Extending the Coverage of a Valency Dictionary

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

Abstract

Information on subcategorization and selectional restrictions 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 to a valency dictionary 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 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 source-language meaning and the same target-language translations. We evaluate the effects of various measures of similarity.

1 Introduction

One of the severest problems facing machine translation between Asian languages is the lack of suitable language resources. Even when word-lists or simple bilingual dictionaries exist, it is rare for them to include detailed information about the syntax and meaning of words.

In this paper we present a method of adding new entries to a bilingual valency dictionary. New entries are based on existing entries, so have the same amount of detailed information. The method bootstraps from an initial hand built lexicon, and allows new entries to be added cheaply and effectively. Although we will use Japanese and English as examples, the algorithm is not tied to any particular language pair or dictionary. The core idea is to add new entries to the valency 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.

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. However, this information is not encoded in normal human-readable dictionaries, and is hard to enter manually. Shirai (1999) estimates that at least 27,000 valency entries are needed to cover around 80% of Japanese verbs in a typical newspaper, and we expect this to be true of any language. Various methods of creating detailed entries have been suggested, such as the extraction of candidates from corpora (Manning, 1993; Utsuro et al., 1997; Kawahara and Kurohashi, 2001), and the automatic and semi-automatic induction of semantic constraints (Akiba et al., 2000). However, the automatic construction of monolingual entries is still far from reaching the quality of hand-constructed resources. Further, large-scale bilingual resources are still rare enough that it is much harder to automatically build bilingual entries.

Our work differs from corpus-based work such as Manning (1993) or Kawahara and Kurohashi (2001) in that we are using existing lexical resources rather than a corpus. 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.

In order to demonstrate 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" [NP-*ga* Cl-*to* V], 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
koku-ou wa kerai ni
king TOP subordinate DAT
出撃 しろ、 と 下命した。
shutsugeki shiro, to kamei-shita.
sally forth QUOT ordered

"The king ordered his follower to sally forth."

with: `The king ordered a follower that sallied forth.`

without: `* ordered to a follower that the king, sallied forth.`

In general, translation tends to simplify text, because the target language will not be able to represent exactly the same shades of meaning as the source text, so there is some semantic loss. Therefore, in many cases, a single target language entry is the translation of many similar source patterns. For example, there are 23 Japanese predicates linked to the English entry *report* in the valency dictionary used by the Japanese-to-English machine translation system **ALT-J/E** (Ikehara et al., 1991).

2 Experimental Method

The approach is based on that of Fujita and Bond (2002). For the explanation we assume that the source language is Japanese, and the target language is English, although nothing depends on this.

2.1 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) a seed valency dictionary (in this case the verbs from **ALT-J/E**'s valency dictionary, ignoring all idiomatic and adjectival entries — this gave 5,062 verbs and 11,214 valency patterns[3]) ; (2) a plain bilingual dictionary which contains word pairs without valency information (in our case a combination of **ALT-J/E**'s Japanese-English word transfer dictionary and EDICT (Breen, 1995)); and (3) a source language corpus (mainly newspapers).

Our method creates valency patterns for words in the bilingual dictionary whose English translations can be found in the valency dictionary. We cannot create patterns for words with unknown translations. Each combination consists of J_U, an Unknown word for which we have no valency information; E, its English translation (or translations); which is linked to one or more valency patterns J_V in the valency dictionary. Figure 1 shows the overall flow of creating candidate patterns.

For each entry in the plain J-E dictionary

- If no entries with the same Japanese (J_U) exist in the valency dictionary
  - For each valency entry (J_V) with the same English (E)
    - Create a candidate pattern consisting of J_V replaced by J_U

Figure 1: Creating Candidate Patterns

The definition of "similar meaning" used to generate new patterns is that they have the same English translation. We had to make this quite loose: any entry with the same En-

[1] We use the following abbreviations: TOP: topic postposition; ACC: accusative postposition; DAT: dative postposition; QUOT: quotative postposition; NP: noun phrase: Cl: clause; V: verb.

