International Journal of Innovative Computing, Information and Control
Volume **4**, Number **6**, June **2008**

ICIC International ©2008 ISSN 1349-4198
pp. 1371–1381

AUTOMATIC SUPER-FUNCTION EXTRACTION FOR TRANSLATION OF SPOKEN DIALOGUE

Manabu Sasayama[1], Fuji Ren[1,2] and Shingo Kuroiwa[1]

[1]Faculty of Engineering
The University of Tokushima
2-1, Minami Josanjima, Tokushima-shi, Tokushima 770-8506, Japan
{ sasayama; ren; kuroiwa }@is.tokushima-u.ac.jp

[2]School of Information Engineering
Beijing University of Posts and Telecommunications
Beijing 100876, P. R. China

Received May 2007; revised September 2007

Abstract. *Extraction of a large number of Super-Function (SF) is the most important factor in realizing SF based machine translation. This paper presents a method for automatic extraction of SF from a Japanese-English bilingual corpus. The extraction process uses a bilingual dictionary to match Japanese and English nouns in each sentence pair. The experimental results using a Japanese-English bilingual corpus show that this method performs very well in automatically extracting SF for machine translation. In addition, we evaluate the extracted SF in SF based machine translation.*
Keywords: Super-function, Automatic extraction, Machine translation

1. **Introduction.** Commercial machine translation systems are useful for understanding the gist of a text in a foreign language. However, commercial machine translation systems often mistranslate spoken dialogue. The reason, for this, is that there are many subject ellipsis in spoken dialogue [1].

Corpus-based machine translation is robust against subject ellipsis. There are two main approaches in corpus-based machine translation. One is statistical machine translation [2]. The other is example-based machine translation (EBMT) [3, 4, 5, 6, 7, 8, 9]. An advantage is that a good translation can be obtained by only performing a search to see if the sentence being translated exists in an exemplary bilingual corpus. But, the coverage of is limited by only using a bilingual corpus. It is inefficient to simply increase the quantity of in-domain bilingual corpus to cover various input sentences.

A method using a language model created from a in-domain corpus is proposed [12]. The method is limited to a condition because the method uses a specific corpus. Super-Function based machine translation (SFBMT) [10, 11] which is a type of EBMT extends coverage by functioning a bilingual corpus However, there are difficulties manually functioning a bilingual corpus, which resulted in only being able to create a limited quantity of SF and creating a practical system.

Other examples of corpus based machine translation are template-based machine translation [6, 8] and EBMT based on syntactic transfer [13]. EBMT based on syntactic transfer is similar to SFBMT. However, this method uses syntactic analysis to translate where SFBMT does not.

As other related research, [3] uses a thesaurus to find a sentence similar to an input sentence. Their research can distinguish a sentence from some example sentences. However, they mistranslate an input sentence using unnecessary nouns (Idiom's or collocation's