Construction of a Bilingual Dictionary Intermediated by a Third Language

Kumiko TANAKA

Division of Engineering, University of Tokyo.
7-3-1 Hongo, Bunkyouku, Tokyo, 113, Japan.
kumiko@ipl.t.u-tokyo.ac.jp

Kyoji UMEMURA

NTT Basic Research Laboratories.
3-9-11 Midori, Musashino, Tokyo, 180, Japan.
umemura@nuesun.NTT.JP

Abstract

When using a third language to construct a bilingual dictionary, it is necessary to discriminate equivalencies from inappropriate words derived as a result of ambiguity in the third language. We propose a method to treat this by utilizing the structures of dictionaries to measure the nearness of the meanings of words. The resulting dictionary is a word-to-word bilingual dictionary of nouns and can be used to refine the entries and equivalencies in published bilingual dictionaries.

1 Introduction

When vocabulary cannot be found in bilingual dictionaries, it is frequently obtained by using a third language as an intermediary. This indicates that supplemental information may lie in other forms in other dictionaries. Here we try using electronic dictionaries which can be reformed on a large scale, to extract this informations so that we can obtain subsidiary data and refine a direct bilingual dictionary.

Looking up words in bilingual dictionaries intermediating the third language is a method often used by people who handle uncommon languages in a specific domain. If this process can be automated, bilingual dictionaries of any kind between any languages may be obtained as long as these concerned languages have dictionaries to a common language. One objective of the research reported here is to establish a first step in automating this process.

To construct a Japanese↔French dictionary, we chose English as the intermediary language because Japanese↔English and English↔French dictionaries exist in electronic forms and because published Japanese↔French dictionaries provide enough vocabulary in comparison with the resulting dictionary.

In Section 2 we describe a method for extracting equivalencies for a given word. Its fundamental concepts are stated in Section 3. The whole procedure used to construct the new dictionary is shown in Section 4 and in Section 5 the resulting dictionary is evaluated.

Japanese-English, English-Japanese, English-French, French-English, Japanese-French and French-Japanese dictionaries are respectively denoted $\text{Dic}_{j\rightarrow e}$, $\text{Dic}_{e\rightarrow j}$, $\text{Dic}_{e\rightarrow f}$, $\text{Dic}_{f\rightarrow e}$, $\text{Dic}_{j\rightarrow f}$, and $\text{Dic}_{f\rightarrow j}$. $\text{Dic}_{x\rightarrow y}$ is called an *inverse dictionary* of $\text{Dic}_{y\rightarrow x}$. Japanese words have information in the following format: pronunciation in romaji, and its equivalence in English. English words are written in this font and French words in this font.

2 Overview of the Method

2.1 Inverse Consultation

The most naive way to use English to obtain French words corresponding to a Japanese word is to look up the Japanese word in a $\text{Dic}_{j\rightarrow e}$ and then look up the resultant English words in a $\text{Dic}_{e\rightarrow f}$. The resulting French words can be regarded as equivalence candidates (ECs) of the original Japanese word. For example, in Fig. 1, ECs for a Japanese word "競争"(kyousou : competition) are **compétition, concours, race** etc. Among these, **race** and **hâte** are inadequate as equivalents of "競争".

As for **race**, the English word race has several meanings with the same spelling: one is *to compete* and another is *human race*. It is *human race* which induces the inadequate EC **race**. As for **hâte**, the English word race has the wider meaning *to hurry* which the original Japanese word "競争" does not. Since **hâte** is a direct translation of *to hurry*, it is inappropriate as an equivalence.

The following three cases generate irrelevant ECs.

1. An English word with the same spelling but with different meanings is intermediated. (**race** in the above example)
2. An English word with a wider meaning than that of original Japanese word is intermediated. (**hâte** in the above example)
3. There are mistakes in dictionaries.

The first two cases are due to the ambiguity in English. An English word with a narrower meaning than the Japanese source may miss some French equivalents. We think that if the original word has ambiguity and several meanings, the dictionary gives the corresponding English words.

We could handle the problem of choosing equivalencies from among ECs by semantic processing, but expressing meaning of vocabulary in dictionaries is a great problem. A simpler way is to look the ECs up in the inverse dictionary. For example, one can consult $\text{Dic}_{f\rightarrow j}$ for **compétition, concours,** and **race** and thereby get "競争", "競
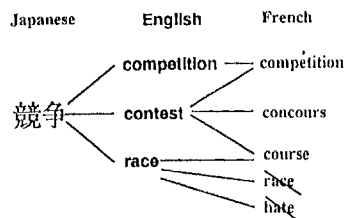
competition —— compétition

競争 —— contest

race —— course

Fig. 1 Equivalence candidates (ECs) for "競争".

