社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
NLC98-13 (1998-07)

動詞多義性解消における語彙交替現象

Timothy Baldwin, 徳永健伸, 田中穂積

東京工業大学大学院情報理工学研究科

概要

本研究では「語彙体系結合価辞書」などのような選択制約ベースの辞書を用いた動詞多義性解消システムを提案する．本手法では，各々の格要素に対し，選択制約がどの程度厳密に働くのかといった度合をシソーラスの木構造により評価する．これにより，格要素を省略された事例に対しても語義確信度を効率良く計量でき，語順・表層格交替にも容易に対応できる．「見る」を中心とする 148 個の動詞辞書を用い，同動詞を含む 291 個の節を多義性解消したところ，92.10% の正解率が得られた．

Lexical Effects in Verb Sense Disambiguation

Timothy Baldwin, Takenobu Tokunaga, Hozumi Tanaka

Tokyo Institute of Technology

{tim,take,tanaka}@cs.titech.ac.jp

Abstract

First, we propose a unified framework for evaluating verb sense in a selectional restriction-based dictionary architecture, including both generalised and fixed verb senses. The proposed methodology is primarily based around the restrictiveness of the selectional restrictions on each case slot, evaluation of which is made through calculation of the average depth of semantic classes governing the case slot in question. As a spin-off of this restrictiveness measure, we develop the ability to predict likelihood of sense in the case of underspecification, and go on to combine the measure with methods to variously handle word order and surface case variation. By way of experimentation, a test dictionary of 148 verb entries lexically deriving from the Japanese verb *mi(-ru)* "(to) see", is used to empirically evaluate the effectiveness of the overall system on an open test set of 291 clauses. The overall accuracy on this test set is 92.10%.

1 Introduction

Word sense disambiguation (WSD) refers to the task of identifying the sense of a term within a given lexical context, from within a set of sense candidates. Our particular interest in WSD lies in verb sense disambiguation (VSD), which in the case of Japanese, is directly equivalent to full WSD in terms of complexity, as finite-inflected verbs cannot occur as any other part-of-speech (POS), hence avoiding overlap with POS disambiguation ("tagging"). The particular task we have focused on is, given a clausal input, identification of the correct verb sense for the main verb, and successful alignment of the *input case slots* within the input, and *target case slots* within the valency frame described in the dictionary entry for the correct verb sense.

In performing any research in WSD, there is clearly a requirement for some means of sense distinction/demarkation, and in our case this is based around the Goitaikei pattern-based valency dictionary (Ikehara *et al.* 1997; Shirai *et al.* 1997b), as developed by NTT for their ALT-J/E machine translation system. A valency frame entry in the Goitaikei valency dictionary is represented as a list of case slots, each of which is provided with a set of class-based selectional restrictions and lexical filler candidates, or alternatively a set of fixed argument-type lexical fillers. As such, the mechanism for determining sense compatibility between an arbitrary input and each individual dictionary entry is encoded within the dictionary entries themselves, removing the need to introduce local preferences between senses; however at the same time, the selectional restrictions and lexical candidates frequently overlap in coverage. This brings about the need for an interface to the dictionary which is able to select between the potentially sense-compatible candidates.

The Goitaikei valency dictionary contains a mixture of both general-sense verb entries where the combined semantics of the verb and its arguments are determined monotonic compositionally from the parts ("generalised entries"), and idiomatic-sense verb entries where the combination of any fixed arguments and the verb produce a particular sense often unaccessible from the parts ("fixed entries"). To give an example of each entry type, *eiga-o mi(-ru)* "film-ACC see"(to) see a film is a generalised entry, and *baka-o mi(-ru)* (idiot-ACC see) "(to) make a fool of oneself" is a fixed entry. Note that the only requirement on fixed entries is that they have at least one fixed argument, that is at least one obligatory argument, fixed as to lexical content, and that non-fixed and fixed case slots generally co-exist within fixed expression valency frames. Whereas verb sense disambiguation techniques to date have largely focused on generalised verb types and semantic effects arising within that restricted context[1], this research attempts to utilise the full coverage of the Goitaikei valency dictionary and propose a common means of selecting not only between generalised entries, but also between the fixed and generalised entry types.

