Automatic Construction of a Transfer Dictionary Considering Directionality

Kyonghee Paik, Satoshi Shirai* and **Hiromi Nakaiwa**
{kyonghee.paik,hiromi.nakaiwa}@atr.jp * sat@fw.ipsj.or.jp

ATR Spoken Language Translation Laboratories
2-2-2, Keihanna Science City Kyoto, Japan 619-0288

*NTT Advanced Technology Corporation
12-1, Ekimaehoncho, Kawasaki-ku, Kawasaki-shi, Japan 210-0007

Abstract

In this paper, we show how to construct a transfer dictionary automatically. Dictionary construction, one of the most difficult tasks in developing a machine translation system, is expensive. To avoid this problem, we investigate how we build a dictionary using existing linguistic resources. Our algorithm can be applied to any language pairs, but for the present we focus on building a Korean-to-Japanese dictionary using English as a pivot. We attempt three ways of automatic construction to corroborate the effect of the directionality of dictionaries. First, we introduce "one-time look up"method using a Korean-to-English and a Japanese-to-English dictionary. Second, we show a method using "overlapping constraint" with a Korean-to-English dictionary and an English-to-Japanese dictionary. Third, we consider another alternative method rarely used for building a dictionary: an English-to-Korean dictionary and English-to-Japanese dictionary. We found that the first method is the most effective and the best result can be obtained from combining the three methods.

1 Introduction

There are many ways of dictionary building. For machine translation, a bilingual transfer dictionary is a most important resource. An interesting approach is the *Papillon Project* that focuses on building a multilingual lexical data base to construct large, detailed and principled dictionaries (Boitet et al., 2002). The main source of multilingual dictionaries is monolingual dictionaries. Each monolingual dictionary is connected to interlingual links. To make this possible, we need many contributors, experts and the donated data. One of the studies related to the *Papillon Project* tried to link the words using definitions between English and French, but the method can be extended to other language pairs (Lafourcade, 2002). Other research that focuses on the automatic building of bilingual dictionaries include Tanaka and Umemura (1994), Shirai and Yamamoto (2001), Shirai et al. (2001), Bond et al. (2001), and Paik et al. (2001).

Our main concern is automatically building a bilingual dictionary, especially with different combinations of dictionaries. None of the research on building dictionaries seriously considers the characteristics of dictionaries. A dictionary has a peculiar characteristic according to its directionality. For example, we use a Japanese-to-English (henceforth, J⇒E) dictionary mainly used by Japanese often when they write or speak in English. Naturally, in this situation, a Japanese person knows the meaning of the Japanese word that s/he wants to translate into English. Therefore, an explanation for the word is not necessary, except for the words whose concept is hard to translate with a single word. Part-of-speech (henceforth POS) information is also secondary for a Japanese person when looking up the meaning of the corresponding equivalent to the Japanese word.

On the other hand, an English-to-Japanese (henceforth E⇒J) dictionary is basically used from a Japanese point of view to discover the meaning of an English word, how it is used and so on. Therefore, explanatory descriptions, example sentences, and such grammatical information as POS are all important. As shown in (2), a long explanation is used to describe the meaning of *tango*, its POS and such grammatical information as singular or plural. Also, an E⇒J dictionary includes the word in plenty of

* Some of this research was done while at ATR.

examples, comparing to a J⇒E dictionary. The following examples clearly show the difference.

(1) J⇒E: タンゴ:《dance》the tango《〜**s**》

(2) E⇒J: tan · go /(**n**. **pl** 〜**s**)
タンゴ:**a**. もと中央アフリカの原住民の舞踏..etc. (trans. tango "a dance of Central African aboriginals,...etc.")**b**. その曲(trans. "its music")**Vi** タンゴを踊る("to dance the tango").

In this paper, we evaluate the effects that occur when we use different combinations of dictionaries and merge them in different ways.

2 Conventional Methods and Problems

The basic method of generating a bilingual dictionary through an intermediate language was proposed by Tanaka and Umemura (1994). They automatically constructed a Japanese-French dictionary with English as an intermediate language and manually checked the extracted results. In this sense, their method is not completely automatic. They looked up English translations for Japanese words, and then French translations of these English translations. Then, for each French word, they looked up all of its English translations. After that, they counted the number of shared English translations (**one-time inverse consultation**). This was extended to "two-time inverse consultation". They looked up all the Japanese translations of all the English translations of a given French word and counted how many times the Japanese word appears. They reported that "comparing the generated dictionary with published dictionaries showed that data obtained are useful for revising and supplementing the vocabulary of existing dictionaries." Their method shows the basic method of building a dictionary using English as an intermediate language. We applied and extended their method in automatic dictionary building especially considering the directionality of dictionaries.