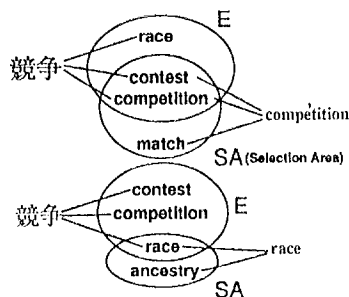Fig. 2 One time inverse consultation (IC_1).

Fig. 3 Two times inverse consultation (IC_2).

争", and "人種" (jinshu : human race) as their respective equivalencies. Since "人種" has nothing to do with "競争", race is excluded. We call this method of looking up ECs in the inverse dictionaries when choosing relevant equivalencies *inverse consultation*, and we call the words obtained by looking up inverse dictionaries the *selection area* (SA). Inverse consultation utilizes the structure of dictionaries to measure the nearness of the meanings of words in different languages.

The simplest application of inverse consultation is to use $Dic_{f \to e}$. In the above example, each EC is looked up in $Dic_{f \to e}$ and the results are compared with the English equivalencies of "競争"(E); namely, competition, contest, and race. The SA of **compétition** is competition, contest and match, which have the elements contest and competition in common with E (Fig. 2). As **compétition** derived from competition, competition should be put aside, but contest is still left as a common element and thus **compétition** is selected as an equivalence of "競争". As for race, the SA of race consists of race and ancestry, whose intersection with E only gives race; so race is judged as an inadequate EC. In short, the number of elements in common between the selection area and E indicates the nearness of the meaning between the EC and the original word.

For the inverse consultation described above, the SA was in English. If we use $Dic_{e \to j}$ as an inverse dictionary successively after consulting $Dic_{f \to e}$, then the SA is in Japanese and we compare "競争" with the SA (Fig. 3). The SA for
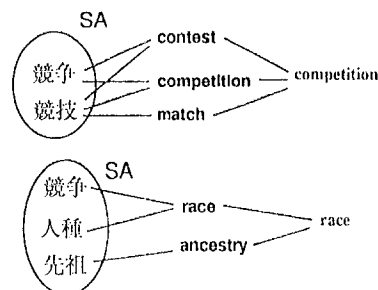
compétition consists of two "競争"s and three "競技"s (kyougi : game). For **race**, the SA has "競争", "人種", "先祖"(senzo : ancestry). Since "競争" appears only once for **race**, we discard the EC **race**.

There can be a infinite number of inverse consultations according to the number of consulted inverse dictionaries. If the inverse dictionaries are consulted n times, we call the method n *times inverse consultation* (IC_n). Which inverse dictionary to use does not always have a unique answer. For IC_2 for example, we may consult $Dic_{e \to f}$ after consulting $Dic_{f \to e}$ with the SA in French.

2.2 Selection Procedure

Once the SA for a given word is obtained, equivalencies are selected by handling two collections of words. We call this process the *selection procedure*.

One way to do this is to count the number of specific elements in the SA. For example, if the SA is in Japanese, the number of the element "競争" itself is counted. Another way is to count how many *parts of words* (PWs) are contained in the SA. For example, the number of "競" and "争" contained in the SA is considered (thus "競" in "競技" is also counted).

If we handle the meaning of words, a third way is to look up "競争" in a Japanese thesaurus and count how many times the synonyms appear in the SA. For example, if "競技" is a synonym of "競争", then the number of appearances of "競技" is added to that of "競争". If we go further to handle the meaning, we might as well process words by semantic processing.

Since a kanji is an ideogram, the second method also handles the meaning. When the selection area is in English or French, the corresponding method is to count morphemes such as "inter" and "national" for the word "international". We are interested in to what extent the method that does not explicitly concern the meaning may be used to handle the meaning of words. In the following, we focus on the former two methods.

3 Fundamental Concepts

3.1 Harmonized Dictionary

A bilingual dictionary forms a graph whose nodes are words and whose branches are correspondences between the words. Branches have directions, which make the graph asymmetric. $Dic^{-1}_{x\to y}$ is a graph with all the branches in $Dic_{x\to y}$ in inverse direction.

