Post-editing User Interface Using Visualization of a Sentence Structure

Yudai Kishimoto† **Toshiaki Nakazawa**‡ **Daisuke Kawahara**† **Sadao Kurohashi**†
†Graduate School of Informatics, Kyoto University, Kyoto 606-8501
‡Japan Science and Technology Agency, Kawaguchi-shi, Saitama 332-0012
† {kishimoto,dk,kuro}@nlp.ist.i.kyoto-u.ac.jp
‡ nakazawa@pa.jst.jp

1 Purpose and characteristics

Translation has become increasingly important by virtue of globalization. To reduce the cost of translation, it is necessary to use machine translation and further to take advantage of post-editing based on the result of a machine translation for accurate information dissemination. Such post-editing (e.g., PET [Aziz et al., 2012]) can be used practically for translation between European languages, which has a high performance in statistical machine translation. However, due to the low accuracy of machine translation between languages with different word order, such as Japanese-English and Japanese-Chinese, post-editing has not been used actively.

We propose a post-editing system based on syntax-based machine translation to deal with different word order. For language pairs with different word order, it is time-consuming for a translator to understand what a machine translation system did. To solve this problem, our system displays the following three portions: a parse of the source language (A), a translation that keeps the word order of the source language (B), and a translation in the word order of the target language (C). This visualization makes the translator efficiently evaluate the quality of the translations and flexibly use the translations in various levels as follows.

1. If the parse or the translation is disorganized, the translator gives up using it and translates the source sentence from scratch. This can be efficiently judged mainly from A. Previous post-editing systems only displayed a final translation and made this judgment difficult.

2. If the translation is partially correct but has errors in word order, the translator changes the word order based on B. To make this process efficient, B is editable and translation blocks can be swapped on GUI.

3. If the translation does not have major errors including errors in word order, the translator makes a few revisions based on C. C is also editable and translation blocks can be swapped.

2 System description

The input of our system is a text file which contains the parse result of an original sentence and a translation result and a translation mapping.[1] This system only uses part of speech tags and dependency relations and can be expanded to many languages if we prepare these data. The output is a JavaScript file, and we can view the system result on a Web browser as in Figure 1.

Figure 1: Sample of post-editing interface

In Figure 1, we display three rectangles explained in section 1: A as a rectangle whose background color is sky blue B as a rectangle whose background color is pink, and C as a rectangle whose background color is orange. We can edit these rectangles in the way described in section 1.

3 Conclusion

We present a post-editing user interface using visualization of sentence structure. This system helps us to analyze the cause of errors more easily and hopefully will improve the efficiency of post-editing.

References

Aziz, W., Castilho, S., and Specia, L. (2012). Pet: a tool for post-editing and assessing machine translation. In Calzolari, N., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3982–3987, Istanbul, Turkey. European Language Resources Association (ELRA). ACL Anthology Identifier: L12-1587.

[1]Sample text is available at `http://lotus.kuee.kyoto-u.ac.jp/~yudaik/zh-ja_sample.txt`