

Bilingual KWIC

– GUI Support Tool for Bilingual Dictionary Compilation –

TOYAMA, Katsuhiko and OGAWA, Yasuhiro

Graduate School of Information Science, Nagoya University

{toyama, yasuhiko}@is.nagoya-u.ac.jp

At present, machine translation systems are employed in a wide variety of fields. In the case of translating technical documents, translation systems need a bilingual dictionary specific to the documents' domain. Since compilation of such a dictionary is expensive, much work is done to acquire a bilingual lexicon automatically from a bilingual corpus. However, the result of such automatic acquisition sometimes includes an incorrect lexicon. In addition, if one word has two or more translation equivalents, users want to know which is the suitable one.

To overcome these problems, we propose a GUI support tool, Bilingual KWIC, to compile a bilingual dictionary. Fig.1 shows its overview. When users input a keyword into the Keyword field and click the "Search" button, Bilingual KWIC displays sentences which contain the keyword from the source language in the left-hand window and translations of the sentences in the right-hand window on each line. At the same, the input keyword is centered in the source texts and its equivalent is also centered in the translations. That is, both texts are displayed in KWIC (KeyWord In Context) format. Bilingual KWIC automatically calculates the equivalent of the input keyword using a technique of word alignment without any dictionaries. The input keywords and calculated equivalents are displayed in blue. For example, "根抵当権" and "floating" are blue in Fig.1. The calculated equivalent is also displayed in the Equivalent field. If users think the equivalent is wrong, they can input a new equivalent into the field and Bilingual KWIC redisplay the sentences that contain it.

Once users have decided which equivalent is suitable for the keyword, they right-click the sentence that contains the equivalent. Then Bilingual KWIC pops up the Dictionary editing window shown in Fig.3 and they can register the keyword and its equivalent into the bilingual dictionary with some information that is useful for proper translation. If users input a registered keyword, its equivalents are displayed in green as shown in Fig.2.

Bilingual KWIC has following features. First, it is **easy to correct errors** of word alignment performed by Bilingual KWIC. In the example of Fig.1, an input keyword in this case is "根抵当権," and Bilingual KWIC says that its equivalent is "floating," though the correct one is "floating mortgage." However, this example enables us to understand that "floating mortgage" is intuitively correct by our observing which words occur before and after "floating" in several sentences.

Secondly, Bilingual KWIC also helps users to **obtain derivational words and its equivalents**. In the example shown in Fig.1, "根抵当権者," a derivational word of "根抵当権," is displayed at the bottom of the left-hand window. In this case, "floating mortgagee" appears in the translated sentences in the right-hand window, clearly indicating that "根抵当権者" is desirable for registration to the dictionary, and that its equivalent is "floating mortgagee."

Thirdly, Bilingual KWIC helps users to **select the suitable equivalent** because it displays the contexts of each word and has sort functions to compare the equivalents. In the example presented in Fig.2, Bilingual KWIC calculated three equivalents "離縁", "解散" and "解消" for the input keyword "dissolution." The figure indicates that the suitable equivalent is "離縁" in the case of "dissolution of adoptive relation," that "解散" is suitable in the case of "dissolution of juridical person," and that "解消" is suitable in the case of "dissolution of marriage."

Finally, Bilingual KWIC easily **deals with other languages**, because it does not perform morphological analysis and it uses only character-level information.

We have opened Bilingual KWIC at <http://www.kl.i.is.nagoya-u.ac.jp/koyori/>, and it has been used in the project of translating Japanese laws at the initiative of the Japanese government in order to compile a standard Japanese-English dictionary in the domain of legal documents.

Bilingual-KWIC

GUI Support Tool for Bilingual Dictionary Compilation

TOYAMA, Katsuhiko and OGAWA, Yasuhiro
Graduate School of Information Science, Nagoya University

Background

High Performance MT

- Applied to various fields
→ limit of performance

Bilingual Dictionary

- Compiled from parallel corpus
→ error of word alignment

Operating Instructions

- Input a keyword or phrase
- If necessary,
 - specify an equivalent
 - sort sentences by the words before or after the keyword
- Right-click and register the keyword and its equivalent (Fig.3)

Features

- Error correction (Fig.1)
- Acquisition of derivations (Fig.1)
- Comparing multiple equivalent (Fig.2)
- Using with a bilingual dictionary (Fig.2)
- Editing a dictionary's entry (Fig.3)
- Application to other languages

