# EXPERIMENTAL STUDY ON SIMILARITY OF DRUG NAMES AND CONFUSION ERRORS

YAMADE Yasuyo, HAGA Shigeru (Rikkyo Univ.), SHIN HongSon (National Research Institute of Earth Science and Disaster Prevention), TSUCHIYA Fumito (Tokyo Medical and Dental Univ.)
yamade@hej.rikkyo.ne.jp

The purpose of this study was to test whether the probability of confusion errors is related to the similarity indices calculated by the formulas proposed by Tsuchiya, et al. (2001): "cos1", "htco", and "edit". The results showed that the rate of confusion errors increased as a function of the degree of similarity. Further analysis of the results suggested the necessity of examining the effects of the coincidence of special features in characters such as "dakuten", and the effects of similarity in the form of characters.

## INTRODUCTION

Reports of errors in medical treatment related to drugs are increasing in an alarming way in Japan. Many are due to medical staff mistaking drugs. One of the factors contributing to these errors is the similarity of the names and appearances of drugs. With this situation in mind, Tsuchiya, et al. (2001) proposed a technique to quantify the similarity of a pair of drug names based upon the N-Grams method (Damashek, 1995). Aiming at testing the validity of these indices, we conducted a laboratory experiment to compare error rates between pairs of drugs with different levels of quantified similarity. The subject saw a drug name very briefly, and then he/she had to answer which drug had been shown from a pair of drugs. At the same time, in order to simulate real-life situations in which medical staff often do not pay full attention to the task, the subject was given another task, i.e. mental calculation, to do simultaneously.

## METHOD

### Subjects

Twenty-seven graduate and undergraduate students (15 males and 12 females) participated.

### Similarity indices

From the similarity indices suggested by Tsuchiya, et al. (2001), we used 3 indices.

*cos1* is calculated on the basis of how many characters are shared by the pair of drug names.

*htco* indicates the number of shared characters at the beginning and end of drug names.

*edit* shows how many editing actions (replacement, insertion or deletion of a character) are required for one drug name to turn into another.

### Selection of drug names and condition

There were about 23,000 drugs approved by the Ministry of Health, Welfare, and Labor in 2002. Among all possible combinations of five-character drug names, we searched for pairs to fulfill the 27 different conditions defined by three levels of similarity in the three similarity indices: *cos1* (high / middle / low) x *htco* (high / middle / low) x *edit* (high / middle / low). Actually only 15 of these combinations exist in Japanese drug names. We selected five pairs in each condition, that is, 75 pairs in total. All the 150 drugs constituting the pairs were different.

### Procedure

The experiment was carried out using a note PC (Fig.1). First, a drug name was shown for 500 ms in the 'question box' on the liquid crystal display as a target stimulus, simultaneously with two numerals, one on top, one below the box. Next, the drug name was masked and a pair of drugs, which included the former drug name, was displayed. The subject's task was first to answer the sum of the figures and then to identify which drug was displayed before. The experiment was over when all the 150 drugs had been shown once in a random order.

## RESULTS AND DISCUSSION

Although 27 students participated in the experiment, seven of them were excluded from the analysis because their error rate in mental arithmetic (summation of two one-digit numbers) exceeded 10 %, from which we suspected that they concentrated only on the drug name at the cost of the secondary task performance. Consequently, only the responses obtained from 20 subjects (9 men, 11 women) were used in the analysis reported below.

**Error rate as a function of the similarities**

The error rate was compared between pairs in different level of similarity, using three different indices independently. A one-way ANOVA detected significant effects of *cos1*, $F(2, 38) = 11.494$, $p<.01$, and *htco*, $F(2,38) = 7.078$, $p<.05$. Although error rate was not significantly different between the three levels of similarity for the *edit* factor, a similar tendency was seen [$F(2,38) = 2.378$, $p=.106$]. In sum, error rate increases as a function of the similarity measured by each of the three indices (Fig.2). Among the three, *cos1* seems to be the most useful predictor of confusion errors.

In addition, the results indicated the necessity to examine individually the effect of the similarity of the beginning (*head*) and ending (*tail*) of drug names. A one-way ANOVA was performed according to the degree of similarity of *head* and *tail* individually. The result showed that the *head* (Fig.3) and the *tail* also affected the error rate.

**Effect of features of characters**

Among the pairs listed in the higher ranks of error rate, some pairs were found not to be predicted by any of the similarity indices. Instead, it seems that confusion was due to the similarity in special features of characters consisting the drug names, such as "dakuten" (an indicator of voicing in consonants), "handakuten" (an indicator changing ‘h’ sound into ‘p’), and "chouon" (an indicator of long tone). (Readers, who are not familiar with Japanese characters, imagine *accents* in French or *umuraut* in German.)