A feature of the Goitaikei valency dictionary is that verb entries are ranked according to likelihood of sense/generality, in an attempt to counter any overlap in coverage between verb senses, and also the tendency for relatively unrestricted argument ellipsis in Japanese. Thus, given a bare verb compound as input, the suggestion is that we should apply this default ranking of entries to select the most plausible sense. While this is certainly efficient and goes part way to solving the various problems arising from zero anaphora in Japanese, our suggestion is that plausibility of sense is inherently encoded within the selectional restrictions associated with each valency frame, in the absence of additional context. Thus, that a Japanese speaker will automatically associate a bare instance of *mi(-ru)* with the 'basic' sense of "(to) see", is symptomatic of this being the broadest, least idiosyncratic sense of *mi(-ru)*, and it is this fact which makes this sense most salient rather than there being some arbitrary cognitive ranking of sense. That is, default sense in the case of underspecification is intrinsically linked to maximal generality of sense of the sense candidates, and this fact should be modelled by any VSD system.

Additionally, each case slot is provided with a list of canonical surface case markers, but no indication of the extent to which surface case alternation is expected to occur. It is perhaps unreasonable to expect the dictionary to contain a full listing of surface case variations, but still we are left with the problem of just how strictly we should adhere to the surface case range specified. There would also appear to be no hard and fast rule to designate exactly how freely case alternation occurs, except that most case markers alternate with the topic (*wa*) and iterative (*mo*) case markers. One observation which can be made is that there seems to be a correspondence between the argument status of the marked case slot (see Section 2.2), and freedom of alternation with different case marker types. For example, for complements, canonical nominative case can frequently be replaced by ablative (*kara*) case, particularly for subjects. However, in the case of a fixed argument (integral complement – see below), little or no surface case alternation may be possible, as is the case with the accusative marking of *asi* "leg" in *asi-o ara(-u)* "to cut off connections with".

One further concern which permeates any Japanese parsing task is the relative freedom of 'word order', or within the terminology of valency frame mapping, freedom of 'case slot order'. That is, all permutations of general-sense case slots are generally grammatical, each with varying degrees of markedness. The most efficient and robust method of dealing with free case slot order is to completely ignore the original (unmarked) ordering within the canonical valency frame. However, we suggest that in cases of verb sense ambiguity, there is a trade-off between departure from canonical case slot ordering, and certainty of sense satisfaction. We model this trade-off through use of the bubblesort algorithm in calculating the degree of markedness as compared to canonical case slot ordering.

To summarise, the main purpose of this research is to present a uniform means for verb sense disambiguation, irrespective of the expressional type of the associated valency frame (fixed or otherwise), and independent of static inter-entry preferences. That is, rather than adopting the conventional preference for fixed entries/idioms over standard entries, and stipulating a default verb sense, we in-build dynamic entry preference determination into the appraisal of the sortal/lexical preferences associated with each case frame (Section 3.1). Further, we introduce a system for penalising non-alignment of the input and target case frames, both in terms of case slot ordering and non-alignment of individual case slots. Finally, we integrate these various proposed measures in a single score for likelihood of sense.

2 Definitions

2.1 Case slot obligatoriness

Case slot obligatoriness in Japanese falls into two categories: *lexical* obligatoriness and *argument* obligatoriness. Lexical obligatoriness indicates the necessity for that case slot to appear lexically, in order for the given sense to be trigerred. Lexically obligatory case slots can be either fixed arguments such as *asi* in *asi-o ara(-u)* (see above), or general-sense arguments such as the direct object position for the verb *nobe(-ru)* "(to) say". Argument obligatoriness is a statement of the inherent requirement for some entity to perform that role within the valency frame. All lexically obligatory case slots are thus also argument obligatory. Most subject case slots

[1] This is with the notable exception of (Ng and Lee 1996), although even here, fixed entries are represented as a fixed lexical pattern.

