A Method of Adding New Entries to a Valency Dictionary by Exploiting Existing Lexical Resources

Sanae Fujita and Francis Bond
NTT Communication Science Laboratories,
Nippon Telegraph and Telephone Corporation
2-4 Hikari-dai, Seika-cho, Soraku-gun,
Kyoto, 619-0237, JAPAN
{sanae, bond}@cslab.kecl.ntt.co.jp

Abstract

Information on subcategorization and selectional restrictions in a valency dictionary is very important for natural language processing in tasks such as monolingual parsing, accurate rule-based machine translation and automatic summarization. However, adding this detailed information is both time consuming and costly.

In this paper we present a method of assigning valency information and selectional restrictions to entries in a bilingual dictionary, based on information in an existing valency dictionary. The method is based on two basic assumptions: words with similar meaning have similar subcategorization frames and selectional restrictions; and words with the same translations have similar meanings. Based on these assumptions, new valency entries are constructed for words in a plain bilingual dictionary, using entries with similar Japanese meaning and the same English translations. The measurement of similarity in Japanese is done using paraphrased examples, so that non-expert native speakers can carry out the task.

An initial evaluation of 171 new patterns showed that adding them to a Japanese-to-English machine translation system improved the translation for 31% of sentences using these verbs, and degraded it for 8%, a clear improvement in quality.

1 Introduction

In this paper we present a method of adding new entries to a bilingual valency dictionary. It is well known that detailed information about verb valency (subcategorization) and selectional restrictions is useful both for monolingual parsing and selection of appropriate translations in machine translation. As well as being useful for resolving parsing ambiguities, verb valency information is particularly important for complicated processing such as identification and supplementation of zero pronouns (Nakaiwa et al. 1996).

Shirai (1999) estimates that at least 27,000 valency entries are needed to cover around 80% of Japanese verbs in a typical newspaper. Various methods of creating detailed entries have been suggested, such as the extraction of candidates from corpora (Haruno & Yamazaki 1996; Manning 1993; Utsuro et al. 1997), and the automatic and semi-automatic induction of semantic constraints (Akiba et al. 1995; Akiba et al. 2000). However, the automatic construction of monolingual entries is still far from reaching the quality of hand constructed resources. Further, large bilingual resources are still rare enough that it is much harder to automatically build bilingual entries.

Table 1: Cover Ratio for Japanese Newspaper (Nihon Keizai Shimbun '95)

	No. of Types	(%)	No. of Tokens	(%)
Pattern Exists	4,773	47.4 %	3,691,572	93.2 %
No Pattern	5,287	52.6 %	267,620	6.8 %
Total	10,060	100 %	3,959,192	100 %

Our work differs from Manning (1993) in that we are using existing lexical resources rather than a corpus, and we are obtaining selectional restrictions as well as subcategorization frames. Thus our method will work for rare words, so long as we can find them in a bilingual dictionary, and know the English translation. It does not, however, learn new frames from usage examples.

Our method adds new patterns by leveraging existing knowledge in the system dictionaries. We will illustrate the method with examples of building a Japanese-English lexicon, but there is nothing in the method itself that is language specific. The basic idea is to add new entries to the pattern dictionary by using Japanese-English pairs from a plain bilingual dictionary (without detailed information about valency or selectional restrictions), and build new entries for them based on existing entries.

The implementation presented in this paper is semi-automatic, but it does not rely on detailed knowledge of the system dictionaries by the analyst: the only judgments required are acceptability judgments of paraphrases. Our method is similar in principle to Ikehara et al. (1995) who add useful information to a user dictionary by comparing input word pairs to existing entries in the system dictionary.

2 ALT-J/E's Valency (Pattern) Dictionary

We use the valency dictionary from the Japanese-to-English machine translation system **ALT-J/E** (Ikehara et al. 1991). The basic structure of a clause comes from the relationship between the main verb and nouns. The structure transfer dictionary provides this basic clause structure.