Tanaka and Umemura (1994) used four dictionaries in two directions (J⇒E, E⇒J, F⇒E and E⇒F). They first harmonized the dictionaries by combining the J⇒E and E⇒J into a single J⇔E and the F⇒E and E⇒F into a harmonized F⇔E dictionary. We followed their basic method without harmonizing the dictionaries to emphasize the influence of directionality.

In general, foreign word entries in a bilingual dictionary attempt to cover the entire vocabulary of the foreign language. However, foreign words that do not correspond to one's mother tongue are not recorded in a bilingual dictionary from one's mother tongue to the foreign language (Hartmann, 1983). A long explanatory phrase is replaced with a word that often does not perfectly correspond to the original.

On the other hand, most of the index words from a foreign language to a mother tongue include many expository definitions or explanations that focus on usage. Such syntactic information as POS and number as well as example sentences are rich compared with a dictionary from mother tongue to a foreign language. These characteristics should be considered when building a dictionary automatically.

Bond et al. (2001) showed how semantic classes can be used along with an intermediate language to create a Japanese-to-Malay dictionary. They used semantic classes to rank translation equivalents so that word pairs with compatible semantic classes are chosen automatically as well as using English to link pairs. However, we cannot use this method for languages with poor language resources, in this case semantic ontology. Paik et al. (2001) improved the method to generate a Korean-to-Japanese (henceforth K⇒J) dictionary using multi-pivot criterion. They showed that it is useful to build dictionaries using appropriate multi-pivots. In this case, English is the intermediate language and shared Chinese characters between Korean and Japanese are used as pivots.

However, none of the above methods considered the directionality of the dictionaries in their experiments. We ran three experiments to emphasize the effects of directionality.[1] There are many approaches to building a dictionary. But our focus will be on the generality of building any pair of dictionaries automatically using English as a pivot. In addition, we want to confirm various directionalities between a mother tongue and a foreign language.

[1]The first two experiments were reported in Shirai and Yamamoto (2001) and Shirai et al. (2001). We present new evaluations in this paper.

3 Proposed Method

We introduce three ways of constructing a K⇒J dictionary. First, we construct a K⇒J dictionary using a K⇒E dictionary and a J⇒E. Second, we show another way of constructing a K⇒J dictionary using an K⇒E dictionary and an E⇒J dictionary. Third, we use a novel way of dictionary building using an E⇒K and E⇒J to build a K⇒J dictionary. However, our method is not limited to building a K⇒J dictionary but can be extended to any other language pairs so long as X-to-English or English-to-X dictionaries exist. These three methods will cope with making dictionaries using any combination.

We assume that the following conditions hold when building a bilingual dictionary: (1) Both the source language and the target language cannot be understood (to build a dictionary of unknown language pairs); (2) Various lexical information of the intermediate language (English) is accessible. (3) Limited information about the source and target language may be accessible.

3.1 Lexical Resources

Our method can be extended to any other language pairs if there are X-to-English and English-to-X dictionaries. It means that there are four possible combinations such as i) X-to-English and Y-to-English, ii) X-to-English and English-to-Y, iii) English-to-X and Y-to-English and iv) English-to-X and English-to-Y to build a X-to-Y dictionary. We tested i), ii) and iv) in this paper and we used the following dictionaries in our experiment:

Type	# Entries	Dictionary
J⇒E	28,310	New Anchor [2]
E⇒J	52,369	Super Anchor[3]
K⇒E	50,826	Yahoo K⇒E [4]
E⇒K	84,758	Yahoo E⇒K [4]

3.2 Linking K⇒E and J⇒E

Our method is based upon a **one-time inverse consultation** of Tanaka and Umemura (1994)(See Section 2.) to judge the word correspondences of Korean and Japanese.

Lexical Resources used here is a K⇒E dictionary (50,826 entries) and a J⇒E dictionary (28,310 entries). There is a big difference in the number of entries between the two dictionaries. This will affect the total number of extracted words.

