

# A Method of Creating New Bilingual Valency Entries using Alternations

Sanae Fujita      Francis Bond

{sanae, bond}@cslab.kecl.ntt.co.jp

NTT Machine Translation Research Group

NTT Communication Science Laboratories

Nippon Telephone and Telegraph Corporation

## Abstract

We present a method that uses alternation data to add new entries to an existing bilingual valency lexicon. If the existing lexicon has only one half of the alternation, then our method constructs the other half. The new entries have detailed information about argument structure and selectional restrictions. In this paper we focus on one class of alternations, but our method is applicable to any alternation. We were able to increase the coverage of the causative alternation to 98%, and the new entries gave an overall improvement in translation quality of 32%.

## 1 Introduction

Recently, deep linguistic processing, which aims to provide a useful semantic representation, has become the focus of more research, as parsing technologies improve in both speed and robustness (Uszkoreit, 2002). In particular, machine translation systems still mainly rely on large hand-crafted lexicons. The knowledge acquisition bottleneck, however, remains: precise grammars need information-rich lexicons, such as valency dictionaries, which are costly to build and extend. In this paper, we present a method of adding new entries to an existing bilingual valency dictionary, using information about verbal alternations.

The classic approach to acquiring lexical information is to build resources by hand. This produces useful resources but is expensive. This is still the approach taken by large projects such as FrameNet (Baker et al., 1998) or OntoSem. Therefore, there is a need to extend these hand-made resources quickly and economically. Another approach is to attempt to learn information from corpora. There has been much research based on this, but due to the inevitable errors, there are few examples of lexicons being constructed fully automatically. Korhonen (2002) reports that the ceiling on the performance of mono-lingual subcategorization acquisition from

corpora is generally around 80%, a level that still requires manual intervention. Yet another approach is to combine knowledge sources: for example to build a lexicon and then try to extend it using corpus data or to enrich mono-lingual data using multilingual lexicons (Fujita and Bond, 2002).

The aim of this research is not to create a lexicon from scratch, but rather to add further entries to an existing lexicon. We propose a method of acquiring detailed information about predicates, including argument structure, semantic restrictions on the arguments and translation equivalents. It combines two heterogeneous knowledge sources: an existing bilingual valency lexicon (the seed lexicon), and information about verbal alternations.

Most verbs have more than one possible argument structure (subcat). These can be regularized into pairs of alternations, where two argument structures link similar semantic roles into different subcats. Levin (1993) has identified over 80 alternation types for English, and these have been extended to cover 4,432 verbs in 492 classes (Dorr, 1997). In this paper, we will consider alternations between transitive (**Vt**) and intransitive (**Vi**) uses of verbs, where the subject of the intransitive verb (**S**) is the same as the object of the transitive verb (**O**) (e.g. *the acid dissolved the metal*  $\Leftrightarrow$  *the metal dissolved (in the acid)*) (Levin, 1993, 26–33)). We call the subject of the transitive verb **A** (ergative) and this alternation the **S=O** alternation.

Figure 1 shows a simplified example of an alternating pair in a bilingual valency dictionary (the valency lexicon from the Japanese-to-English machine translation system **ALT-J/E** (Ikehara et al., 1991)). This includes the subcategorization frame and selectional restrictions. As shown in Figure 1, Japanese, unlike English, typically morphologically marks the transitivity alternation.

