

# The Strategic Supply of Reference Resources and Lookup Environments via an Online Translation Hosting Site

Masao Utiyama<sup>§</sup>, Takeshi Abekawa<sup>#</sup>, Kyo Kageura<sup>¶</sup> and Eiichiro Sumita<sup>§</sup>

<sup>§</sup>National Institute of Information and Communication Technology, Japan  
({mutiyama, [eiichiro.sumita](mailto:eiichiro.sumita@nict.go.jp)}@nict.go.jp)

<sup>#</sup>National Institute of Informatics, Japan (abekawa@nii.ac.jp)

<sup>¶</sup>University of Tokyo, Japan (kyo@p.u-tokyo.ac.jp)

## Abstract

In this demonstration, we introduce the strategic supply of a range of reference resources for translators via Minna no Hon'yaku (MNH: <http://trans-aid.jp/>), a translation hosting site with integrated translation-aid mechanisms, which was made publicly available in April 2009. As of June 2011, nearly 1700 users have registered with MNH and more than 7000 documents have been translated, of which over 2800 have been published on the site.

## 1. Introduction

This demonstration introduces the translation hosting site Minna no Hon'yaku (MNH: translation of/for/by all) accessible at <http://trans-aid.jp/>, placing special emphasis on the role and status of the range of reference tools provided via the site. MNH was made public on April 7, 2009. As of June 22, 2011, 1669 users have registered with MNH, 7023 documents have been translated using the translation-aid functions provided by MNH, and of these 2808 have been published on the MNH site. Currently MNH accommodates English-to-Japanese, Japanese-to-English, English-to-Chinese, Chinese-to-English, Japanese-to-Chinese, Chinese-to-Japanese, and English-to-Catalan language pairs. In this presentation we focus on English-to-Japanese translation, as reference resources are most complete in the English-to-Japanese translation environment on MNH and constitute a model case for the future development of reference resources for other language pairs. Section 2 provides a brief overview of MNH, while section 3 sets out general desiderata for reference resources in translation. Section 4 outlines the reference resources provided by MNH together with their conceptual status and importance.

## 2. A brief overview of MNH

MNH consists of two main parts: (a) the MNH platform, which manages all the information related to users and translations, and (b) QRedit, which is an advanced translation-aid editor facilitating translation itself. The basic functions provided by the MNH platform can be summarized as follows (Utiyama, et. al. 2009):

1. Users can register with MNH anonymously, and are provided with their own personal space;
2. Users can publish their translations on the MNH site, when copyright permits;
3. Users can take advantage of a variety of social networking functions, including social



(b) The state of existing reference resources. The number of reference resources, especially lexical and terminological resources used by translators, is finite at any given time for a given language or language pair.

(c) The common perception of and attitude towards resources by translators and users. Checking certain reference resources, especially those used for checking TL expressions, are obligatory for translators, such as the authentic terminological lexicon of a given domain.

From consideration of these factors, the concepts of “(relative) comprehensiveness” and “normativeness” can be derived, and very rough desiderata for reference resources can be prescribed on the basis of these two concepts (Kageura & Abekawa, 2011):

(a) With regard to general dictionaries (which are used for looking up SL meanings), at least one of the most comprehensive dictionaries available should be provided;

(b) With regard to resources used for looking up TL expressions, such as terminological dictionaries, proper name dictionaries and translation memories, those resources accepted as necessary among users, even if they are not comprehensive, should be provided. If authentic resources are not available (which is often the case with terminological dictionaries and other resources), the system need to provide resources that are relatively more comprehensive than others. Also, irrespective of whether authentic resources are available or not, resources that may be less authentic but more comprehensive should be provided as secondary resources;

(c) With regard to online textual resources (corpora) that are used as a surrogate for comprehensive library searches, those accepted as an essential resource should be provided. Currently, this means Google search should be provided from within the system.

