

Source Language Effect on Translating Korean Honorifics

Kyonghee Paik¹, Kiyonori Ohtake¹, Francis Bond²,
Kazuhide Yamamoto^{1,3}

¹ ATR Spoken Language Translation Research Laboratories
2-2-2 Hikaridai “Keihanna Science City”, Kyoto 619-0288 JAPAN
{kyonghee.paik, kiyonori.ohtake}@atr.co.jp

² Nippon Telegraph and Telephone Corporation
2-4 Hikaridai, “Keihanna Science City”, Kyoto 619-0237 JAPAN
bond@cslab.kecl.ntt.co.jp

³ Nagaoka University of Technology
1603-1 Kamitomioka, Nagaoka City, Niigata 940-2188 JAPAN
yamamoto@fw.ipsj.or.jp

Abstract. This paper investigates honorific phenomena on two variants of Korean translation corpus, based on translations from Japanese and English. One surprising result is how different the corpora were, even after normalizing orthographic differences. Translations are dependent not just meaning, but also on the structure of the source text.

1 Introduction

This paper investigates the effect of source language on honorifics in translations using two variants of a Korean translation corpus. The original corpus was compiled from Japanese-English parallel sentences collected from phrase books for Japanese traveling abroad. The corpora used for this research consist of 324,616 Korean sentences. Half of the Korean sentences (162,308 sentences: henceforth, K_J) were translated from Japanese and the other half (henceforth, K_E) have been translated from English sentences which match the Japanese.

Although the two Korean corpora are generally equivalent in meaning, they have different characteristics, since they were translated from such different languages as English and Japanese. We show that the differences between English and Japanese lead to different results in the Korean translation, even though the original source texts were matching Japanese-English translation pairs.

Our ultimate goal is to find the differences in translations by comparing two variants with different source texts in order to improve the quality of machine translation. So far such comparison has not been extensively studied.

2 Honorifics in Japanese and Korean

Korean is an honorific language with an extremely systematic grammaticalization. A sentence cannot be uttered without the speaker’s knowledge of their