

Construction of a Dictionary for Translating Japanese Phrases into One English Word

Satoshi Shirai, Kazuhide Yamamoto and Kazutaka Takao

ATR Spoken Language Translation Research Laboratories
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, JAPAN
E-mail: {shirai,yamamoto,ktakao}@slt.atr.co.jp

Abstract

In translation between languages that have different linguistic characteristics like Japanese and English, there are many cases in which contents are not correctly transmitted in the substitution from word to word. A method known to be effective as a measure for this is to determine the translations of verbs and nouns by using valency pattern pairs, which describe the semantic co-occurrences of verbs and nouns as valency patterns, and to pair them in the source language and the target language. However, this does not eliminate the problem of expressions regarded as ungrammatical (from the viewpoint of translation) being translated literally. In this research, we carried out analyses on expressions of compound Japanese nouns and verbs in correspondence with English words, by focusing on Japanese equivalent phrases to English words described in an English-Japanese dictionary. Consequently, there is hope that many cases of expressions created from one Japanese case element and verb corresponding to one English word can be obtained.

Keywords: Japanese-English machine translation, Collocation expression, English-Japanese dictionary, Valency pattern

1 Introduction

In the semantic analysis of machine translation, there is a need to correctly understand the co-occurrence relationships of words. One approach known to be effective, especially in translation between languages that have different linguistic characteristics like Japanese and English, is the use of valency pattern pairs, which direct attention to the meaningful co-occurrences of nouns and verbs to form correspondences of the basic structures of expressions in the source language and the target language. In the collection of such pattern pairs, however, there are concerned with problems both the description accuracy and the collection method.

Concerning the former problem, it is known that if the semantic attributes of nouns used as case elements were to be classified according to about 2,000 (or more) types of resolution accuracies for Japanese-English machine translation, valency pattern pairs having properly translated Japanese verbs could be described, excluding idiomatic expressions or special expressions [1]. Concerning the latter problem, we could refer to examples in dictionaries for human readers or self-observations, but it would be necessary to eventually collect about 25,000 examples of Japanese-English valency patterns [2]¹.

Basically speaking, there are limitations in improving the translation quality with a method making English expressions correspond to Japanese words, even if basic structures were changed from the viewpoint of valency. On the other hand, if there were many cases where the Japanese translation of an English word were equivalent to multi-

ple clauses, or contrarily, if it were possible to effectively correspond an expression comprising multiple clauses of Japanese to one English word, it would be possible for natural English sentences to be obtained.

Accordingly, in this paper, we aim at the construction of a dictionary that matches multiple clauses of Japanese with an English word, and as a part of this, with an English-Japanese dictionary, we carry out analyses on expressions where an English word and multiple clauses of Japanese correspond. In particular, we aggressively investigate cases where a case element and a verb in Japanese correspond to one English word, with the aim of enriching valency pattern pairs.

2 Word Composition of Japanese Translation Expressions to an English Word

In this work, we attempted analyses on Japanese translation expressions given by a key word in English, in an electronic English-Japanese dictionary (65,500 key words) [4]. Concretely speaking, we carried out morphological analyses on Japanese translations of English words, and then, we classified the part-of-speech sequences in those Japanese translations comprising multiple clauses and arranged the type and number of appearances. The total number of English-Japanese translation pairs recorded in the English-Japanese dictionary was 147,807 examples, and among them, 43,120 examples were analyzed as expressions of multiple clauses (Table 1).

In the classification of the part-of-speech sequences of the Japanese translation expressions, we decided to grasp a rough tendency first. For example, although the structures of word sequences differ from compound nouns and single nouns, it is easier to understand the tendency if we

¹To date, a little more than 14,000 examples have been disclosed [3].

Table 1: Number of clauses of Japanese translation expressions for an English word.

Number of Clauses	1	2	3	4	5	6	7	8
Number of Appearances	104,687	31,225	8,594	2,267	665	226	79	35
Number of Clauses	9	10	11	12	13	14	15	Sum total
Number of Appearances	13	8	2	2	1	1	2	147,807

Table 2: Part-of-speech sequences for higher numbers of appearances.