ALT-J/E provides 13,000 patterns for the common structure transfer dictionary and 3,000 patterns for the idiomatic structure transfer dictionary. In the common structure transfer dictionary, there is an average of 2.3 patterns for each verb.

The coverage of verbs appearing in one year of a Japanese newspaper (Nihon Keizai Shimbun '95) is given in Table 1. The coverage of tokens is high, but many rare verb types are missing from our system (over 52% of verb types have no pattern). A lot of verbs which appear about 100 times have no pattern. But many common words only appear once, for example 復讐する *fukushuu-suru* "revenge", 微笑する *bishou-suru* "smile", 暖まる *atatamaru* "get warm", 決闘する *kettou-suru* "duel", etc. And because there is an average of 3.1 verbs for each sentence, one verb in every 5 sentences has no pattern. In order to reduce the number of unknown verbs to one in 10 sentences, we need to add valency information to 2,647 verbs.

Figures 1 and 2 show examples of the common structure transfer dictionary. Each

predicate has one or more arguments associated with it, and links between the two languages are marked with indices (N1, N2, ...). Each case-slot has information such as grammatical function, case-marker, case-role (the index number gives the case role), semantic restrictions on the filler and default order (not all the features are shown in the examples). Most arguments are NPs or PPs, but it is possible to have a sentential argument, as in Figure 2, where it is marked with S10. The arguments correspond between Japanese and English, thus giving the backbone of the transfer. It is possible for an argument to only appear on one side, this is useful for verbs in one language that incorporate information given explicitly in the other.

The formalism does not allow the alternation of a NP and sentential argument, so verbs of information-transfer will typically have two similar patterns.

```
Pattern ID: 202969              U_SENT  (action)
┌ N1 (agent) "ga"               ├ PREDICATE - VERB "report"
├ N2 (abstract) "o"             ├ CASE S N1
├ N3 (agent) "ni"              ├ CASE DO N2 OBJ-form
└ 上申する joushin-suru          └ CASE PP
                                  └ U_PP "to" N3 OBJ-form
```

Figure 1: Lexical Entry for the verb *joushin-suru* ⇔ *report* No.1 (SVOP)

```
Pattern ID: 202970              U_SENT  (action)
┌ N1 (agent) "ga"               ├ PREDICATE - VERB "report"
├ N3 (agent) "ni"              ├ CASE S N1
├ S10 (*) "to"                 ├ CASE PP
└ 上申する joushin-suru          │ └ U_PP "to" N3 OBJ-form
                                └ CASE DO
                                  └ CLAUSE that
                                    └ S10
```

Figure 2: Lexical Entry for the verb *joushin-suru* ⇔ *report* No.2 (SVPC)

We call the combination of case-role and case-marker the **slot-type**. A verb's basic argument type is given by the combination of slot-types it allows. For example: N1:**agent**+ga is one slot-type, N2:**object-1**+o is another, and their combination is the basic transitive **frame-type**: N1:**agent**+ga, N2:**object-1**+o.

Yokoo et al. (1994) organized the English structure into a skeleton (roughly corresponding to a case-frame) and its flesh (which includes the word specific information). For example, the verb *report* in Figure 1 reduces to the skeleton `ditransitive-1` plus the flesh {*report*, N1, N2, *to*, N3}.

Thus the frame-types effectively correspond to English skeletons, although they are slightly more restrictive, as they include the case-roles. In **ALT-J/E**'s common structure dictionary, there are 201 slot-types, 569 frame-types and 493 skeleton types.

Table 2: Cover Ratio for Patterns of Skeleton and Frame-Types

Cover Ratio(%)	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Skeleton Types	1	1	2	3	4	5	8	15	41	493
Frame Types	1	1	2	2	3	4	7	17	50	569

The percentage of patterns that can be described with a given number of skeleton or frame types is given in Table 2. Over 50% of the verbs can be covered by a few subcategorization patterns (mainly basic transitive and intransitive). At the other end of the scale, 317 patterns have unique frame-types, and 251 patterns have unique skeletons.