	Integral Complement	Complement	Middle	Adjunct
CASE ALTERNATION	almost none	high	medium	low
CASE SLOT ORDER	highly restricted	free, but produces markedness	relatively free	free
OBLIGATORINESS	lexically oblig.	argument oblig. or opt.	argument opt.	argument opt.

Table 1: Argument status matrix

Argument type (a)	Canononical case marker (cm_i)	Input case marker					
		TOP	NOM	GEN	ACC	ABL	...
Integral complement	TOP	1	0	0	0	0	
	NOM	0	1	1	0	0	
⋮							
Complement	TOP	1	0	0	0	0	
	NOM	1	1	1	0	1	
⋮							

Table 2: A fragment of the case marker alternation matrix

are argument obligatory but not lexically obligatory.

Goitaikei relies heavily on lexical obligatoriness, to the extent that each case slot has a flag to indicate its lexical obligatoriness. We apply this information directly in ensuring that such lexical obligatory items are instantiated in mappings onto that verb sense. This information is maintained independently of argument status, however, and in implementation terms, fixed expressions are treated no differently from other lexically obligatory case slots. Indeed, no special treatment is applied to lexically obligatory case slots other than a check for lexical instantiation.

2.2 Argument status

This research draws partially on the argument status of case slots in predicting aptitude for surface case alternation, and also in penalising occurrences of non-aligning case slots (Section 3.2) and determining freedom of case slot placement (Section 3.2.3). The range of argument types utilised is a simplified version of the sixfold scale proposed by Somers (1984, 1987), and consists of *integral complements*, *complements*, *middles* and *adjuncts*. Integral complements correspond to fixed arguments in the Goitaikei framework, are highly restricted as to scope for surface case alternation and case slot order alternation, and are lexically obligatory; complements, as major verb constituents, generally display high scope for surface case alternation, produce marked semantics upon case slot permutation, and can be either argument obligatory or optional; middles form an 'in-between' category between complements and adjuncts, are generally reasonably restricted in terms of surface case alternation, are relatively free in case slot positioning, and are argument optional; and adjuncts are the most peripheral argument type, being highly restricted in terms of surface case alternation, insertable freely into any position within the case frame, and argument optional. A summary of the characteristics of each argument type is given in Table 1.

Argument status was not indicated within the original Goitaikei representation, but was instead deduced from the grammatical relation mark-up of each case slot.

3 Overall system

Our VSD system is founded on the precept of the restrictiveness of each case slot, a measure which is used to multiplicatively weight alignments of that case slot, and penalise non-alignments. Initially, a lattice is produced, describing all possible mappings between input and valency frame case slots, with a plausibility attached to each. These preliminary mappings are determined without regard to correspondence of case marker type, and subsequently verified according to the combination of the input case marker, the argument status of the target case slot, and the associated set of canonical case markers. The full set of isomorphic mappings between input and valency frame case slots is then extracted, and each candidate mapping penalised for non-aligned input and target case slots, respectively, and disparity in canonical case slot ordering.

3.1 Case slot restrictiveness

All non-fixed case slots are encoded with a set of selectional restrictions, described as a set of semantic classes indexed to the Goitaikei thesaurus (Ikehara *et al.* 1997). The Goitaikei thesaurus is set up in a conventional tree structure, of non-uniform depth, and with lexical entries contained at both leaf and non-terminal node positions. So as to establish a measure of restrictiveness of each position in the thesaurus structure, the depth of the subtree subsumed under each node was calculated recursively, with leaves assigned a depth of one.

Details of the tree structure are given below. Note that lexical ambiguity is modelled within the thesaurus by multiple occurrences of the lexical item within the thesaurus structure, with the average number of nodes per lexical item as indicated.