We chose the **S=O** alternation because it is one

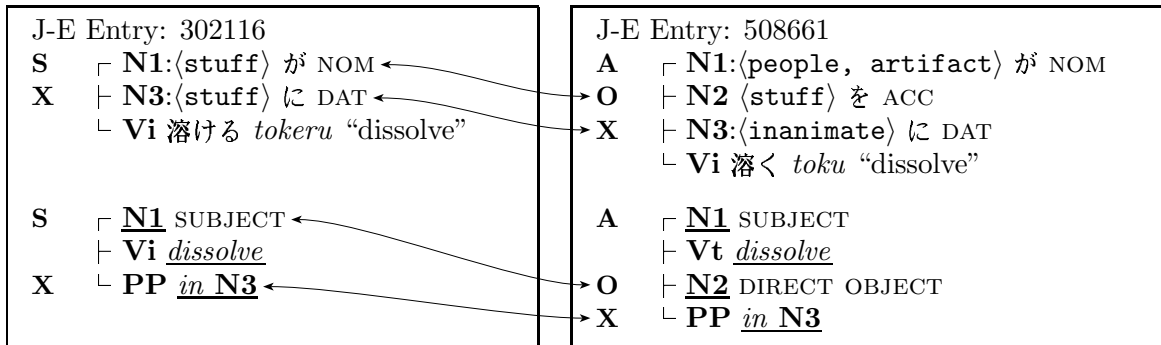


Figure 1: **Vi** 溶ける *tokeru* “dissolve” ↔ **Vt** 溶く *toku* “dissolve”

of the most common types of alternations, making up 34% of those discovered by Bond et al. (2002) and has been extensively studied. The method we present, however, can be used with any alternation for which lists of alternating verbs exist.

## 2 Resources

We use two main resources in this paper: (1) a seed lexicon of high quality hand-made valency entries; and (2) lists of verbs that undergo one or more **S=O** alternations.

The alternation list includes 449 native Japanese verbs that take the **S=O** alternation, based on data from Jacobsen (1981), Bullock (1999) and the Japanese/English dictionary EDICT (Breen, 1995). Each entry consists of a pair of Japanese verbs with one or more English glosses. Expanding out the English results in 839 Japanese-English pairs in all. Some examples are given in Table 1.

Intransitive		Transitive		
Ja	En	Ja	En	
溶ける	tokeru dissolve	溶く	toku	dissolve
泣く	naku cry	泣かす	nakasu	make cry
上がる	agaru rise	上げる	ageru	lift

Table 1: Verbs Undergoing the **S=O** Alternation

As a seed lexicon, we use the valency dictionary (Ikehara et al., 1997) from the Japanese-to-English machine translation system **ALT-J/E**. It consists of linked pairs of Japanese and English verbs. There are 5,062 Japanese verbs and 11,214 entries (ignoring all idiomatic and adjectival entries). Verb entries in both languages have information about the argument structure (subcat) of the verb. In addition to the core arguments, adjunct cases are added to many patterns to help in disambiguation.<sup>1</sup> The Japanese side has selec-

tional restrictions (SR) on the arguments. The arguments are linked between the two languages using case-roles (**N1**, **N2**, ...).

The seed lexicon covered 381 out of the 449 linked Japanese pairs (85%). In the next section, in order to examine the nature of the alternation we compare the case roles and translation of the linked valency pairs.

## 3 The Nature of the **S=O** Alternation

### 3.1 Comparing Selectional Restrictions of **A**, **O** and **S**

In alternations, a given semantic role typically appears in two different syntactic positions: for example, the **DISSOLVED** role is the subject of intransitive *dissolve* and the object of the transitive. Baldwin et al. (1999) hypothesized that selectional restrictions (SRs) stay constant in the different syntactic positions. Dorr (1997), who generates both alternations from a single underlying representation, implicitly makes this assumption. In addition, Kilgarriff (1993) specifically makes the **A** (+sentient, +volition), while the **O** is (+changes-state, +causally affected).

However, we know of no quantitative studies of the similarities of alternating verbs. Exploiting the machine translation lexicon for linguistic research, we compare the SRs of **S** with both **A** and **O** for verbs that take the **S=O** alternation.

The SRs take the form of a list of semantic classes, strings or \*. Strings only match specific words, while \* matches anything, even non-nouns. The semantic classes are from the GoiTaikei ontology of 2,710 categories (Ikehara et al., 1997). It is an unbalanced hierarchy with a maximum depth of 12. The top node (level 1) is **noun**. The lower the level, the more specialized

<sup>1</sup>This is common in large NLP lexicons, such as COMLEX (Grishman et al., 1998). For example, the COMLEX 3.0 entry for *gather* notes that it cooccurs with PPs headed by *around*, *inside*, *with*, *in* and *into*.

<sup>1</sup>This is common in large NLP lexicons, such as COM-

the meaning, and thus the more restrictive the SR.