For Evaluation, we use a similarity score S_1 for a Japanese word j and a Korean word k is given in Equation (1), where $E(w)$ is the set of English translations of w. This is equivalent to the Dice coefficient. The extracted word pairs and the score are evaluated by a human to keep the accuracy at approximately 90%.

$$S_1(j,k) = \frac{2 \times |E(j) \cap E(k)|}{|E(j)| + |E(k)|} \qquad (1)$$

The most successful case is when all the English words in the middle are shared by K⇒E and J⇒E. Figure 1 shows how the link is realized and the similarity scores are shown in Table 1. The similarity score shows how many English words are shared by the two dictionaries: the higher the score, the higher possibility of successful linking. However, as Table 1 shows, we have to sort out the inappropriately matched pairs by comparing the S_1 score of equation (1) against a threshold τ. The threshold allows us to exclude unfavorable results. For example, for words having one shared English translation equivalent, we have to discard the group (3) in Table 1.

When the words translated from English match completely, the accuracy is high. And if the number of shared English translated words ($|E(J) \cap E(K)|$) is high, then we get a high possibility of accurate matching of Korean and Japanese. However, accuracy deteriorates when the number of the shared English translated words (shown by the threshold) decreases as in (2) and (3) of Table 1. We solved this problem by varying the threshold according to the number of shared English equivalents. The value of the threshold τ was determined experimentally to achieve an accuracy rate of 90%.

Result: Linking through English gives a total of 175,618 Korean-Japanese combinations. To make these combinations, 28,479 entries out of 50,826 from the K⇒E dictionary and 17,687 entries out of 28,310 from the J⇒E dictionary are used. As a result, we can extract 25,703 estimated good matches with an accuracy of 90%.

[2](Yamagishi et al., 1997) [3] (Yamagishi and Gunji, 1991) [4] `http://kr.engdic.yahoo.com`

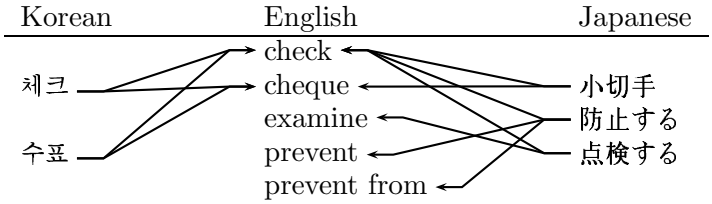


Figure 1: Linking through English translation equivalents (K⇒E, J⇒E)

	Shared Eng.	τ	Korean ⇒ English	Japanese ⇒ English
(1)	2	1.000	(체크check;cheque)	(小切手check;cheque)
	2	1.000	(수표 check;cheque)	(小切手check;cheque)
(2)	1	.667	(체크 check;cheque)	(照合check)
(3)	1	.500	(체크check;cheque)	(点検するcheck;examine)
	1	.400	(체크check;cheque)	(防止するprevent from;prevent;check)
	1	.333	(수표check;cheque)	(預けるleave;deposit;check;entrust)

Table 1: Example of linking through English translations

Shared Eng[5]	Extracted	τ	Good matches
7	1	0	1
6	1	0	1
5	16	0	16
4	165	0	165
3	1,325	0.4	1,206
2	12,037	0.5	7,401
1	161,863	0.667	16,790
Total	175,408		25,580

Table 2: Matching words by K⇒E +J⇒E

3.3 Linking K⇒E and E⇒J

Method: We investigated how to improve the extraction rate of equivalent pairs using an **overlapping constraint** method here. To extract Korean-Japanese word pairs, we searched consecutively through a K⇒E dictionary and then an E⇒J dictionary. We take English sets corresponding to Korean words from a Korean-English dictionary and Japanese translation sets for each English words from an E⇒J dictionary. The overlap similarity score S_2 for a Japanese word j and a Korean word k is given in Equation (2), where $E(w)$ is the set of English translations of w and $J(E)$ is the bag of Japanese translations of all translations of E.

$$S_2(j,k) = |j|, j \in J(E(k)), \qquad (2)$$

After that, we test the narrowing down of translation pairs by extracting the overlapped words in the Japanese translation sets. See Figure 2.

Lexical Resources: We used a K⇒E dictionary (50,826 entries), the same as the one used in section 3.2 and a E⇒J dictionary (52,369 entries). Compared to the resources used in our first method, the number of entries are well balanced.

