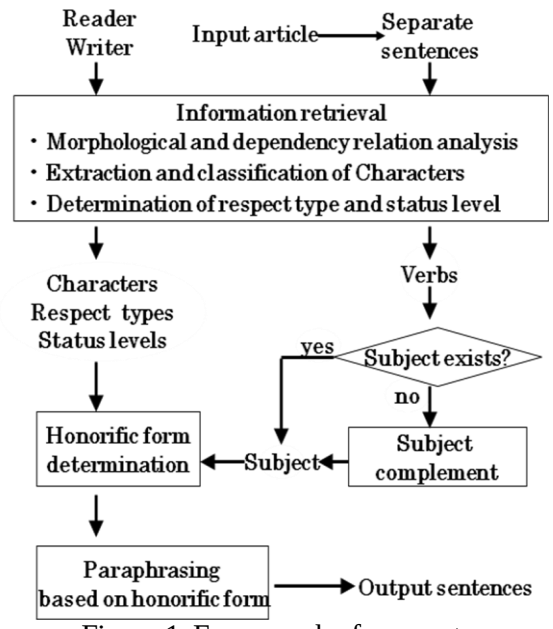

Figure 1: Framework of our system

Before running the system, the user (the person who has composed an e-mail and wants to check the honorifics with the system) is recommended to input the names and the positions or statuses of both himself and the person he is going to contact by e-mail, represented as the *Writer* and the *Reader* in Figure 1. This optional function is to help the system make more precise judgment on the hierarchical relations among characters in the e-mail article, and hence make more reasonable decision on respect type and status level which will be used in Honorific-form-determination Unit.

The procedure will be repeated until all sentences in the input article are processed. We describe the main parts next in section 3, 4, 5, and 6 in detail.

3 Information Retrieval

Information-retrieval Unit is the first and most essential part in our system. We first retrieve basic information including parts of speech and dependency relations among constituent words from a sentence through a free morphological and dependency parsing software, Cabocha[1].

Then based on the basic information obtained above, the system attempts to extract nouns or pronouns representing characters from the sentence, using a Japanese concept thesaurus: EDR Concept Dictionary[2]. The extracted nouns will be divided into three categories: first-person group, second-person group, and third-person group, by checking them against a first-person noun list and a second-person noun list we have made beforehand. The identification results will be used later in the Honorific-form-determination Unit.

Finally, the system assigns a respect type and a status level to each character that is appearing together with nouns showing duty positions or social statuses. Respect type reflects the degree of respect. A larger number indicates a character with higher position, suggesting that a higher honorific form with more regard to the character should be appropriate. Status level has a similar nuance with respect type. It breaks each respect type down into several positions and ranks them according to tiny difference among them.

4 Subject Complement

The system could not determine which kind of honorific form should be applied to a verb with the information on characters only. We need to know the subject of each verb as well. Generally, the subject of a predicate is identified through the dependency analyzing process. However, in case the subject of a verb is omitted, we have to find the subject to help determine the honorific form as described later in Section 5.

In our system, we employ five factors to help recognize the subject for a verb. In this section, we first explain the factors and then describe the method for complementing subjects based on the five factors.

4.1 Nonhuman-behavior Verbs

Our final purpose with this system is to transform a verb into an appropriate kind of honorific form to show the writer's regard or respect to the reader. Situation will be different when the subject of a verb is not a person or character. No respect is needed to be paid to a thing.

In our system, before supplementing the subject, we check the verb against the EDR

[1] http://chasen.org/~taku/software/cabocha/

[2] http://www2.nict.go.jp/r/r312/EDR/J_index.html

Dictionary to see whether the verb represents a nonhuman-behavior. For example, the Japanese verb "沸騰する" meaning *boil*, will never appear with a person as its subject. In this case, the system will not supplement the subject for the verb, but leave a check-mark here to change the verb into a polite speech later in Section 6.

4.2 Expressions for Estimation

There are a number of expressions in Japanese following verbs and implying estimation or hearsay. For example, "だろう" or "でしょう" indicates possibility but uncertainty. In case one of these expressions appears following a verb in a subject-omitted sentence, the subject of the verb tends to be the second person or the third person. We prepare a list containing these expressions and supplement the subject as non-first-person if we find such an expression after the verb.

4.3 Auxiliary Predicates

Expressions following the predicate and appearing in the end of a sentence are called as auxiliary predicates (Kudo et al. 1993). They help predicates describe the modality of a sentence, and at the same time contain the information on subjects. For instance, "~(し)たい" represents the desire of the writer, while "~ても構わない" meaning that *it is all right (for somebody) to do (something)*, indicates that *somebody* here should be the second person or the third person.

4.4 Expressions of Internal Feeling

Internal feeling means the emotion or feeling in the back of one's mind, implying that no one could understand or represent your feeling except yourself. In Japanese, we use adjectives or adjective verbs to express internal feelings, and the second person or the third person will never act as the subject of such an adjective or adjective verb. Here is an example. "楽しい" meaning *happy*, is a frequently used adjective. But different from *happy* which can be used for anybody, "楽しい" in Japanese is used only for the first person.

This fact helps us supplement the subject in a sentence with the first-person noun that we have extracted in the Information-retrieval Unit. We use a Japanese lexicon, Goi-Taikei (Ikehara et al. 1999) as the data source, and employ all the adjectives or adjective verbs in the category of *state of mind of a person* in our system.

