Design and Construction of a machine-tractable Japanese-Malay Dictionary

Francis Bond,* **Ruhaida Binti Sulong**,† **Takefumi Yamazaki**‡ and **Kentaro Ogura***
* NTT Communication Science Laboratories <{bond, ogura}@cslab.kecl.ntt.co.jp>
† NTT MSC <ruhaida@nttmsc.com.my>
‡ NTT Cyber Space Laboratories <yamazaki@nttnly.isl.ntt.co.jp>

Abstract

We present a method for combining two bilingual dictionaries to make a third, using one language as a pivot. In this case we combine a Japanese-English dictionary with a Malay-English dictionary, to produce a Japanese-Malay dictionary suitable for use in a machine translation system. Our method differs from previous methods in its use of semantic classes to rank translation equivalents: word pairs with compatible semantic classes are preferred to those with dissimilar classes. We also experiment with the use of two pivot languages. We have made a prototype dictionary of over 75,000 pairs.

Keywords

bilingual lexicon, lexicon construction, machine translation, Japanese, Malay

1 Introduction

We present a method for combining two bilingual dictionaries to make a third, using one language as a pivot. The aim of our research is to create a dictionary to be used in the machine translation system **ALT-J/M**: the Automatic Language Translator — Japanese-to-Malay (Ogura et al., 1999). We take the Japanese-to-English dictionary used in the machine translation system **ALT-J/E** (Ikehara et al., 1991) and cross it with a Malay-English dictionary to produce a Japanese-Malay dictionary.

The reasons we wish to do this are two-fold. Firstly, there are no large-scale Japanese-Malay dictionaries available, either for human use or machine tractable. The largest lexicons we could find had between 6,000 (Onozawa, 1990) and 7,000 head words (Nagata, 1994), and fewer than 15,000 translation pairs. This is too few for an unrestricted machine translation system. The second is that we need to build a dictionary that has not only Japanese words and their Malay equivalents, but also semantic and syntactic information. By using our existing Japanese-English Dictionary, we can exploit the semantic information it contains, transferring as much as possible to the new dictionary.

The **ALT** systems are semantic transfer systems, and rely on having nouns marked with appropriate semantic classes (from our ontology of roughly 3,000 classes). These semantic classes are then used to describe the selectional restrictions of predicate-frames.

Clearly different senses of the same noun can be differentiated because they will appear in different semantic classes, for example, *seal* ⇔ あざらし *azarashi* ⟨**animal**⟩ vs *seal* ⇔ 印 *in* ⟨**tool**⟩. We will refer to such clearly distinct senses as homonyms. In a machine translation system, homonyms can be translated correctly if they have the correct semantic classes marked.

Finer grained variations, such as the difference between *doves* and *pigeons* (both 鳩 *hato* in Japanese) are harder to distinguish using semantic classes. Instead, collocation and usage information is necessary. Various methods exist to distinguish between such variants in machine translation, including the use of domain information, noun-modifier collocation, n-grams and other statistical information. The fall-back method for distinguishing between similar variants is frequency: which of a set of translation equivalents occurs most often. In our system, this is implemented as a preference value: if the semantic classes are the same, in the absence of other restrictions, choose the translation candidate with the highest preference.

For example *dove* and *pigeon* are both potential translations of Japanese 鳩 *hato*, with the same basic meaning.[1] In the absence of other information, **ALT-J/E** will always translate 鳩 *hato* as *pigeon* because it is the preferred translation.

When translating, it is essential to distinguish between homonyms, in order to faithfully convey the sense of a text. It is less important to distinguish between variations, and indeed often impossible: different languages make different distinctions. Because of this, when building our dictionary, it is essential to distinguish homonyms correctly, and our method aims to do this.

1.1 Previous work

Tanaka and Umemura (1994) and Tanaka et al. (1998) used English as an intermediate language to link Japanese and French. Their method relies on inverse consultation. To find suitable equivalents for a given Japanese word, they

[1] *Pigeon*: "wild and domesticated birds having a heavy body and short legs" ; *dove* "any of numerous small pigeons" (WordNet, 1997).

first look up its English translations, and then the French translations of these English translations, giving a set of French equivalence candidates (ECs) of the original Japanese. For each French word, they then look up all of its English translations, and see how many match the English translations of the original Japanese word. The more matches there are, the better the candidate is. They call this "one time inverse consultation". This can be extended further, by looking up all the Japanese translations of all the English translations of a given French word and seeing how many times the Japanese word appears; this is "two times inverse consultation".

