English Bilingual Lexicography in Japan:
Meeting Serious Challenges

Yukio TONO
Meikai University

Abstract
This paper discusses the serious challenges that English bilingual lexicography in Japan is facing now, which are threefold: corpus lexicography, misconceptions about user-friendliness, and changing modes of dictionaries. A new paradigm in the interaction between dictionaries and users has been emerging, which makes research on dictionary use even more essential.

1. Introduction

In a general sense, dictionary-making is a history of copying and adaptation. Describing word usage with illustrative examples, categorizing different patterns of usage using signs and symbols, signposting different senses of words, are among the many features which were invented by our predecessors and copied or adapted in subsequent dictionaries. Japanese bilingual lexicography is no exception; it has benefited considerably from what came before. During the Meiji and Taisho eras, for example, English bilingual dictionaries in our country were heavily influenced by both British (e. g. the *Concise Oxford Dictionary*) and American lexicography (e. g. the *Webster's* or the *Century English Dictionary*). Pedagogical dictionaries in Japan also owe their successful development to the ground-breaking work by Harold E. Palmer and A. S. Hornby.

Whilst we owe much to the heritage of English lexicography in Western countries, Japanese bilingual lexicography from the 1960s to the 1980s made tremendous progress by harnessing homegrown skills and creativity. Unique features such as frequency symbols, transparent verb pattern codes, cross-linguistic usage information, learner error

notes, vocabulary-building columns, among others, made dictionaries more user-friendly. Many of these innovative ideas were later adapted and adopted by monolingual learner's dictionaries published in the United Kingdom.

Despite its huge success in the second half of the 20th century in terms of its market growth and the sophistication of its content, current bilingual lexicography in Japan is facing serious challenges. There are at least three different challenges that will possibly determine the future directions of bilingual dictionary-making in our country.

2. First challenge: corpus lexicography

It was not until the late 1980s that bilingual lexicography in Japan, which was famous for its originality and high standards, was "put to the test" in the light of a new lexicographical paradigm. The publication of the *COBUILD English Dictionary* (1987) challenged lexicographical conventions by introducing the notion of a "corpus-based" lexicography. In *COBUILD*, every illustrative example was taken from actual corpus data. The sense order of entries was re-examined in the light of frequencies of grammatical collocational patterns. These new ideas had never been tried out in Japanese bi-, lingual dictionaries.

After the *COBUILD English Dictionary*, all other major UK publishers followed, and revised their dictionaries using corpora. In 1995, the so-called Big Four (*OALD, LDOCE, CIDE and COBUILD*) were either revised or overhauled, and all began claiming that they were "corpus-based". This marked the beginning of a new era of corpus lexicography, and a widening gap began to form between Europe and Japan in terms of the use of corpora for dictionary-making.

One of the reasons UK publishers made the significant shift to the use of corpora is that they were focused on making better dictionaries for foreign language learners. For general-purpose monolingual dictionaries, the main focus is on meaning and spellings. Other types of information are of secondary importance because native speakers know how to construct a phrase or a sentence intuitively. Learner's dictionaries, however, need to pay attention to encoding information such as collocations, an area where corpus data is extremely valuable. Whilst pioneering work for a dictionary of collocations was completed by Senkichiro Katsumata in the 1930s (*Kenkyusha's Dictionary of*

English Collocations, 1939), it never occurred to Japanese publishers to store the language resources in-house and used corpora in a more sophisticated manner. Most Japanese publishers had no experience of keeping full-time lexicographers in-house. Instead, they recruited linguists and language teachers to form the editorial team, and got them to design and write the most important entries or certain featured columns only. In the meantime, the rest of the dictionary entries would be subcontracted to another dictionary-production company. As a result, no resources, including word frequency lists and example data processed from corpora, were stored by publishers. The subcontractors kept all those resources and sometimes re-used the data for the works by rival companies.

It was not until the 21st century that Japanese publishers started to produce dictionaries based on corpus data. In 2003, the *Wisdom English-Japanese Dictionary* was published by Sanseido, which claimed to be the first corpus-based medium-sized learner's dictionary in Japan. Shogakukan followed, producing two corpus-based or corpus-informed dictionaries: one was a semi-bilingual dictionary entitled the *Shogakukan-Cambridge Learner's Dictionary Semi-bilingual Version* (2004) and the other was a specialized dictionary called the *Shogakukan Corpus-Based Dictionary of English Synonyms* (2005).

