

Hosting Volunteer Translators

Masao Utiyama MASTAR Project NICT Keihanna Science City 619-0288 Kyoto, Japan mutiyama@nict.go.jp	Takeshi Abekawa National Institute of Informatics 2-1-2, Hitotsubashi 101-8430 Tokyo, Japan abekawa@nii.ac.jp	Eiichiro Sumita MASTAR Project NICT Keihanna Science City 619-0288 Kyoto, Japan eiichiro.sumita@nict.go.jp	Kyo Kageura Tokyo University 7-3-1 Hongo, Bunkyo-ku 113-0033 Tokyo, Japan kyo@p.u-tokyo.ac.jp
---	---	--	--

Abstract

We have developed a web site called *Minna no Hon'yaku* (“Translation for Everyone by Everyone”), which hosts online volunteer translators. Its core features are (1) a blog-like look and feel; (2) the legal sharing of translations; (3) high quality, comprehensive language resources; and (4) the translation aid editor QRedit. Translators who use QRedit daily reported an up to 30 per cent reduction of the overall translation time. As of 3 July 2009, there are about 600 users and 4 groups registered to MNH, including such major NGOs as Amnesty International Japan and Democracy Now! Japan.

1 Introduction

Online volunteer translators, who are involved in translating online electronic documents in their free time, translate a variety of documents every day, such as blogs, Wikipedia articles, open source software manuals, documents on nongovernmental organization (NGO) activities, and so on.

These translations are read for pleasure, for practical purposes, for language learning, and many other reasons.

Needless to say, volunteer translators contribute a great deal to the sharing and spreading of information around the world. Consequently, supporting their activities is a very important research issue.

Volunteer translators translate a large number of documents everyday. However, they lack proper translation support tools (Abekawa and Kageura,

2007a). Thus, providing a good supporting environment should be of great assistance in improving volunteer translators’ efficiency and increasing the level of enjoyment they experience in translating. This is the motivation for our work in this paper.

2 Hosting volunteer translators

Abekawa and Kageura (2007a) have developed a translation aid editor, *QRedit*, which has been experimentally provided to a limited number of volunteer translators. They report that QRedit is very effective for aiding their work, as described in Section 8.

Based on the success of this translation aid editor, we have developed a web site called *Minna no Hon'yaku* (MNH, “Translation for Everyone by Everyone”) where everyone uses QRedit and other translation support tools. In addition, documents translated on MNH are open to the public because these translations are assigned open licenses such as Creative Commons licenses. A screenshot of MNH is shown in Figure 1.

Currently, MNH hosts volunteer translators who translate Japanese (English) documents into English (Japanese), as QRedit currently only supports Japanese-English and English-Japanese translation directions. We plan to extend QRedit and MNH to other language pairs in our future work. We also plan to make QRedit and the software system of MNH open source.

There are three reasons why we have developed MNH.

First, we were inspired by the success of hosting services for open source software such as `sourceforge.net`. These services help engi-



Figure 1: Screenshot of “Minna no Hon’yaku” site (<http://trans-aid.jp/>)

neers develop and distribute open source software. Our hope is that hosting services for volunteer translators could have a similar impact on the creation and distribution of translations.

Second, hosting volunteer translators enables them to use translation support tools without installation efforts (or fees) because support tools such as QRedit are built-in functions of MNH (and everyone uses MNH for free). This situation sharply contrasts to the usual situation of volunteer translators in which they do not use translation aid systems for a variety of reasons, such as the fact that they are too expensive for personal use. (Abekawa and Kageura, 2007a).

Third, hosting volunteer translators means that their documents are saved and published on MNH. These translations are shared by translators and used to make a parallel corpus. Translators can search this parallel corpus for expressions translated by other translators. This is an important benefit of hosting volunteer translators, for if a large number of translators use MNH, a large parallel corpus will become available on the site. Consequently, translators will be able to share a rich parallel corpus for their translation activities.