4.5 Property of Case

In most situations, if a character or person noun has been used as a surface case with some certain particles in a sentence, the character will seldom act as other surface cases in the same sentence (Isozaki et al. 2006). Along this idea, we avoid supplementing subjects with non-first-person characters or person nouns if they have appeared as other surface cases. Here the reason we exclude the first-person characters from applying the rule lies in the fact that some first person characters do act as multiple surface cases although not that frequently.

4.6 Subject-complement Procedure

Our system tries to supplement the subject for a verb in a sentence utilizing all the previously described factors in a comprehensive manner. At first, every rule is checked to see whether it is applicable or not. Then we generate a slot containing four bits representing nonhuman, the first person, the second person, and the third person respectively for each rule.

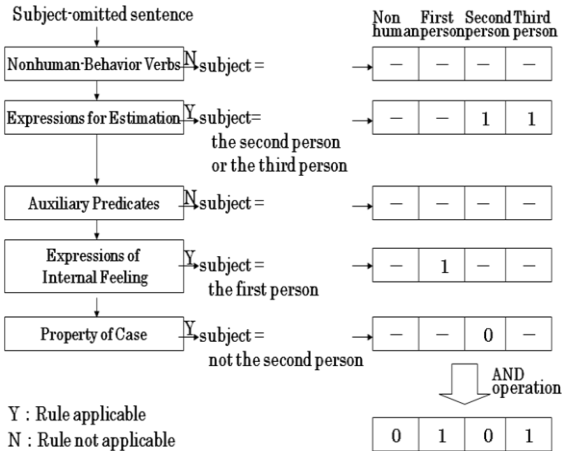

Figure 2: An example of subject complement

According to the applying result of each rule, each slot is updated with 1 or 0 representing possibility and impossibility at the appropriate bit. At last, we carry out an And Operation

with all slots and get the final answer. Figure 2 is an example of subject complement.

If we get multiple candidates for the omitted subject, we have to determine the final one based on the priority order: nonhuman > the first person > the second person > the third person as shown in Figure 2. We have established the above priority order from the result of a preliminary experiment.

Here in this example, the system will supplement the subject of the corresponding verb in the sentence with the first-person noun.

5 Honorific Form Determination

In this section, we describe the method of determining the honorific forms for verbs. We have obtained the respect types, the status levels, and have supplemented the subjects for verbs in Information-retrieval Unit and Subject-complement Unit respectively. Now, the system will determine the honorific form for each verb according to the following rules (R1~R4). Here, the signals *sub*, n^{th} *P*, and SL_n indicates the subject, the nth person, and the status level of the nth person.

R1. If ((*sub* = 2^{nd} *P*) and ($SL_1 < SL_2$)) or
((*sub* = 3^{rd} *P*) and ($SL_1 < SL_3$) and ($SL_2 < SL_3$))
Then Respectful Speech
R2. If ((*sub* = 1st *P*) and ($SL_1 < SL_2$))
Then Humble Speech
R3. If ((*sub* = 1st *P*) or (*sub* = 3^{rd} *P*)) and ($SL_1 < SL_2$) and ($SL_3 < SL_2$)
Then Teichogo Speech
R4. Otherwise Polite Speech

The formula $SL_m < SL_n$ means that the nth person has a higher position than the mth person.

6 Paraphrasing

In accordance with the results of honorific-form determination, we transform verbs in each sentence into their corresponding speeches. There are two types of transformation. One is with most normal verbs based on general paraphrasing rules and the respect levels that we have got in Section 3, such as the verb "働く" meaning *work*, and "書く" meaning *write*.

Another transformation is more complicated. We have to convert the original verb into some particular form first, and then inflect the new form according to the same general paraphrasing rules as those being used for normal verbs. Here is an example. The verb "行く" meaning *go*, holds a particular form: "いらっしゃる" for expressing respect, and "参る" for expressing modesty.

Besides, we have added some exception processing into our system to cope with individual or isolated cases.

7 Conclusions

We have conducted a questionnaire to examine the practicality of our system. Participants in the questionnaires include 5 Japanese college students. They are told to evaluate the naturalness and correctness of a set of transformed articles from our system in 3 levels: 2 for good, 0 for bad, and 1 for the intermediate level between good and bad: not good but acceptable. The average evaluation result is 1.32 showing the effectiveness of our system. We believe that the system could be utilized in situations of creating business documents or learning honorifics in Japanese.

References

Ikehara, S., Miyazaki, M., Shirai, S., Yokoo, A., Nakaiwa, H., Ogura, K., Ooyama, Y., and Hayashi, Y. 1999. *Goi-Taikei - A Japanese Lexicon*, Iwanami Shoten, Tokyo. (in Japanese)

Isozaki, H., Kazawa, H., and Hirao, T. 2006. Japanese Zero Pronoun Resolution Based on Lexicographical Ordering of Penalties. *IPSJ Trans.* 47(7):2279-2294. (in Japanese)

Kudo, I., and Tomokiyo, M. 1993. An Ellipsis-Resolution Mechanism by Using Japanese Predicate Particularity. *IEICE Trans.* J76-D-II(3):624-635. (in Japanese)

Noguchi, S., Nanjo. H., and Yoshimi, T. 2007. Doushi No Tsujohyogen Kara Keigohyogen Eno Kangen. *Proc. of the 13th Annual Meeting of the Association for Natural Language Processing*, pages 978-981. (in Japanese)

Tazoe, T., Watanabe, C., Shiino, T. 2005. Development of a Computer Model for Translating in Respect Language. *IPSJ SIG Notes* 2005(94):1-6. (in Japanese).