Despite these attempts, Japanese bilingual lexicography is falling behind in the important field of utilizing corpus resources for dictionary-making. The use of corpora for monolingual learner's dictionaries has become even more sophisticated these days, as shown in the publication of the *Macmillan English Dictionary* (*MED*; 2002) or the *Longman English-Japanese Dictionary* (*LEJD*; 2006). In *MED*, they used a lexical database called the Sketch Engine developed by Adam Kilgarriff and his team, which enables users to view a summary ("word sketch") of a word's behaviour with respect to grammatical relations (the unary and binary relations for the search word, together with relevant statistical measures of confidence). This is a good example of the new "second-generation" use of corpora for lexicographical purposes: lexicographers do not have to search and browse the raw corpus data directly any more, but can view processed results summarizing the word behaviour. This makes it easier for lexicographers to judge what should be in the entry. In *LEJD*, they used both English and Japanese corpora in order to improve the quality of translation equivalents and

bilingual examples (see Ikegami, this volume). This was the first attempt in Japan to exploit bilingual corpora in the production of bilingual dictionaries.

It is still an open question whether Japanese bilingual dictionary publishers can catch up with UK and European publishers in terms of effective corpus use. The prospects, however, do not look very bright. Japanese publishers still have a long way to go toward the second-generation use of corpus data as mentioned above. We might have to depend on the resources built for monolingual learner's dictionaries for at least another decade before bilingual lexicography in Japan can stand on its own feet in the field of corpus exploitation for dictionary-making.

3. Second challenge: misconceptions about *user-friendliness*

It is fair to say that the quality of English bilingual lexicography in Japan is extremely high compared with average lexicographical standards in other countries. Whilst most English bilingual dictionaries in other countries are based on the framework of monolingual dictionaries produced by UK or American publishers, Japanese publishers have been producing original work. They take great pride in their professional craftsmanship.

At the same time, however, they are faced with the dilemma of choosing between clarity and elaboration. English–Japanese dictionaries are famous for their meticulous usage notes, which give the impression that they are almost like specialized reference works on grammar and usage. The inclusion of many special columns of notes makes English–Japanese dictionaries very colourful and detailed. In order to compete with other domestic products, publishers have to find ways to appear cutting-edge while including all the information that is already covered in existing dictionaries. Paper dictionaries end up being unnecessarily big and fat, and give the impression of being somewhat extravagant.

Ironically, this trend has been in response to language teachers' misconceptions about user-friendliness. In Japan, students do not usually choose dictionaries on their own, but rely on their teacher's recommendations. Teachers are supposed to take into account the proficiency levels of students in choosing dictionaries. There is, however, a strong tendency for teachers to choose dictionaries which satisfy their needs, not that of their learners, and they tend to favour dictio-

náries which contain more entries and more information (especially grammar and usage information), and, as a result, their favorite dictionaries are very often too difficult for learners to use. Also, users themselves make the assumption that all-inclusive dictionaries are better, and prefer to buy electronic dictionaries which contain more than twenty or thirty different titles all in one. Therefore, it is often the case that high school students use all-inclusive general-purpose dictionaries which, in most cases, are too complicated and not appropriate for that level of learners.

This current state of affairs tells us that people have only a vague idea about what a good dictionary is, and just follow their teachers' advice or brand image. There is a serious mismatch between the types of dictionaries learners have and their actual proficiency levels. We should therefore educate both language teachers and learners on the importance of dictionary use in language learning, and inform them about how to select the most appropriate dictionary for the right context and type of learner. The whole idea of "user-friendliness" needs to be reconsidered.

In this sense, bilingual lexicography in Japan is facing another challenge: the content in learner's dictionaries is fast reaching saturation point. We need to sort out necessary from unnecessary information, and tune the amount of information to the level of prospective users. Japanese publishers have continued to add information to dictionaries, but need to stop and think whether all the information is really needed. To do this, we will have to discuss the purposes of dictionary use in the process of language learning: whether its intended use is for encoding or decoding, for fluency or accuracy. As Tono (2001) suggested, more empirical research will be needed in order to fully understand what components of dictionary information are significant for developing what aspects of language proficiency.