Evaluation: After extracting the overlapped words in the Japanese translation sets, the words were evaluated by humans. The main evaluation was to check the correlation between the overlaps and the matches of Korean and Japanese word pairs. Table 3 shows the overlapped number of shared English words and the number of index words of the K⇒E dictionary.

Overlaps	Num of entries in K⇒E
4 or more	1,286
3	3,097
2	13,309
1-to-1 match	1,315
Subtotal	19,007
Other match	8,832
No Match	22,987
Total	50,826

Table 3: The number of entries in K⇒E dictionary according to overlapped English words

Result: Entries with a 1-to-1 match have $|E(K)| = |E(J)| = 1$. These are generally good matches (90%). If more than two overlaps occur, then the accuracy matching rate is as high as 84.0%. It means that the number of useful entries is the sum of the 1-to-1 matches and 2 or more overlaps: 19,007 (37.4% of the K⇒E entries) with 87% accuracy. However, using K⇒E and E⇒J there is a problem of polysemy in English words. For example, *clean* has two different POSs, adjective and verb in a K⇒E dictio-

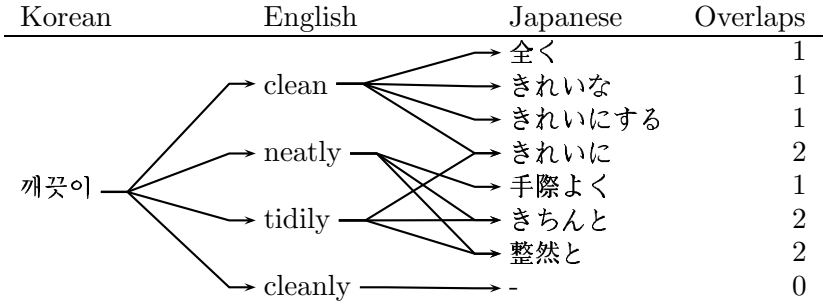


Figure 2: Overlapping Translation equivalents (K⇒E, E⇒J)

nary. Unfortunately, this information cannot be used effectively due to the lack of POS in K⇒E when linking them to a E⇒K dictionary. On the other hand, *clean* using E⇒J can be translated into either きれいな, an adjective or きれいにする, a verb. This makes the range of overlap score widely distributed as shown in Figure 2. This is the reason using K⇒E and E⇒J is not as good as using K⇒E and J⇒E. We will discuss this more in section 4.

3.4 Linking E⇒K and E⇒J

As we have discussed in earlier sections, the characteristics of dictionaries differ according to their directionality. In this section, we introduce a novel method of matching translation equivalents of Korean and Japanese. From the Korean speaker's point of view, the E⇒K dictionary covers all English words, includes explanatory equivalents, and example sentences showing usage. The same thing is true for the E⇒J dictionary from a Japanese speaker's point of view. In this respect, we expect that the result of extraction is not as effective as the other combinations such as K⇒E +J⇒E and K⇒E +E⇒J. On the other hand, we think that there must be other ways to exploit explanatory equivalents and example sentences.

Method: First, we linked all the Korean and Japanese words if there is any shared English words. Then, we sorted them according to POSs to avoid the polysemous problem of POS. The left hand side of Figure 3 shows how we link Korean and Japanese pairs.

Lexical Resources: We used a E⇒K dictionary (84,758 entries) and a E⇒J dictionary (52,369 entries). Both of the dictionaries have many more entries than the ones used in the previous two methods.

Evaluation: We use similarity score S_3 in Equation (3) as a threshold which is used to extract good matches.

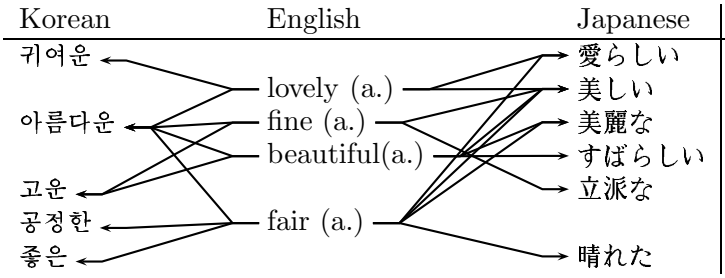
$$S_3(k,j) = \frac{|K(E(k) \cap E(j))| + |J(E(k) \cap E(j))|}{|E(k) \cap E(j)|} \quad (3)$$

