

An Algorithm for Determining DingYu Structural Particle using Grammar Knowledge and Statistical Information

Fuji Ren

Faculty of Engineering, Tokushima University
Tokushima, 770-8506 Japan
ren@is.tokushima-u.ac.jp

Abstract. In a machine translation system from one language to Chinese, it is difficult to decide whether there is a structural particle "DE" between the "DingYu" and the "ZhongXinCi". The DingYu is a term in Chinese grammar which resembles the modifier and the attributive words or phrases in English or Japanese, but not the same. The ZhongXinCi is referred to the word modified by DingYu. Nowadays a practical Japanese-Chinese machine translation system based on translation rules has been implemented. However, the current system lacks the ability for resolving the problem mentioned above. To resolve this problem, this paper presents an algorithm for determining DingYu structural particle using the grammar knowledge and statistical information. We first collect a large number of grammar items from Chinese grammar books, and obtain some elementary judgment rules by classifying and inducing the collected grammar items. Then we put these judgment rules into use in actual Chinese language, and modify the rules by checking their results instantly. Lastly we check and modify the rules by using the statistical information from a actual corpus. An experiment system based on the proposed algorithm has been constructed and an experiment is carried out. The result shows the effectiveness of the presented method.

1 Introduction

The machine translation for practical use has been developed rapidly these days[2, 3]. The Japanese-Chinese and English-Chinese translations using machines are also being studied actively for the last several years[6, 8–11]. However, there are many unsolved problems that need to be further explored for Chinese processing. We are studying and developing of Japanese-Chinese machine translation based on the linguistic characteristics of the languages. This paper discusses a problem which will occur in case of machine translation for Chinese language, presents an algorithm for determining Chinese "DingYu"¹ structural

¹ In this paper, Roman alphabet is used to indicate the Chinese character. For example, *ji* indicates a Chinese character, and *ji suan ji* indicate three Chinese characters (meaning "computer"). Notice that the first *ji* and the last *ji* are written in different Chinese characters, although they are written same in the Roman alphabets.