Argument Status in Japanese Verb Sense Disambiguation

Timothy Baldwin and Hozumi Tanaka
Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 JAPAN
{tim,tanaka}@cs.titech.ac.jp

Abstract

This research aims to incorporate argument status-based modelling within an otherwise selectional constraint-based system of verb sense disambiguation, to capture effects such as underspecification, surface case alternation and semantic backing-off. The proposed implementation hinges around a description of the general behavioural characteristics of integral complements, complements, middles and adjuncts through a pre-determined weighting schema. On limited evaluation, the resultant system returned an accuracy of over 83%, and was further shown to significantly outperform baseline methods.

1 Introduction

The purpose of this paper is to devise a robust verb sense disambiguation (VSD) technique based around selectional constraints which, for a given clausal input, both determines the correct sense of the main verb and aligns *input case slots* from the input, with *target case slots* within the valency frame described for the chosen verb sense. The driving mechanism behind the proposed system is **argument status,** that is an expanded model of complementhood/adjuncthood as laid out by Somers (1984) incorporating 'integral complements', 'complements', 'middles' and 'adjuncts'. Our particular interest in argument status stems from the fact that the properties of each category of argument status can be generalised to provide a surprisingly accurate model of such effects as surface case alternation, scrambling, and propensity for semantic backing-off. Note that although this paper is centered around a VSD system for Japanese, the system mechanism is transferable to other languages for which the notion of argument status is equally well defined.

In performing any research in VSD, there is clearly a requirement for some means of sense distinction/demarcation, and in our case this is based around the Goi-Taikei pattern-based valency dictionary (Ikehara et al. 1997; Shirai et al. 1997), as developed by NTT for their ALT-J/E machine translation system (Ikehara et al. 1991). A valency frame entry in the Goi-Taikei valency dictionary is represented as a list of case slots, each of which is provided with a set of class-based selectional constraints and/or lexical filler candidates.

By nature, selectional constraints commonly overlap between individual case slots for a given valency frame, and also between valency frames; the co-existence of lexical filler candidates for many verb senses further complicates things. This brings about the need for an interface to the dictionary which is able to select between the potentially sense-compatible candidates, and at the same time have recourse to relax the selectional constraints appropriately in the case of over-restriction.

A robust VSD system must be able to cope with non-canonical input. This correlates to being able to model semantic effects such as metonymy (see (Mahesh et al. 1997b)),

	Integral Complement	Complement	Middle	Adjunct
OBLIGATORINESS	Lexically oblig.	Lexically/argument oblig.	Argument opt.	Argument opt.
SEMANTIC DEMARKATION	High (absolute)	Low	Medium	High
SCOPE FOR CASE ALTERNATION	Almost none	High	Medium	Low
CASE SLOT ORDER	Highly restricted	Free, but produces markedness	Relatively free	Free

Table 1: Argument status matrix

sense extension (see (Mahesh et al. 1997a)) and occurrences of unknown words, and lexical effects such as (in the case of Japanese) surface case alternation, zero anaphora, and scrambling. While we make no attempt to individually analyse the various types of semantic effects, there is a clear need to be able to back-off from the level of selectional constraint described in Goi-Taikei **(semantic backing-off).** This need is met through the application of argument status in balancing the trade-off between over-relaxation of selectional constraints bringing incompatible senses into play, and over-restriction discounting the correct analysis.

To summarise, the main intent of this research is to apply argument status in scoring the various semantic and lexical idiosyncrasies faced in robustly disambiguating verb sense. Section 2 defines argument status and sets the scene for its later system application; Section 3 then goes on to describe the overall system architecture and various scoring schema utilised therein, followed by a basic evaluation of the system on Japanese inputs in Section 4.