$K(W)$: bag of Korean translations of set W
$J(W)$: bag of Japanese translations of set W
$E(w)$: set of English translations of word w

$|K(E)|$ means the number of Korean translation equivalents, and$|J(E)|$ means the number of Japanese translation equivalents. The sum of the numbers is divided by the number of intermediate English words. It is used to reduce the polysemous problem of English words. It is because it is hard to decide which translation is appropriate, if an English word has too many translation equivalents in Korean and Japanese. The value of threshold (S_3) is shown in Table 4. We vary the threshold according to $N = |E(j) \cap E(k)|$ to maximize the number of successful matches experimentally. N represents the number of intermediate English words. For N=1, we only count one-to-one matches, which means one Korean and one Japanese are matched through only one English. The following are examples of being counted when N is 1-to-1: e.g. 자기 암시-autosuggestion(n.)- 自己暗示, 당구(용)의-billiard(a.)-玉突きの, etc. We may lose many matching pairs by this threshold, but the accuracy rate for 1-to-1 is very high (96.5%). To save other matches when N=1, we need to examine further. In our experiment, 귀여운⇔ 愛らしいis rejected because lovely has two Korean translations and two Japanese translations; the match 귀여운⇔ 愛らしいis not 1-to-1. We postpone this part to further research.

N	Extracted	Matched	Good	S_3	Extracted	Matched	Good
24-6	438	422	96.3%	any	438	422	96.3%
5	313	301	96.2%	≤ 35	302	293	97.0%
4	790	698	88.3%	≤ 25	661	601	90.9%
3	2,432	1,960	80.7%	≤ 10	634	586	92.4%
2	12,862	(6,784)	(52.8%)	≤ 10	3,613	(3,150)	(87.2%)
$*$1[-to-1]	4,712	(4,547)	(96.5%)	2	4,712	(4,547)	(96.5%)
	21,547	(14,712)	(68.3%)		10,360	(9,599)	(92.7%)

Table 4: Summary of matching words by E⇒K and E⇒J
N: Number of total English translation equivalents
$*$: We only count word pairs under the condition of 1-to-1 match.



Examples	N	S_3	Matches
귀여운⇔愛らしい	1	(2+2)/1=4.0	N
귀여운⇔美しい	1	(2+2)/1=4.0	N
아름다운⇔美しい	4	(9+11)/4=5.0	Y
아름다운⇔美麗な	2	(5+7)/2=6.0	Y
공정한⇔晴れた	1	(3+4)/1=7.0	N

Figure 3: An example of matching E⇒K and E⇒J

Result: Table 4 shows the extracted 21,564 pairs of Korean and Japanese words. On average, 14,712 pairs match with a 68.3% success rate. The numbers in parentheses are estimated.

As expected, by setting this threshold we get fewer extracted words such as 10,360 words as shown in Table 4. However, the accuracy of the matched word pairs averages 92.7%.

Comparison: To compare the three methods, we randomly chose 100 Korean words from a K⇒J dictionary[6] which could be matched through all three methods. The number of extracted matches was 28 using K⇒E and J⇒E, 34 using K⇒E and E⇒J, and 13 using E⇒K and E⇒J. For K⇒E and E⇒J method, 21 out of 34 K⇒J pairs were found only in K⇒E and E⇒J method but not in K⇒E and J⇒E method. Among the 21 new K⇒J word pairs, only one pair is an error (not a good match). One new pair was found in E⇒K and E⇒J method. Therefore, combining all three methods gave 49 (28+20+1) different K⇒J pairs, a better result than any single method. These results are shown in Table 5. Clearly the dictionaries used greatly affect the number of matches. The number of matches could be improved by considering English derived forms (e.g. matching *confirmation* with *confirm*).

	K⇒E +J⇒E	K⇒E +E⇒J	E⇒K +E⇒J
Total	28	34	13
Good	28	33	10
Error	0	1	3

Table 5: Comparison of the Proposed Methods

4 Discussion

We have shown the results of different matching metrics for different dictionary directions. Directionality is an important matter for building dictionaries automatically. In a K⇒E (or J⇒E) dictionary an index word contains non-conjugated forms whereas an index word in E⇒K (or E⇒J) dictionary contains POS and conjugated forms. Therefore we expect the combination of K⇒E and J⇒E to be better than K⇒E and E⇒J since we can avoid the mismatch of POS.

