

QRedit: An Integrated Editor System to Support Online Volunteer Translators

Takeshi Abekawa (abekawa@p.u-tokyo.ac.jp)

University of Tokyo

Kyo Kageura (kyo@p.u-tokyo.ac.jp)

University of Tokyo

1. Introduction

We are currently developing an English-to-Japanese computer-aided translation (CAT) system with the aim of aiding online volunteer translators, who are involved in translating online electronic documents in their free time. In the following, we will first describe the characteristics of target users, and discuss the basic concept of the system within which the choice of functions and system specifications has been made. Then we will describe the prototype CAT system which we are now experimentally providing to a limited number of online translators.

2. Background and basic concepts

Many CAT systems have been developed to date [see Kay 1997 for background information and Gow 2003 for a comparison of several CAT systems]. While some existing CAT systems have proved useful for some translators and translation companies, volunteer translators who work online do not use these systems [Kageura, et. al. 2006; see also Fulford and Granell Zafra 2004 in relation to the situation of freelance translators in Britain], for a variety of reasons: they are too expensive for personal use (except for Omega-T, which is free [Omega-T 2007]); they have more functions than are necessary; such translators are not under pressure from managers to keep a translation log in order to control quality, etc. This shows that some sort of CAT-lite system is required by online translators, which shares some basic functionalities with existing commercial CAT systems but with a different emphasis and overall design principle [Boitet, et. al. 2005].

We have therefore been engaged in developing a CAT system for online volunteer translators, based on the principle of maximally aiding the translators' work flow by removing existing obstacles, rather than providing extra functions which

translators have never used. Thus we are taking a bottom-up approach in developing system functions, reflecting the concrete requirements of translators. Another essential principle we have adopted is that it should be translators who make decisions, not the system.

After consulting some 15 volunteer translators working online, we identified the following issues as being particularly important:

1. Most English-to-Japanese volunteer translators do not have native level command in the source language (English). As a result, unlike such CAT systems as TransType [Macklovitch 2006] which assume that the target users are bilingual, well-trained and professional translators, and therefore try to reduce the time used for inputting the target language text, the main point to be tackled in our environment is making the reference lookup process as easy as possible.
2. When selecting translation expressions, translators examine various possibilities. When necessary, they look up more than one dictionary and check other information resources such as print encyclopaedias or Wikipedia, etc. The CAT system should facilitate the translation process by reducing the effort involved in looking up multiple resources. As it is not the mission of the system to decide on behalf of translators, good output for the system is considered to be a range of candidates and information that translators can take into account in making decisions, and not simply the same expression as translators chose.
3. Some online translators do not use online dictionaries, because doing so breaks the rhythm of translation and hinders the construction of the target text. Partly because of this, some prefer using print or standalone electronic dictionaries. A few use dictionaries built into the text editor, but none of them are fully satisfied with the dictionary look-up environment.
4. There is a pressing need for improved idiom and phrasal look-up. The importance of this function derives from two factors: (a) many translators, even experienced ones, have relatively less knowledge of idioms than of words, and (b) some idioms may not be identified as such by translators, because they make sense without an idiomatic interpretation. This leads to translation mistakes.
5. As our system does not aim to choose and restrict information on behalf of translators by using "sophisticated" NLP techniques, the amount of information it provides will naturally tend to increase. As a result, the user-interface becomes an important issue.

3. The prototype system QRedit

Based on these concepts and requirements, we are currently developing a prototype system that supports online translators. The system runs on the TOMCAT server and users access the system through a Web browser. The overall image of QRedit is shown in the figure. The browser screen is divided into two areas: (i) the source text area, and (ii) the target text area. The users can choose between horizontal and perpendicular division of the areas, i.e. source text area on the bottom and target language area on the top, or source text area on the left and target language area on the right. The two areas are linked with a synchronised scroll function.