We opened MNH to the public on April 8 2009. We also asked several volunteer translation groups to use MNH as their translation platform in order to get feedback for improving MNH. We are now running MNH to see if MNH can attract and host many volunteer translators.¹

¹Currently, MNH only has a Japanese interface. We plan to

In the following sections, we describe the details of MNH.

3 MNH core features

MNH has the following four core features: (1) a blog-like look and feel; (2) the legal sharing of translations; (3) high quality, comprehensive language resources; and (4) the translation aid editor QRedit.

4 Blog-like look and feel

We designed the look and feel of MNH to be similar to those of standard blogs, to make it as easy as possible for a wide variety of volunteer translators to use.

For example, the top page of MNH (Figure 1) lists new translations, active translators, translations of Wikipedia articles, and so on. The three lines at the top part of the page list tags that are assigned by translators to their translations. By clicking on these tags, the corresponding translations are listed in the top page. The box in the upper right corner is the search box for documents, tags, translators, and so on. MNH also provides a place where translators can ask and answer questions on any topic. This mechanism is very useful for sharing translation experiences.

When a translator, *taro*, logs in to MNH, he is delivered to his personal page, as shown in Figure 2, where he edits his translations. To translate a new document, he enters the URL of the document into the long box at the top of the page and clicks the button on the right side. This launches the translation aid editor QRedit, which he uses to translate that document. The details of QRedit are described in Section 7. When he finishes his translation, the translation is saved to his personal space. He may open his translation to the public after assigning a proper license to the document. Open translations are shared by translators (and others) as described in Section 5.

Translators also share term lists which they enter into MNH. A screenshot of a term list is shown in Figure 3. These terms are treated as public domain contents as described in the terms of service of MNH. Thus, anyone can use these term lists for any purpose.

add an English interface by the end of this summer.



Figure 2: Screenshot of a personal page



Figure 3: Screenshot of a term list

5 Legal sharing of translations

Sharing translations legally is the third reason for the development of MNH as described in Section 2. To achieve this goal, MNH requires translators to assign open licenses to their translations if they make them public.

Translators cannot make their translations public without the permission of the original authors of translated texts. Thus, MNH first asks translators to confirm that the authors of original texts allow this. If they do not, MNH asks translators not to open their translations to the public and only use them for personal purposes.

If translators have the right to make their translations public, MNH next asks them to assign open licenses to their translations. This allows anyone to modify the translations and make derivative works public under certain conditions.

Specifically, MNH uses four Creative

Commons licenses. Brief descriptions of how these licenses work are set out below (<http://creativecommons.org/about/licenses/>).

Attribution This license lets others distribute, remix, tweak, and build upon the translators work, even commercially, as long as they credit the translator for the original creation.

Attribution Share Alike This license lets others remix, tweak, and build upon the translator's work even for commercial reasons, as long as they credit the translator and license their new creations under the identical terms.

Attribution Non-Commercial This license lets others remix, tweak, and build upon the translator's work non-commercially, and although their new works must also acknowledge the translator and be non-commercial, they don't have to license their derivative works on the same terms.

Attribution Non-Commercial Share Alike This license lets others remix, tweak, and build upon the translator's work non-commercially, as long as they credit the translator and license their new creations under identical terms.

Translators can also use other licenses that are similar to the Creative Commons licenses listed above.

In this way, translators can legally share their translations on MNH. These shared translations are used to make a parallel corpus by using a sentence alignment method (Utiyama and Isahara, 2003). Currently, MNH has a simple bilingual concordancer as shown in Figure 4. We plan to extend this concordancer in our future work because a bilingual concordancer is a very important tool for translation (Macklovitch et al., 2009).

6 High quality, comprehensive language resources

Dictionaries and the web are the two main language resources that online volunteer translators use during translation.