An example of one time inverse consultation, between Japanese and Malay, is given in Figure 1. There are three translations of the Japanese word 印 *in* "seal", and four translations of its equivalence candidate *tera* "seal". There are two shared translations (underlined in the figure). To normalize the score, it is multiplied by two (thus if all words match the score will be one). This gives a score of 0.57.

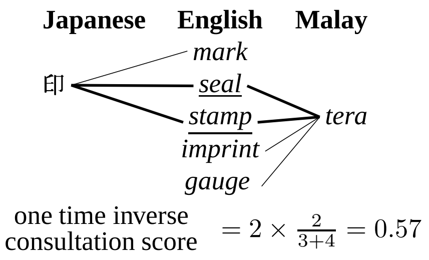
Japanese
English
Malay
印
mark
seal
stamp
imprint
gauge
tera
one time inverse consultation score $= 2 \times \frac{2}{3+4} = 0.57$

Figure 1: One Time Inverse Consultation

Tanaka et al. (1998) were able to find translation equivalents not found in equivalent Japanese-French dictionaries by mMatching published Japanese-English and English-French dictionaries against each other. Evaluating the results for one time inverse consultation gave recall of 44% and precision of 76% for nouns, down to 15% and 65% for adjectives.

Shirai and Yamamoto (2001) also use one time inverse consultation to create a Japanese-Korean Dictionary, using English as the pivot language. By limiting the types of matching allowed, they were able to increase precision to as high as 82.6%, but at the cost of greatly reducing the number of pairs found.

Sanfilippo and Steinberger (1997) suggest the use of a thesaurus in the pivot language. This is used to map the source and target languages to senses in the thesaurus, and the sense-differentiated links are used to join the language words (in their case English between German and Italian). In theory, this should eliminate matching through inappropriate homonyms, unfortunately, they did not give an evaluation of bilingual linking using this method.

Lafourcade (1997) also used English as an intermediate language, in his case to construct a multi-lingual French-English/Malay/Thai on-line lexicon, aimed at human users.[2] Malay and Thai entries were added to an existing French-English dictionary by linking entries in Malay-English and Thai-English dictionaries. There was no automatic filtering of the crossed results; instead emphasis was placed on producing a productive environment for human dictionary editors. In particular, human editors were found to prefer word-processor based environments to data-base interfaces.

One shared characteristic of these approaches is the use of English as the pivot language. This is because, in general, there are more bilingual resources available with English as one of the languages. None of the previous work uses semantic information or matches through two or more languages.

2 Creating a Japanese-Malay Dictionary

In this section we first describe the Japanese-English and Malay-English dictionaries we use, and then how we combine them.

2.1 The Japanese-English dictionary: Goi-Taikei

For the Japanese-English dictionary, we are using the dictionaries developed for the machine translation system **ALT-J/E** (Ikehara et al., 1991), a subset of which has been published as Goi-Taikei (**GT**) — a Japanese lexicon (Ikehara et al., 1997).

GT consists of three main components: (i) an ontology, (ii) a semantic word dictionary, and (iii) a semantic clause structure dictionary which includes subcategorization frames for predicates.

2.1.1 Ontology

GT's ontology classifies concepts to use in expressing relationships between words. The meanings of common and proper nouns are given in terms of a semantic hierarchy of 2,710 nodes. Each node represents a semantic class. Edges in the hierarchy represent **is-a** or **has-a** relationships, so that the child of a semantic class related by an **is-a** relation is subsumed by it. For example, **`nation`** **is-a** **`organization`**.

2.1.2 Semantic Transfer Dictionary

The semantic transfer dictionary includes roughly 380,000 Japanese-English word-pairs.

Each record specifies an index form (Japanese), translation (English), preference ranking, English syntactic information and a set of semantic classes. Optionally there may be more detailed selectional restrictions, domain and genre information and so on.

In the noun dictionary, there are 63,926 Japanese index words. 90% have only one translation, 8.5% have two, 2%

[2] An on-line version of the FEM dictionary can be found at `http://www-clips.imag.fr/geta/services/fem/`.

have three. The maximum number of translations is 12, the average is 1.12, for a total of 71,818 Japanese-English pairs. There is a tendency for many Japanese words to be translated into the same English translation, there are only 49,205 different English entries (many of them are multi-word expressions).

