

## How to calculate degree of politeness for Japanese sentence generation

2 Q - 1

Ichiro KOMIYA, Takeyuki KOJIMA, Masaya YAMAGUCHI, Nobuo INUI, Yoshiyuki KOTANI  
Department of Computer Science, Tokyo University of Agriculture and Technology

### 1. Introduction

Polite expressions in Japanese have one of the most complicated rules and are said to be difficult to generate appropriately even for natives. This is because they are not strictly specified by the rule of the society, however the language could be powerful enough to judge the relationship between the speaker and the listener handling grammar rules.

A number of sentence generation in Natural Language Generation concerning the polite expressions has been attempted before [1]. However, it turned out to be necessary to consider the problems with another point of view to make it more accurate. This paper introduces a new method as a model which enables the system to select appropriate words according to the degree of politeness. The table is statistically created by the results of a questionnaire to help the sentence generators construct the right expression. With the proposed method, the expressions become much more considerate and precise in spite of the unspecified rules.

### 2. What the linguists say

According to [2] and [3], the linguists say that they need five factors about both the speaker and the listener for making a right polite expression, which are practical relationship, closeness, sex, occupation, and the age difference. However, they do not specify how those five relate to each other. Instead, they state that by using the five factors as parameters, it is possible to draw a relationship table between the speaker and the listener to show the politeness, which is shown in table 1. The relationship of the two is very much connected to the polite expressions.

table 1: The relationship table by the linguists

The speaker is in the centre of the table, and the position of the listener is situated by considering the two dimensional table. The first two factors determine the position of the horizontal axis, called inside-outside. The other three factors determine the vertical axis, known as upper-lower axis.

The table divides into four areas, and each area has own rule. Also, if the position of the listener is in the shaded area, then it implies that the speaker is supposed to use polite expressions. For example, if the listener is a speaker's doctor, older, and not very close to the speaker, then the position of the listener would be in II. Therefore, the speaker has to use a polite expression.

### 3. Degree of politeness

However the positions of the listener and the speaker is specified, the concept of having the two-dimensional table using the inside-outside and the upper-lower axis could well cause a number of problems when it comes to the generation, due to the lack of rules.

This paper proposes the same task, but with another point of view called the degree of politeness. Diagram 1 shows a

structure of the polite expression generating system.

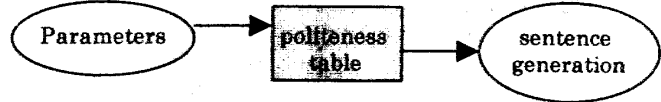


diagram 1: The sentence generation system

First of all, the parameters are given, which are the factors that the linguists stated. These data are passed on to the politeness table. At the same time, each word containing polite words lie on the same table, and does the calculation to select necessary words considering the relationship between the listener and the speaker. Then this structure becomes as a basic construction to complete in the sentence generation.

### 4. Method

The politeness table was designed based on the results of a questionnaire. All the factors including the five items stated by the linguists were already set. Also, a wide range of occupations of the listener was prepared in order to cover the most of people in the society. The examinees were asked to generate two sentences and supposed to think in the student point of view.

#### 4.1 Sentence "Do you always read the newspaper?"

The first sentence is a simple do-question. There was a number of limitations made and hence 64 possible ways. The most important part of the polite expressions is thought to be verbs. The sentence contains either "お読み" or "読む". Also, it is essential to include "れる・られる" expressions. Finally, any of particles such as "ます", "の", "です" and "か" acts as an important key for the generation.

The structural diagram for generating possible sentences was designed by using BNF, and is shown in diagram 2.

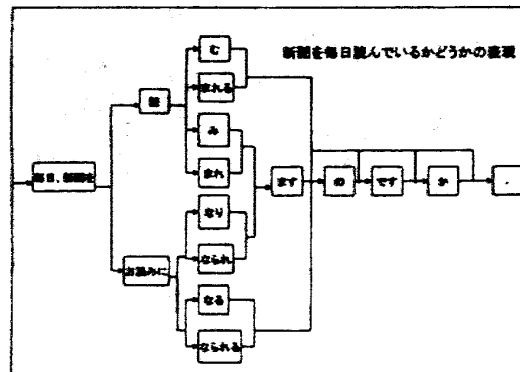


Diagram 2: Structural possible sentences

#### 4.2 Sentence asking for the book

The second sentence concerns about the speaker asking the listener for the book which is practically close to him/her, which allowed 163 possible sentences. The sentence is a little more complicated than the first one as it covers the questions of requests, possibilities, permission and showing the desire. The question of request is represented by "ください" or "くれ". The question of permission contains one of "かまわない", "よろしい", "良い" and "いい". The question showing the desire contains "ほしい" or "たい". Finally, the question of

possibility uses “取って” or “取ることができ”. Also affirmative questions containing “ます” or “る” and negative questions like “ません” or “る” were considered as important factors. Similarly, words containing “だ” or “です” were considered as factors as well. The verbs “もらう” and “いただく” mean the same, but as far as the level of politeness is concerned, they are thought to be completely different.

4.3 Data analysis

With the gathered data, I used quantification theory type III to statistically analyze and put both all the factors and the words onto the same axis on a graph in order to create the politeness table. Quantification theory type III deals with the data without criterion variables. This method creates a number of new axes as a ruler to find out certain subjects. We then assume that either the first or the second eigenvalues of the category is the degree of politeness.