Since the purpose of bilingual dictionaries is to denote the correspondences of words that have the same meaning [Har83], it is natural that branches are bidirectional. We therefore design symmetrical dictionaries and we denote a dictionary from language x to y as $D_{x\to y}$,calling it a *harmonized dictionary*. When $D_{x\to y}$ and $D_{y\to x}$ are constructed from the same dictionaries, $D_{y\to x} = D^{-1}_{x\to y}$ holds. We remove the overlaps of branches.

3.2 Syntactic Selection Procedure

A multiset here is a set in which each element has a weight that is a natural number. The weight of an element is defined as the number of times it appears when looking up words in dictionaries. In the example shown in Fig. 3, the multiset SA for **compétition** consists of "競爭" with weight 2 and "競技" with weight 3. We denote the weight of element x in multiset X as $\delta_a(X, x)$; for instance, $\delta_a(SA, 競爭) = 2$. Using the same notation, $\delta_a(X, Y)$ is defined as follows when X and Y are multisets:

$$\delta_a(X, Y) = \sum_{y \in Y} \delta_a(X, y)$$

When a multiset Z consists of "競爭" and "競技", then $\delta_a(SA, Z) = 5$.

The notation $\delta_b(X, x)$ represents the sum of the weights of the elements that contain PWs of x in multiset X. For instance, if PWs are defined as kanji, $\delta_b(SA, 競爭)$ is 7 by adding 5 (sum of weights of elements in the SA that have kanji "競") and 2 (sum of weights of elements in the SA that have kanji "爭"). Using the same notation, $\delta_a(X, Y)$ is defined as follows when X and Y are multisets:

$$\delta_b(X, Y) = \sum_{y \in Y} \delta_b(X, y)$$

For instance, $\delta_b(SA, Z) = 12$; that is 7 plus 5 ($= \delta_b(SA, 競技)$).We use the notation δ as a parameter for δ_a or δ_b.

3.3 Properties of Inverse Consultation

In the following, we use $D_{f\to e}$ and $D_{e\to j}$ as inverse dictionaries when starting from Japanese words and we use $D_{j\to e}$ and $D_{e\to f}$ as inverse dictionaries when starting from French words. French ECs for a Japanese word j form a multiset F expressed as

    F = $D_{e\to f}D_{j\to e}j$.
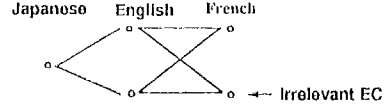
French equivalencies selected by IC_1 form a

Fig. 4 A structure that δ_a is inapplicable.

multiset whose element f satisfies

    $f \in F$ and $\delta(D_{f\to e}f, D_{j\to e}j) > 1$.

French equivalencies selected by IC_2 form a multiset whose element f satisfies

    $f \in F$ and $\delta(D_{e\to j}D_{f\to e}f, j) > 1$.

In the following, we focus on the IC_1 and IC_2 described above and examine their properties.

Property 1

$$\delta_a(D_{e\to j}D_{f\to e}f, j) = \delta_a(D_{j\to e}j, D_{f\to e}f)$$
$$= \delta_a(D_{e\to f}D_{j\to e}j, f)$$

This property indicates that when using selection procedure δ_a, equivalencies selected by IC_1 and IC_2 are exactly the same. Moreover, it is sufficient to choose ECs whose weights are greater than 1 in F. The proof of Property 1 is shown in Appendix.

Property 2 *If δ is δ_a, the resulting Japanese-French dictionary is a harmonized dictionary. This is not always true for δ_b.*

This property is clear with symmetrically structured dictionaries. When δ_b is used, the resulting dictionary depends on how PW is defined and does not always become a harmonized dictionary.

Property 3 *If there is a structure such as shown in Fig. 4, δ_b must be used as δ to exclude irrelevant ECs.*

When δ_a is used, the SA for two ECs will be exactly the same, which makes it impossible to discard inappropriate ECs. Although this kind of structure seems to be rare, it can exist because of the historical transition of words. When a single English word is intermediated, inappropriate ECs cannot be discarded by using δ_a for IC_1.