We calculate the similarity between two SRs as the minimum distance (MD), measured as links in the ontology. If the SRs share at least one semantic class then the MD is zero. In this case, we further classified the SRs which are identical into “0 (Same)”. For example, in Figure 1, the MD between **S** and **O** is “0 (Same)” because they have the same SR:  $\langle \text{stuff} \rangle$ . The MD between **A** and **S** is two because the shortest path from  $\langle \text{artifact} \rangle$  to  $\langle \text{stuff} \rangle$  traverses two links ( $\text{artifact} \subset \text{inanimate} \subset \text{stuff}$ ).<sup>2</sup>

Figure 2 shows the MD between **O** and **S**, and **A** and **S**. The selectional restrictions are very similar for **O** and **S**. 30.1% have identical SRs, distance is zero for 27.5% and distance one is 28.3%. However, for **A** and **S**, the most common case is distance one (26.7%) and then distance two (21.5%). Although **O** and **S** are different syntactic roles, their SRs are very similar, reflecting the identity of the underlying semantic roles.

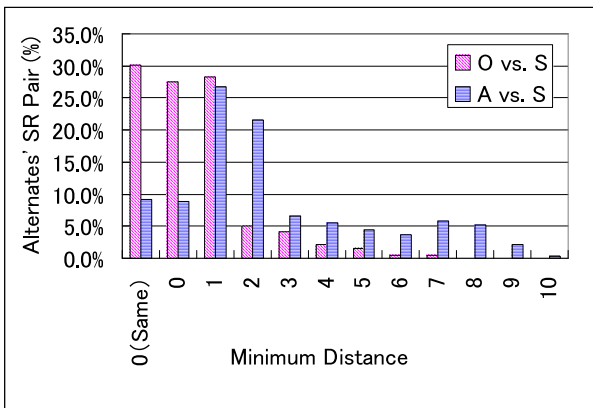


Figure 2: The Minimum Distance of Selectional Restrictions

Next, we examine whether **A**, **O**, and **S** are  $\langle +\text{sentient}, +\text{volition} \rangle$  or not. In the GoiTaikei hierarchy, semantic classes subsumed by **agent** are  $\langle +\text{sentient}, +\text{volition} \rangle$ . **A** was very agentive, with 60.1% of the SRs being subsumed by **agent**. The most frequent SR for **A** is  $\langle \text{agent} \rangle$  itself (41.4%). **S** and **O** are less agentive, with 13.9% and 14.1% of their respective selectional restrictions being agentive. This data supports the hypothesis in Kilgarriff (1993).

<sup>2</sup>There is some variation due to lexicographer’s inconsistencies. For example **X**’s SR is  $\langle \text{stuff} \rangle$  in the intransitive and  $\langle \text{inanimate} \rangle$  in the transitive entry. It should be  $\langle \text{stuff} \rangle$  in both entries.

In summary, the SRs of **S** and **O** are not identical, but very similar. In comparison, **A** is more agentive, and not closely linked to either.

### 3.2 Comparison of Japanese and English

From the point of view of constructing bilingual lexical entries, if the English main verb can translate both Japanese entries, then it is possible to automatically construct a usable English translation equivalent along with the Japanese alternation. In order to see how often this is the case, we compare Japanese and English alternations and investigate the English translations in the alternation list.

We divide the entries into five types in Table 2. The first three are those where the main English verb is the same. The most common type (30.0%) is made up of English unaccusative verbs which also undergo the **S=O** alternation [**s=0**]. The next most common (19.8%) is entries where the Japanese intransitive verb can be translated by making the transitive verb’s English translation passive [**passive**]. In the third type (6.5%) the English is made transitive synthetically [**synthetic**]: a control verb (normally *make*) takes an intransitive verb or adjective as complement. The last two are those where either different translations are used (42.8%), or the same English verb is used but the valency change is not one of those described above.