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ N-no / N /	3,784	自生の動物 (N, aboriginal), トウガラシの実 (N, capsicum)
2	/ N-o / V /	1,762	境を接する (I, abut), 目を覚ます (I, wake)
3	/ V / N /	1,487	対の一方 (N, doublet), 高めること (N, elevation)
4	/ N-ni / V /	1,184	空中にある (A, aerial), 憂うつにする (T, depress)
5	/ A / N /	1,132	注意深さ (N, carefulness), まぶしい光 (N, glare)
6	/ N-no / A /	916	害のない (A, harmless), 運のいい (A, lucky)
7	/ N-na / N /	864	すてきな人 (N, peach), 不可解な事 (N, riddle)
8	/ F / V /	706	そっとのぞく (I, peek), 必ず起こる (A, certain)
9	/ N-no / V /	550	品位のある (A, dignified), 気品のある (A, graceful)
10	/ N / V / N /	486	装飾すること (N, embellishment), 世界飛び歩き男 (N, globetrotter)
11	/ V-te / V /	402	すりむけて痛む (I, chafe), ひどく衰弱する (I, emaciate)
12	/ N-o / V / N /	376	影響を与えるもの (N, influence), 気持ちを和らげる音 (N, lull)
13	/ A / V /	336	詳しく述べる (T, enlarge), 強く触れる (T, hit)
14	/ F / V-ta /	308	きちんとした (A, methodical), まるまるした (A, plump)
15	/ N / A /	302	しつと深い (A, jealous), 申し分なく (F, satisfactorily)
16	/ N-ni / V-ta /	271	悲しみにくれた (A, disconsolate), ルールにかなった (A, fair)
17	/ V / A /	244	切りやすい (T, cut), 扱いにくい (A, ticklish)
18	/ N-ga / V /	231	耳が聞こえる (I, hear), 葉が作る (A, leafy)
19	/ N-no / N-na /	217	なぞのような (A, dark), 真綿のような (A, flossy)
20	/ V-ta / N /	215	困った事態 (N, mess), とがった岩 (N, needle)
21	/ N-o / V-ta /	202	上薬をかけた (A, glazed), 腹を立てた (A, ratty)

Table 2 (Supplement): Display of parts-of-speech.

Symbol	POS	Symbol	POS
N	Noun	T	Transitive Verb
V	Verb	I	Intransitive Verb
A	Adjective	(others)	Particle or Auxiliary verb
F	Adverb		

treat them equally in terms of whole nouns. Accordingly, when the same part-of-speech continued, it was collectively considered as one. However, we used the input declarations without modification, since features are lost when abstraction is done at the part-of-speech level with particles or auxiliary verbs. In the morphological analyses [5], supplementary expressions [in brackets] were excluded.

We sorted the obtained morphological analysis results and summed those part-of-speech sequences with a large number of appearances. Table 2 shows the summation results. The table shows a clause boundary with a “/”.

From Table 2, we can see a tendency of parts-of-speech with a large number of appearances. In the next section,

we look at cases including nouns and verbs from which highly frequent results appeared. The possibility is high that expressions containing nouns and verbs can be case elements and verbs respectively, which can potentially be considered as new valency pattern pairs.

3 Classification of Expressions Containing Nouns and Verbs

In this section, we take up the situation where nouns and verbs are included in the Japanese translations of English words. Considerations are made in the case where the case element modifies the verb and in the case where the verb modifies the noun. The former is a valency pattern itself,

and the latter can be treated based on a valency pattern since it can generally reduce a modified noun to a case element.

Among these, we classified the results with large numbers of cases into seven categories as shown in Table 3, and each tendency was explored. Then, we selected the part-of-speech sequences consisting of one verb and one or more nouns. Below, Tables 4 to 10 show the top results of each classification.