On the other hand, a dictionary E⇒K or E⇒J contains less uniform information such as long expository terms, grammatical explanations and example sentences. Especially, POS is far more detailed than the dictionaries of the

[6]We used **Korean-Japanese dictionary** (Shogakukan: 1993) for the sampling that includes 110,000 entries, many of which are used infrequently.

other direction. These all contribute to fewer good matching words.

As for the better result using K⇒E and J⇒E, we cannot overlook language similarity: Korean and Japanese are very similar with respect to their vocabularies and grammars. This must have result in sharing relatively more appropriate English translations and further matching more appropriate Korean and Japanese translation equivalents.

In the combination of K⇒E and E⇒J, the common English translations are reduced due to the characteristics of K⇒E and E⇒J. A K⇒E dictionary from the Korean speaker's point of view tends to have relatively simple English equivalents and normally POS is not shown. On the other hand, an E⇒J dictionary shows such complicated equivalents as explanation of the entry **a**, a piece of translation equivalent **b** and grammatical information as shown in (2) in Section 1. Therefore, it is natural that the matching rate is far less than the combination of K⇒E and J⇒E. Considering the size of dictionaries used in K⇒E and J⇒E (estimated maximum matches: 28,310 K⇒J pairs) and the one used in K⇒E and E⇒J (estimated maximum matches: 50,826 K⇒J pairs), we extrapolate from Table 5 that the method using K⇒E and J⇒E is better than the method using K⇒E and E⇒J.

We concluded that: K⇒E + J⇒E outperforms K⇒E + E⇒J which outperforms E⇒K + E⇒J. The following briefly summarizes the three methods.

- K⇒E + J⇒E:
  - Equal characteristics of the dictionaries
  - The meaning of the registered words tends to be translated to a typical, core meaning in English
  - Synergy effect: Korean and Japanese are very similar, leading to more matching.
- K⇒E + E⇒J:
  - The combination of different characteristics of dictionaries makes automatic matching less successful.
  - A core meaning is extended to a peripheral meaning at the stage of looking up E⇒J. (See Figure 2.)
- E⇒K + E⇒J:
  - There are far fewer matches.
  - We can take advantage of example sentences, expository terms, and explanations to extract functional words.
  - We can improve accuracy by including English POS data.

Even though we expected that the combination of dictionaries between E⇒K and E⇒J will not provide a good result, it is worthwhile to know limits. After analyzing all of the result, we found that there is the effect of dictionary directionality. Also, we confirm that if we can use all the methods and combine them, we will get the best result since the output of the three dictionary combinations do not completely overlap.

Future Work

Our goal is not restricted to making a Korean-Japanese dictionary, but can be extended to any language pair. We assume that we do not know the source and target languages so well that it is not easy to match just the content words. Instead, we need to match automatically any kind of entries, even such functional words as particles, suffixes and prefixes. We think that it is best to extract these functional words by taking advantage of the characteristics of the E⇒K and E⇒J dictionaries. For example, one of the merits of using E⇒K and E⇒J is that we can get conjugated forms such as the Korean adjective **아름다운** which matches the English adjective **beautiful**; it is normally not registered in a K⇒E dictionary because **아름다운** is an adjective conjugated form of the root **아름답다**. Only the root forms are registered in an X-to-English dictionary. Also for verbs, we can get non finite forms using E⇒K and E⇒J dictionaries. As index word, the non-conjugated forms are registered in a J⇒E dictionary such as **きれいだ** meaning *beautiful* or *clean*. However, by using E⇒J, we can get conjugated forms such as **きれいに**, **きれいな** and so forth. Registering all conjugated forms in a dictionary simplifies the development of a machine translation system and further second language acquisition.

The direction from English-to-X contains a lot of example sentences. So far, the idea of using example sentences and idiomatic phrases for dictionary construction has not been adopted. To check the possibility of extracting functional words, we extracted example sentences and idiomatic phrases from E⇒J and E⇒K dictionaries based upon the number of shared English words and look into the feasibility of using them to extract functional words.