Total no. nodes: 2715
Total no. non-terminal nodes: 794
Total no. leaves: 1921
Ave. depth at root: 8.26
Ave. subtree depth: 1.42
Ave. nodes per lexical item: 1.42

The degree of case slot restrictiveness (CSR) of case slot x is estimated as the inverse of the average depth of the set of thesaurus classes SR_x associated with x.

$$CSR(x) = \frac{n}{\sum depth(t \in SR_x)} \quad (1)$$

Hence, CSR for a leaf is one, while CSR for a case slot of unrestricted selectional scope (i.e. with selectional restriction set to the root node) is $\frac{1}{8.26} \approx 0.12$; CSR is set to one for fixed case slots and optional case slots devoid of selectional restrictions.

Satisfaction of selectional restriction

In the current formulation, we give consideration only to verb sense disambiguation, and choose not to disambiguate other lexical items. In terms of determining the satisfaction of selectional restrictions, this translates to comparing all possible senses of the input lexical item in question against the set of selectional restrictions, and determining what proportion of the possible senses fall within the scope thereof. Clearly, it would be possible to perform more exacting VSD if we were able to disambiguate the sense of case fillers either prior to or as a component of the VSD process, and this is left as a matter for future research.

The degree of satisfaction of selectional restriction (SS) can be formalised as in (2), where lexical item l with senses $s_{1...n}$ is fed into the binary function *class_sat*, which determines whether each sense s_i falls within the sub-tree(s) subtended by the selectional restriction set associated with case slot SR_x.

$$SS(l, SR_x) = \frac{\sum_{i=1}^{n} class\_sat(s_i, SR_x)}{n} \quad (2)$$

where

$$class\_sat(s, SR) = \begin{cases} 1 & \text{if } \exists t \in SR\{s \in t\} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Valency frame mapping determination

In determining mappings between the input case slots and target valency frame for each verb sense, the input case slots are individually evaluated against the selectional restrictions applied to each target case slot, and the resultant degree of satisfaction of selectional restriction (SS) determined. Figure 1 depicts this process, with all linked input/target case slot pairs indicating a positive SS value for that edge. The lattice is then used to extract all isomorphic mappings between input and target case slots, such that there is a maximum of one edge connected to each case slot in the input and target valency frame. These mappings are then considered in turn, to determine the lexical correspondence between case markers for aligned case slots.

For Figure 1, thus, extracted isomorphic mappings would include $(<1,\beta>, <2,\alpha>)$, $(<1,\alpha>)$ and $()$ – the empty string. Note that these mappings describe only cases of case slot alignment between the input and target valency frame, and thus that the empty string indicates that no case slots align.

Case marker alternation

Freedom of surface case alternation varies according to the canonical case marker type(s), and argument status of the target case slot. Ideally, we would like to be able to weight case alternations accordingly, but in the current implementation, we simply provide a matrix of possible case marker alternations for each canonical case type, for a given argument type. This is based around the relative degrees of freedom of case marker alternation for each argument type outlined in Table 1, in that whereas a complement-type nominative case marker (*ga*) can commonly alternate to any of a range of case markers including the topic (*wa*), genitive (*no*) and ablative (*kara*) markers[2], an integral complement type nominative case marker can freely alternate only to the genitive marker, and potentially to the topic marker. A fragment of the connection matrix which documents such potential case marker alternations is provided in Table 2.

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

$$calt(cm_i, \{cm_x\}) = \begin{cases} 1 & \text{if } \exists cm_j \in \{cm_x\}, (connect(cm_j, cm_i)) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

[2]Nominative–ablative case alternation occurs readily for subjects.