In Table 4, the Japanese case element and verb correspond to one English word. In a narrow sense –although there is a difference that when the English word is a verb, the Japanese verb is an inflection, and when the English word is an adjective, the Japanese verb is attributive– we believe that it is important to effectively grasp the relationship between the case element and verb in either case. Basically speaking, we want to describe them as a valency pattern pair. According to a trial describing case elements and verbs with 100 samples, we found that about half of them could be described as valency patterns. We estimated that we could get 4,000 new valency patterns.

In Table 5, many of the results seem to have problems in the translations of the English-Japanese dictionary from the viewpoint of Japanese-English translation. At a glance, the translations are explanation-like, and this is because many of the expressions do not appear in typical Japanese sentences. Consequently, these words would have little effect even if they were to be registered in a dictionary for Japanese-English translation. If they were registered, measures such as substitution to an expression likely to actually be used would be required.

In Table 6, the English part-of-speech is, on the surface, an adjective and an adverb, and the corresponding Japanese expression is attributive and continuous (particularly adverbial). When a modality has been added to an expression comprising a case element and verb, although we want to register a type where the adjective at the end is attributive in our valency pattern pair dictionary, with the present framework of pattern pairs, it is difficult to describe expressions accompanied by a subordinate clause. Future work will investigate correspondences to these.

In Table 7, the English part-of-speech is an adjective in general, and the Japanese verb is attributive. Like in Table 4, the Japanese case element and verb correspond to one English word. Basically speaking, we also want to describe them as a valency pattern pair. We estimated that we could get 600 new valency patterns.

In Table 8, similar to Table 7, the English part-of-speech is an adverb, and the Japanese auxiliary verb “-*naï*” comes to have the form of a clause. Like in Tables 4 and 6, the Japanese case element and verb correspond to one English word. Basically speaking, we also want to describe them as a valency pattern pair. We estimated that we could get 400 new valency patterns.

In Table 9, the English part-of-speech is an adverb, and the Japanese expression comes to have the form of a clause. The description of such an expression is difficult since the description of a valency pattern pair is considered to be at

the simple sentence level. For example, another framework (like [6]) might be required.

If we look at Table 10, it seems that there are a lot of cases capable of applying the views explained in Tables 4 to 9. Nos. 2, 4, 9, 10, and 15 are similar to the results ending with an auxiliary verb “-*ta* (-*da*)” in Table 7. However, because individually there are also complex word constructions and they are quite unnatural as Japanese expressions, future work will also investigate the handling of these.

4 Conclusion

By morphologically analyzing Japanese translations of an English-Japanese dictionary, we observed whole tendencies. When Japanese expressions corresponding to an English word were constructed from multiple clauses, we found quite a number of expressions that included a case element and verb. By processing these, we can expect the generation of valency patterns. According to the described trial, we will get 5,000 new valency patterns. In addition, by applying morphological analysis, we can automatically eliminate explanatory translations to some degree.

In the future, we plan to proceed with investigations on these and investigate methods of semi-automatically obtaining effective valency pattern pairs in Japanese-English translation from an English-Japanese dictionary.

References

- [1] Ikehara, S., M. Miyazaki & A. Yokoo (1993). “Classification of language knowledge for meaning analysis in machine translation”. *Trans. of Information Processing Society of Japan*, Vol. 34, No. 8, pp. 1692–1704 (in Japanese).
- [2] Shirai, S., S. Ikehara, A. Yokoo & H. Inoue (1995). “The quantity of valency pattern pairs required for Japanese to English machine translation and their compilation”. *Proc. of NLPRS-95*, pp. 443–448.
- [3] Ikehara, S., M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama & Y. Hayashi eds. (1997), and CD-ROM version (1999). “Goi-Taikai: A Japanese Lexicon”. Iwanami Shoten Publisher (in Japanese).
- [4] Yamagishi, K., T. Kodama & C. Kaise eds. (1996). “The Super Anchor English-Japanese Dictionary”. Gakken (in Japanese).
- [5] Matsumoto, Y., A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda & M. Asahara (1999). “Japanese Morphological Analysis System ChaSen version 2.0 Manual 2nd edition”. *NAIST-IS-TR99009*, Nara Institute of Science and Technology. URL=<http://chasen.aist-nara.ac.jp/>. (At present, a new version is available.)
- [6] Matsuo, Y., S. Shirai & S. Ikehara (1995). “Changing syntactic classes in transfer-based machine translation”. *Proc. of NLPRS-95*, pp. 432–437.