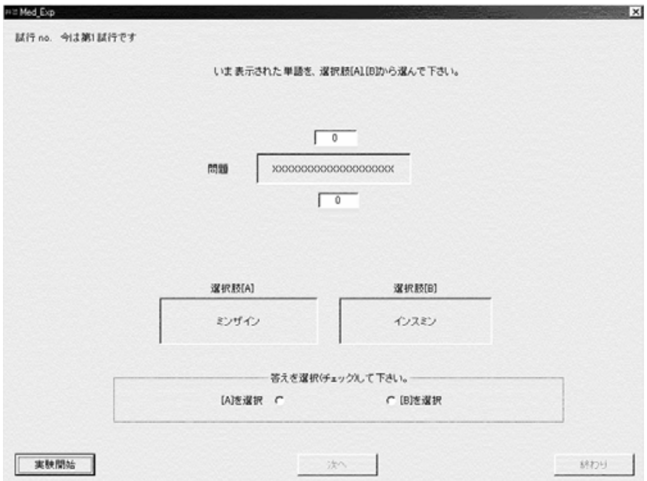


**Fig.1 Experiment display**

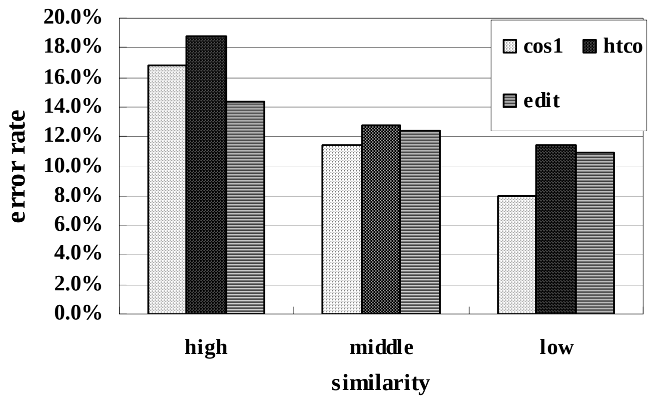


**Fig.2 Error rate of every index**

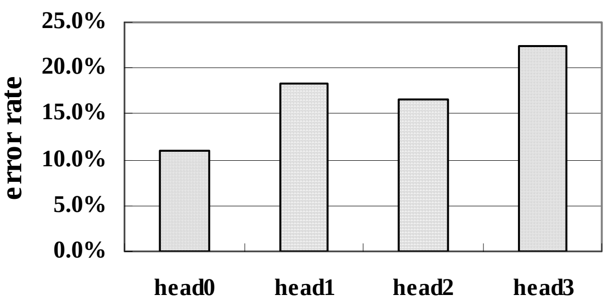


**Fig.3 Error rate of "head"**

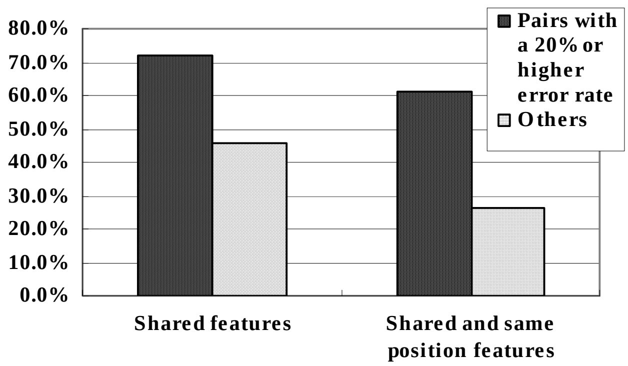


**Fig.4 Rate of sharing of features and position of features**

Shared features : a pair sharing features irrespective of feature position

Shared and same position features : a pair sharing features in the same position

To confirm this speculation, we classified all the pairs presented in the experiment into two groups: the more-confusing pairs, whose error rates were 20% or more, and the less-confusing pairs. Then we examined whether one or more of the special features existed in common in the tallies, and if so, whether their position in the spellings was the same or not. The result of this additional analysis is shown in Fig. 4. More pairs shared the features in the 'more-confusing' group, and more pairs in the group had the features in the same position.

Due to the secondary task requirement, the subject must read the drug name and numerals together. Consequently he/she might utilize the special features in the stimulus as a means of discrimination. Thus it is possible that, when the characters are similar, confusion arises more easily as a result of paying more attention to those special characteristics and less to the character itself. Future research in needed to clarify this.

## CONCLUSION

The similarity indices created by Tsuchiya, et al. (2001) predict probability of confusion errors in general. However, some pairs of drugs showed a high rate of confusion despite a lower level of similarity as measured by the existing indices. This result suggested, the necessity of examining the effects of the special features of characters and the appearance of characters. We will carry out further experiments, taking these factors into consideration, aiming at proactive safety regulations for drug names and the elimination of one of the major risk factors in medical treatment.

## REFERENCES

Fumito TSUCHIYA, Noboru KAWAMURA, Chie OH & Akihiro HARA 2001 Standardization and similarity deliberation of Drug-names *Japan Journal of Medical Informatics*, **21**(1), 59-67

Mark Damashek. 1995 Gauging Similarity with N-Grams : Language-Independent Categorization of text. Science, **267**, 843-848