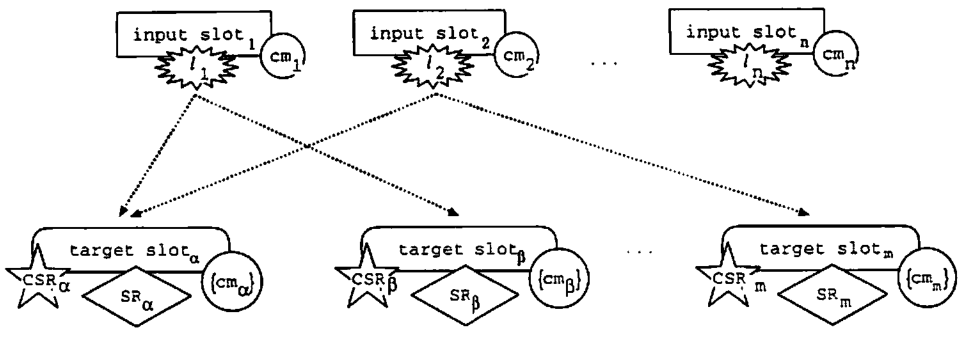

Figure 1: Input/valency frame case slot lattice

Scoring individual case slot alignments

The score *align* for each case slot alignment $< i, x >$, for input case slot i and target case slot x, is determined via the product of CSR, SR and *calt*.

$$align(i, x) = \\ CSR(x) \cdot SS(l_i, SR_x) \cdot calt(cm_i, \{cm_x\}) \quad (5)$$

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 Non-alignment of input case slots

Potential adjuncy

One concern which inevitably arises when attempting to capture the valency content of a verb sense, is the handling of adjuncts. The methodology employed in the development of the Goitaikei valency dictionary has been to describe all complement and middle case slots, and largely avoid overt description of the broad range of adjunct types. This leads to complications in mapping case slots onto the valency frame, as we are invariably left with a residue of adjunct case slots. While this poses no immediate threat to the determination of verb sense through instantiation of the case slots described in the target valency frame, it is necessary to verify the oblique argument status of any non-aligned input case slots, and penalise any unaligned instances of what would appear non-adjunct case slots.

Verification of adjunct status is carried out by way of a series of lexical and semantic filters developed in (Baldwin 1998). The basic methodology employed is to assume that adjuncts occur in temporal or locative sense, or as adverbs, and that restrictions on acceptable lexical fillers for the different types are basically consistent across all verb types. Additionally, while there is slight case marker variation within each adjunct category, case marker alternation generally takes the form of a topic or iterative marker suffix on, rather than replacement of, the basic case marker. This makes it possible to recover the basic case marker content of adjuncts.

Potential adjuncy is detected prior to case slot alignment. In this, the combination of the lexical content l and case marker cm for case slot j is tested by way of the separate adjunct filters, and the combined score recorded. The temporal and adverb type filters are regular expression-based and return a binary judgement of whether a match was found or not, through functions *temp* and *adv* respectively. The locative type filter, on the other, is selectional restriction-based and indexed to a set of Goitaikei thesaurus classes LOC; input case fillers are evaluated by way of $SS(l, LOC)$. Each of the individual scores from the filters are then multiplied with binary case marker compatibility values (with case marker compatibility evaluated through lexical pattern matching), and added to return the potential adjuncy (PA) of that case slot. A ceiling of one is imposed on PA to produce the combined potential adjuncy (CPA).

$$\begin{aligned} PA(l, cm) = {} & temp(l) \cdot tcase(cm) + \\ & adv(l) \cdot adv\_case(cm) + \\ & SS(l, LOC) \cdot loc\_case(cm) \quad (6) \end{aligned}$$

$$CPA(l, cm) = \begin{cases} 1 & \text{if } PA(l, cm) > 1 \\ PA(l, cm) & \text{otherwise} \end{cases} \quad (7)$$

Approximation of other argument types

While it is possible to apply a uniform treatment to detect the general adjunct types, the same does not hold for middles and complements where the only real indication we have of argument status is the case marker. Penalties for each input case marker type were thus established, determined by the rough distribution of usage of each case marker. The nominative case marker, for example, occurs only with complement case slots, whereas the accusative can occur in either complement or adjunct usages.[3] Determination of an average penalty (*inpen*) for each case marker was carried

[3]The accusative marks the perlative locus for motion-based verbs.

out by analysing the basic distribution of case marker usage, and multiplying the relative occurrence in the context of each argument type by a constant penalty (*argpen*) for that type.