²Examples following the semicolon are examples checked off in Tables 4 to 10.

Table 3: Cases where nouns and verbs are included in the Japanese translations of English words.

No.	Category (Examples of J-POS Sequences) ²	Number of Cases (selected)	
1	Ending with a verb (e.g. /N-o/V/, /N-no/N-o/V/ ; /N/V/)	16,250	(8,018) see Table 4
2	Ending with a noun (e.g. /V/N/, /N-ni/V/N/ ; /N-no/N/)	12,858	(7,356) see Table 5
3	Ending with an adjective (e.g. /V/N-ga/A/, /N-o/V/A/ ; /N-no/A/)	2,472	(263) see Table 6
4	Ending with an auxiliary verb “-ta (-da)” (e.g. /N-ni/V-ta/, /N-o/V-da ; /F/V-ta/)	1,616	(1,168) see Table 7
5	Ending with an auxiliary verb “-nai” (e.g. /N-o/V-nai/ ; /F/N-de-nai/)	933	(697) see Table 8
6	Ending with a particle “-te (-de)” (e.g. /N-o/V-te/, /N-ni/V-te/ ; /F/V-te/)	576	(476) see Table 9
7	Others (e.g. /V/N-no/, /N/V-zu-ni/ ; /N-no/N-na/)	3,142	(1,372) see Table 10
Total		37,847	(19,350)

Table 4: Ending with a verb (frequency of 50 or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ N-o / V /	1,762	境を接する (I, abut), 目を覚ます (I, wake)
2	/ N-ni / V /	1,184	空中にある (A, aerial), 憂うつにする (T, depress)
3	/ N-no / V /	550	品位のある (A, dignified), 気品のある (A, graceful)
4	/ N-ga / V /	231	耳が聞こえる (I, hear), 葉が作る (A, leafy)
5	/ N / V-te / V /	190	警戒している (A, leery), 確信している (A, positive)
6	/ N-de / V /	176	道楽でする (A, amateur), 量で示す (T, quantify)
7	/ N-o / N / V /	153	酸を中和する (A, antacid), 本を拾い読みする (I, browse)
8	/ N-ni / N / V /	129	盲目的に心酔する (A, idolatrous), メートル法に移行する (I, metricize)
9	/ N-to / V /	128	ニャーと鳴く (I, mew), 不利となる (A, prejudicial)
10	/ N-no / N-o / V /	119	他の陰口を言う (T, backbite), 遺憾の意を表わす (A, regretful)
11	/ N-o / V-te / V /	99	顔を赤らめている (A, blushing), 足を引きずって歩く (T, drag)
12	/ N-ni / V-te / V /	82	地球に向かってる (A, earthbound), 薄片になってはげ落ちる (I, flake)
13	/ F / N / V /	74	よく成育する (I, flourish), ちょっと失敬する (T, pilfer)
14	/ V / N-o / V /	65	忍び込み窃盗を働く (I, burgle), 進むのをいやがる (A, restive)
15	/ N-na / N-o / V /	64	卑屈な態度をとる (I, crawl), ばかなことをする (I, fool)
16	/ N-no / N-ni / V /	60	進取の気性に富む (A, go-ahead), 人のために尽くす (I, oblige)
17	/ N-ni / N-o / V /	60	土中に穴を掘る (I, burrow), 社会に害毒を流す (A, pestilent)
18	/ F / N-o / V /	55	カタカタ音を立てる (I, clatter), ぐるりと向きを変える (I, double)