Object

GUI Support Tool

- Word alignment +
- KWIC(KeyWord In Context)

Technical Details

- No need for dictionary
- No morphological analysis
- Using Dice coefficient

$$\text{Dice}(x, y) = \frac{2 \cdot \text{freq}(x, y)}{\text{freq}(x) + \text{freq}(y)}$$

- Implemented by Ruby/Tk works on MS-Windows, Linux

Bilingual-KWIC

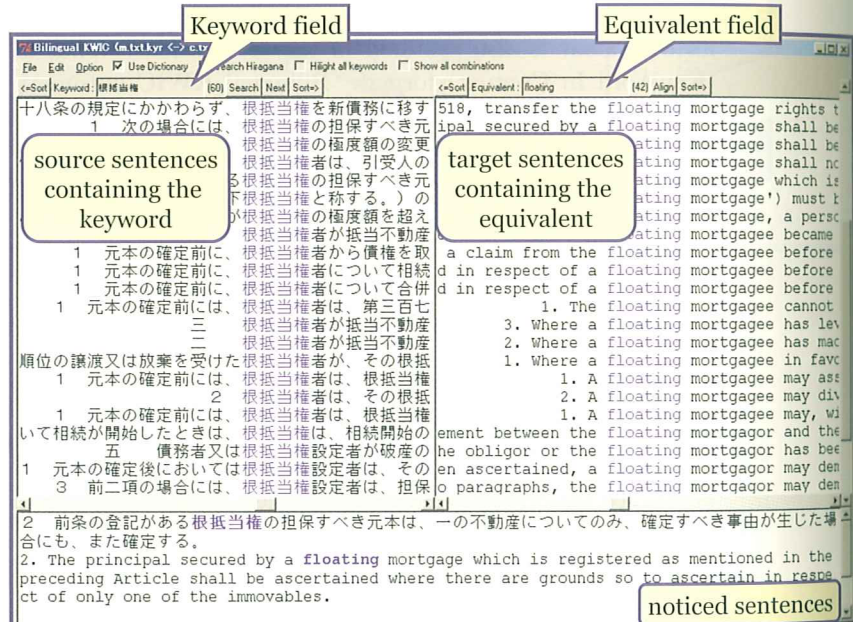


Fig.1. Overview of Bilingual KWIC

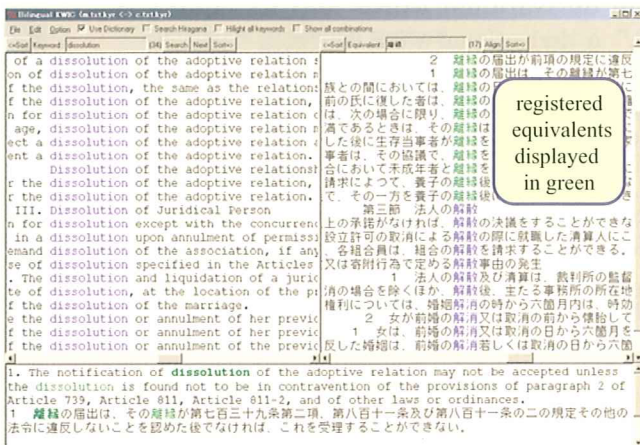


Fig.2. Multiple equivalents for "dissolution"

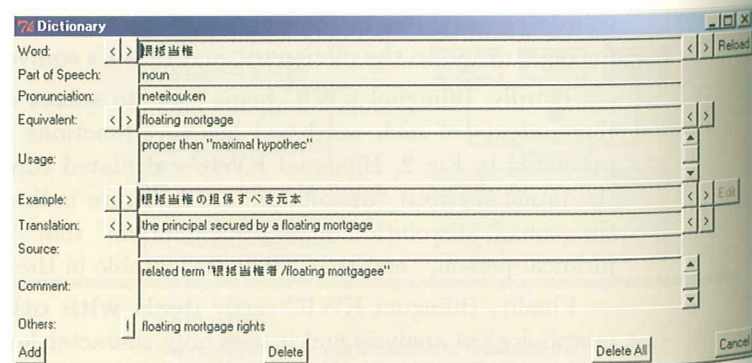


Fig.3. Dictionary editing window

Future Tasks

- Speeding up
- Using with morphological analyzer

Open Source

<http://www.kl.i.is.nagoya-u.ac.jp/koyori/>