The first three rows of Table 2 show the verbs whose alternate can be created automatically, 56.3% of the total. This figure is only an approximation, for two reasons. The first is that the translation may not be the best one, most verbs can have multiple translations, and we are only creating one. The second is that this upper limit is almost certainly too low. For many of the alternations, although our table contained different verbs, translations using identical verbs are also acceptable. In fact, most transitive verbs can be made passive, and most intransitive verbs embedded in a causative construction, so this alternative is always possible (and is also possible for Japanese). However, if the Japanese uses a lexical alternation, it is more faithful to link it to an English lexical alternation when possible.

## 4 Method of Creating Valency Entries

In this section we describe how we create new alternating entries. Given a verb, with dependents  $N_i$ , and an alternation that maps some or all of the  $N_i$ , we can create the alternate by analogy with existing alternating verbs. The basic flow of

Japanese		English Translation		English Structure		Type	No. (%)
Vi	Vt	Vi	Vt	Vi	Vt		
弱まる	弱める	S <u>weaken</u>	A <u>weaken</u> O	S Vi	A Vt O	S=0	138 30.0
漏れる	漏らす	S be <u>omitted</u>	A <u>omit</u> O	S be Vt-ed	A Vt O	passive	91 19.8
泣く	泣かす	S <u>cry</u>	A <u>make</u> O <u>cry</u>	S Vi/be Adj	A Vc O Vi/Adj	synthetic	30 6.5
亡くなる	亡くす	S <u>pass away</u>	A <u>lose</u> O	S Vi	A Vt O	Diff Head	197 42.8
じゃれる	じゃらす	S <u>play</u>	A <u>play</u> with O	S Vi	A Vt prep O	Diff Struct	4 0.9

Vc is control verb such as *make, get, let, become*. Many entries also include information about non-core arguments/adjuncts.

Table 2: Classification of English Translations of the S = O Alternation List (Reference Data)

creating valency entries is as follows.

- For each dependent  $N_i$ 
  - if  $N_i$  participates in the alternation
  - if  $N_i$  has an alternate in the target then map to it
  - else delete  $N_i$
  - else transfer [non-alternating dependent]
- If the alternation requires a dependent not in the source
  - Add the default argument

We use the most frequent argument in existing valency entries as a default. Specific examples of creating S = O alternations are given in the next section.

Although we only discuss the selectional restrictions and subcat information here, we also map the verb classes (given as verbal semantic attributes (Nakaiwa and Ikehara, 1997)). The mapping for the dependents in the alternation can be taken from existing lexical resources (Dorr, 1997), learned from corpora (McCarthy, 2000) or learned from existing lexicons (Bond et al., 2002).

#### 4.1 Target

In this experiment, we look at one family of alternations, the S = O alternation. The candidate words are thus intransitive verbs with no transitive alternate, or transitive entries with no intransitive alternate. Alternations should be between senses, but the alternation list is only of words. Many of the candidate words (those that have a entry for only one alternate) have several entries. Only some of these are suitable as seeds. We don't use entries which are intransitive lemmas but have an accusative argument, which are intransitive (or transitive) lemmas but have an transitive translation (or intransitive), or which have both topic and nominative, such as (1), where the nominative argument is incorporated in the English translation.

- (1) N1:<animals> は N3:<"力"> が  
*N1 ha N3:"chikara" ga*  
 N1 TOP N3:power NOM  
 抜ける  
*nukeru*  
 lose  
 N1 lose N1's energy

There are 115 entries (37 lemmas) which have only intransitive entries and 81 entries (25 lemmas) which have only transitive entries which are in our reference list of alternating verbs. We create intransitive entries using the existing transitive entries, and transitive entries using the existing intransitive entries.

#### 4.2 Creating the Japanese subcat and SRs

In creating the intransitive entries from the transitive entries, we map the O's SRs onto the S's SRs, and change the case marker from accusative to nominative. We delete the A argument, and transfer any other dependents as they are.

In creating the transitive entries, we map the intransitive S's SRs onto the new O's SRs, and give it an accusative case-marker. If the intransitive entry has a demoted subject argument (where the Japanese case-marker is *ni* and the English preposition is *by*), we promote it to subject and use its SR for A. Otherwise we add a causative argument as ergative subject (A) with a default SR of <agent><sup>3</sup> and a nominative case-marker. We show an example in Figure 3.