Because the Japanese slot-type combinations have not been treated as fixed case-frames, there are many minor variations, such as N1+ga with N3+ni and N1+ga with N3+ni/e which are treated as different. In most cases, these are unmotivated distinctions, and it would be advantageous to combine them, as suggested by Nomura & Muraki (1996) and Baldwin et al. (1999). This would serve to reduce the number of different frame-types.

As an example of the utility of the valency information, we give an example of a sentence translated with the system default information (basically a choice between transitive and intransitive), and the full valency information. The verb is 下命する *kamei-suru* "order", which takes a sentential complement. In (1)[1] the underlined part is the sentential complement. The verb valency entry is the same as 上申する *joushin-suru* "report" given in Figure 2, except with the clause marked as *to*-infinitival.[2] The translation with the valency information is far from perfect, but it is comprehensible. Without the valency information, the translation is incomprehensible.

(1) 国王　　は　家来　　　に　<u>出撃しろ、</u>　　と　　下命した。
    koku-ou wa kerai   ni shutsugeki shiro, to kamei-shita.
    king   TOP subordinate DAT sally forth   QUOT ordered

"The king ordered his follower to sally forth."

with: `The king ordered a follower` <u>`that sallied forth`</u>`.`

without:　`* ordered to a follower that the king,` <u>`sallied forth`</u>`.`

In general, translation tends to simplify text. Therefore, in many cases, a single English entry is the translation of many similar Japanese patterns. For example, there are 23 Japanese predicates linked to the English entry *report*. Six of these have the same frame-type as that shown in Figure 1. Five patterns have the frame-type shown in Figure 2. Three more link to a variation of that in Figure 1 where N3+ni is replaced by N3+$ni/e/made$. Collapsing such minor variations, 11 are of one type, 7 of the

---

[1]We use the following abbreviations: TOP: topic postposition; DAT: dative postposition; QUOT: quotative postposition.

[2]The subordinate class is incorrectly translated as a *that*-clause. This is a bug in the English generation, the Japanese parse and semantic structure are correct.

other, and only 2 are genuinely different. Therefore, in order to make new frames for predicates that translate into English *report*, we need to add only two patterns, one of the types in Figure 1 and one in Figure 2. Ideally, we should combine these into a single **alternation** and link to that as suggested in Baldwin et al. (1999).

3   Experimental Method

3.1   Method of Making New Patterns

Our method is based on two facts: (1) verbs with similar meanings typically have similar valency structures; (2) verbs with identical translations typically have similar meanings.

We use three resources: (1) **ALT-J/E**'s valency dictionary; (2) a plain Japanese-English bilingual dictionary which contains Japanese-English word pairs without valency information (such as **ALT-J/E**'s word dictionary); and (3) a Japanese monolingual corpus (such as newspaper text).

Our method creates valency patterns for words in the word dictionary (or any bilingual dictionary) whose English translations can be found in the valency dictionary. We extract these combinations. Each combination consists of J_U, an Unknown word for which we have no valency information, E, its English translation (or translations), for which we have one or more valency patterns linked to J_V Valency patterns in the valency dictionary.

Next, we extract example sentences which use J_V from our corpus and replace J_V with J_U. We also do the reverse, replacing J_U with J_V.

Third, we compare the original sentence with the paraphrased sentences. This requires a monolingual human operator. The analyst must decide: "Are the paraphrased sentences grammatical?", that is do the verbs have the same valency; "Do both sentences have the same meaning?", that is, do the verbs have equivalent meanings; if the meaning has changed, we ask "Are the meanings close?".

If the paraphrased sentences are grammatical and have identical or close meanings, we make a similarity link between J_V and J_U and make J_U's pattern from J_V's pattern. The new pattern for J_U has same Japanese/English structure as that for J_V.