Table 5: Ending with a noun (frequency of 50 or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ V / N /	1,487	対の一方 (N, doublet), 高めること (N, elevation)
2	/ N / V / N /	486	装飾すること (N, embellishment), 世界飛び歩き男 (N, globetrotter)
3	/ N-o / V / N /	376	影響を与えるもの (N, influence), 気持ちを和らげる音 (N, lull)
4	/ V-ta / N /	215	困った事態 (N, mess), とがった岩 (N, needle)
5	/ N-ni / V / N /	175	夜間にうろつく人 (N, nightwalker), 頼みにする人 (N, reliance)
6	/ F / V / N /	142	当然伴う属性 (N, appanage), ずきずきする痛み (N, smart)
7	/ N / V-ta / N /	140	矛盾した行動 (N, inconsistency), 修復した建物 (N, renovation)
8	/ N-no / V / N /	116	矢の届く距離 (N, bowshot), つるのある植物 (N, vine)
9	/ N-ni / V-ta / N /	56	即座に作ったもの (N, improvisation), 真珠に似たもの (N, pearl)
10	/ V / A / N /	54	変わりやすさ (N, inconstancy), だましやす人 (N, pushover)
11	/ N-to / V / N /	52	名誉となる人 (N, credit), 中心となる人 (N, soul)
12	/ V-no / N /	52	作り替えの部分 (N, interpolation), 残りの物 (N, remainder)
13	/ V-te / V / N /	52	知っていること (N, acquaintance), 聞いてもらうこと (N, hearing)
14	/ F / V-ta / N /	51	ゆったりしたさま (N, easiness), よりすぐれた人 (A, superior)

Table 6: Ending with an adjective (frequency of ten or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ N / V / A /	45	操縦しやすい (A, handy), 名状しがたい (A, nameless)
2	/ V / N-no / A /	44	飽くことのない (A, insatiable), 並ぶものがない (A, only)
3	/ V / N / A /	15	抜けめなく (F, cutely), 尽きることなく (F, unfaillingly)
4	/ N / V-no / A /	10	肉づきのよい (A, beefy), 塩けのない (A, fresh)

Table 7: Ending with an auxiliary verb “-ta (-da)” (frequency of ten or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ N-ni / V-ta /	271	悲しみにくれた (A, disconsolate), ルールにかなった (A, fair)
2	/ N-o / V-ta /	202	上薬をかけた (A, glazed), 腹を立てた (A, ratty)
3	/ N-no / V-ta /	126	せっけんのはいった (A, soapy), 線の付いた (A, veined)
4	/ N-de / V-ta /	57	図案で表わした (A, figured), 単純でばかげた (A, naive)
5	/ N-to / V-ta /	39	朗々とした (A, sonorous), 断固とした (A, emphatic)
6	/ N-ga / V-ta /	39	芝居がかった (A, melodramatic), 用意ができた (A, ready)
7	/ N-ni / V-da /	28	創意に富んだ (A, inventive), 機知に富んだ (A, salty)
8	/ N-ni / N / V-ta /	19	十分に成長した (A, full-grown), 自説に固執した (A, opinionated)
9	/ F / N / V-ta /	17	やや格式ばった (A, semiformal), しっかり固定した (A, firm)
10	/ N-no / N-o / V-ta /	16	人間の姿をした (A, incarnate), 多くの辺を持った (A, multilateral)
11	/ N-o / V-da /	15	金を含んだ (A, auriferous), 人目を盗んだ (A, furtive)
12	/ N-no / N-ni / V-ta /	13	一定の方式によった (A, methodical), 父の死後に生まれた (A, posthumous)
13	/ N-no / N / V-ta /	12	筋肉の発達した (A, muscular), 戦争の結果生じた (A, war)
14	/ N-kara / V-ta /	12	聖書から出た (A, biblical), 歴史から見た (A, historical)
15	/ A / N / V-ta /	11	長い間続いた (A, agelong), 細かくカールした (A, crisp)
16	/ N-no / F / V-ta /	10	輪郭のはっきりした (A, clear-cut), 身なりのきちんとした (A, wellgroomed)