2.1.3 Semantic Structure Dictionary

The basic structure of a clause comes from the relationship between the main verb and nouns. **GT**'s structure transfer dictionary, designed for machine translation applications, provides this basic clause structure. **GT** has over 15,000 patterns.

2.2 The Malay-English dictionary

We use the Malay-English Dictionary **KAMI**: KAmus Melayu-Inggeris. This is a dictionary being compiled by NTT-MSC, based on a dictionary produced by a translation company (Quah et al., 2001). The dictionary currently has 67,658 Malay words with English translations. 79% have only one translation, 14% have two, 4.1% have three; the average number of translations is 1.35, giving 91,426 Malay-English pairs.

Each entry in the dictionary consists of the following fields: (1) Malay index word; (2) Malay root word; (3) Malay part of speech (POS); (4) detailed syntactic features; (5) semantic classes; (6) English translation; (7) English gloss and comments; (8) Chinese translation. All entries have values for fields 1,2 and 3; most have syntactic features. Only 30.4% have semantic classes from the **GT** ontology, 25% have Chinese translations. We also use a variety of meta-codes, to show other relevant information such as honorific use, origin, and register. English and Chinese translations and comments are provided for use in a machine translation system, as well as an aid for non-Malay speakers. One of the most useful syntactic features is the numeral classifier (or classifiers) most commonly used to count the noun. Uncountable nouns (most common nouns in Malay) must use a classifier when they are being counted. For example, *two tigers* becomes *dua ekor harimau* "two-CL tiger". Listing of classifiers is common in dictionaries for foreign learners, and can also be found in the CICC Malay dictionary (CICC, 1994b).

When entering the index words, the wide variety of spelling variation in Malay was particularly problematic. To deal with this we have allowed a single record to have multiple index forms, with the preferred form, as judged by native speakers of Riau (standard) Malay, listed first. There are currently 1,039 such entries in our dictionary, for example *hasab;hisab* "calculation".

Semantic classes were entered in four ways: (1) the original dictionary we purchased had some syntactic-semantic codes (10,000 entries). These were mapped to the **GT** semantic classes by hand. (2) The CICC Indonesian dictionary has semantic classifications (CICC, 1994a). As Malay and Indonesian share much of their vocabulary, we looked up Malay-English pairs in the CICC Indonesian-English dictionary, and took used the semantic classes in the matching Indonesian pairs (14,784 entries). These were then mapped to the **GT** semantic classes using a simple hand mapping. (3) Because some classifiers select for the meanings of their targets (Bond and Paik, 2000), we could use the classifiers to predict the semantic class of their targets (18,303 entries). For example, anything counted by *orang* is **`human`**, anything counted by *ekor* is **`animal`**,[3] anything counted by *pokok* is **`plant`** and so on. Shape classifiers (such as *batang* "long thing") could not be used for this, as they select for physical shape, not semantic class. (4) We added semantic classes by checking against known word lists such as the ISO 639 language names and the ISO 4217 currency names (a few hundred entries). Finally, (5) we added some semantic classes to some words by hand, although not in any systematic way. Because of the overlap between the four classes described above, we only have semantic classes for around 20,000 entries (30%).

2.3 Crossing the dictionaries

Building the Japanese-Malay dictionary involves two steps: creating Japanese-Malay equivalence candidates, and then filtering and ranking the candidates. The overall flow is shown in Figure 2.

- For each pair in the Japanese-English dictionary
  - Look up the Malay equivalent of the English
    if an entry with the same POS exists
    - Create a Japanese-Malay pair
    - Store the intermediate English
    - Calculate scores
      - one time inverse consultation
      - semantic matching
      - second-language matching
  - else mark the Japanese-English pair
- For each Japanese index in the Japanese-English dictionary
  - Output any Japanese-Malay pairs ranked by total score
  - Output marked Japanese-English pairs ranked by preference

Figure 2: Crossing of the dictionaries

Our crossing process is opportunistic: taking immediate advantage of any circumstance of possible benefit. Ideally we will only apply it once, and then check all entries by hand. Because of this, we do minimal filtering, prefer-

[3] Strictly speaking, *ekor* is also used to count ghosts and spirits, but there are so few such entries that we could fix them by hand.

ring instead to maximize the number of equivalence candidates. English entries were normalized, in particular articles (*a/an, the*) and infinitival *to* were stripped from the beginning of noun and verb entries respectively.