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

[3] We call an entry in the valency dictionary (consisting of source and target language subcategorization information and selectional restrictions on the source side) a valency pattern.

glish head. Therefore *give up* and *give back* are counted as the same entry. This allows for minor inconsistencies in the target language dictionaries. In particular the valency dictionary is likely to include commonly appearing adjuncts and complements that do not normally appear in bilingual dictionaries. For example: *iku* "go" is translated as *to go* in EDICT, *go* in the **ALT-J/E** word transfer dictionary and *NP go from NP to NP* in the **ALT-J/E** valency dictionary (among other translations). To match these entries it is necessary to have some flexibility in the English matching.

In order to filter out bad candidates, we compare the usage of J_V with J_U using examples from a corpus. Two judgments are made for each paraphrase pair: is the paraphrase grammatical, and if it is grammatical, are the meanings similar? This judgment can be done by monolingual speakers of the source language. This is done in both directions: first we find example sentenes using J_U, replacing J_U with J_V and compare the paraphrased sentences, then we find sentences for valence patterns using J_V, replace with J_U and judge the similarity. Figure 2 shows the comparison using paraphrases.

In the implementation, we added a pre-filter: `reject`, for verb pairs that obviously differed in meaning. This allowed the analysts to immediately reject verb pairs (J_U-J_V) that were obviously not similar, and speeded up things considerably (e.g., J_U 決着する *ketchaku-suru* "settle on" has translation E *settle*; both J_{V1} 落ち着く *ochitsuku* "calm down" and J_{V2} 定住する *teijuu-suru* "settle in" are candidates but J_{V2} appear because of the polysemy of E.[4]).

The three grammaticality classes are: `grammatical`, `ungrammatical`, `grammatical in some context`.[5] Semantic similarity was divided into the following classes:

- `same`: J_U 脅す *odosu* "threaten" and J_V 威す *odosu* "threaten"
- `close`: J_U 具申する *gushin-suru* "report" and J_V 上申する *joushin-suru* "report"

[4] J_{V1} is equivalent to wordnet sense 8 "become quiet", J_{V2} is sense 4 "take up residence and become established" (Fellbaum, 1998).

[5] The analysts also rejected 7.9% of the example sentences as irrelevant. These were sentences where the verb did not actually appear, but that had been selected due to errors in the morphological analysis.

For each candidate pattern J_U-E (from J_V-E)

- If J_U is obviously different to J_V
  `reject`
- Extract 5 sentences using J_U from the corpus
  For each sentence
  - Replace J_U with J_V
  - Classify the paraphrased sentence into 3 grammaticality classes
    if the class is `grammatical`
    - Classify the semantic similarity into 6 classes
- Extract 5 sentences using each pattern of J_V from the corpus
  - Replace J_V with J_U
  - Test as above

Figure 2: Paraphrasing Check

- [J_U] `broader`: J_U 作り出す *tsukuri-dasu* "create" and J_V 発明する *hatsumei-suru* "invent"
- [J_U] `narrower`: J_U 再婚する *saikon-suru* "remarry" and J_V 結婚する *kekkon-suru* "marry"
- `different nuance`: J_U 押収する *oushuu-suru* "expropriate" and J_V 取り上げる *toriageru* "confiscate" (J_U is more formal than J_V.)
- `different`: J_U 立ち向かう *tachi-mukau* "confront" and J_V 反論する *hanron-suru* "argue against" (their meanings overlap so they are classified into other classes in some context.)

Next, we give an example of the paraphrasing; 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 5 sentences from our corpus which use J_U, for example (2; slightly simplified here), and replace J_U with J_V (3).

(2) 罰則 を 重く する 必要 は
bassoku o omoku-suru hitsuyou wa
penalty ACC increase need TOP
ない と 具申 した。
nai to gushin-shita.
nothing QUOT reported.

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

(3) 罰則 を 重く する 必要 は ない
bassoku o omoku suru hitsuyou wa nai
と 上申 した。
to jyoushin-shita.

The paraphrase (3) is `grammatical` and the pair (2, 3) have `close` meanings. This is done for all five sentences containing J_U and then done in reverse for all 5 sentences matching the pattern for J_V.