4 Experiment

4.1 Dictionary Data

The dictionaries used in the experiment are $Dic_{j\to e}$ [Ich90], $Dic_{e\to j}$ [Koi90], $Dic_{e\to f}$ [For82], and $Dic_{f\to e}$ [Led82]. The whole experimental procedure is shown in Fig. 5. Word-to-word dictionaries are first extracted from each dictionary. All words are nouns; in particular, they are one word nouns in English and French. Since the dictionary syntax was not always consistent, word-to-word dictionaries contain some mistakes (inadequate correspondences).

Harmonized dictionaries are then constructed from the word-to-word dictionaries as follows:

    $D_{e\to j} = Dic^{-1}_{j\to e} \cup Dic_{e\to j}$
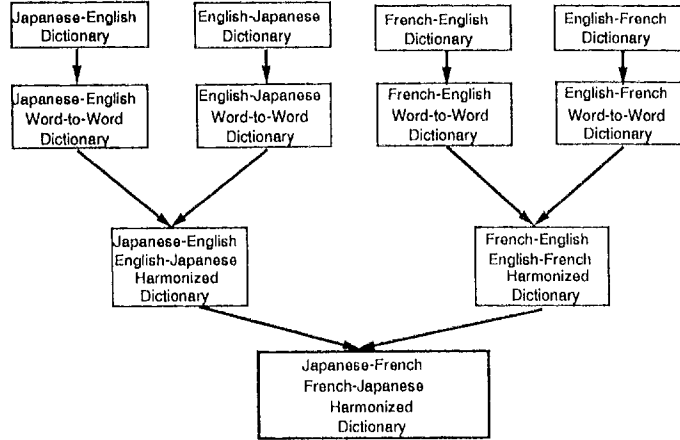
Fig. 5 Whole procedure.

$$D_{j\to e} = D_{e\to j}^{-1}$$
$$D_{e\to f} = Dic_{f\to e}^{-1} \cup Dic_{e\to f}$$
$$D_{f\to e} = D_{e\to f}^{-1}$$

Although there are other ways to symmetrize dictionaries, (for example, by removing all branches that are not bidirectional), we chose the above procedures for the lexicographical reason described below [Har83].

There are two kinds of bilingual dictionaries, one is from a foreign language (f) to the mother language ($Dic_{f\to m}$), and the other is from a mother language (m) to a foreign language ($Dic_{m\to f}$). In $Dic_{f\to m}$, when there are no equivalencies in m for a foreign word, the dictionary gives its definition or explanation of the word in m. Therefore, all the foreign words can be contained in the dictionary. In $Dic_{m\to f}$ on the other hand, if there are no equivalencies in f for a word of m, the word itself is often dropped from the dictionary. The words contained in the dictionary are therefore a part of m, and $Dic_{m\to f}$ lacks many entries. A harmonized dictionary is a solution to this problem because it contains equivalencies of $Dic_{f\to m}$ as entries of $Dic_{m\to f}$.

4.2 Procedure of Inverse Consultation

From Property 1, we use δ_a with IC_1 and we use δ_b with IC_2. The PW for δ_b are defined as follows:

- Japanese   6353 kanjis.
- French   morphemes [Mau85].
  1151 prefix and 710 suffix.

From Property 2, inverse consultation is applied to both Japanese and French entries, and then the results are put together to construct a harmonized dictionary. We denote it as $D_{j\to f}$ or $D_{f\to j}$.

Each entry within $D_{j\to e}$ and $D_{f\to e}$ is classified into one of five types according to the procedure used to select its equivalencies from ECs.

- Type A   A single EC exists and is selected unconditionally as the equivalence.
- Type B   Equivalencies by IC_1 exist; all are selected and the rest of the ECs are discarded.
- Type C   There are no equivalencies by IC_1 but there are by IC_2. One third (empirically decided) of ECs by IC_2 are selected according to the value of the following fraction (larger ones are selected):

$$\frac{\text{Value of function } \delta_b}{\text{SA (byte)}}$$

- Type D   No equivalencies by IC_1 nor by IC_2 appeared but there are several ECs. For this entry, it is impossible to select the relevant equivalencies.
- Type E   There are no ECs.

Entries of Type A acquire more appropriate equivalencies than do entries of Type B, and entries of Type B acquire more appropriate equivalencies than do entries of Type C.

5 Evaluation of Experiment

5.1 Result of the Example

Equivalencies obtained for the Japanese word "競争" are

concours, rivalité, compétition,
course, concurrence, émulation

and the intermediated English words are

competition, contest, rival, rivalry, race.