Argument status a	Penalty $argpen(a)$
Complement	1
Middle	0.5
Adjunct	0

A selection of the *inpen* penalties for the various case markers are provided below.

Case marker c	Penalty for non-alignment $inpen(c)$
NOM	1.0
ACC	0.6
ITER	0.6
LOC	0.3
nil	0

Combined penalty for non-alignment of input case slots

CPA and *inpen* are combined in determination of the combined penalty PEN_{in} for all non-aligned input case slots i by way of (8).

$$PEN_{in} = \sum_i \left(1 - CPA(l_i, cm_i)\right) \cdot inpen(cm_i) \quad (8)$$

3.2.2 Non-alignment of target case slots

Non-alignment of target case slots is penalised through application of our CSR measure. In this, however, we have to be careful not to over-penalise unaligned adjunct and middle case slots, and to account for the possibility of high levels of argument ellipsis in optional case slots, irrespective of argument status. This is achieved by multiplying CSR for each unaligned case slot x with the *argpen* value (see above) corresponding to the argument status of x. The combined penalty for non-alignment of target case slots (PEN_{targ}) is set to the maximum such value.

$$PEN_{targ} = \max_x argpen(x) \cdot CSR(x) \quad (9)$$

By virtue of the combination of taking the maximum here and providing a trade-off between restrictiveness and argument status, we avoid penalising valency frames with higher numbers of case slots, in cases of argument ellipsis or underspecification. Consider the case, however, that a valency frame v_1 comprises a proper subset of valency frame v_2, and is identical in restrictional content; this occurs with 'conflated transitive' verbs such as ***tabe(-ru)*** which have both a transitive and intransitive sense, with the intransitive valency frame (v_1) being identical to the transitive valency frame (v_2) in all respects except that it lacks a direct object. Here, we will always penalise against v_2 in favour of v_1 given full instantiation of v_1, despite the fact that the direct object may simply have been ellipted due to discourse salience. We provide no means of avoiding this situation in the system as it exists, except to say that recourse to discourse would easily solve this problem, as plugging a suitable argument into the direct object case slot would produce an inflated overall score for v_2 over v_1.

3.2.3 Case slot ordering

Penalty for alternation in case slot ordering is facilitated through the bubblesort algorithm, in which we were inspired by the example-based machine translation algorithm of Shirai *et al.* (1997a). Essentially, we bubble-sort those input case slots that are aligned, to produce the ordering as stipulated in the target valency frame. No additional penalisation of non-aligned case slots occurs in this process, and any exchanges of case slots while sorting are penalised according to the respective argument status of the involved case slots. Degree of exchange penalisation is set in accordance with the various effects predicted for case slot order variation of the different argument types (see Table 1), such that any exchange involving an adjunct goes unpenalised while an exchange between complements is maximally penalised. Here, again, we rely on the *argpen* of the various argument types, in formulating the degree of penalisation between input case slots i and j by way of *swap_pen*:

$$swap\_pen(i,j) = \frac{1}{1 + (argpen(i) \cdot argpen(j))} \quad (10)$$

The penalty returned from $swap\_pen(i,j)$ is multiplied with the current value of $align(i,x)$ on occurrence of an exchange, where x is the target case slot for i. Note that the value for $align(j,y)$, where y is the target case slot for j, remains unchanged in this process.

The final value for each $align(i,x)$ is returned as $ralign(i,x)$.