Table 8: Ending with an auxiliary verb “-nai” (frequency of ten or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ N-o / V-nai /	109	人を疑わない (A, confiding), 統制を受けない (A, free)
2	/ N-no / V-nai /	99	我慢のならない (A, impossible), 気の進まない (A, reluctant)
3	/ N-ni / V-nai /	85	問題にしない (T, sink), 目につかない (A, unnoticed)
4	/ N / V-te / V-nai /	46	活動していない (A, inactive), 熟練していない (A, unskilled)
5	/ N-ga / V-nai /	26	口がきけない (A, dumb), 目が見えない (A, sightless)
6	/ V / N-no / V-nai /	16	抑えることのできない (A, irrepressible), 欠くことのできない (A, needful)
7	/ N-o / V-te / V-nai /	13	帽子をかぶっていない (A, bareheaded), 公判を受けていない (A, untried)
8	/ N-ni / N / V-nai /	12	十分に活動しない (A, sluggish), 環境に適応しない (A, maladjusted)
9	/ N-mo / V-nai /	10	物も言えない (A, dumb), 想像も及ばない (A, inconceivable)

Table 9: Ending with a particle “-te (-de)” (frequency of ten or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ N-o / V-te /	156	いたずらをして (F, mischievously), 目的を持って (F, purposefully)
2	/ N-ni / V-te /	136	夢中になって (F, absorbedly), 海に向かって (F, seaward)
3	/ N-ga / V-te /	27	火がともって (A, alight), 気が狂って (F, dementedly)
4	/ N-no / N-o / V-te /	10	進取の気性を持って (F, enterprisingly), 多額の費用をかけて (F, expensively)

Table 10: Others (frequency of ten or more).

No.	J-POS Sequence	Freq.	Japanese Translation Example (E-POS, English Word)
1	/ V / N-ni /	94	むさぼるように (F, gluttonously), 至る所に (F, passim)
2	/ V / N-na /	93	疑うような (A, incredulous), 裂くような (A, splitting)
3	/ V / N-no /	79	荒れ模様の (A, dirty), なるつもりの (A, would-be)
4	/ N-o / V / N-na /	32	身を切るような (A, cutting), 耳をつんざくような (A, shrill)
5	/ V / N-de-aru /	32	刺すようである (I, prickle), 起こりそうである (I, threaten)
6	/ N-ni / V-ba /	28	適切に言えば (F, appropriately), 率直に言えば (F, candidly)
7	/ N / V / N-ni /	26	魅惑するように (F, bewitchingly), 浸透するように (F, pervasively)
8	/ N / V / N-no /	24	しりごみしがちの (A, backward), 矯正するための (A, reformatory)
9	/ F / V / N-na /	21	ぞっとするような (A, creepy), いまにも起こりそうな (A, momentary)
10	/ N / V / N-na /	16	ほれほれするような (A, adorable), 刺激するような (A, provoking)
11	/ V / N-de /	15	だますつもりで (F, deceitfully), 至る所で (F, over)
12	/ V-ta / N-ni /	14	ばかげたことに (F, absurdly), 聞いたそうに (F, inquiringly)
13	/ N-o / V-nai-de /	11	時機を得ないで (F, malapropos), 家賃を払わないで (F, rent-free)
14	/ N-o / V / N-ni /	10	疑惑を招くように (F, suspiciously), 危機を告げるように (F, threateningly)
15	/ N-no / V / N-na /	10	気のめいるような (A, oppressive), 幽霊の出そうな (A, spooky)
16	/ N-ni-mo / V-zu /	10	それにもかかわらず (F, notwithstanding), それにもかかわらず (F, still)
17	/ V-ta / N-no /	10	閉じたままの (A, unopened), 青みがかった黒色の (A, blue-black)