We classify the similarity links into the following: `same`, `close`, [J_U] `narrower`, [J_U] `broader`, `overlapping`. If both J_V to J_U and J_U to J_V are grammatically interreplaceable and have the same meaning, the link is `same`. If both J_V to J_U and J_U to J_V can grammatically replace each other and have close meanings the link is `close`. If J_U can grammatically replace J_V and has nearly the same meaning, but J_V cannot replace J_U, the link is (J_U is) `narrower`. If J_V can grammatically replace J_U and has nearly the same meaning but J_U cannot replace J_V, the link is (J_U is) `broader`. If J_V and J_U can have the same meaning in some specific contexts, then the link is `overlapping`.

For example, for the unknown Japanese word J_U 具申する *gushin-suru* "report" we look at the existing word J_V 上申する *joushin-suru* "report" which exists in the valency dictionary, with the same English translation.

We extract some sentences from our corpus which use J_V, for example (2; slightly simplified here), and replace J_V with J_U (3).

(2) 経営　　　　トップ に　この こと　を　上申し、　　ＯＫ が　　出た。
Keiei　　　　toppu ni　 kono koto o　 jyoushin-shi, OK ga　 deta.
management top　　DAT this　thing ACC report,　　　ok　NOM came-out

"I <u>reported</u> this to the top management and they OKed it."

(3) 経営　トップ に この　こと　を 具申し、　　ＯＫ が 出た。
Keiei toppu ni kono koto o gushin-shi, OK ga deta.

Similarly, we extract some sentences from our corpus which use J_U, for example (4), and replace J_U with J_V (5).

(4) 罰則　　　　　　を　重く　する 必要　　は　 ない　　と　 具申した。
bassoku　　　　 o　 omoku suru hituyou ha　 nai　　 to　 gushin-shita.
penal-regulations ACC severe do　 need　　TOP nothing QUOT reported.

"I <u>reported</u> that there is no need to make the penal regulations more severe."

(5) 罰則　　を 重く　　する 必要　　は ないと 上申した。
bassoku　o　omoku suru hituyou ha nai　to jyoushin-shita.

Both paraphrases (3 and 5) are grammatical and both pairs (2, 3) and (4, 5) have very similar meanings. So we make a close link between 上申する *joushin-suru* "report" and 具申する *gushin-suru* "report", and make a new pattern for 具申する *gushin-suru* "report" from the pattern of 上申する *joushin-suru* "report".

In this paper, we attempt to build patterns for all four link types, although we realize that the restrictions will be too loose in some cases. In particular, when J_U is narrower or broader than J_V, there are problems in treating them as the same.

3.2 Experiment

In **ALT-J/E**'s Japanese-English word dictionary, there are 55,615 J-E pairs whose Japanese part of speech is adjective, adjectival noun or verb. There are a total of 20,925 distinct Japanese entries. However, due to the cost of making detailed entries, only those 4,937 entries have patterns: 15,988 entries have no pattern. Of the 55,615 J-E pairs, 35,999 have no entry in the pattern dictionary. Our method is applicable to 13,408 of these pairs: their English entry has an entry in the pattern dictionary.

In Table 1, we showed that 5,287 kinds of verbs have no pattern. Of those, 2,165 (40.9 %) verbs have the possibility of adding valency information by the proposed method using **ALT-J/E**'s Japanese-English word dictionary.

For our initial evaluation, we restricted ourselves to looking at verbs that matched verbs which take clause complements in the valency dictionary. We took them as our test case because they are particularly hard to translate correctly without the detailed valency information. There are 268 such verbs in the valency dictionary. These were not particularly frequent verbs, they were only used in 2,502 sentences in the 1995 Nihon Keizai Shimbun. On the other hand, a native speaker of Japanese judged that

none of them were particularly rare, an educated native speaker would be expected to know all of them.

It took the lexicographer about 7 minutes per verb to judge the fitness of the paraphrases, all the rest of the construction was automatic. This a significant speed up over the 30 minutes normally taken by an expert lexicographer. We show the number of patterns made in Table 3. Our basic aim was to make patterns with a sentential complement, where possible we also made patterns with an equivalent NP complement. Note that all the patterns are made by copying from existing entries, at no time do we build new frame-types.