3.3 Scoring and ranking valency frame mappings

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

$$score(M) = \sum_{<i,x>\in M} \left(ralign(i,x)\right) - PEN_{in} - PEN_{targ} \quad (11)$$

The various mapping candidates are then ranked in descending order according to their respective scores, with **complexity of inflectional content** (Baldwin *et al.* 1998:p. 5) employed as a secondary ascending measure in cases of equal score. Complexity of inflectional content is a count of the component inflectional morphemes in a given verb parse, not including the verb stem. *Mi-a-u* "to see-MUTUAL-PRES", for example, has a complexity of two (the mutual/reciprocal and non-past inflectional morphemes), whereas *mia-u)* "to correspond-PRES" has a complexity of one (the non-past inflectional morpheme. As such, the second parse type of *mia-u* would be ranked above the first given an identical *score* for these two verbs.

4 Evaluation

We evaluated our system on all verbs found in the Goitaikei valency dictionary which derive from the verb *mi(-ru)* "(to) see". By this is meant that, in addition to all dictionary entries for the base verb *mi(-ru)*, all verbs containing a kanji '見' prefix in their stem, were considered. Examples of verbs included in this set are *mimawar(-u)* "(to) inspect/make a round", *miayamar(-u)* "(to) mistake", *mie(-ru)* "can see/(to) be visible" and *minaos(-u)* "(to) reconsider/re-evaluate". The dictionary composition is as follows:

Total no. verbs in dictionary: 148
No. distinct verb stem types: 54
Average entries per verb stem: 2.72

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 *et al.* 1998) produced by varying combinations of verb stems and auxiliary verbs. In the case of the verb set in question, full and partial verb homophony occur when verbs with distinct stem content coincide in lexical form due to combination with auxiliary verb suffixes. This occurs between *m-a(-u)* "(to) see-MUTUAL", with stem *mi-*, and *mia(-u)* "(to) correspond/(be) commensurate", with stem *mia-*.

The 148 verb senses were used in a verb sense evaluation task on a set of 291 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 Goitaikei valency dictionary, such that all 291 clauses had to be manually annotated for both verb sense and case slot alignment. 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. At the same time, however, the evaluation was an open one, and hence results are suggested to be representational of the expected performance of our VSD framework on larger input sets.

Additionally, the reader is cautioned that, in this evaluation task, all inputs were both segmented and clustered into single-level case-marked phrases ('*bunsetsu*') according to the original EDR mark-up. This was in an attempt to minimise noise in the task arising from pre-processing, and attain evaluation data which truly reflects the inherent performance of the VSD system proposed herein.

Total no. clauses in corpus: 291
Total no. fixed sense clauses: 21
Coverage of distinct verb senses: 58
Average no. *bunsetsu* per clause: 1.46

In terms of determining the semantic head of each phrase for calculation of *class_sat*, we used the maximum case filler suffix found in the thesaurus; for unknown words, *class_sat* was set at 0.5, irrespective of the *CSR* of the case slot in question.

In evaluation, solutions were ranked based upon the *score* for that mapping, with outputs sub-ranked through the notion of complexity of inflectional content (see above) in the instance of equality of *score*.[4] An output was adjudged to be correct if and only if it coincided both in verb sense and case slot mapping, with those annotated for the input in question. Thus, in the case of the system identifying verb sense correctly but misaligning input and target case slots within that sense, that solution was marked as incorrect.

Evaluation of the corpus according to the manner given above produced the results given in Table 3, in which BEST-*n* indicates the number and percentage of clauses for which the correct solution was found in the top *n* ranked outputs. The average number of outputs returned for each clause was 5.84, with 1.39 distinct verb parses on average. The correct solution was located at a mean rank of 1.19.

	No. clauses	Proportion of total
BEST-1	268	92.10%
BEST-2	278	95.53%
BEST-3	284	97.59%
BEST-4	285	97.94%
BEST-6	287	98.63%
BEST-7	289	98.97%
BEST-8	290	99.66%

Table 3: Empirical results

The most striking result from this table is that the correct solution was returned with the highest score for over 92% of clauses, and furthermore that approximately 98% of correct analyses were returned within the top three outputs.

Close inspection of the data reveals that for one clause input, the correct solution was missing from the top eight ranked outputs. Indeed, for this clause, the correct solution was not contained among the list of successful parses of the clause, as case slot lexical obligatoriness constraints were not met. This represents an over-constraint in the original Goitaikei dictionary, and is not seen as a flaw in the existing framework.