MNH, in cooperation with Sanseido, provides the "Grand Concise English Japanese Dictionary"

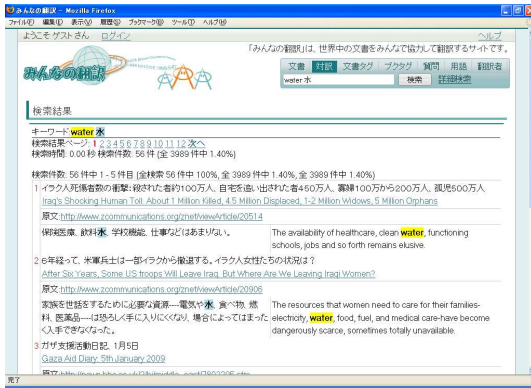


Figure 4: Screenshot of bilingual search results

(Sanseido, 2006) to translators. It has about 360,000 entries and is socially accepted as a standard and comprehensive dictionary. Consequently, translators will not need to use other dictionaries in most cases. We also plan to incorporate a comprehensive Japanese-English dictionary of names, collected from the web and consisting of about 400,000 entries (Sato, 2009), in order to handle person names that are not covered in usual bilingual dictionaries.

MNH also provides seamless access to the web, because the web is a very important resource for checking factual information.² For example, MNH provides a dictionary that was made from the English Wikipedia. This enables translators to reference Wikipedia articles during the translation process as if they are looking up dictionaries. MNH also provides a seamless connection to web searches as described in the next section.

7 Translation aid editor: QRedit

7.1 About QRedit

QRedit is a translation aid system which is designed for volunteer translators working mainly online (Abekawa and Kageura, 2007a). Volunteer translators involved in translating online documents have a variety of backgrounds. Some are professional translators, some translate documents about topics they are interested in, while others translate as a part of their NGO activities. They nevertheless share a few basic characteristics:

²The fact check function was traditionally provided by libraries, but online translators normally use the web instead of going to libraries.

- (1) They are native-speakers of the target language (TL).
- (2) Most of them do not have a native-level command in the source language (SL).
- (3) They do not use a translation aid system or machine translation (MT) system.
- (4) They want to reduce the burden involved in the process of translation.
- (5) They spend a great deal of time looking up reference sources.
- (6) The smallest basic unit of translation is the paragraph and “at a glance” readability of the SL text is very important.

The philosophy of QRedit reflects these characteristics and can be summarized as follows:

- (R1) To reduce the time it takes for translators to do what they are currently doing, rather than to add new functions;
- (R2) To provide sufficient information sources so that translators do not have to look elsewhere;
- (R3) To provide information to facilitate decision-making by translators;
- (R4) To provide information that triggers translators’ creativity;
- (R5) To make the interface as simple as possible.

These requirements stem from the requests made by volunteer translators (Abekawa and Kageura, 2007a). QRedit was designed to meet these requirements. For example, it uses the high quality, comprehensive language resources described in Section 6, based on R2. It also provides a seamless connection to web searches, based on R4. Further, based on R1 and R3, QRedit does not provide an MT function, for the following reasons:

- (1) In terms of the quality of translation, MT results are far behind human translations for Japanese-English (or English-Japanese) translation. As a result, using MT results does not contribute to the reduction of the overall translation time.³

³For some language pairs such as English-French, MT systems can be used to prepare rough translations. In that case, we can easily extend QRedit to incorporate MT. For example, we can regard MT as an extended word lookup function, so that MT results are presented to users in addition to dictionary entries as shown in Figure 5.



Figure 5: Screenshot of QRedit

(2) Given input texts, MT systems produce output translations without human intervention. This situation contradicts R3. Even if MT outputs are good translations, for now most translators are not ready to accept MT results as reliable, because they do not regard computers as communication partners. Thus, they want to decide on translations and control translation process by themselves.

Finally, we made the interface of QRedit as simple as possible to meet R5.

In the following sections, we introduce major features of QRedit.

7.2 Automatic word lookup

When a URL of an SL text is input into QRedit, it loads the corresponding text into the left-hand panel, as shown in Figure 5. (Users can also copy-and-paste the SL text.) Then, QRedit automatically looks up all words in the SL text using the dictionaries described in Section 6. When a user clicks on an SL word, its translation candidates are displayed in a pop-up window. The user can paste a translation candidate into the right-hand panel, which is used for writing the translation, as described in Section 7.3.