Table 3: Number of New Patterns

Verb Type	No. of patterns	No. of distinct Japanese verbs
With sentential complement	92	65
With NP complement	79	49
Total	171	65

There are 7 frame-types with no sentential complements: 30 plain ditransitives, 18 transitives, 5 ditransitives with a different case-role marking, and four other types.

4   Evaluation and Results

For the evaluation, we picked two sentences at random for each J_U in the 1995 Nihon Keizai Shimbun corpus. This corpus had not been used in the paraphrasing stage, i.e., all the sentences were unknown. We then translated them with and without the new valency entries. We only found test sentences for 58 of the 65 Japanese words for which we created patterns (we only found one sentence for some verbs).

The two translated results were compared by a native speaker of English who is fluent in Japanese. The differences were placed into six categories: *improved, slightly improved, no change, equivalent, slightly degraded, degraded*. All the judgements were based on the change in translation quality, not the absolute quality of the entire sentence. We evaluated on raw newspaper text, and as we were looking at verbs with sentential complements, most sentences were long and complex: the kind of sentences that really stress a machine translation system.

The results of the evaluation are given in Table 4. A majority of sentences showed *no change*, followed by `equivalent` (61.3%). 30.6% of the translations were either *improved* or *slightly improved*. Only 8.1% of the sentences were *degraded* or *slightly degraded*. There was a clear improvement in the overall translation quality.

5   Discussion

In this section we discuss the reasons for the improvements and degradations.

First the improvements. In general, the parsing accuracy improved. There was a major improvement in the handling of zero pronouns. Most sentences in newspaper

Table 4: Results of Translating with Newly Created Valency Entries

Judgement	No.	%	Explanation
improved	10	9.0%	improved structure, tense, choice of translation
slightly improved	24	21.6%	improved supplementation of zero pronouns
no change	41	36.9%	no change
equivalent	27	24.3%	changed, but same quality
slightly degraded	8	7.2%	better parse but worse choice of translation
degraded	1	0.9%	worse choice of translation
Total	111	100.0%	

articles lack at least one element, and the rules **ALT-J/E** uses to supplement them, or choose defaults, rely heavily on accurate valency information.

We show some examples of the changed translations, using simplified example sentences.

In (6) the English valency information supplies the subcategorized preposition *for* in *wish for*. The default translation makes the argument a plain direct object, which is ungrammatical for *wish*.

(6) 国民　　の 大半　 が 平和　 を 欲し、 その ため の 危険　 を 負う 覚悟　　 が
kokumin no taihan ga heiwa o hosshi, sono tame no kiken o ou   kakugo ga
ある と 信じて　 きた。
aru to shinjite kita.

without: It was believed that national most <u>wished</u> peace and that there
was the preparedness that we owe danger for that purpose to.

with: It was believed that national most <u>wished for</u> peace and that
there was the preparedness that we owe danger for that purpose to.

In (7), the translation selected is an improvement.

(7) ＮＡＴＯ は セルビア人 に 宣戦　　 を 布告した。
NATO   ha Serbia-jin ni sensen o fukoku-shita.

without: NATO <u>decreed</u> a declaration of war to Serbia person.

with: NATO <u>announced</u> a declaration of war to Serbia person.

Finally, in (8) the supplementation, parse, and adverb scope are all improved.

(8) 静か　　 に 外交　　 当局　　　 の 間　 で 話し合い　 を 進めて　　 ほしい、 と
shizuka ni gaikou toukyoku no aida de hanashiai o susumete hoshii   to
ロシア の 外相　　 に 下命して　　 いた。
Russia no gaisyou ni kamei-shite ita.

without: <u>*</u> ordered to Russia's Minister of Foreign Affairs that it
proceeded with a conference <u>quietly</u> between diplomatic authorities.