[4] In terms of ranking the correct solution, this double-measure ranking of outputs proved sufficient to produce a discrete rank in all observed instances.

One effect not apparent from Table 3 is the effectiveness of our system on modelling default sense in cases of underspecification. We thus further evaluated the influence the number of input case slots had on the average rank, with results given in Table 4. Here, it becomes plain that all cases of complete underspecification (zero input case slots) were disambiguated correctly, and that, perhaps surprisingly, we are able to disambiguate cases of complete underspecification more successfully than partial underspecification (one and two input case slots). With the exception of this jump up in average rank between zero and one input case slots, the more complete the input case frame is, the more accurate our method would appear to become.

No. input case slots	No. clause instances	Ave. rank
0	13	1.00
1	174	1.30
2	87	1.02
3	15	1.00
4	2	1.00

Table 4: Average rank vs. input case slot number

An additional item worthy of verification is the ability of the system to correctly discriminate between fixed and general sense. Analysis of the 21 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.

As a final point, we divided clauses into those potentially displaying verb homophony, to check for disparity between results for homophonous and non-homophonous verbs:

Verb homophony?	No. clause instances	Ave. rank
yes	35	2.03
no	256	1.07

Clearly, the system performs at about half accuracy on homophonous verbs as compared to non-homophonous verbs, although the results for homophonous verbs are still relatively high. The largest single source of error for homophonous verbs was the parser incorrectly analysing *mie-ru* "to be visible/able to see-PRES" as the potential compound *mi-e-ru* "to see-POT-PRES". This points to the need for some additional mechanism to penalise unlikely morpheme combinations, possibly cross-referenced to the lexicon.

5 Conclusion

The verb sense disambiguation method presented here was devised around the Goitaikei valency dictionary, as a means of overcoming problems related to underspecification, case marker alternation, word order, and fixed expressions, primarily through analysis of the selectional preferences encoded within the original dictionary. Techniques employed include evaluating the restrictiveness of individual clase slots to use in case slot alignment weighting, penalising non-alignment of both input and target case slots, and calculating deviation from standard case slot ordering. One key element of the proposed system is the argument status (integral complement/complement/middle/adjunct) of each case slot, which is used throughout the penalisation process.

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

References

BALDWIN, T., 1998. *The Analysis of Japanese Relative Clauses.* Master's thesis, Tokyo Institute of Technology.

——, T. TOKUNAGA, and H. TANAKA. 1998. Relative clause coordination and subordination in Japanese. In *Proceedings of the Australian Natural Language Postgraduate Workshop*, 1-10.

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

IKEHARA, S., M. MIYAZAKI, A. YOKOO, S. SHIRAI, H. NAKAIWA, K. OGURA, Y. OOYAMA, and Y. HAYASHI. 1997. *Nihongo Goi Taikei - A Japanese Lexicon.* Iwanami Shoten. (In Japanese).

NG, H. T., and H. B. LEE. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *the Annual Meeting of the ACL*, 40-7.

SHIRAI, S., F. BOND, and Y. TAKAHASHI. 1997a. A hybrid rule and example-based method for machine translation. In *Proceedings of the 4th Natural Language Processing Pacific Rim Symposium 1997 (NLPRS'97)*, 49-54.

——, S. YOKOO, H. INOUE, H. NAKAIWA, S. IKEHARA, and A. YAGI. 1997b. Nichi-ei kikai-honyaku ni-okeru imi-kaiseki no tame no kobun jisho. In *Proceedings of the Third Annual Meeting of the Japanese Association for Natural Language Processing*, 153–6. (In Japanese).

SOMERS, H. L. 1984. On the validity of the complement-adjunct distinction in Valency grammar. *Linguistics* 22.507–30.

—— 1987. *Valency and Case in Computational Lingustics.* Edinburgh University Press.