2 Argument status

The argument status of a case slot is defined as the degree of boundedness of that case slot within the valency frame, or a measure of the intrinsic association between an argument and the predicate. The system of argument types utilised herein is a simplified version of the sixfold scale proposed by Somers (1984, 1987), and consists of integral complements, complements, middles and adjuncts. **Integral complements** (i.e. fixed arguments) are highly restricted as to scope for surface case alternation and case slot order alternation, are fixed as to lexical content, and are lexically obligatory (must have a lexical spell-out); **complements,** as major verb constituents, generally display high scope for surface case alternation, produce marked semantics upon case slot permutation, are generally not heavily demarcated semantically, and are argument obligatory (zero instances must be co-indexed to discourse entities or otherwise instantiated through deixis – see (Kameyama 1985)); **middles** form an 'in-between' category between complements and adjuncts, and are generally reasonably restricted in terms of surface case alternation, relatively free in case slot positioning, generally well-demarcated semantically, and argument optional (zero instances indicate the non-instantiation of that case-role); and **adjuncts** are the most peripheral argument type, being highly restricted in terms of surface case alternation, freely insertable into almost any position within the case frame, strongly demarcated semantically, and argument

197

Argument status	*arg_weight*
Integral complement	1
Complement	1
Middle	0.5
Adjunct	0

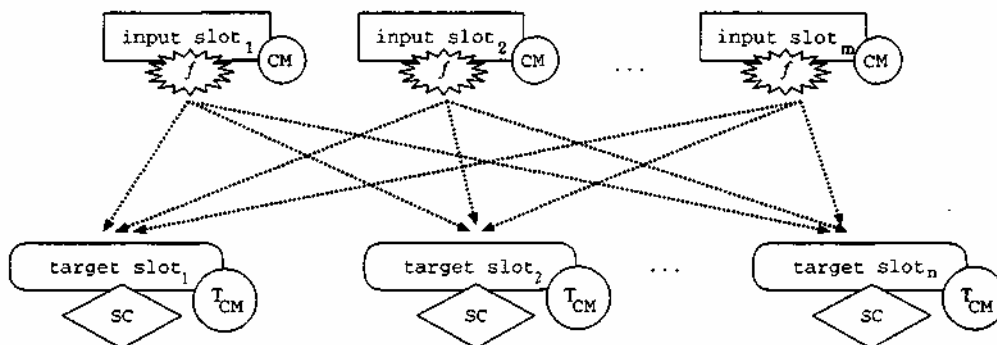Table 2: Weights for the various argument statuses

Figure 1: Input/valency frame case slot lattice

optional. A summary of the characteristics of each argument type is given in Table 1.[1]

Argument status is not indicated within the basic Goi-Taikei framework, and was thus predicted from the case-role mark-up of each case slot and post-checked manually.

The compatibility with different lexical effects for complements, middles and adjuncts is translated into the *arg_weight* function, with integral complements conditionally equivalent to complements in weight.

3   Basic system framework

The basic means of analysis is to produce a lattice of all possible mappings between input and valency frame case slots (see Fig. 1). Each edge is then scored through evaluation of the relative satisfaction by the input case filler of the selectional constraints for the target case slot. These preliminary mappings are subsequently verified for surface case marker compatibility, the full set of 1-to-l mappings from input to target case slots is extracted for each verb sense (including null mappings where permissible), and each candidate mapping is scored accordingly.

3.1   Case slot restrictiveness

All non-fixed case slots are encoded with a set of selectional constraints indexed to the Goi-Taikei thesaurus (Ikehara et al. 1997) and/or a list of lexical fillers. The Goi-Taikei thesaurus is set up in a conventional tree structure, of non-uniform depth and incorporating lexical items at all levels (average depth from the root $\approx$ 8.26). So as to establish a measure of restrictiveness of each subtree (node) in the thesaurus

---

[1] See Meyers et al. (1996) for a discussion of additional semantic factors affecting the basic complement/adjunct distinction.

structure, the naive notion of **case slot restrictiveness** is introduced, which draws on the intuition that, in cases of selectional overlap between verb senses such as between the "read" and "look at" senses of *miru* with *sinbun-o miru* "newspaper-ACC read/look at",[2] the more highly specialised (restrictive) sense of "read" is inherently preferred in the absence of other distinguishing information.

The degree of case slot restrictiveness *(CSR)* of a given node *x* in the thesaurus is estimated as the inverse of the mean subtree depth of all leaf nodes $l_{1..n}$ subsumed by *x*, with the *tree_depth* function defined as the number of nodes between the subtree root *x* and leaf l_i, inclusive.

$$CSR(x) = \frac{n}{\sum_{i=1}^{n} tree\_depth(x, l_i)} \qquad (1)$$

Hence, *CSR* for a leaf is one. while *CSR* for a case slot of unrestricted selectional scope (i.e. with selectional constraint set to the root node) is $\frac{1}{8.26} \approx 0.12$. Lexical fillers (e.g. fixed arguments) are treated as leaf nodes, and hence have *CSR* of one.