However, we wish to use the dictionary immediately, as thorough checking may take several person-years. Therefore, it is important to get as good a translation as possible in the top ranked position.

The crossing process produces a dictionary with as many Japanese-Malay entries as possible, followed by English entries as a last resort. We deliberately kept the English entries, both as a guide to the lexicographers to identify possibly missing senses; and as default translations: most Malaysians speak more English than Japanese, so it is better to translate to English than to leave unknown words in Japanese.

Pairs were only crossed if they had the same part of speech (using a small set of coarse categories: **common-noun, proper-noun, verb, adjective, adverb, pronoun, auxiliary, preposition**). We actually used the English part of speech in the J-E dictionary, and the Malay part of speech in the M-E dictionary. Ideally we would like to use English part of speech for both lexicons, if available. Matching only compatible parts of speech cut down greatly on the number of false matches. Crossing to different parts of speech only increased the number of new Japanese matches by 2%, at the cost of increasing the number of equivalence candidates by 15%, most of which were spurious.

The scores were calculated as follows: The one time inverse consultation score for Japanese word J and Malay word M is given in Equation (1), where $E(W)$ is the set of English translations of W:

$$\begin{array}{c}\text{one time inverse}\\ \text{consultation score}\end{array} = \frac{2 \times (|E(J) \cap E(M)|)}{|E(J)| + |E(M)|} \qquad (1)$$

The semantic matching score was the number of times a semantic class of J was compatible with a semantic class of M, where two classes are compatible if one semantic class subsumes the other, or visa versa. For example, **animal** is compatible with **living-thing**. Only nouns have semantic classes in our lexicons, so this score is only applicable to nouns.

The second-language matching score used Chinese as a second intermediate language. Our Malay-English dictionary also has Chinese entries for 21,190 of its entries (25%). If a matched Malay entry had a Chinese translation, then we checked to see whether the Japanese and Chinese pair could be found in a Japanese-Chinese dictionary of some 83,000 entries (Shogakukan and Peking Shomoinshokan, 1987). We assume that anything that matches through two different languages (Japanese to Malay through English and Chinese) should be a good match. In particular, we expect different homonyms in different languages, so using two pivot languages should be effective in distinguishing between them. We give an example of a match through two languages in Figure 3. Here *tera* "seal" matches through both English and Chinese, so is a good match. The entry *mohor* "seal" matches through two English words, so is a reasonable match, and *anjing laut* "seal" matches through only one word, so is a bad match.

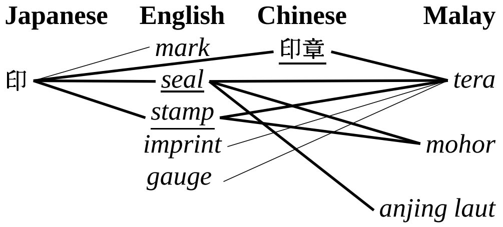

Figure 3: Matching through two languages

The total score is a combination of the semantic matching score, the original preference of the Japanese-English pair, and the one time inverse consultation score, combined so that the Chinese matches come first, followed by the semantic matches, followed by high ranked pairs; within the same ranking, pairs are ordered by one time inverse consultation score. There is no mechanism in our algorithm for deleting candidates, that is left to the lexicographers.

Consider the following simplified example.

- Japanese-English pair (Input)

Japanese	あざらし *azarashi*
English	*seal*
Sem Classes	⟨**animal**⟩

- Malay-English pairs (Input)

Malay	*anjing laut*
English	*seal*
Classifier	*ekor*
Sem Classes	⟨**animal**⟩

Malay	*tera*
English	*seal*
Chinese	印章
Sem Classes	⟨**stationery**⟩

Malay	*mohor*
English	*seal*
Classifier	*buah*
Sem Classes	⟨**tool**⟩

- Japanese-Malay pairs (Output)

Japanese	あざらし *azarashi*
Malay	*anjing laut*
Rank	1
English	*seal*
Sem Classes	⟨**animal**⟩

$$\begin{bmatrix} \text{Japanese} & \text{あざらし } \textit{azarashi} \\ \text{Malay} & \textit{tera} \\ \text{Rank} & 2 \\ \text{English} & \textit{seal} \\ \text{Sem Classes} & \langle - - - \rangle \end{bmatrix}$$

$$\begin{bmatrix} \text{Japanese} & \text{あざらし } \textit{azarashi} \\ \text{Malay} & \textit{mohor} \\ \text{Rank} & 3 \\ \text{English} & \textit{seal} \\ \text{Sem Classes} & \langle - - - \rangle \end{bmatrix}$$

In this small example, there are three potential translations for あざらし *azarashi* “seal”. English *seal* is homonymous, the correct sense here is the “marine animal” sense, which corresponds to *anjing laut*. The translations *tera* and *mohor* are variations of the sense that means “a stamp used to authenticate documents”.