2.2 Experiment

To test the useful range of our algorithm, we considered the coverage of **ALT-J/E**'s valency dictionary on 9 years of Japanese newspaper text (6 years of Mainichi and 3 years of Nikkei) (see Table 1). The valency dictionary had Japanese entries for 4,997 verb types (37.5%), which covered most of the actual words (92.5%). There were 8,304 verbs with no Japanese entry in the valency dictionary. Of those, 4,129 (49.7 %) verbs appear in **ALT-J/E**'s Japanese-English transfer dictionary or EDICT and have a pattern with the same translation in the valency dictionary.[6] Most of these, 3,753 (90.9 %) have some examples that can be used to check the paraphrases. We made candidate patterns for these verbs.

Table 1: Cover Ratio for Japanese Newspapers (9 years)

In lexicon	No. of Types (%)		No. of Tokens (%)	
Jp exists	4,997	37.5	24,656,590	92.5
En exists	4,129	31.0	1,355,552	5.1
No entry	4,175	32.4	645,158	2.4
Total	13,301	100.0	26,657,300	100.0

[6]EDICT typically translates Japanese verbal nouns as nouns, without giving a separate verb entry: e.g., 共同 *kyōdō* "cooperation". We used **ALT-J/E**'s English morphological dictionary and the EDICT part-of-speech codes to create 10,395 new verb entries such as: 共同する *kyōdō-suru* "cooperate".

For the 3,753 target verbs, we did the check using the pre-filter and paraphrasing. The original number of candidates was enormous: 108,733 pairs of J_U and J_V. Most of these were removed in the pre-filtering stage, leaving 2,570 unknown verbs matching 6,888 verbs in the valency dictionary (in fact, as the pre-filter check doesn't need the valency patterns, they can be made after this stage). When these were expanded into patterns, they made a total of 8,553 candidate patterns (3.3 patterns/verb) whose semantic similarity was then checked using paraphrases.

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 is a significant speed-up over the 30 minutes normally taken by an expert lexicographer to construct a valency entry.

3 Evaluation and Results

We evaluated the effect on translation quality for each new pattern that had at least one paraphrase that was `grammatical`. There were 6,893 new patterns, for 2,305 kinds of verbs (3.0 patterns/verb). For each verb (J_U) we picked two shortish sentences (on average 81.8 characters/sentence: 40 words) from a corpus of 11 years of newspaper text (4 years of Mainichi and 5 of years Nikkei). This corpus had not been used in the paraphrasing stage, i.e., all the sentences were unknown. We tried to get 2 sentences for each verb, but could only find one sentence for some verbs: this gave a total of 4,367 test sentences.

Translations that were identical were marked `no change`. Translations that changed were evaluated by Japanese native speakers who are fluent in English. The new translations were placed into three categories: `improved`, `equivalent` and `degraded`. All the judgments were based on the change in translation quality, not the absolute quality of the entire sentence.

We compared the translations with and without the valency patterns. There were two setups. In the first, we added one pattern at a time to the valency dictionary, so we could get a score for each pattern. Thus verbs with more than one pattern would be tested multiple times. In this case we tested 13,140 different sentence/pattern combinations. In the sec-

ond, we added all the patterns together, and let the system select the most appropriate pattern using the valency information and selectional restrictions. The results of the evaluation are given in Table 2.

Table 2: Evaluation of New Valency Entries

	Each pattern		All patterns	
Judgment	No.	%	No.	%
`improved`	4,536	34.5 %	1,636	37.5 %
`no change`	3,238	24.6 %	1,063	24.3 %
`equivalent`	3,465	26.4 %	1,115	25.5 %
`degraded`	1,901	14.5 %	552	12.6 %
Total	13,140	100.0%	4,366	100.0 %

As can be seen in Table 2, most sentences `improved` followed by `equivalent` or `no change`. `Degraded` is a minority. There was a clear improvement in the overall translation quality.

In particular, the result using all patterns (which is the way the dictionary would normally be used) is better than using one pattern at a time. There are two reasons: (1) Many verbs have different alternations. When we used all patterns, we covered more alternations, therefore the system could select the right entry based on its subcategorization. (2) The entries also have different selectional restrictions for different translations. When we used all patterns, the system could select the best valency entry based on its selectional restrictions. Even without using the full paraphrase data, only the pre-filter and the grammatical judgments, 37.5% of translations improved and only 12.6% degraded, an overall improvement of 24.9%.

