Automatic Synonym Acquisition
Based on Matching of Definition Sentences
in Multiple Dictionaries

Masaki Murata[1], Toshiyuki Kanamaru[2], and Hitoshi Isahara[1]

[1] National Institute of Information and Communications Technology,
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan,
{murata,isahara}@nict.go.jp,
WWW home page: http://www.nict.go.jp/jt/a132/members/murata/
[2] Kyoto University,
Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501, Japan,
kanamaru@hi.h.kyoto-u.ac.jp

Abstract. Studies on paraphrasing are important with respect to various research topics such as sentence generation, summarization, and question-answering. We consider the automatic extraction of synonyms (which are a kind of paraphrase) through the matching of word definitions from two dictionaries, and describe a new method for extracting paraphrases. Higher precision was obtained than with a conventional frequency-based method. The new method provided a precision rate of 0.764 for the top 500 data pairs and 0.220 for 500 randomly extracted data pairs when only synonyms were considered a correct answer. It provided a precision rate of 0.974 for the top 500 data pairs and 0.722 for 500 randomly extracted data pairs when hypernyms and similar expressions were also considered correct answers. Our method should be useful for other studies on paraphrase extraction.

1   Introduction

Studies on paraphrasing [6, 2] have had important consequences in various domains such as sentence generation, summarization, and question-answering [3, 12]. Likewise, studies on paraphrase extraction are also important. In this paper, we discuss the automatic extraction of synonym expressions which can be considered a kind of paraphrase. We extract synonym expressions by matching definitions of the same word from two dictionaries. In this work, we studied the extraction of synonym expressions in the Japanese language.

For example, we examined the definition sentences for the word *abekobe* meaning "reverse". Two Japanese dictionaries gave the definitions shown in Figure 1 for the word. We expected to extract pairs of expressions having the same meaning when we compared the two definitions, since they both defined the same word and thus had the same meaning. We compared the two definition