#### 4.3 Creating the English Equivalent

The English translation can be divided into three types: S=0, passive and synthetic. Therefore it is necessary to judge which type is appropriate for each entry, and then create the English. This judgement is shown in Figure 4. To judge whether an English

<sup>3</sup><agent> is the most frequent SR for transitive verbs undergoing this alternation as shown in § 3.1.

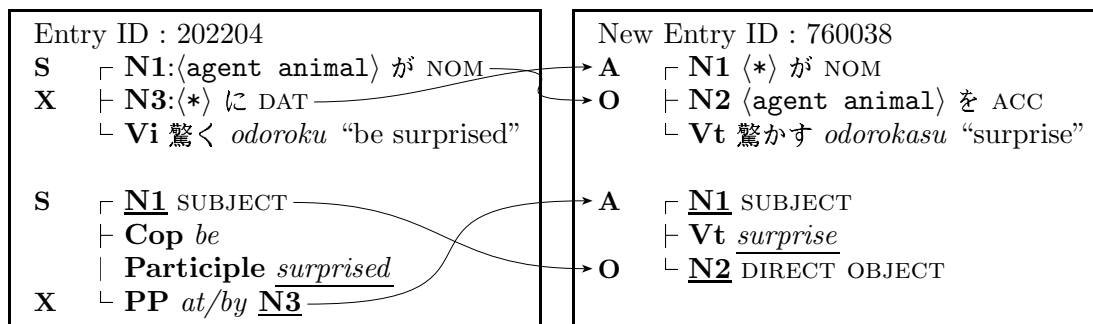


Figure 3: Seed: **Vi** 驚く *odoroku* “be surprised”  $\Rightarrow$  New entry: **Vt** 驚かす *odorokasu* “surprise”

verb could undergo the **S = O** alternation we used the LCS Database (EVCA+) (Dorr, 1997, [http://www.umiacs.umd.edu/~bonnie/LCS\\_Database\\_Documentation.html](http://www.umiacs.umd.edu/~bonnie/LCS_Database_Documentation.html)).

## 5 Evaluation

A total of 196 new entries were created for 62 verbs (25 **Vi** + 37 **Vt**) using the method outlined in § 4. We evaluated the quality by using the new entries in a machine translation system.

### 5.1 Translation-Based Evaluation

We evaluated the quality of the created entries in a translation-based regression test. We got two example sentences using each verb from Japanese newspapers and web pages: this gave a total of 124 test sentences. We translated the test sentences using **ALT-J/E**, both with (**with**) and without (**w/out**) the new entries.

Translations that were identical were marked **no change** (the system translates with a simple word dictionary if it has no valency entry). Translations that changed were evaluated by people fluent in both languages (two thirds by Japanese native speakers and one third by an English native speaker, not the authors). The translations were randomly presented to the evaluators labeled by A and B. Therefore evaluators did not know whether a translation is **with** or **w/out**. The translations were placed into three categories: (i) A is better than B, (ii) A and B are equivalent in quality, and (iii) A is worse than B. For example in (2), the evaluation was (iii). In this case A is **w/out** and B is **with**, so the new entry has improved the translation.

- (2) 塩田 喜代子 さんは、毛布 に  
*Shioda Kiyoko san wa, moufu ni*  
 Shioda Kiyoko Ms. NOM blanket in  
 くるまり ながら。  
*kurumari nagara.*  
 wrapped while.

(A) Ms. Kiyoko Shioda is wrapped  
up to a blanket.

(B) Ms. Kiyoko Shioda is wrapped  
in a blanket.

Table 3 shows the evaluation results, split into those for transitive and intransitive verbs. The most common result was that the new translation was **better** (46.0%). The quality was **equivalent** for 13.7% and **worse** for 14.5%. The overall improvement was 31.5% (46.0 – 14.5). Extending the dictionary to include the missing alternations gave a measurable improvement in translation quality.