The semantic class of *anjing laut* matches with あざらし *azarashi* “seal”, so it is listed first: it is the only correct translation. The other two translations are listed according to their one time inverse consultation scores. There are no matches on the Chinese: あざらし *azarashi* “seal” links to 海豹 *hǎibào* “seal” in our Japanese-Chinese dictionary, but that was not entered in our Malay dictionary.

3 Results and Discussion

In this section we report on crossing the Japanese-English common-noun dictionary with the Malay-English dictionary.

22,658 out of 63,926 Japanese words were linked to 16,974 Malay words. There were 32.7% with one translation, 19.5% with two and 11% with three. The average number of translations was 3.4 for a total of 75,872 pairs. Clearly, we have introduced many spurious translations: the average number of translations is almost triple that of the original dictionaries.

However, we do not consider this a serious problem for the following reasons. The main reason is that, most of the time, only the first translation is output by the machine translation system. Therefore, as long as our ranking is correct, the spurious translations will be invisible to the user. Another important reason is that it is far quicker to delete a spurious entry than add a new one. Our lexicographers prefer to be presented with a large list to be whittled down, rather than having to add translations from scratch.

A preliminary evaluation of 65 randomly selected Japanese index words with 232 translations gave the following result: 65% of translations were acceptable. The results are summarized in Table 1. Concentrating only on the highest ranked translation (the translation most likely to be used), 80% of the translations were acceptable (Table 2).

93 (40%) of the translations were judged to be good translations, usable in any context. 58 (25%) were judged to be usable in some contexts, and thus acceptable as dictionary entries, but not ideal as translation equivalents. 81 (35%) were judged to be inappropriate translations. Of these, just over a third (28) were due to errors in KAMI, our Malay-English dictionary. If the dictionary were perfect, the results would be around 77%.

Evaluation	Number	Percentage
Good translation	93	40.1
Acceptable translation	58	25.0
Bad (error in ME dic)	28	12.1
Bad (link mismatch)	53	22.8
Total:	232	100.0

Table 1: Results (all pairs)

To evaluate the results of our ranking, we also looked at the highest ranked pairs; shown in Table 2.

Evaluation	Number	Percentage
Good translation	30	46.2
Acceptable translation	22	33.8
Bad (error in ME dic)	6	9.2
Bad (link mismatch)	7	10.8
Total:	65	100.0

Table 2: Results (highest ranked pairs)

Our ranking has successfully increased the percentage of good pairs to 46%, and acceptable pairs to 34%. This means that 80% of the translations provided by our machine translation system will be good, even with no manual revision.

24 of the entries had a single equivalence candidate (that is there was a single Japanese-English pair matching a single English-Malay pair with the same part of speech). In this case, 11 (46%) were good, 12 (50%) were acceptable, and only one was bad (due to an error in the ME lexicon). In applications which want to avoid any erroneous translations, one strategy would be to only take such single matches.

5,238 pairs matched using both English and Chinese as the intermediate language. We checked a sample of 100 pairs and found 84 good translations, 13 acceptable translations and only three errors: 97% were good. This shows clearly that matching through two languages improves accuracy, as predicted. The number of pairs is reduced: only 5,238 out of 75,872. However, these still cover almost one in four of the 22,658 Japanese index words matched.

By using all the information we could, we have been able to automatically build a reasonably accurate large-scale dictionary Japanese-Malay dictionary with the information required by a semantic transfer-based machine translation system.[4] One of the arguments against transfer-based sys-

[4] Slightly different versions of the dictionaries discussed in this paper are on-line at `http://sangenjaya.arc.net.my/`.

tems has been that it is hard to add new language pairs. However, as we show here, new pairs can be effectively bootstrapped from existing resources.

3.1 Further Work

We are now concentrating on improving the environment for our lexicographers. As the dictionary has grown to several tens of thousands of entries, many of them will be unfamiliar, even to an educated native speaker. It is thus useful to make it easy to look up monolingual dictionaries with as few key strokes as possible, allow browsing of the semantic classes, and present examples of words in context.