The entry "競争" is of Type B. The number of ECs is 41 (including overlaps), and 13 of them are selected by IC_1. The number of ECs in each category of irrelevant words described in Section 2.1 is listed in Table 1, which indicates that inverse consultation can detect the relevant words even when there are mistakes in the dictionaries. The word which should not be dropped was joute.

Equivalencies for "競争" in [Tam85] are

Table 1 Details of ECs of "競争".

Selected equivalencies	13
Different meaning in English	9
Meaning extended in English	9
Mistakes of dictionaries	9
Words which must not be dropped	1

Table 2 Classification of entries.

	$D_{j \to f}$		$D_{f \to j}$	
Total	42190	(100.0%)	23710	(100.0%)
Type A	1600	(3.8%)	3852	(16.2%)
Type B	8179	(19.4%)	7397	(31.3%)
Type C	24047	(57.0%)	6452	(27.2%)
Type D	1514	(3.6%)	2988	(12.6%)
Type E	6850	(16.2%)	3021	(12.7%)

Table 3 Entries with no EC.

category	example	English equivalence
Cultural words	お年玉 (otoshidama : Our tradition to give money to children on New Year)	handsel
	cédille	cedilla
Technical terms or proper nouns	ガウス (gaus : gauss)	gauss
	Cicéron	Cicero
Borrowed words	アペリチフ (aperichihu : apetizer)	apéritif
	tee-shirt	teeshirt

concours, rivalité, compétition, course, concurrence.

Our result contains émulation (which means *rivalry*) in addition to the entries in the published dictionary.

5.2 Evaluation of Entries

Table 2 lists the each number of entries belonging to the Types A~D (defined in Section 4.2). Type D consists of the following entries.

- One English word is intermediated and several ECs appear.
- The entry contains no PW.

The Type D percentage difference between $D_{j \to f}$ and $D_{f \to j}$ shows that the number of entries of Type D depends on the number of registered PWs.

No EC appears when an English word to be intermediated does not exist as the entry of $D_{e \to f}$ or $D_{e \to j}$. Such entries can be categorized as in Table 3.

Original words are apt to be translated into uncommon English words, so they normally do not appear as entries if the same kind of words do not exist in the objective langauge. Technical terms and proper nouns depend very much on culture.

Table 4 Evaluation of equivalencies.

	$D_{j \to f}$		$D_{f \to j}$	
rate	R1	R2	R1	R2
80%~100%	58	56	18	58
60%~80%	4	14	8	15
40%~60%	9	13	13	9
20%~40%	8	10	24	4
0%~20%	22	7	37	14

Many French placenames, for instance, are Type E. Borrowed words are expressed in inconsistent spellings, and which of them are to be found in the dictionary also depends on the culture (apéritif is the equivalence in $Dic_{j \to e}$).

Since a harmonizing dictionary augments the entries, the resulting dictionary contains entries that are not in the published dictionary. As explained in Section 4.1, this phenomenon is conspicuous when we compare $D_{j \to f}$ with the published dictionary [Suz70]. These entries can be categorized as follows:

1. Colloquial words.
   Ex. わんちゃん (wanchan : puppy)
2. Technical terms or proper nouns.
   Ex. アスベスト (asubesuto : asbeste)
3. Compound nouns.
   Ex. 風化作用 (huukasayou : disintegrate)
   物価安定政策 (bukkaanteiseisaku : valorization)

Harmonizing dictionaries help to gather correspondences between the mother language and a foreign language and are useful in revising published dictionaries.

5.3 Evaluation of Equivalencies

We evaluated the equivalencies of resulting dictionaries by comparing them with those of published dictionaries [Tam85] [Suz70]. For random 100 entries in both dictionaries, the following two percentages (calculated manually) are listed in Table 4:

- R1   Fraction of equivalencies in the published dictionary which were also found by this method.
- R2   Fraction of equivalencies found by this method which were judged appropriate.

Note that entries with greater R2 contain appropriate equivalencies on higher rate. This is not true for R1, since R1 indicates the discrepancy of equivalencies between the resulting dictionary and published dictionaries.

Entries of R2=100% can be classified as follows:
1. Entries of Type A or B. (See Section 4.2.)
2. Entries that have less than three ECs.

Entries with less than 3 ECs amount to 29.0% for $D_{j \to f}$ and 34.2% for $D_{f \to j}$. This indicates that words with specific meanings are apt to acquire appropriate equivalencies.