	Vi Created		Vt Created		Total	
	No.	%	No.	%	No.	%
<b>better</b>	19	38.0	38	51.4	57	46.0
<b>equivalent</b>	5	10.0	12	16.2	17	13.7
<b>no change</b>	18	36.0	14	18.9	32	25.8
<b>worse</b>	8	16.0	10	13.5	18	14.5
Change		+22.0		+37.9		+31.5
Total	50	100.0	74	100.0	124	100.0

Table 3: Results of Translation-based Evaluation

### 5.2 Lexicographer’s Evaluation

A manual analysis of a subset of the created entries was carried out by expert lexicographers familiar with the seed lexicon (not the authors). They found three major source of errors. The first was that alternation is a sense based phenomenon. As we built alternations for all patterns in the seed dictionary, this resulted in the creation of some spurious patterns. An example of an impossible entry is 捕られる *torawareru* “be caught”, translated as *be picked up* with the inappropriate semantic restriction **<concrete,material-phenomenon>** on the subject. However, another good entry was cre-

---

### Creating Intransitive entries:

if the original subcat has a control verb  
( $Vc \in \{make, have, get, cause\}$ )

- $A Vc O Vi/Adj \Rightarrow S Vi/be Adj$  [synthetic]  
( $A make O cry \Rightarrow S cry$ )

else (original head is  $Vt$ )

- if  $Vt$  undergoes the  $S = O$  alternation
  - $A Vt O \Rightarrow S Vi$  [S=0]  
( $A turn O \Rightarrow S turn$ )
- else
  - $A Vt O \Rightarrow S be Vt-ed$  [passive]  
( $A injure O in X \Rightarrow S be injured in X$ )

We made a special rule for the English  $Vt$  *have*. In this case the intransitive alternation will be *There is*: for example, 「及ぼす」  $A have O on X \Rightarrow$  「及ぶ」  $There be S on X$ .

### Creating Transitive Entries :

If the original subcat is:

- $S Vi$ 
  - if  $Vi$  undergoes the  $S = O$  alternation
    - $\Rightarrow A Vt O$  [S=0]  
( $S spoil \Rightarrow A spoil O$ )
  - else  $\Rightarrow A Vc^\dagger O Vi$  [synthetic]  
( $S rot \Rightarrow A make O rot$ )
- $S be Adj \Rightarrow A Vc^\dagger O Adj$  [synthetic]  
( $S be prosperous \Rightarrow A make O prosperous$ )
- $S be Vt-ed \Rightarrow A Vt O (by A)$  [passive]  
( $S be defeated (by A) \Rightarrow A defeat O$ )

<sup>†</sup> We use *make* as the control verb,  $Vc$

Figure 4: Method of Creating English Side

---

ated, with the translation *be caught* and SRs  $\langle \text{people, animal, artifact} \rangle$ , and this was judged to be good.

The second source of errors was in the selectional restrictions. In around 10% of the entries, the lexicographers wanted to change the SRs. The most common change was to make the SR for  $A$  more specific than the default of **agent**.

The third source of errors was in the English translation, where the lexicographers sometimes preferred a different verb as a translation, rather

than a regular alternation.

## 6 Discussion and Future Work

The above results show that alternations can be used to create rich and useful bilingual entries. In this section we discuss some of the reasons for errors, and suggest ways to improve and expand our method.

### 6.1 Rejecting Inappropriate Candidates

To make the construction fully automatic, a test for whether the Japanese side of the entry is appropriate or not is required.

One possibility is to add a corpus based filter: if no examples can be found that match the selectional restrictions for an entry, then it should be rejected. This could be done for each language individually. The problem with this approach is that many of the entries we created were for infrequent verbs. The average frequency in 16 years of Japanese newspaper text was only 173, and 22 verbs never appeared, although all were familiar to native speakers. We can, of course, use the web to alleviate the data sparseness problem.

### 6.2 Improving the English Translations

In this section we compare the distribution of the different types of translations for the reference data (§ 3.1) and the entries created by our method (§ 3.2). The breakdown is shown in Table 4. The first three rows show entries with the same English main verb.