Now we analyze the reasons for the `improved` and `degraded` translations. The reasons for `improved`: (1) The system was able to translate previously unknown words. The translation may not be the best but is better than an unknown word. (2) A new pattern with a better translation was selected. (3) The sentence was translated using the correct subcategorization, which allowed a zero pronoun to be supplemented or some other improvement. The reasons for `degraded`: (1) the detailed nuance was lost. For example, 撫で上げる *nade-ageru* "brush up" became simply *brush*. (2) A new pattern was selected whose translation was less appropriate.

Further Refinements

We then examined the paraphrase data in an attempt to improve the quality even further by filtering out the bad entries. To examine this, we defined scores for the evaluation categories: `improved` is +1, `no change` and `equivalent` are +0.5 and `degraded` is -1. Because we used up to two evaluation sentences for each J_U, the evaluation score varies between -2 and 2.

We expected that restricting the patterns to those with `grammatical` paraphrases and `same` or `close` meaning would improve translation quality. However, as can be seen in Table 3, the distribution of improvements in translation quality did not change significantly according to the percentage of paraphrases that were either `same` or `close`.

One reason for the lack of correlation of the results is that change in translation quality of an example sentence is a very blunt instrument to evaluate the fitness of a lexical entry. Particularly for complicated sentences, the parse may improve locally, but the translation degrade overall. In particular, a translation that was incomprehensible could become comprehensible but with the wrong meaning: the worst possible result for a translation system. However, even allowing for these imperfections, the result still holds: a test for paraphrasability on a small set of sentences is not a useful indicator of whether two verbs have the same valency. One reason for the lack of utility of the paraphrase tests is that the example sentences were chosen randomly: there is no guarantee that they show either typical usage patterns or the full range of use.

We were actually suprised by these results, so we tried various other combinations of grammaticality and semantic similarity, and found the same lack of effect. We also tried a monolingual similarity measure based on word-vector distances taken from word-definitions and corpora (Kasahara et al., 1997). This measure was also not effective. We then looked at our inital grammaticality filter (at least one paraphrase that was `grammatical`). Evaluating a small sample of verbs that failed this test (evaluated on 50 sentences), we found that only 16% improved and 24% were degraded. Therefore this test was useful. However, it only removed 265 verbs (less than 10%). If we had left them

Table 3: `grammatical` and either `same` or `close`

% `same` or `close`	Change in Translation Quality												
	-2,-1	%	-0.5	%	0	%	0.5, 1	%	1.5	%	2	%	All
- 10	248	8	364	11	148	4	1,160	36	740	22	571	17	3,231
- 20	0	0	0	0	0	0	1	100	0	0	0	0	1
- 30	32	6	58	10	23	4	161	30	142	26	117	21	533
- 40	6	6	12	12	4	4	40	40	16	16	21	21	99
- 50	22	9	31	12	13	5	71	28	56	22	57	22	250
- 60	10	5	25	13	17	9	81	43	25	13	30	15	188
- 70	14	6	27	11	12	5	77	33	52	22	52	22	234
- 80	3	3	9	9	12	12	35	36	21	21	17	17	97
- 90	20	9	24	10	12	5	71	32	57	25	41	18	225
- 100	91	5	149	8	90	4	682	37	437	23	376	20	1,825
Total	446	7	699	10	331	4	2,379	36	1,546	23	1,282	19	6,683

in, the extrapolated result (testing each pattern individually) is an improvement of 32% versus a degradation of 16%, which should improve further if all patterns are tested together.

Next, we looked at the target language translation (used to judge that the meanings are similar). We made the conditions used to match the English stricter (e.g., head match plus n words different for various n), and found no useful difference.

Finally, we looked at the source of the English used to find the candidate patterns. Translations with a very low preference value in our system dictionary (i.e., the 12th best or lower translation for a word with 12 translations or more) were significantly worse. However there were only 4 such verbs, so it is not a very useful filter in practice. An interesting trend we did find was that translations found in EDICT were significantly better than those found in **ALT-J/E**'s transfer dictionary (see Table 4).