7.2.1 Idiom lookup

In addition to the single word lookup method, QRedit has a flexible multi-word unit lookup function (Takeuchi et al., 2007). For example, QRedit automatically looks up the dictionary entry “with one’s tongue in one’s cheek” for the expression “He said that *with his big fat tongue in his big fat cheek*”

or “head screwed on *right*” for “head screwed on *wrong*”.

This function is a major advantage of QRedit compared to other MT systems or computer-aided translation (CAT) systems. Indeed, this function has not been realized in any English-Japanese MT system we have checked, and while some CAT systems realize similar functions through approximate matching, they do not specifically target the look-up of idioms with their variations.

The lack of flexible idiom lookup in other systems does not mean that this function is not needed. In fact, there is a pressing need for it. 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.

The flexible multi-word unit lookup function of QRedit helps translators identify idioms, and thus reduce translation mistakes.

7.2.2 Stratified term emphasis

Because idioms are often missed by translators, QRedit notifies translators of the existence of idioms or terms that have special meanings.

To notify users of the existence of such terms, it is good to highlight the terms with text decoration. But too many highlighted terms reduces the readability of source language texts dramatically. This is a serious problem, because the richer the reference sources become, the greater the number of candidates for notification.

To resolve this problem, QRedit adopts a stratified term emphasis method, which distinguishes three user awareness levels depending on the type and nature of the reference unit, or the candidate term for notification. These awareness levels are reflected in the way the reference units are displayed, such as a change of background color or underlining. These levels are decided according to a four criteria: “composition,” “difficulty,” “specialty” and “resource type.” See Figure 5 for an example and refer to (Abekawa and Kageura, 2007b) for details of this method.

7.3 Lookup that doesn't disturb translation

When users click on a term in the left-hand panel, its translation candidates are displayed in a pop-up window, as shown in Figure 5. This action does not affect keyboard operation. That is, users can still write their translations continuously without moving the mouse cursor back to the right-hand panel, as whatever they do with the mouse, the keyboard cursor always stays in this panel.

This function helps translators concentrate on translation, as the interviews in Section 8 show. If translators had to reset the cursor every time they looked up the dictionaries, it would break the rhythm of their work. QRedit helps them maintain their rhythm and focus on their translation activities.

7.4 Functions associated with lookup

Translators can call up the following four functions for each term and each translated candidate in a pop-up window, as shown in Figure 5.

- **Paste**
Paste a term or translation term to the right-hand panel. This helps the translator avoid misspelling of long named entities or numerical representations.
- **Detailed lookup**
Display complete dictionary entries for the term: attributes (domain, nuance, etc.), relative information (derivative words, antonyms, etc.), and example sentences. These entries are displayed in another window because of the large amount of information.
- **Web search**
Display search results from a web search engine. This help translators confirm in which context the term is used or how to use the term.
- **Register term**
Register a term to user's term list. Note that users can look up registered terms in QRedit.

7.5 Lookup by keyboard

To display a pop-up window, translators click on terms with their mouse. But mouse operation can be an obstacle to efficient input of TL texts. Therefore, in QRedit, translators can look up terms by using only their keyboards. While inputting TL

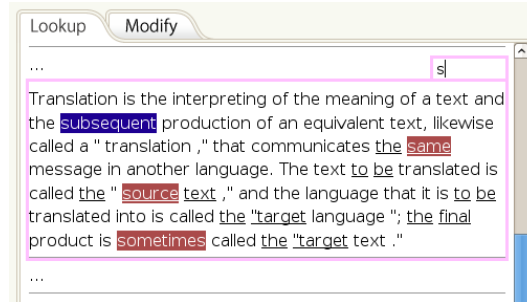


Figure 6: Lookup by incremental search

texts, translators can activate a keyboard incremental search method. Figure 6 shows an example of an incremental search when the user has input the first character 's'.