#### **4. The nature of the reference resources provided by MNH**

##### **4.1 General comprehensive dictionaries commonly used by translators**

In MNH, we provide two of the most commonly used and most comprehensive English-to-Japanese dictionaries available in Japan, i.e. *Sanseido's Grand Concise English-Japanese Dictionary* (Sanseido, 2001) and *Eijiro* (Eijiro, 2010). While there are other high-quality, comprehensive dictionaries on the market, we chose Sanseido (2011) because it has good coverage of topics frequently dealt with online, and is thus likely to be of particular use to online translators. Sanseido (2001) supports both normativeness and relative comprehensiveness, while Eijiro (2010) supports comprehensiveness.

##### **4.2 Google Web and dictionary search**

Users can trigger Google Web search and dictionary search in a seamless manner from within Qredit. By means of the Google Web search function, MNH provides a comprehensive online archival search on the Internet. On the other hand, Google dictionary search can be less normative than that provided by Sanseido (2001), but it contributes collectively to the comprehensiveness of ordinary dictionaries provided by MNH.

##### **4.3 Translation memories provided by users**

Some group users (NGOs) such as Amnesty International Japan or DemocracyNow! Japan provide translation memories that could potentially be recycled for checking target language expressions. As this class of information depends on users or user groups, MNH provides a platform and mechanism through which users who want to look up translations they made in the past can register their data as translation memories. In addition, the translations made using MNH and made public are automatically incorporated into \*\*\*the

translation memory database/OR/translation memory databases ?? \*\*\*. These resources constitute reference information that is “comprehensive” in a narrow sense, i.e. it is comprehensive in terms of the task carried out by those users who provide the translation memory data.

#### **4.4 User-registered terminologies**

Like translation memories, users can register terminological lists maintained specifically by them. The status of this terminological data is essentially the same as that of translation memories, the only difference being the units of expressions being looked up.

#### **4.5 Automatically collected resources**

A semi-automatically constructed proper name bilingual lexicon (Sato, 2009) and a large-scale, comprehensive semi-automatically compiled terminological lexicon (Abekawa & Kageura, 2009) are to be provided as well. While we do not have space here to explain the technical details, these two lexicons are constructed on the basis of a strategy completely different from most existing corpus-based methods. The methods adopted in generating the bilingual proper name lexicon and terminological lexicon simulate what translators do online in looking for translations of proper names and terms. Thus they guarantee that what users can find online can be found in the lexicons, while simultaneously reducing the burden on translators of the current trial-and-error method of looking for necessary information using Google search. These lexicons currently run as independent systems but they will be incorporated into MNH in the near future.

#### **Acknowledgements**

This work is partly supported by the Japan Society for the Promotion of Sciences (JSPS) grant-in-aid (A) 21240021 “Developing an integrated translation-aid site which provides comprehensive reference sources for translators” and by the 2011 National Institute of Informatics (NII) research cooperation project “Construction and use of practical terminological resources from a variety of information sources.”

#### **Bibliography**

- Abekawa, T. and Kageura, K. (2007). “A translation aid system with a stratified lookup interface.” *Proceedings of ACL 2007 Demo and Poster Sessions*. 5-8.
- Abekawa, T. and Kageura, K. (2009). “QRpotato: A system that exhaustively collects bilingual technical term pairs from the Web.” *Proceedings of the 3rd International Universal Communication Symposium*. 115-119.
- Eijiro*, Tokyo: Aruku, 2010.
- Kageura, K. and Abekawa, T. (2011). “On the concept of 'comprehensiveness' in information services,” *Proceedings of A-LIEP 2011*. 496-505.
- Sanseido's Grand Concise English Japanese Dictionary*. Tokyo: Sanseido, 2001.
- Sato, S. (2009). “Crawling English-Japanese person-name transliterations from the Web.” *WWW 2009*. 1151-1152.
- Takeuchi, K., Kanehira, T., Hilao, K., Abekawa, T. and Kageura, K. (2007). “Flexible automatic look-up of English idiom entries in dictionaries.” *MT Summit XI*. 451-458.
- Utiyama, M.; Abekawa, T.; Sumita, E.; Kageura, K. (2009). “Hosting volunteer translators.” *MT Summit XII*. No pagination.