We are now in a position to posit a metric for the degree of **satisfaction of selectional constraint** (*SS*) of target case slot *t* with selectional constraint *c*, by sense f_s of input case filler *f*. In the formulation presented in (2), $\succeq$ is the subsumption operator (a $\succeq$ b $\Rightarrow$ 'a subsumes b'), *sub(a,b}* returns the least common hypernym node subsuming both *a* and *b,* and *rdepth(a)* is a statement of the inclusive path length from the thesaurus root to *a*. *SS* essentially allows for semantic back-tracking up the thesaurus structure until *c* has been relaxed sufficiently to subsume f_s. As we ascend the thesaurus structure, the degree of *SS* is diminished through a deflated *CSR* value, as well as the *rdepth* ratio of the relaxed selectional constraint to the original selectional constraint being accentuated.[3] Clearly in the case that *c* subsumes f_s, the *rdepth* ratio becomes 1, and the *CSR* remains unchanged. On occurrence of semantic back-tracking, *SS* is additionally weighted by way of the *arg_weight* of target case slot *t,* in line with the observations on semantic demarcation presented above. This acts so as to uphold adjunct constraints *(arg_weight = 0)* and penalise the relaxation of middle constraints *(arg_weight = 0.5)*, but permit the relaxation of complement constraints without additional penalisation *(arg_weight = 1)*. Note that integral complements are never represented by way of selectional constraints.

$$SS(f_s, c) = \begin{cases} CSR(c) & \text{if } c \succeq f_s \\ \frac{rdepth\left(sub(f_s, c)\right)}{rdepth(c)} \cdot CSR\left(sub(f_s, c)\right) \cdot arg\_weight(t) & \text{otherwise} \end{cases} \qquad (2)$$

Due to the potential for multiple senses of the input case filler, and multiple sets of selectional constraints/candidate fillers for the target case slot, we score each edge in the alignment lattice as the maximum value of *SS* (i.e. SS_{max}) for the given combination of filler *f*, with senses $f_{1..m}$, and target case slot *t*, with selectional constraints $c_{1..n}$:

$$SS_{max}(f, t) = \max_{i,j} SS(f_i, c_j) \qquad (3)$$

---

[2] Here, ACC stands for the accusative case.

[3] Parallels can be drawn between this general method and the conceptual similarity metric of Palmer & Wu (1995).

Argument type (a)	Canonical case marker (i_{CM})	Input case marker				
		wa	*ga*	*no*	*o*	...
Integral complement	*wa*	1	0	0	0	
	ga	0	1	1	0	
		⋮				
Complement	*wa*	1	0	0	0	
	ga	1	1	1	0	
		⋮				

Table 3: A fragment of the case marker alternation matrix

Note that in the current formulation, we give consideration only to verb sense disambiguation and choose not to independently disambiguate case fillers. Clearly, there is scope to augment this unidirectional verb-driven approach and potentially improve the precision of VSD through intra-case slot *local* and intra-clausal *topical* disambiguation techniques (Yarowsky 1994; Ng & Lee 1996; Leacock et al. 1998), as well as through consideration of domain (Wilks & Stevenson 1998), which are left as topics for future research.

Case marker alternation

Surface case alternates derive from the canonical case marker type(s), and argument status of the target case slot. In the current implementation, this is represented simply as a matrix of possible case marker alternations for each canonical case marker, for a given argument type. This matrix has been developed based around the relative degrees of freedom of case marker alternation outlined in Table 1, in that whereas a complement-type nominative case marker *(ga)* can commonly alternate with any of a range of case markers including the topic *(wa),* and genitive *(no)* markers, an integral complement type nominative case marker can generally alternate only with the genitive marker. A fragment of the connection matrix which documents such potential case marker alternations is provided in Table 3.