7.6 Customization

The environment for inputting TL texts can greatly impact the efficiency of translation work. In addition, an optimal environment differs from translator to translator. So, a flexible customization method that meets the needs of various translators is required. In QRedit, users can customize the following factors in their translation input environments.

- Placement of the SL area [left, right, top, bottom]
- Width or height of the SL area
- Placement of translation candidate display [inside, left, right] of the SL area
- Synchronized scroll [both directions, source→target, target→source, none]

Figure 7 shows an example of a customized QRedit window. (See Figure 5 for a default window.)

8 User response

As of 3 July, 2009 – three months after we made MNH and QRedit publicly available – there are about 600 users and 4 groups registered to MNH, including such major NGOs as Amnesty International Japan and Democracy Now! Japan.

As quantitatively evaluating the benefit of using translation aid systems is technically a difficult task (cf. (Macklovitch, 2006)), and as we are dealing



Figure 7: Customized QRedit window

with volunteer translators who are not working on a “time is money” basis but rather wish to reduce the subjective burden of translation, we carried out qualitative evaluations through e-mail and phone interviews with three users. They are trial users who have been using QRedit since before it was made public, thus have enough experience to give informed judgment on the system.

According to them, without QRedit, the division of time for the different elements of translation is roughly: 10 to 30 per cent for dictionary lookup, 10 to 40 per cent for searching the web to obtain information, 20 to 50 per cent for draft translation, and 0 to 10 per cent for revision and editing. All in all, around one fourth to half of the time is used for dictionary and web lookup.

Translator A, a novice translator with two years’ experience, said that she feels it takes her an average of 20 per cent and a maximum of 30 per cent less time to complete translations using QRedit, with the clear additional benefit that the quality of draft translations is improved because she can concentrate on context/paragraph reading rather than dictionary lookup in making draft translations. Translator B, a middle level translator with three years’ experience, also reported an average 25 per cent to 30 per cent reduction in the overall translation time, with the same effect on quality as translator A. Translator C, who is an expert translator, reported that she prefers the environment she is familiar with for translating easier texts, but can reduce translation time by 20 per cent when dealing with other texts. She said that QRedit enables her to tackle a wider variety of texts

than before.

We have not yet been able to obtain responses from group users, because they have only started using QRedit recently and thus have not accumulated sufficient experience to give an informed judgment on the system.

We also have not been able to evaluate the usability of MNH because it is still under development and we are now improving its usability based on comments and suggestions from users.

9 Related work

Related work can be roughly classified into three types, i.e. projects that aim at translation and publishing translated documents; work that is concerned with hosting translated documents, often multilingually; and work that is addressed at aiding translators and translation communities.

There are too many joint or collaborative online translation projects to mention here. GlobalVoices Online⁴ is perhaps one of the most well known, along with TUP⁵ (Translators United for Peace). Most projects do not provide translation aid facilities or collaborative working environments. They are rather projects defined by interested groups of people, using existing facilities.

An example of the second category is Yakushite.net (Shimohata et al., 2001). It provides a collaborative translation environment in which users can use MT for translation, while contributing to collaborative terminology augmentation for the improvement of MT. Except for providing the MT engine, the translation aid functions are weak. Worldwide Lexicon (McConnell, 2007) is another example. Within the project a variety of mechanisms are provided that facilitate the sharing of translated documents world wide, with which one can (i) detect translated texts, if there are any; (ii) translate by oneself; (iii) subscribe to an RSS feed for translation; and (iv) use machine translation. The system, however, does not provide rich facilities to aid human translations (step (ii)). As such, the system is more a hosting service rather than a translation aid system.

⁴<http://globalvoicesonline.org/>

⁵<http://groups.yahoo.co.jp/group/TUP-Bulletin/>

We are also witnessing the rapid growth of Wiki-based platforms to facilitate collaborative translations, such as Traduwiki (related projects are listed in <http://wiki-translation.com/tiki-index.php>) and BEYtrans (Bey et al., 2008). They provide functions that support collaboration among translators, as well as providing a translation memory function, thus having the features of translation aid systems as well as translation hosting functions.