Table 5 French entries whose R1 are 0 %.

entry	$D_{f \to j}$	Published dictionary [Tam85]
distique	行連句,対句 (gyourenku, tuiku : the terminology for Japanese and Chinese poems of same kind)	2 行 詩 (2 gyou-shi : coinage for the term for European poems)
pull	セーター (seetaa : sweater written with Japanese letters. This word is common)	プルオーバー (puru obar : pull-over written with Japanese letters. This term is not common in Japan)
boulier	算盤 (sanban : Abacus in general)	そろばん, 数え玉 (soroban, kazoedama : Japanese abacus)

Table 6 Japanese entries whose R1 are 0 %.

entry	$D_{j \to f}$	Published dictionary [Suz70]
つぐみ (tsugumi : thrush)	grive	merle
洒落 (share : witticism, joke, jest, pun)	astuce	calembour
	badinage	
	drôlerie	
	facérie	
	farce	
	plaisanterie	

The entries tend to have either R1=80%~100% or R1=0%~20%. Of the latter, some examples are listed in Table 5 and 6. For distique in Table 5, which is a term in French literature, $D_{f \to j}$ translates it into the term for the corresponding kind of Japanese literature. Although it resembles to the direct translation of distique, it is only an analogy. On the other hand, since the direct translation of distique in the published dictionary adopts the concept of French poems, it is uncommon and cannot be understood by most Japanese readers. The same is true for boulier except that the common Japanese is indicated in a published dictionary. Pull is borrowed from the English word pull-over whose direct translation is contained in a published dictionary, and it is not a common word in Japanese.

In the first example in Table 6, grive is the generic name equivalent to thrush, whereas merle is a kind of thrush. The second example shows that more equivalencies are found in $D_{j \to f}$.

To sum up, the resulting dictionary can be utilized in conjunction with the published dictionaries as follows:

- To revise the equivalencies.
- To supplement the equivalencies.

Sections of the resulting dictionaries is listed in Table 7. List 1 is $D_{j \to f}$ and List 2 is $D_{f \to j}$.

For each list, entries are in the first row and their equivalencies are in the third row. Symbols in the second row indicate how appropriate each equivalence is. (Refer to the notes beside.)

6 Related Work

The use of a third language English as an intermediary in the construction of a bilingual dictionary was tested manually on a large scale on editing the Spanish-Japanese dictionary [Knw90]. It is now a representative middle-sized dictionary having a large quantity of information.

Tokunaga and Tanaka [Tok90] tried to extract a conceptual dictionary from Japanese-English and English-Japanese dictionaries. Although they used a concept similar to ours that is the graph structure of a dictionary related to meaning of words, their frameworks and final product differ from ours.

7 Conclusion and Future Works

The proposed method for using a intermediate language to construct a bilingual dictionary utilizes the structure of dictionaries and morphemes and can choose appropriate equivalencies for most entries. Comparing the resulting dictionary with published dictionaries showed that data obtained are useful for revising and supplementing the vocabulary of existing dictionaires.

To increase the accuracy with which equivalencies can be selected, mistakes in word-to-word dictionaries must be corrected even if our method may detect appropriate equivalencies. One way to do this would be to use thesaurus to check whether the extracted correspondences are relevant.

Nouns were taken into consideration in this research, and the next step will be to apply the proposed method to other parts of speech. We also need to establish a way to handle compound words in European languages.

References

[Bog89] Boguraev, B. et al. (1989). *Computational Lexicography for Natural Language Processing.* Longman.

[Fil85] Fillmore, C. (1985). Frames and the semantics of understanding. *Quaderni di Semantica, Vol.6, No.2.*

[Fil92] Fillmore, C. and Atkins, B. (1992). Toward a Frame-Based Lexicon. *The semantics of RISK and its Neighbors.* Academic Press.

Table 7 A part of resulting Japanese↔French dictionary.