From the case marker alternation matrix, we define the binary function *calt* for input case marker i_{CM} and the canonical case marker set T_{CM} for target case slot *t*, such that *calt(i_{CM} ,T_{CM})* is 1 if *connect(i_{CM},j)* = 1 for some case marker $j \in T_{CM}$, and 0 otherwise.

$$calt(i_{CM}, T_{CM}) = \begin{cases} 1 & \text{if } \exists j \in T_{CM}, (connect(i_{CM}, j)) \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

Scoring individual case slot alignments

The score *align* for each case slot alignment *(i,t),* incorporating input case slot *i* (filler i_f and case marker i_{CM}) and target case slot *t* (selectional constraint set t_c and case marker set T_{CM}), is determined via the product of SS_{max} and *calt.*

$$align(i, t) = SS_{max}(i_f, t_c) \cdot calt(i_{CM}, T_{CM}) \qquad (5)$$

3.2 Penalising non-alignment

Equally important as scoring case slot alignments is the enforcement of penalties on unaligned case slots. In this, we treat input and target case slots distinctly.

3.2.1 Potential adjuncthood of unaligned input case slots

When mapping case slots onto the valency frame, we are invariably left with a residue of adjunct case slots. We thus devised a classification of nuclear adjunct types *(temporal=TEMP, locative=LOC* and *adverbial=ADV[4])* and, under the assumption that all adjuncts of same basic type coincide in semantic demarkation, we developed a fixed set of lexical and semantic filters for each. These filters and the associated set of canonical case markers for each adjunct type, are employed as meta-case slots to match non-aligning input case slots against, so as to distinguish between 'dangling' input case slots not aligning at any level of processing, and adjunct case slots which are simply not described within the valency frame proper.

The scoring of adjunct case slot alignment is performed identically to that for other case slots, with each adjunct type being described by a pre-defined region within the Goi-Taikei thesaurus. The final score for the **potential adjuncthood** *PA* of a given input case slot *i* is given as the maximum score for the three given adjunct types.

$$PA(i) = \max_i \Big(align(i, t_{TEMP}), align(i, t_{LOC}), align(i, t_{ADV}) \Big) \tag{6}$$

3.2.2 Non-alignment of target case slots

Non-alignment of target case slots is penalised by way of the mean *CSR* measure for the case slot in question. In this, however, we have to be careful not to over-penalise unaligned adjunct and middle case slots, which we would expect to be readily omissible. This is achieved by multiplying the mean *CSR* value for each unaligned target case slot *t* with the *arg_weight* value corresponding to the argument status of *t,* and taking the maximum such value as the combined penalty for non-alignment of target case slots *(PEN_{targ}).*

$$PEN_{targ} = \max_t arg\_weight(t) \cdot \sum_{i=1}^{n} \frac{CSR(t_i)}{n} \tag{7}$$

Our motivation for taking the maximum here is that we want to identify that unaligned target case slot which is most characteristic (i.e. highly constrained) of the given verb sense, while respecting the potential for that case slot to be genuinely uninstantiated in the case of middles and adjuncts (hence the *arg_weight* factoring), and avoiding over-penalisation of valency frames with higher numbers of case slots.

3.3 Scoring and ranking valency frame mappings

The scores and penalties detailed above are added to return a single combine *score* for each mapping *M*.

$$score(M) = \sum_{\langle i,x \rangle \in M} \Big(align(i, x) \Big) + \sum_{\langle i,\_ \rangle \notin M} \Big(PA(i) \Big) - PEN_{targ} \tag{8}$$

---

[4] Strictly speaking, all three types can also occur as 'extra-peripherals' within the original Somers nomenclature, and what we define as adjuncts corresponds to an amalgam of Somers' adjuncts and extra-peripherals.

The various mapping candidates are then ranked in descending order according to their respective scores.

4 Evaluation

We evaluated our system on all verbs found in the Goi-Taikei valency dictionary which derive from the verb *miru* "to see". By this is meant that, in addition to all dictionary entries for the base verb *miru*. all verbs containing the kanji '見' prefix in their stem were considered. Examples of verbs included in this set are *mieru* "can see/to be visible" and *minaosu* "to reconsider/re-evaluate". The motivation for this seemingly arbitrary choice of verbs is the high degree of lexical ambiguity that exists between them, in the form of full and partial verb homophony (Baldwin 1998). The dictionary composition is as follows:

> Total number verb senses in dictionary: 148
> Number distinct verb stem types: 54
> Average entries per verb stem: 2.72

The 148 verb senses were used in a verb sense evaluation task on a set of 289 simplex clauses extracted from the EDR corpus (EDR 1995), with each extracted clause having a main verb lexically matching one or more verbs in the dictionary. Unfortunately, the verb sense indices used in the EDR corpus do not align well with those designated in the Goi-Taikei valency dictionary, such that all 289 clauses had to be manually annotated for both verb sense and case slot alignment. Because the Goi-Taikei verb senses are linked to unique English verb translations, appropriateness of the English translation was used as the sole criterion in sense annotation. Hence, verb sense annotation was construed as an interlingual lexical selection task. The labour overhead involved in this annotation severely restricted the size of the test corpus, and results given below should be interpreted in light of the limited scope of the evaluation task.