One major discrepancy is in the frequency of the control verb construction. In  $Vi$ , no original transitive entry used control verbs. In general, when lexicographers create an entry, they prefer a simple entry to a synthetic one. Looking at the linguists' reference data, about 6.5% of the examples used control verbs. In the constructed data, 66.1% (77 entries) use the control verb *make*, more than any other category. For example, when the original intransitive entry is  $N1 be exhausted$ , *exhausted* is defined as adjective in the existing dictionary. So we create a new entry  $N1 make N2 exhausted_{adj}$ . However, there is a transitive verb *exhaust*, and it was preferred by the lexicographers:  $N1 exhaust N2$ . The algorithm needs to optionally convert adjectives to verbs in cases where there is overlap between the adjective and past participle.

Finally, we consider those Japanese alternations where the transitive and intransitive alternatives need translations with different English main verbs. A good example of this is  $Vi$  亡くな

Type	English Structure		Reference Data (Table2)		Vi Created		Vt Created	
	Vi	Vt	No.	(%)	No.	(%)	No.	(%)
S=0	<i>S Vi</i>	<i>A Vt O</i>	138	30.0	9	11.1	24	21.7
passive	<i>S be Vt-ed</i>	<i>A Vt O</i>	91	19.8	71	87.7	14	12.2
synthetic	<i>S Vi/be Adj</i>	<i>A Vc O Vi/Adj</i>	30	6.5	0	0	76	66.1
Different Head			191	41.5	0	0.0	0	0.0
Different Structure			10	2.2	1	1.2	0	0.0
Total			460	100	81	100	115	100

Table 4: A Comparison of Reference Data with Created Alternations

る *nakunaru* “S pass away” and Vt 亡くす *nakusu* “A lose O”.<sup>4</sup> These are impossible to generate using our method. Even with reliable English syntactic data, it would be hard to rule out *pass away* as a possible transitive verb or *lose* as an intransitive. They can only be ruled out by using data linking the subcat with the meaning, and this would need to be linked to the Japanese verbs’ meanings. This may become possible with larger linked multi-lingual dictionaries, such as those under construction in the Papillon project,<sup>5</sup> but is not now within our reach.

In summary, we could improve the construction of the English translations by using richer English information, especially about past-participles or verb senses.

### 6.3 Usage as a Lexical/Translation Rule

Although we have investigated the use of alternations in lexicon construction, the algorithms could also be used directly, either as lexical/translation rules or to generate transitive and intransitive entries from a common underlying representation. For example, Shirai et al. (1999) uses the existing entries and lexical rules deploying them to translate causatives and passives (including adversative passives) from Japanese to English. Trujillo (1995) showed a method to apply lexical rules for word translation. That is, they expand the vocabulary using prepared lexical rules for each language, and create links for translation between the lexical rules of a pair of languages. Dorr (1997) and Baldwin et al. (1999) generate both alternates from a single underlying representation.

Our proposed method could partially be implemented as a lexical or a translation rule. But not all the word senses alternate (§ 4.2), and not all the target language entries are regularly translated by the same head (§ 3). Further many of the

rules mix lexical and syntactic information, making them quite complicated. Because of that, it is easier to expand out the rules beforehand and enter them into the system.

### 6.4 Further Work

In this paper, we targeted native Japanese verbs only. ALT-J/E already has a very high coverage of native Japanese verbs. However, even in this case, we could increase the cover of this alternation from 85% to 98% (442 out of 449 alternation pairs now in the dictionary). Most valency dictionaries or new language pairs have less cover, and so will get more results. It is also possible to use this method so as to only create half the entries by hand, and then to automatically make the alternating halves (although not all the created entries will be perfect).

In addition to the native Japanese verbs, there are many Sino-Japanese verbal nouns that undergo S=O alternation (For example, (3) ↔ (4)).

(3) 店 が 製品 を 完売した  
*mise ga seihin o kanbai-shita*  
shop NOM products ACC sold out  
The shop sold out of the products.

(4) 製品 が 完売した  
*seihin ga kanbai-shita*  
products NOM sold out  
The products are sold out.

ALT-J/E’s Japanese dictionary has about 2,400 verbal nouns which have usage as both transitive and intransitive. Of these only 536 are in the valency dictionary. Our next plan is to add them all to the valency dictionary, using alternations to make the process more efficient and consistent.