Turning our eyes to fully-fledged translation aid systems, there are several commercial and non-commercial systems. SDL Trados⁶ is one of the most well known and widely used. SDL also developed Idiom WorldServer system⁷, an online multilingual document management system with translation memory functions. There are many other systems such as TransType (Macklovitch, 2006) and the free translation memory and terminology management system Omega-T⁸. Though it is now a little dated, Gow (2003) evaluates different translation aid systems.

The functions that MNH provides are closer to those provided by Idiom WorldServer or Wiki-based collaborative translation aid systems, but MNH provides a high-quality bilingual dictionary and functions for seamless Wikipedia and web searches within the integrated translation aid editor Qredit, thus providing connections to the existing reference information infrastructure that online translators use. This reflects the fact that the main target of MNH is online documents available under the Creative Commons licenses. On the other hand, collaborative, project-oriented document management functions are weak in MNH, because MNH basically assumes use by individual translators or smaller groups. Implementing this functionality will be a future task.

10 Conclusion

We have developed a web site called *Minna no Hon'yaku* (MNH, “Translation for Everyone by Everyone”), which hosts online volunteer translators. Its core features are (1) a blog-like look and feel;

(2) the legal sharing of translations; (3) high quality, comprehensive language resources; and (4) the translation aid editor; Qredit. Translators who use Qredit daily reported an up to 30 per cent reduction of their overall translation time. We are now running MNH to see if it can attract and host many volunteer translators. Please join MNH!

References

- Takeshi Abekawa and Kyo Kageura. 2007a. Qredit: An integrated editor system to support online volunteer translators. In *Digital humanities*, pages 3–5.
- Takeshi Abekawa and Kyo Kageura. 2007b. A translation aid system with a stratified lookup interface. In *ACL Poster and Demo*, pages 5–8.
- Youcef Bey, Kyo Kageura, and Christian Boitet. 2008. BEYTrans: A Wiki-based environment for helping online volunteer translators. Yuste, E. ed. *Topics in Language Resources for Translation and Localisation*. Amsterdam: John Benjamins. p. 139–154.
- Francie Gow. 2003. *Metrics for Evaluating Translation Memory Software*. Ph.D. thesis, University of Ottawa.
- Elliott Macklovitch, Guy Lapalme, and Fabrizio Gotti. 2009. *TransSearch: What are translators looking for?* In *AMTA*, pages 412–419.
- Elliott Macklovitch. 2006. TransType2: The last word. In *LREC*, pages 167–172.
- Brian McConnell. 2007. The worldwide lexicon: Adding collaborative translation to your site. <http://www.oreillynet.com/pub/a/etel/2007/09/27/the-worldwide-lexicon-adding-collaborative-translation-to-your-site.html>, <http://worldwidelexicon.appspot.com/>.
- Sanseido. 2006. *Grand Concise English Japanese Dictionary*. Tokyo, Sanseido.
- Satoshi Sato. 2009. Crawling English-Japanese person-name transliterations from the web. In *WWW 2009 Poster Sessions*, pages 1151–1152.
- Sayori Shimohata, Mihoko Kitamura, Tatsuya Sukehiro, and Toshiki Murata. 2001. Collaborative translation environment on the Web. In *MT Summit*, pages 331–334.
- Koichi Takeuchi, Takashi Kanehira, Kazuki Hilao, Takeshi Abekawa, and Kyo Kageura. 2007. Flexible automatic look-up of English idiom entries in dictionaries. In *MT Summit*, pages 451–458.
- Masao Utiyama and Hitoshi Isahara. 2003. Reliable measures for aligning Japanese-English news articles and sentences. In *ACL*, pages 72–79.

⁶<http://www.translationzone.com/>

⁷<http://www.idiominc.com/en/>

⁸<http://www.omegat.org/>