The main reason that EDICT gave such good results was that words with no entry in **ALT-J/E**'s transfer dictionary could not be translated at all by the default system: the translated system included the Japanese verb as is. Building new patterns from EDICT allowed the system to find a translation in these cases.

4 Discussion and Future Work

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. We were able to add entries for 2,305 different verbs to an initial lexicon with 5,062 verbs.

The results of our examination of various filters on the constructed patterns were negative. Contrary to the claims of Fujita and Bond (2002), using paraphrasing as a filter did not help to improve the quality. However, their central claim, that words with similar meaning have similar valency, and that words with the same translations often have similar meanings was validated.

From a practical point of view the results are encouraging: we can build useful new patterns with only a simple monolingual judgment as pre-filter: "are these verbs similar in meaning", and these patterns improve the quality of translation. Even this crude filter produces improvements in 32% of sentences versus degradations in only 16%. Adding a check on grammaticality of paraphrases gives an overall improvement in translation quality for these verbs of 24.9% (37.5% improved, 12.6% degraded). Further evaluation of the semantic similarity of the paraphrases did not improve these results.

EDICT, which is built by the collaboration of many volunteers, had a wider coverage than the **ALT-J/E** system dictionary. This shows the quality of such cooperatively built dictionaries. New projects to build lexical resources for Asian languages should take heed and (a) make the resources freely available and (b) encourage people to contribute to them. Projects already embracing this approach include SAIKAM (Thai-Japanese), PAPILLON (Japanese-French-...) and LERIL (English-Hindi/Tamil/...).

Table 4: Effect of Translation Source

Translation Source	Change in Translation Quality												
	-2,-1	%	-0.5	%	0	%	0.5, 1	%	1.5	%	2	%	All
ALT-J/E	312	9	481	13	228	6	1,284	36	793	22	498	13	3,596
ALT/EDICT	72	6	128	10	44	3	501	43	256	21	168	14	1,169
EDICT	110	4	194	7	103	3	894	33	681	25	739	27	2,721
Total (%)	494	7	803	10	375	5	2,679	36	1,730	23	1,405	18	7,486

We therefore propose a method of building information-rich lexicons that proceeds as follows: (1) build a seed lexicon by hand; (2) extend it semi-automatically using a simple prefilter check; (3) release the resulting lexicon so that people can use it and provide feedback to remove bad entries. People will not use a lexicon if it is too bad, but an error rate of 12.6% should be acceptable, especially as it can be easily decreased with feedback.

5 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 less than 7 minutes per pattern. An evaluation of 6,893 new patterns showed that adding them to a Japanese-to-English machine translation system improved the translation for 37.5% of sentences using these verbs, and degraded it for 12.6%, a substantial improvement in quality.

Acknowledgments

The authors would like to thank the other members of the NTT Machine Translation Research Group, Satoshi Shirai and Timothy Baldwin. This research was supported by the research collaboration between the NTT Communication Science Labs and CSLI.

References

Yasuhiro Akiba, Hiromi Nakaiwa, Satoshi Shirai, and Yoshifumi Ooyama. 2000. Interactive generalization of a translation example using queries based on a semantic hierarchy. In *ICTAI-00*, pages 326–332.

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

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Sanae Fujita and Francis Bond. 2002. A method of adding new entries to a valency dictionary by exploiting existing lexical resources. In *Ninth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-2002*, pages 42–52, Keihanna, Japan.

Satoru Ikehara, Satoshi Shirai, Akio Yokoo, and 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*, pages 101–106, Washington DC. (`http://xxx.lanl.gov/abs/cmp-lg/9510008`).

Kaname Kasahara, Kazumitsu Matsuzawa, and 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).

Daisuke Kawahara and Sadao Kurohashi. 2001. Japanese case frame construction by coupling the verb and its closest case component. In *Proceedings of First International Conference on Human Language Technology Research (HLT 2001)*, pages 204–210, San Diego.

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

Satoshi Shirai. 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`).

Takehito Utsuro, Takashi Miyata, and Yuji Matsumoto. 1997. Maximum entropy model learning of subcategorization preference. In *Proc. 5th Workshop on Very Large Corpora*, pages 246–260.