We would also like to extend the number of matches by improving our normalization of entries. First, we would like to add a British/American spelling converter. Our Malay-English dictionary uses mainly British spelling, but our Japanese-English dictionary uses mainly American spelling, so currently words such as *armor/armour* don't match. Second, we would like to do some lemmatization, at the least converting plural noun forms to singular before matching.

4 Conclusion

It is possible to create a useful bilingual dictionary by matching two or existing bilingual dictionaries and linking them through a pivot language. The accuracy of the resulting dictionary can be improved by (1) using semantic classes, and (2) matching through a different pivot language. Using these techniques we have made a prototype Japanese-Malay dictionary of over 75,000 words.

Acknowledgments

The authors would like to thank Akinori Abe, Chiew Kin Quah, Siti Aishah Salleh, Kyonghee Paik, Satoshi Shirai and the members of the NTT Machne Translation Research Group for their valuable comments and discussion.

References

Francis Bond and Kyonghee Paik. Re-using an ontology to generate numeral classifiers. In *18th International Conference on Computational Linguistics: COLING-2000*, pages 90–96, Saarbrücken, 2000.

CICC. Research on Indonesian dictionary. Technical Report 6—CICC—MT53, Center of the International Cooperation for Computerization, Tokyo, 1994a.

CICC. Research on Malaysian dictionary. Technical Report 6—CICC—MT54, Center of the International Cooperation for Computerization, Tokyo, 1994b.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. *Goi-Taikei — A Japanese Lexicon*. Iwanami Shoten, Tokyo, 1997. 5 volumes/CDROM.

Satoru Ikehara, Satoshi Shirai, Akio Yokoo, and Hiromi Nakaiwa. Toward an MT system without pre-editing – effects of new methods in **ALT-J/E**–. In *Third Machine Translation Summit: MT Summit III*, pages 101–106, Washington DC, 1991. (`http://xxx.lanl.gov/abs/cmp-lg/9510008`).

Mathieu Lafourcade. Multilingual dictionary construction and services. In *3rd Pacific Association for Computational Linguistics Conference: PACLING-97*, pages 173–181. Meisei University, Tokyo, 1997.

Hideo Nagata, editor. *Japanese-Malay-English Dictionary*. TK Kenkyusha, Tokyo, 1994.

Kentaro Ogura, Francis Bond, and Yoshifumi Ooyama. **ALT-J/M**: A prototype Japanese-to-Malay translation system. In *Machine Translation Summit VII*, pages 444–448, Singapore, 1999.

Jun Onozawa, editor. *Marēshia-go jōyō 6,000-go [6,000 Malaysian Common Words]*. Taigaku Shorin, Tokyo, 1990.

Chiew Kin Quah, Francis Bond, and Takefumi Yamazaki. Design and construction of a machine-tractable Malay-English lexicon. In *Asialex-2001*, Seoul, 2001. (to appear).

Antonio Sanfilippo and Ralf Steinberger. Automatic selection and ranking of translation candidates. In *Seventh International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-97*, pages 200–207, Santa-Fe, 1997.

Satoshi Shirai and Kazuhide Yamamoto. Linking English words in two bilingual dictionaries to generate another language pair dictionary. In *19th International Conference on Computer Processing of Oriental Languages: ICCPOL-2001*, pages 174–179, Seoul, 2001.

Shogakukan and Peking Shomoinshokan, editors. *Ri-Zhong Cidian [Japanese-Chinese Dictionary]*. Shogakukan, 1987.

Kumiko Tanaka and Kyoji Umemura. Construction of a bilingual dictionary intermediated by a third language. In *15th International Conference on Computational Linguistics: COLING-94*, pages 297–303, Kyoto, 1994. (`http://xxx.lanl.gov/abs/cmp-lg/9410020`).

Kumiko Tanaka, Kyoji Umemura, and Hideya Iwasaki. Construction of a bilingual dictionary intermediated by a third language. *Transactions of the Information Processing Society of Japan*, 39(6):1915–1924, 1998. (in Japanese).

WordNet. *WordNet - a Lexical Database for English*. Cognitive Science Laboratory, Princeton University, 221 Nassau St., Princeton, NJ 08542, 1997. Version 1.6, `http://www.cogsci.princeton.edu/~wn/`.