The reader is additionally cautioned that, in this evaluation task, all inputs were both automatically clustered into case-marked phrase units (i.e. input case slots) and segmented (for automatic head determination purposes) according to the original EDR mark-up.

> Total number clauses in corpus: 289
> Total number fixed sense clauses: 23
> Coverage of distinct verb senses: 58
> Average number case slots per clause: 1.46

In terms of determining the semantic head of each phrase for calculation of *SS*, we identified the maximum segment-preserving suffix of the overall case filler which matches with a thesaurus entry; for unknown words, *SS* was set to 0.1.

In evaluation, solutions were ranked based upon the overall *score* for that mapping, with solutions of equivalent score down-ranked. An output was adjudged to be correct if and only if it coincided both in verb sense and case slot alignment, with those annotated for the input in question.

Evaluation of the test corpus in the manner described above, with *PA* and PEN_{targ} variously activated and deactivated *(±PA* and *±TARG,* respectively), produced the results given in Fig. 2. Note that the *arg_weight* schema given in Table 2 was applied consistently for all given system configurations. BEST-*n* in Fig. 2 indicates the

percentage of clauses for which the correct solution was found in the top *n* ranked outputs.

As a point of reference, the native ALT-J/E parser was run over the same corpus and *sense-level* accuracy (ignoring correctness of case slot alignment) calculated on the same BEST-*n* scale. The ALT-J/E verb sense scoring mechanism can be likened to a coarse-grained version of our *CSR,* weighted according to the part-morphological, part-semantic "case-role" of the target case slot (Bond & Shirai 1997). Unfortunately, we were unable to get ALT-J/E to work on the pre-parsed input format required by our system, such that inputs had to be given as unformatted raw text. This tainted performance in that any internal errors in morphological and syntactic analysis led to noise during verb sense scoring. Having said this, the restricted length of the simplex clause inputs is thought to have minimised the scope for parsing errors.
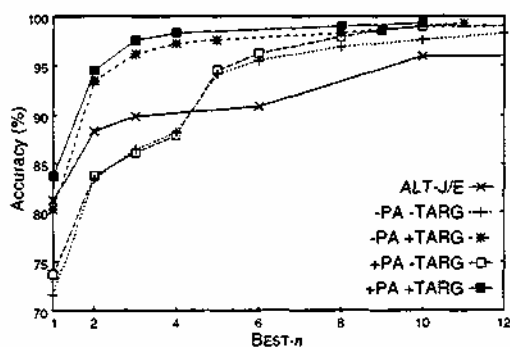

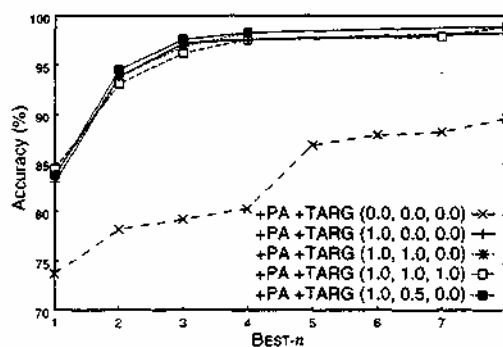Figure 2: BEST-*n* accuracy over different system configurations

Figure 3: System accuracy for differing *arg_weight* values

The correct solution (over both verb sense and case slot alignment) was located at a mean rank of 1.26 out of an average of 6.84 outputs, with the correct verb sense (irrespective of case slot alignment) identified at a mean rank of 1.22. The results show that we are able to identify the correct case slot alignment for the correct verb sense in over 83% of cases, in the instance that both *PA* and PEN_{targ} are activated. Further, in nearly 98% of cases, the correct alignment and verb sense were contained in the top three ranking system solutions. Close inspection of the relative accuracy with component scoring mechanisms deactivated, reveals that the use of PEN_{targ} avails a performance gain around 10% for $n \leq 5$. Similarly, *PA* produces significant gains, particularly when employed in tandem with *PENtarg.*

Despite the overall high accuracy rates attained, there were two clauses for which the correct analysis could not be produced due to lexical obligatoriness constraints on target case slots not being met. This represents an over-constraint on the part of the Goi-Taikei dictionary, and is not perceived as a threat to the integrity of our system.

