296 An idea on using hierarchical word meaning classification inside the example-based machine translation

Tantely Andriamanankasina†
Yoshikazu Miyanaga†
†Hokkaido University

Kenji Araki††
Koji Tochina†
††Hokkai-Gakuen University

1 Introduction

This paper proposes an idea of using word meaning classification tree inside the example-based machine translation [1]. It is expected to improve the quality of the translation in the selection of the best matching sentence. In this paper, the concept of the idea and the approach are presented. The system is implemented for French-Japanese translation, expecting that the proposed idea would give an interesting result to this unexplored field.

2 Motivation

The idea came from the fact that word meaning should be considered to improve the existing example-based machine translation system, as the following example illustrates:

Input sentence: j'ai un cahier (I have a notebook)
Example base:
(1) j'ai un stylo - watashi wa borupen wo motte iru
(2) j'ai un ami - watashi wa tomodachi ga iru
Both examples match with the input sentence, which translation will be used must be decided. Word meaning is generally used to solve this problem. The classification proposed here does not stay only in a simple level (human, thing) but goes as far as possible (place, time,...). An attempt to classify words hierarchically has been proposed [3], the present idea concerns its application in the example-based machine translation. Other work tried to solve the word-selection problem by taking word similarity value from a thesaurus [2], but building a good thesaurus still remains a problem. The idea presented here is to use the hierarchy number inside the word meaning classification tree to evaluate the distance of the example from input sentence.

3 Structure of the example base

The example base is a collection of couples "source sentence/target sentence", with the words correspondence. Every word is followed by its grammatical class (noun, verb,...) and its meaning class (place, time,...). The example base are constructed manually. In this first step, only classifications for nouns are considered. The general structure of the examples is the following:

Source: Je(pr) vais(v) a(pp) le(art) ecole(n,place).
Target: Watashi(pr) wa(pp) gakko(n) ni(pp) iku(v).
Correspondence: ([1],[1]), ([2],[5]), ([5],[3]).

4 Algorithm

Since the attention is especially focused on the effect of the word meaning classification, the matching algorithm was simplified by considering only the exact match at the level of grammatical class. For example, "Tu es belle" (You are beautiful) will match with "Il est malade" (He is ill) because both have the structure (pron) (be-present) (adj). After, hierarchical word classification is used to choose the best match among selected sentences and then the transfer is performed to get the target sentence.

5 Conclusion

Although some problems might appear because of wide variety of language phenomena, hierarchical word meaning classification is expected to improve the result of the selection of the best matching sentence, since it details the classification. Extension of the classification to verbs and using a complex matching algorithm considering this classification would be very interesting subject to deal with, depending on the results of this study.

References