We extracted a total of 1,033 paraphrasing sentence pairs between Korean and Japanese with five or more shared English words. Among them, 465 sentences (45%) matched all the English exactly (=), and 373 sentences (36.1%) almost (≈) matched. We give examples below:

= (10) "as for me, give me liberty or give me death." 私としては自由が得られなければ死んだほうがましだ.
"as for me, give me liberty or give me death." 나에게는 자유가 아니면 죽음을 달라.

≈ (8) "he is taller than any other boy in the class." 彼はクラスのだれよりも背が高い.
"Tom is taller than any other boy in his class." 톰은 반에서 누구보다도 키가 크다.
(extracted from E⇒K and E⇒J)

The numbers in parentheses in the above examples represent how many English words are shared between E⇒K and E⇒J. Using these paraphrasing sentences we will examine the effective way of extracting functional words.

Finally we would like to apply our method to open source dictionaries, in particular EDICT (J⇒E, Breen (1995)) and `engdic` (E⇒K, Paik and Bond (2003)). This would make the results available to everyone, so that they can be used in comparative evaluation or further research.

5 Conclusion

We have shown three major combination of dictionaries to build dictionaries. These methods can be applied to any pairs of language; we used a K⇒E dictionary, a J⇒E, an E⇒K dictionary and an E⇒J to build a K⇒J dictionary using English as a pivot.

We applied three different methods according to different combination of dictionaries. First, a one-time look up method (Tanaka and Umemura, 1994) is tried using K⇒E and J⇒E. Second, an overlapping constraint method in one direction is applied using K⇒E and E⇒J. Finally, a novel combination for building a dictionary is attempted using E⇒K and E⇒J. We found that the best result is obtained by the first method. However, by combining all methods we can extract far more entries since the results from the three method do not overlap. Our result shows that 60% of word pairs in the second method are not found in the first or the third method. For the third method (using E⇒K and E⇒J), we could not extract as many matched pairs, but it is potentially useful for extracting conjugated forms and functional words.

Acknowledgments
This research was supported in part by the Ministry of Public Management, Home Affairs, Posts and Telecommunications. We would also like to thank Francis Bond for his comments and discussion.

References

Christian Boitet, Mathieu Mangeot, and Gilles Sérasset. 2002. The Papillon Project: cooperatively building a multilingual lexical data-base to derive open source dictionaries and lexicons. *The 2nd Workshop NLPXML-2002*, pages 93–96, Taipei, Taiwan.

Francis Bond, Ruhaida Binti Sulong, Takefumi Yamazaki, and Kentaro Ogura. 2001. Design and construction of a machine-tractable Japanese-Malay dictionary. In *MT Summit VIII*, pages 53–58, Santiago de Compostela, Spain.

Jim Breen. 1995. Building an electronic Japanese-English dictionary. Japanese Studies Association of Australia Conference.

Reinhard Rudolf-Karl Hartmann. 1983. *Lexicography: Principles and Practice*. Academic Press.

Mathieu Lafourcade. 2002. Automatically populating acception lexical database through bilingual dictionaries and conceptual vectors. In *Papillon 2002 Seminar(CD-Rom)*, Tokyo, Japan.

Kyonghee Paik and Francis Bond. 2003. Enhancing an English and Korean dictionary. In *Papillon-2003*, pages CD–rom paper, Sapporo, Japan.

Kyonghee Paik, Francis Bond, and Satoshi Shirai. 2001. Using multiple pivots to align Korean and Japanese lexical resources. In *NLPRS-2001*, pages 63–70, Tokyo, Japan.

Satoshi Shirai and Kazuhide Yamamoto. 2001. Linking English words in two bilingual dictionaries to generate another language pair dictionary. In *ICCPOL-2001*, pages 174–179, Seoul.

Satoshi Shirai, Kazuhide Yamamoto, and Kyonghee Paik. 2001. Overlapping constraints of two step selection to generate a transfer dictionary. In *ICSP-2001*, pages 731–736, Taejon, Korea.

Kumiko Tanaka and Kyoji Umemura. 1994. Construction of a bilingual dictionary intermediated by a third language. In *COLING-94*, pages 297–303, Kyoto.

Katsuei Yamagishi and Toshio Gunji, editors. 1991. *The New Anchor Japanese-English dictionary*. Gakken.

Katsuei Yamagishi, Tokumi Kodama, and Chiaki Kaise, editors. 1997. *Super Anchor English-Japanese dictionary*. Gakken.