Looking again to Fig. 2, we notice that our system outperformed the ALT-J/E parser despite the more stringent criterion of both alignment and sense accuracy imposed on it (vs. simple sense accuracy for ALT-J/E). Indeed, the relative performance of the two systems is perhaps better reflected in the figures of 86.51% vs. 81.22% — the simple sense BEST-1 accuracy for our system against that for ALT-J/E. As an additional measure of true performance gain, we additionally calculated the accuracy of a naive "first-sense" technique on the test corpus. Here, we simply return the most common/prominent verb sense lexically matching with the input, as defined within

Goi-Taikei. This produces a baseline accuracy of 81.66%, underlining the significance of the 86.51% sense accuracy for our system.

While these figures are promising, they do not directly verify the validity of our claim that argument status aids VSD. We thus additionally tested the system (*+PA + TARG*) with various *arg_weight* parameter settings, indicated in Fig. 3 as the triple (*complement_weight, middle_weight, adjunct_weight*), with the inequality

$$1 \geq complement\_weight \geq middle\_weight \geq adjunct\_weight \geq 0$$

upheld in all cases. The most striking result in Fig. 3 is the degradation in performance for the parameter triple (0.0,0.0,0.0), that is where all case slots are treated as adjuncts. The principle reason for this fall-off in performance is the blocking of semantic backing-off for complements and middles. There is little separating the remaining parameter combinations, although the proposed (1.0,0.5,0.0) setting did marginally outperform other settings at all positions other than BEST-1. This suggests that we get better performance when we treat middles as an in-between category than when conflating them with either complements or adjuncts, and provisionally backs up Somers' claim as to a need for this extra category. Part of the reason for the minimal differentiation between the different *middle_weight* scores is that the particular verb types targeted in evaluation tended not to collocate with middle case slots, suggesting the need for further evaluation on 'middle-heavy' verbs.

One effect not apparent above is the relation between the number of input case slots and accuracy in analysis. We thus further evaluated the influence the number of input case slots had on the average rank, with results given below. Interestingly enough, we get a 100% success rate for input case frames devoid of case slots, pointing to an ability to cope admirably with underspecification. There is then a drop in accuracy for a single case slot in the input, with gradually recovers with increasing numbers of input case slots.

No. input case slots	No. clause instances	Ave. rank
0	14	1.00
1	172	1.33
2	87	1.20
3	14	1.21
4	2	1.00

An additional item worthy of verification is the ability of the system to correctly discriminate between fixed and general sense. Analysis of the 23 clauses containing a verb of fixed sense revealed that the system returned the correct analysis with the highest rank in all cases, and that there were no instances of a fixed sense solution being returned for a general-sense verb. While this certainly bodes well, these results should perhaps be played down, as all clauses of fixed sense were very clearly so. The performance of the system on more borderline examples of fixed expressions thus remains to be determined.

5 Discussion

In this research, we distance ourselves from much of conventional VSD research in that we specifically set out to handle cases of underspecification and non-canonical input. This is particularly salient in the case of Japanese due to the high levels of zero anaphora

and commonality of case marker alternation, as are not found in English. It is thus difficult to identify comparable results in the VSD literature, although Kurohashi & Nagao (1994) and Fujii (1998) cite average sense-level accuracies of 76.5% and 82.3%, respectively, the former on underspecified and the latter on fully instantiated inputs. In this respect, our performance levels would appear to improve on previous research.

One other important distinction between our work and much VSD research is that we do not explicitly employ the case-role, case marking or grammatical relation of a case slot to weight it. While we do utilise argument status to weight occurrences of semantic backing-off and penalise non-alignment of target case slots, we do not differentiate between aligned case slots other than implicitly through the *CSR* score. In this, we distance ourselves from the heuristical, static formulation of ALT-J/E and also the observations of Yarowsky (1993) that certain syntactic relations derive more disambiguating evidence than others (e.g. objects are better disambiguators of verbs than subjects). We do claim, however, that *CSR* is able to dynamically capture the types of phenomena targeted in these handlings.