3.1 Functions in the source text area

After a translator inputs the URL of a Web page (in which case the system analyses the tags and extracts the textual area automatically) or copies and pastes text into the source text area, the system activates the dictionary look-up functions. When the user clicks on a particular word in the source text area with the mouse, the system shows the translation candidates in a pop-up window.

Dictionary look-up functions

The system displays translation candidates from the dictionaries incorporated into the system [Sanseido 2004; Eijiro 2006]. The system does not only incorporate simple word look-up functions but also incorporates flexible idiom look-up functions. The idiom look-up functions can match such idiom occurrences as "He said that with his big fat tongue in his big fat cheek" with the dictionary entry "with one's tongue in one's cheek." This function has not been realised in any English-Japanese MT systems we have checked, and while some CAT systems realise similar functions through approximate matching, they do not specifically target the look-up of idioms with their variations. The system alerts users to idioms by marking them with an underline.

Displaying translation candidates

The system can display information in two ways. One is a small pop-up window displayed within the source text area, which includes only target word candidates. The other is a large pop-up window that displays all the information given under the headings in the dictionaries. The latter is particularly useful when translators wish to examine related information in detail before deciding on a translation, while the former is convenient for more straightforward dictionary look-up.

3.2 The connection between the source text area and the target text area

Translators can paste a selected expression from the list of candidates in the pop-up area to the target text area by clicking the mouse. The system also provides automatic transformation of numerical expressions to Japanese conventions and the pasting of HTML tags. This mouse operation does not affect the keyboard operation, and whatever the user does with the mouse, the keyboard cursor always stays on the target text editing area, so that the user can input translated text continuously.

3.3 Target text area

The target text area consists of the open source Web editor FCKeditor, which offers WYSIWYG textual decoration and editing, as well as saving and loading functions. The text can be saved in HTML as well as in basic textual format. The target area is split into paragraph spaces which correspond to the source text.

4. Conclusions

We are currently running a prototype system of QRedit that incorporates these functions, and accumulating feedback from a few online volunteer translators. The feedback that we have obtained so far can be categorised into two types:

1. The need to refine existing functions and improve basic usability, e.g., to improve the accuracy of the extraction of the textual area when the user specifies an URL for the source text area.
2. The need to incorporate higher-level functions, e.g. to allow the user to specify the register of the text that (s)he is translating, in order to have the system block irrelevant dictionary information from being displayed.
3. In addition, we are developing a module that automatically compiles bilingual technical terminologies from the Web, a module that detects existing translation pairs and recycles the translation information, and a system that detects candidates for transliterated expressions of proper names from the Web, all of which will be integrated into the QRedit environment.

Bibliography

Boitet, Christian, Youcef Bey, and Kyo Kageura. "Main Research Issues in Building Web Services for Mutualized,

Non-commercial Translation ." *Proceedings of the SNLP: 6th Symposium on Natural Language Processing* . 2005.

Eijiro. Accessed 2006-10-31. <<http://www.eijiro.jp>>

FCKeditor. Accessed 2006-10-31. <<http://www.fckeditor.net>>

Fulford, Heather, and Joaquin Granell Zafra. "The Uptake of Online Tools and Web-based Language Resources by Freelance Translators: Implications for Translator Training, Professional Development, and Research ." *Proceedings of the Second International Workshop on Language Resources for Translation Work, Research and Training* . 2004. 37-44.

Kageura, K. et. al. "Improving the Usability of Language Reference Tools for Translators ." *Proceedings of the 10th of Annual Meeting of the Japan Association for Natural Language Processing* . 2006. 707-710.

Kay, Martin. "The Proper Place of Men and Machines in Language Translation." *Machine Translation* 12.1-2 (1997): 3-23.

Macklovitch, Elliott. *TransType2: The Last Word*. LREC, 2006. 167-172.

Omega-T. 2007. <http://www.omegat.org/omegat/omegat_en/omegat.html>

Sanseido. *Grand Concise English Japanese Dictionary*. Tokyo: Sanseido, 2006.