> with: <u>It was</u> ordering Russia's Minister of Foreign Affairs <u>quietly</u>
> that it wanted to be proceeded with a conference between diplomatic
> authorities.

Next, we discuss the 8.1% of sentences which degraded in quality. Most of the degradations were due to differences in the selection of the predicate to be translated. When there is no pattern available, **ALT-J/E** uses a translation from its word dictionary. As there is little information available to choose between alternatives, the first listed translation is used (the first listed translation is meant to be the most general translation). However, when we made patterns, we looked at all listed translation equivalents. 51% of the time we were able to make a pattern with the first listed translation, 27% with the second, 11% with the third, and 11% with the fourth, fifth or sixth.

When a pattern exists **ALT-J/E** uses it in preference to an entry in the word dictionary. Therefore, the translation changed for many entries. Sometimes the new translation was an improvement, but sometimes it was not. For example, 口答えする *kuchi-kotae-suru* "answer back" had two translations in the word dictionary: (1) *answer back* and (2) *retort*. We could only find a pattern for *retort* however, and so this became the system's choice. However, *answer back* was in fact a better translation in the examples we saw.

Overall, our method of fully exploiting existing resources was able to create many useful valency entries, but caused some degradations where we could not add entries for the most appropriate translations.

6 Future Work

We would like to build patterns for the remaining 13,140 pairs in our word dictionary. Given enough resources, we would like to also look at adding words from other plain bilingual dictionaries. A preliminary examination of **edict** (a downloadable J-E lexicon: Breen (1995)) found over 25,000 entries which our method is applicable to.

There are three major ways to improve our method of adding new entries: (1) fully automating it; (2) adding alternations to the existing valency dictionary so we can construct multiple patterns in one step; (3) further refining the new patterns by tuning their selectional restrictions.

If we want a tool to quickly add new lexical entries, even though the user is not a system expert, then we should try to fully automate the process. To do this, we need a method of judging the similarity of two words. Kasahara et al. (1997) judge similarity using a concept-base built from both machine readable dictionaries and corpora. This could potentially be used to automate the selection currently being done by a lexicographer judging paraphrases.

At the moment, there are gaps in our existing lexicon where one pattern of an alternation exists, but the other does not. For example, for verbs that can be translated as *entreat*, collapsing minor variations, 3 patterns had the same frame-type as Figure 1, and 2 patterns had the same frame-type as Figure 2. 嘆願する *tangan-suru* "entreat" and 懇願する *kongan-suru* "entreat" have both frame-types, but 懇望する *konmou-suru* "entreat" has only the frame-type of 1. This is a gap in our lexicon, ideally 懇望す

る *konmou-suru* "entreat" should also have a frame-type like in Figure 2. We should fill in these gaps before using these verbs as exemplars to add new entries. The most obvious way to do so would be to reorganize the lexicon into the alternation based architecture proposed by Baldwin et al. (1999). Further, if we could properly identify these as alternations, then there would be no need to build two patterns, just a single alternation. We are currently investigating ways to do this automatically (Baldwin & Bond 2002).

In the current implementation, we make no use of the lexicographers' different kinds of similarity judgments. However, if a verb has a broader or narrower meaning, then the selectional restrictions need to be adjusted to fit this. Akiba et al. (2000) has shown that it is more efficient to adjust selectional restrictions from a reasonable start than to start from the most or least possible restrictions. By choosing a similar verb's entry as starting point, and handing the constructed pattern on to a another round of semi-automatically adjusting the restrictions, we can further improve the quality of the system.

Finally, our method currently uses a very restrictive definition of "similar meaning": having the same English translation. We could try to replace this with some other similarity measure, either based on usage (if we can find enough examples), or on direct similarity in the source language, target language or both. This would increase our coverage, although almost certainly at the cost of lowering the quality of the created patterns.

7   Conclusion

In this paper we present a method of assigning valency information and selectional restrictions to entries in a bilingual dictionary. The method exploits existing dictionaries and is based on two basic assumptions: words with similar meaning have similar subcategorization frames and selectional restrictions; and words with the same translations have similar meanings.