List1		
わら屋根	0*	chaume
	2	crinière
	2	tifs
わるさ	1	gaminerie
	0*	malice
わんちゃん！	0	chien-chien
	0	chienchien
	0	toutou
わんわん	0	chien-chien
	0	chienchien
	0*	toutou
！アーチ	0	arc
	0	arche
	0	voute
！アート	0	art
	2	japonisant
！アービュタス	0	arbousier
！アームロック	1	couette
	0	enclenchement
	0	fermeture
	0	immobilisation
	1	platine
	0	serrure
	0	trapillon
アーモンド	0*	amande
	1	dragée
勧進元	0*	promoteur
	0	fondateur
勧誘	1	discussion
	0	invite
	1	pointage
	0*	sollicitation
	2	vérification
巻尺	1	ruban-cache
	2	centimètre
	2	magnétophone
！巻き添え	2	existence
！巻き物	0	rouleau
	0	volute

List2		
Zambie	0*	ザンビア
Zaïre	0*	ザイール
Zimbabwe	0*	ジンバブウェ
!Zoroastre	0	ゾロアスター
abaissement	0	引き下げ
	0	下押し
	0	下降
	0	下落
	0	減額
	0*	減少
	0	減退
	0	降下
	0	生活水準の低下
	0	脱落
	0	沈下
	0	墜落
	0*	低下
	0	落下
	1	落下距離
	1	落差
	1	限石の落下
abandonnement	0*	放擲
	0	諦め
abandon	0*	遺棄
	0	屈服
	0	降参
	0	怠慢
	0	退会
	0	退職
	0	脱退
	0	諦め
	0	忍従
	0*	放棄
abaque	0	グラフ
	0	図
	0	図式
	0	図表
	0	早見表
	0	表

0: Judged appropriate (counted for R2).

1: Significance slightly different.

2: Significance completely different.

*: Found common in two dictionaries (counted for R1).

Mark "!" attached to entries means that the entries were not found in the published dictionaries.

[For82] Forbes, P. et al. (1982). *Shorter English-French Dictionary*. Harrap Limited.

[Har83] Hartmann, R. (1983). *Lexicography: principles and practice*. Academic Press.

[Ich90] Ichikawa, S. (1990). *New Japanese-English Dictionary*. Kenkyuusha.

[Ide93] Ide, N. and Véronis, J. (1993). Extracting knowledge basis from machine readable dictionaries. *KB&KS, JIPDEC*.

[Koi90] Koine, Y. (1990). *New English-Japanese Dictionary*. Kenkyuusha.

[Kuw90] Kuwana, K. (1990). *Spanish-Japanese Dictionary*. Shougakukan.

[Led82] Ledésert, M. et al. (1982). *Shorter French-English Dictionary*. Harrap Limited.

[Mau85] Maubourguet, P. et al. (1985). *Grand Dictionnaire Encyclopédique LAROUSSE*. Larousse.

[Mil90] Miller, G. et al. (1990). Five Papers on WordNet. *CSL Report 43*. Cognitive Science Laboratory, Princeton University.

[Oht78] Ohtsuki, T. et al. (1978). *Crown French-Japanese Dictionary*. Sanseido.

[Suz70] Suzuki, S. (1970). *Standard Japanese-French Dictionary*. Taishuukan.

[Tam85] Tamura, T. (1985). *Dictionnaire Français-Japonais ROYAL*. Hakusuisha.

[Tok90] Tokunaga, T. and Tanaka, H. (1990). The Automatic Extraction of Conceptual Items from Bilingual Dictionaries. *PRICAI*.

Appendix

The following two lemmas are needed to prove Property 1.

Lemma 1-1 $x \in D_{y\to x}x \iff y \in D_{x\to y}x$

This is clear from the symmetric structure of the harmonized dictionary.

Lemma 1-2 *If X is a set (every element has a weight of 1) then $\delta_a(D_{x\to y}X, y) = \delta_a(X, D_{y\to x}y)$*

Proof: ($|X|$ denotes the number of elements in X)

$$\delta_a(D_{x\to y}X, y)$$
$$= |\{x | x \in X \land y \in D_{x\to y}x\}|$$
$$= |\{x | x \in X \land x \in D_{y\to x}y\}|$$
$$= \delta_a(X, D_{y\to x}y) \quad \square$$

As branches do not overlap, $D_{f\to e}f$ is a set. From Lemma 1-2, the proof for Property 1 is given as follows:

$$\delta_a(D_{e\to j}D_{f\to e}f, j)$$
$$= \delta_a(D_{f\to e}f, D_{e\to j}^{-1}j)$$
$$= \delta_a(D_{f\to e}f, D_{j\to e}f)$$
$$= \delta_a(D_{j\to e}f, D_{f\to e}f) \quad \square$$

The proof for the second equation is the same.