Another extension is to apply the method to other alternations, using either linguists’ data or automatically acquired alternations (Oishi and Matsumoto, 1997; Furumaki and Tanaka, 2003;

<sup>4</sup>My friend passed away ↔ I lost my friend.

<sup>5</sup><http://www.papillon-dictionary.org/>

McCarthy, 2000). In particular, **S = O** alternations make up only 34% of those discovered by Bond et al. (2002), we intend to investigate the alternations that make up the remainder.

## 7 Conclusion

We presented a method that uses alternation data to add new entries to an existing translation lexicon. The new entries have detailed information about argument structure and selectional restrictions. We were able to increase the coverage of the **S=O** alternation to 98%, and the new entries gave an overall improvement in translation quality of 32%.

## References

- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics: COLING/ACL-98*, Montreal, Canada.
- Timothy Baldwin, Francis Bond, and Ben Hutchinson. 1999. A valency dictionary architecture for machine translation. In *Eighth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-99*, pages 207–217, Chester, UK.
- Francis Bond, Timothy Baldwin, and Sanae Fujita. 2002. Detecting alternation instances in a valency dictionary. In *8th Annual Meeting of the Association for Natural Language Processing*, pages 519–522. The Association for Natural Language Processing.
- 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](http://www.csse.monash.edu.au/~jwb/jsaa_paper/hpaper.html)).
- Ben Bullock. 1999. Alternative sci.lang.japan frequently asked questions. <http://www.csse.monash.edu.au/~jwb/afaq/jitadoushi.html>.
- Bonnie J. Dorr. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12(4):271–322.
- 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.
- Hisanori Furumaki and Hozumi Tanaka. 2003. The consideration of <n-suru> for construction of the dynamic lexicon. In *9th Annual Meeting of The Association for Natural Language Processing*, pages 298–301. (in Japanese).
- Ralph Grishman, Catherine Macleod, and Adam Myers, 1998. *COMLEX Syntax Reference Manual*. Proteus Project, NYU. (<http://nlp.cs.nyu.edu/comlex/refman.ps>).
- 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-1g/9510008>).
- Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997. *Goi-Taikai — A Japanese Lexicon*. Iwanami Shoten, Tokyo. 5 volumes/CDROM.
- Wesley Jacobsen. 1981. *Transitivity in the Japanese Verbal System*. Ph.D. thesis, University of Chicago. (Reproduced by the Indiana University Linguistics Club, 1982).
- Adam Kilgarriff. 1993. Inheriting verb alternations. In *Sixth Conference of the European Chapter of the ACL (EACL-1993)*, pages 213–221, Utrecht. (<http://acl.ldc.upenn.edu/E/E93/E93-1026.pdf>).
- Anna Korhonen. 2002. Semantically motivated subcategorization acquisition. In *Proceedings of the ACL Workshop on Unsupervised Lexical Acquisition*, Philadelphia, USA.
- Beth Levin. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, London.
- Diana McCarthy. 2000. Using semantic preferences to identify verbal participation in role switching alternations. In *Proceedings of the first Conference of the North American Chapter of the Association for Computational Linguistics. (NAACL)*, Seattle, WA.
- Hiromi Nakaiwa and Satoru Ikehara. 1997. A system of verbal semantic attributes in Japanese focused on syntactic correspondence between Japanese and English. *Information Processing Society of Japan (IPSJ)*, 38(2):215–225. (In Japanese).
- Akira Oishi and Yuji Matsumoto. 1997. Detecting the organization of semantic subclasses of Japanese verbs. *International Journal of Corpus Linguistics*, 2(1):65–89.
- Satoshi Shirai, Francis Bond, Yayoi Nozawa, Tomiko Sasaki, and Hiromi Ueda. 1999. One method of fitting valency patterns to text. In *5th Annual Meeting of the Association for Natural Language Processing*, pages 80–83. The Association for Natural Language Processing.
- Arturo Trujillo. 1995. Bi-lexical rules for multi-lexeme translation in lexicalist MT. In *Sixth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-95*, pages 48–66, July.
- Hans Uszkoreit. 2002. New chances for deep linguistic processing. In *19th International Conference on Computational Linguistics: COLING-2002*, pages XIV–XXVII, Taipei.