To sum up, the verb sense disambiguation method presented here was devised around the Goi-Taikei valency dictionary, as a means of overcoming problems related to underspecification, case marker alternation, word order, and fixed expressions, primarily through the application of argument status-based weighting and analysis of the selectional constraints encoded within the original dictionary. Argument status was successfully applied in semantic backing-off techniques, penalisation of input and target case slot non-alignment, and prediction of case marker alternation. The system returned an accuracy of over 83% on a limited test set of 289 simplex clauses.

Areas of further research include the need to expand evaluation of the system, possible integration with disambiguation of case fillers, expansion of the handling of fixed expressions, and consideration/handling of the effects of scrambling.

References

Baldwin, T.: 1998, 'Relative clause coordination and subordination in Japanese', in *Proceedings of the Australian Natural Language Processing Postgraduate Workshop,* pp. 1-10.

Bond, F. & S. Shirai: 1997, 'Practical and efficient organization of a large valency dictionary', Handout at the *Workshop on Multilingual Information Processing, held in conjunction with NLPRS'97.*

EDR: 1995, *EDR Electronic Dictionary Technical Guide,* Japan Electronic Dictionary Research Institute, Ltd., (In Japanese).

Fujii, A.: 1998. 'Corpus-based word sense disambiguation', Ph.D. thesis, Tokyo Institute of Technology.

Ikehara, S., M. Miyazaki. A. Yokoo, S. Shirai, H. Nakaiwa, K. Ogura, Y. Ooyama & Y. Hayashi: 1997, *Nihongo Goi Taikei – A Japanese Lexicon,* Iwanami Shoten. 5 volumes. (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).

Kameyama, M.: 1985, 'Zero anaphora: The case of Japanese', Ph.D. thesis, Stanford University.

Kurohashi, S. & M. Nagao: 1994, 'A method of case structure analysis for Japanese sentences based on examples in case frame dictionary', *IEICE Transactions on Information and Systems,* **E77-D**(2): 227-39.

Leacock, C., M. Chodrow & G.A. Miller: 1998, 'Using corpus statistics and WordNet relations for sense identification', *Computational Linguistics,* **24**(1): 147-65.

Mahesh, K., S. Nirenburg & S. Beale: 1997a, 'If you have it, flaunt it: Using full ontological knowledge for word sense disambiguation', in *Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation (TMI '97),* pp. 1-9.

Mahesh, K., S. Nirenburg, S. Beale, E. Viegas, V. Raskin & B. Onyshkevych: 1997b, 'Word sense disambiguation: Why statistics when we have these numbers?' in *Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation (TMI '97),* pp. 151-9.

Meyers, A., C. Macleod & R. Grisham: 1996, 'Standardization of the comple-ment/adjunct distinction', in *Proceedings of the Seventh Euralex International Congress,* (ftp://cs.nyu.edu/pub/html/comlex.html/adj-compl.ps).

Ng, H. T. & H. B. Lee: 1996, 'Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach', in *Proceedings of the 34th Annual Meeting of the ACL,* pp. 40-7.

Palmer, M. & Z. Wu: 1995, 'Verb semantics for English-Chinese translation', *Machine Translation,* **10**: 59-92.

Shirai, S., S. Yokoo, H. Inoue, H. Nakaiwa, S. Ikehara & A. Yagi: 1997, 'Nichi-ei kikai-hon'yaku ni-okeru imi-kaiseki no tame no kōbun jisho [A structural dictionary for semantic analysis in Japanese-English machine translation]', in *Proceedings of the Third Annual Meeting of the Japanese Association for Natural Language Processing,* pp. 153-6, (In Japanese).

Somers, H. L.: 1984, 'On the validity of the complement-adjunct distinction in Valency gram-mar', *Linguistics,* **22**: 507-30.

Somers, H. L.: 1987, *Valency and Case in Computational Linguistics,* Edinburgh University Press.

Wilks, Y. & M. Stevenson: 1998, 'Word sense disambiguation using optimised combinations of knowledge sources', in *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguis-tics: COLING/ACL-98,* pp. 1398-1402.

Yarowsky, D.: 1993, 'One sense per collocation', in *Proceedings of ARPA Human Language Technology Workshop,* pp. 266-71.

Yarowsky, D.: 1994, 'Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French', in *Proceedings of the 32nd Annual Meeting of the ACL,* pp. 88-95.