4. Third challenge: changing modes of dictionaries

Dictionary user interfaces have been changing radically. There are at least four different modes of dictionary interface currently available:

    (a) paper dictionaries
    (b) CD-ROM dictionaries
    (c) Web-based dictionaries
    (d) pocket dictionaries (hand-held)   (Tono 2004: 16ff)

The last three are electronic, but have different interfaces. CD-ROM dictionaries are usually installed onto a single computer and work on a stand-alone basis. Web-based dictionaries, on the other hand, are accessed via the Internet. The data is usually stored on a server and client machines need to be connected to the server in order to retrieve the information. Pocket hand-held dictionaries are a recent development in Japan. Many major manufacturers (CASIO, SONY, SHARP, SEIKO, CANON, among others) are involved in this business and have produced literally hundreds of different models. They have a very large product range: student—adult, general—specific (e. g. medical content), monolingual—multilingual, non-professional—professional (e. g. versions for researchers and teachers of English). CASIO's EX-Word XD-ST6300, for example, contains a hundred different titles, ranging from the unabridged Japanese monolingual dictionary, *Kojien*, English monolingual dictionaries such as *OALD* (7th edition), medical dictionaries, and various specialized dictionaries and phrase books for English conversation.

As electronic dictionaries grow in popularity, the way users look up words has also been changing. CD-ROM dictionaries, for example, have become multifunctional, integrating a thesaurus, grammar/usage reference notes, and vocabulary exercises all in a single interface. They also provide a pop-up mode in addition to the main interface, to allow users to invoke them from within any program (most usefully from within a word processor). Web-based dictionaries are also improving, such that dictionary look-up is now integrated with web browsing and e-mail. All you have to do now is to select the unknown word with a mouse while surfing and the Google dictionary, for instance, will display the definition or meaning in a pop-up window in your language of choice.

With rapidly changing user interfaces, users have great difficulty adjusting themselves to the different modes and functions. In a sense, dictionary search has become very easy and quick, thanks to the intelligent search functions. On the other hand, some people are dubious about whether electronic dictionaries are always a positive thing. Hulstijn and Laufer (2001) have written about the Involvement Load Hypothesis, which claims that vocabulary learning takes place most effectively when the three elements, *need, search*, and *evaluation*, come together: only then will learners process the lexical knowledge at a deeper level. If this hypothesis is correct, a cautionary note must be

sounded about the use of electronic dictionaries because the ease of using them can possibly make the task "too easy" for learners and they may not undergo a deeper processing of the lexical information. The "micro" nature of electronic dictionaries, with their limited windows sizes also makes it difficult for learners to access the right kinds of information. Because electronic dictionaries tend to differ from paper dictionaries, proper training in the necessary retrieving skills is needed.

This brings us to the third challenge for lexicographers in Japan: how to train dictionary users to adapt to the new modes of electronic dictionaries. This may not be a problem specific to bilingual lexicography in Japan, but we should bear in mind that things are more complicated in our country because of the huge success of pocket electronic dictionaries. More and more high school students buy pocket electronic dictionaries without even knowing how to use paper dictionaries. The distinction between microstructure and macro-structure and the meaning of "browsing the entries" are changing. This can be viewed as a very exciting paradigm shift, but on the other hand it can be seen as a complicating change.

Research on dictionary use needs to be re-framed as different types of dictionary interfaces came about. Much more empirical research needs to be done on the outcomes of using e-dictionaries in comparison with paper dictionaries. It is a welcome change to see more and more young researchers becoming interested in research on dictionary use as a result of the increasing popularity of electronic dictionaries. It seems to be the case, however, that many are not fully aware of the wealth past research in the field (see Tono 2001 for a review), some of which can still provide valuable insights even today.

5. Conclusion

This short paper has discussed the paradigm shift in lexicography in terms of three dimensions: corpus lexicography, misconceptions about user-friendliness, and changing modes of dictionaries. I have argued that these are new challenges that bilingual lexicography in Japan is facing or will be facing in the near future.

We are experiencing an exciting period of technological innovations and language learners should be able to take advantage of this change. Yet a lot of research and empirical tests need to be conducted before we can generalize about the applications of the new modes of dictionaries to language learning and teaching. As shown in this paper,

we should investigate the effects of the use of dictionaries with different interfaces on the performance of various language tasks in order to better understand their effectiveness. At this stage, it is crucial to emphasize the theoretical, empirical, and practical groundwork that must be laid first. I hope that the three challenges mentioned above will turn out to be golden opportunities for English bilingual lexicography to flourish again.

References

Hulstijn, J. and B. Laufer (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning* 51: 539-558.

Tono, Y. (2001). *Research on Dictionary Use in the Context of Foreign Language Learning*. Tübingen: Max Niemeyer Verlag.

Tono, Y. (2004). Research on the use of electronic dictionaries for language learning: methodological considerations. In Mari Carmen Campoy and Pilar Safont (eds.). *Computer-Mediated Lexicography in the Foreign Language Learning Context* (pp. 13-28). Castelló de la Plana: Universitat Jaume I.