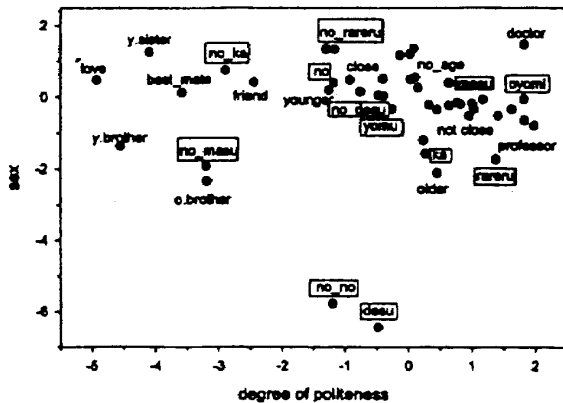
5. Results

5.1 Sentence 1

From 1140 samples, with 47 factors allowing 37 eigenvalues, the first eigenvalue came to 0.275, which we assume as a degree of politeness. The second eigenvalue we assume happened to be sex. We take x-axis as politeness, and the y-axis as sex. The category scores of x is plotted against the category scores of y for each factor, which is shown in graph 1 as a politeness table. The factors in boxes are expressing words, and the rest is factors on people. The higher the x value is, the more the speaker should be polite.

As the table shows clearly, the word “お読み” is in a higher position than the word “読む”. Also, “られる” is in a more polite position than “no\_られる”.

According to the table, professors and doctors are in the highest positions as far as the degree of politeness from the student point of view is concerned. Similarly, the closeness and the age differences are evidently located appropriately. This graph suggests that a student should use suitable words depending on the listener. For instance, when a student asks his professor a simple do-question, then s/he should use “ます”, “れる・られる” form, and “お読み” instead of “読む”.



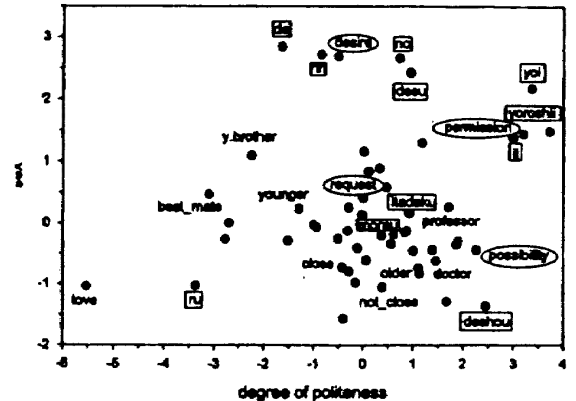
graph 1: politeness table for sentence 1

5.1 Sentence 2

The first eigenvalue out of 47 came to 0.442, which happened to be sex, and the second one was 0.34 which we assume as politeness. It enables us to create a graph in the same method as the previous graph.

The factors in boxes are expressing words, the ellipse shaped ones are the types of the expressions, and the rest is factors on people. The graph shows clearly that questions of asking for permission, and the possibilities situate in higher places than the questions of showing the desires and the requests. It also shows that the negative questions using

words such as “ない” or “ません” are in higher position than affirmative questions using words such as “ます” or “る”. The positions of the verbs “いただく” and “もらう” slightly differ from each other. Similarly, it shows that a students should respect people such as doctors, professors, but not their boy/girlfriends or their best friends. Therefore, the graph suggests that students should follow certain points for speaking politely. For instance, when a student asks the question to someone who is close and younger, then s/he should use showing the desire or the request form, rather than asking for the possibility or permission. S/he should not use “いただく” instead of “もらう”. S/he should also use “だ” instead of “です”.



graph 2: politeness table for sentence 2

6. Evaluation

To test the accuracy of the model, some of exactly the same situations (i.e. factors), but the similar expressions in conversations were gathered from a number of novels [4-6]. The rule is that the factors whether the words from the books are correctly used according of the listener is applied into the graph, and a certain type or words are selected, according to the degree of politeness table. Then compare to see if some words are correctly picked and to the rule stated in BNF.

For the do-questions, 34 out of 40 sentences were correctly selected, and for the second type of question, there were 29 correct sentences out of 38 trials.

7. Conclusion

This paper managed to prove that the polite expressing generation based on the politeness table was successful although a number of points should be discussed.

One of the reasons for the sentences which did not succeed to obtain the right expressions is that there was a large number of questions asking for permission no matter what the factors came to. Similarly, there was also a large number of questions showing the desire even though the listener was a doctor. It is possible to assume that there is another axis from one of the other 45 eigenvalues which might tell the relationship of question types.

As only sentences involving only two people were considered, it is now necessary to make a model involving a third person containing sentences with “giving” and “receiving”.

References

- [1] S. Akamine, Y. Ishihara, A. Okumura, K. Muraki, Generation of polite expressions in Multilingual Translation, Proc NLPRS, (1991) pp153-159
- [2] kokuritsu kokugokenkyujo, (1994), basic problems of learning honorific expressions, Okurashou
- [3] bunka chou, honorific expressions, (1993), Okurashou
- [4] Shinichi Hoshino, (1994), Welcome earth, shinchou-bunko
- [5] Yukio Mishima (1992), blue period, Kadokawa
- [6] Kenjiro Haitani (1991), next to children, Shinchou-bunko