A prototype system allowed new patterns to be built at a cost of 7 mintues per pattern. An initial evaluation of 171 new patterns showed that adding them to a Japanese-to-English machine translation system improved the translation for 31% of sentences using these verbs, and degraded it for 8%, a substantial improvement in quality.

References

Akiba, Yasuhiro, Megumi Ishii, Hussein Almuallim & Shigeo Kaneda: 1995, 'Learning English verb selection rules from hand-made rules and translation examples', in *Sixth International*

Conference on Theoretical and Methodological Issues in Machine Translation: TMI-95, Leuven, pp. 206–220.

Akiba, Yasuhiro, Hiromi Nakaiwa, Satoshi Shirai & Yoshifumi Ooyama: 2000, 'Interactive generalization of a translation example using queries based on a semantic hierarchy', in *ICTAI-00*, pp. 326–332.

Baldwin, Timothy & Francis Bond: 2002, 'Alternation-based lexicon reconstruction', in *Ninth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-2002*, Keihanna, Japan, (this volume).

Baldwin, Timothy, Francis Bond & Ben Hutchinson: 1999, 'A valency dictionary architecture for machine translation', in *Eighth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-99*, Chester, UK, pp. 207–217.

Breen, Jim: 1995, 'Building an electronic Japanese-English dictionary', Japanese Studies Association of Australia Conference (`http://www.csse.monash.edu.au/~jwb/jsaa_paper/hpaper.html`).

Haruno, Masahiko & Takefumi Yamazaki: 1996, 'High-performance bilingual text alignment using statistical and dictionary information', in *34th Annual Conference of the Association for Computational Linguistics*, pp. 131–138.

Ikehara, Satoru, Satoshi Shirai, Akio Yokoo, Francis Bond & Yoshie Omi: 1995, 'Automatic determination of semantic attributes for user defined words in Japanese-to-English machine translation', *Journal of Natural Language Processing*, **2**(1): 3–17, (in Japanese).

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

Kasahara, Kaname, Kazumitsu Matsuzawa & Tsutomu Ishikawa: 1997, 'A method for judgment of semantic similarity between daily-used words by using machine readable dictionaries', *Transactions of IPSJ*, **38**(7): 1272–1283, (in Japanese).

Manning, Christopher D.: 1993, 'Automatic acquisition of a large subcategorization dictionary from corpora', in *31st Annual Meeting of the Association for Computational Linguistics: ACL-93*, pp. 235–242.

Nakaiwa, Hiromi, Satoshi Shirai & Satoru Ikehara: 1996, 'Anaphora resolution of Japanese zero pronouns with deictic reference', in *16th International Conference on Computational Linguistics: COLING-96*, Copenhagen, pp. 812–817.

Nomura, Naoyuki & Kazunori Muraki: 1996, 'An empirical architecture for verb subcategorization frame', in *16th International Conference on Computational Linguistics: COLING-96*, Copenhagen, pp. 640–645.

Shirai, Satoshi: 1999, 'Tanbun-no ketsugō patān-no mōra-teki shūshū-ni mukete [towards a comprehensive cover of sentence valency patterns]', in *NLP Symposium*, (In Japanese: `www.kecl.ntt.co.jp/icl/mtg/members/shirai/nlpsym99.html`).

Utsuro, Takehito, Takashi Miyata & Yuji Matsumoto: 1997, 'Maximum entropy model learning of subcategorization preference', in *Proc. of the 5th Workshop on Very Large Corpora*, pp. 246–260.

Yokoo, Akio, Hiromi Nakaiwa, Satoshi Shirai & Satoru Ikehara: 1994, 'Skeleton-flesh type semantic structure dictionaries for Japanese-to-English machine translation', in *48th Annual Convention of the IPSJ*, vol. 6Q-8, pp. 3:139–140, (In Japanese).