

An Empirical Architecture for Verb Subcategorization Frame

- a Lexicon for a Real-world Scale Japanese-English Interlingual MT

NOMURA, Naoyuki and MURAKI, Kazunori

NEC Corporation

Miyazaki 4-1-1, Miyamae-ku, Kawasaki-city, JAPAN

nomura@hum.cl.nec.co.jp

Abstract

The verb subcategorization frame information plays a major role of disambiguations in many NLP applications. Japanese, however, imposes difficulties of subcategorizing in part because it allows arbitrary ellipses of case elements. We propose a new type of verb subcategorization frame code set that combines the verb's surface case set and the deep case set, as a solution to the difficulties of empirical researches on Japanese. The lexicon developed by this design has comprehensive information on the correspondences between the surface case frame and the deep case frame, and yet restrains the potential combinatorial explosion of the number of verb subcategorization frames by carefully identifying superficially different frames with an idea of *alternative case markers and semantic roles*, and by introducing the notion of *surface case frame permutations*. The number of different surface/deep case mapping types is 250, after we completed the new subcategorization frame code development for 30,000 verbs and adjectives.

1. Introduction

An NLP system is supposed to be able to recognize the differences and commonalities between almost the same set of words in two different syntactic structures, such as "Mary bit the dog" vs "Mary was bit by the dog." Considering this requirement, the contents in verb subcategorization frames play a major role of disambiguations in many NLP systems.

These frames should reflect linguistic facts of lexical information being concentrated on verbs. There have been, on the other hand, some researches that aim to equally treat lexical information of nouns and that of verbs (see e.g. EDR94). Pustejovsky91 stepped further into formulating useful, intrinsic information in nouns with the notion of co-compositionality among others, so as to recover from an elliptical sentence such as 'he began the book' a default verb such as 'reading.' However, in order to

develop a working NLP system, even these recent researches may presuppose the use of exhaustive coding of verb subcategorization frame knowledge to let the new lexical features be automatically extracted and fully functional in their systems.

Ellipses appear to be a much more serious problem with Japanese than with English because all the supposedly obligatory case elements are virtually free to be dropped or to be placed anywhere in the sentence except for the predicate position at the end. These phenomena seem to have imposed difficulties upon the design of the lexicon so that no list of Japanese verb classes and types comparable to Grishman94, Levin93 or Hornby75 for English was readily available when our project started¹. Thus, we decided to make one by ourselves by a bootstrapping method: to make the initial list of the classification and to make it to grow by developing a working lexicon for MT systems. Upon this empirical study and development of over 30,000 Japanese verbs and adjectives, we propose an architecture for verb subcategorization that represents the mapping information between surface case frame and deep case (thematic role) frame.

The proposal is to serve as a solution to the empirical difficulties with Japanese verbs and case elements described above. The lexicon by this design has comprehensive information on both the surface frame and the deep frame, and the correspondences between them, which are embedded in a code. Yet, the number of codes has been controlled under a manageable figure of several hundreds so that the coding system could evade the potential combinatorial explosion. This is to be done by identifying superficially different case patterns with an idea of *alternative case markers and semantic roles*, and by largely extending the notion and the formulation of voice conversion for Japanese auxiliary verbs and equivalents.

The developed lexicon is adopted in a real world scale interlingua-based MT system that translates between English and Japanese (Muraki87). Our aim

¹ Martin75 & FM&T85 contain some lists, but is too partial for the purpose of developing an MT system.

here is to show an empirical result of the development and analysis of the lexicon from the point of view of space complexity order (cf. Jackendoff 1990&93). In the following section are described the major linguistic requirements of the architecture, the case elements of which are free of word ordering and can increase in number when their voice is converted. The architecture that combines the verb surface case frame and deep case frame is described in section 3, followed by extended mechanisms for applying what we generalized from voice conversion phenomena triggered by auxiliary verbs. We, then, describe the lexicon structure for ambiguity representations in relation to word senses. Finally, we present some statistic figures from the results of the lexicon development and confirm that the proposed architecture and the code system can empirically constrain the potential combinatorial explosions of the verb subcategorization frame representation varieties.

2. Linguistic Requirements of Verb subcategorization Frames

The most notable syntactic phenomenon of Japanese is so-called scrambling. Any verb-modifying NP in a simple sentence in Japanese can appear at any position or does not have to appear at all, regardless of its surface case and deep case. In other words, word ordering is almost free for the major syntactic elements in a Japanese simple sentence except for the predicate itself, which is to be placed at the end of the sentence. Not the word ordering but case postpositions mark the case of these NPs in relation to the main verb. All the examples in e.g. 2-1 lead to the same event structure interpretation, which is shared by the English translation "X gives Y to Z."²

- e.g. 2-1
- a. ok "X-ga Y-wo Z-ni age-ru"
 - b. ok "Y-wo X-ga Z-ni age-ru"
 - c. ok "Y-wo Z-ni X-ga age-ru"
 - d. ok "X-ga Z-ni Y-wo age-ru"
 - e. ok "Z-ni X-ga Y-wo age-ru"
 - f. ok "Z-ni Y-wo X-ga age-ru"

Particles "-ga", "-wo" and "-ni" are the most frequently used case postpositions that often mark nominative, accusative, and dative cases, respectively³. There are various other alternative postpositions that can replace or be added to these case postpositions. Discourse representing postpositions such as "-ha", which had been erroneously treated as general subject

² The differences concern the emphasis and scope that are to be handled by a pragmatics system.

³ "-ga" can sometimes mark accusative case and "-ni" can mark various other cases and semantic roles including "locative" and "time."

marker, could replace "-wo" and "-ga", and could be added to "-ni". These replacement or addition do not alter the basic event structure of the sentences (2-1 a., b., ..., f.), but sometimes just add ambiguities to the syntactic structure as is shown in e.g. 2-2.

- e.g. 2-2
- a. ok "X-ha Y-ha Z-ni age-ru"
 - b. ok "Y-ha X-ha Z-ni age-ru"

Practical Japanese sentence analyzers would need some semantic inference and default inference to plausibly identify X and Y using the semantic restrictions on each case element and the standard word ordering. Most nominative cases in Japanese verbs including "age-ru" ('to give') have strong preference for human/animate attribute so that a meaningful difference between semantic similarity of X to animate object and the similarity of Y to other kind of concrete object leads to allocate nominative case on X and accusative case on Y in either 2-2 a. or 2-2 b. If there is no such difference in the semantic restriction score, the standard word ordering "-ga," "-wo," and "-ni" seems to let the listener to interpret 'e.g. 2-2 a' as 'e.g. 2-1a'.

Japanese not only has typical voice conversions such as passivization but also appears to have similar phenomena that alter the surface case markings such as the cases with causative construction. This variation roughly corresponds to the variety of English auxiliary verbs and higher predicate verbs such as "dekiru (can)," "rareru (be *pp./I* can)," "kotoga-dekiru (be able to)," "tai (want to)," "seru (make/let)," "garu (feel *complement*)" etc.

- e.g. 2-3
- a. X-ga Y-wo tabe-ru. b. X-ni Y-ga tabe-rareru.
 - X-NOM Y-ACC eat. X-DAT Y-NOM eat-PASSIVE
 - "X eats Y" "X can eat Y" xor "X is eaten by Y"

As is observed in e.g. 2-3, the nominative case marker "-ga" turns into dative case marker "-ni" and the accusative case marker "-wo" turns into nominative case marker "-ga" when the *passive/potential* auxiliary verb "rareru" is attached.

Multiple voice conversions can often occur for a single verb phrase as is shown in e.g. 2-3c.

- e.g. 2-3 c. X-ga Y-wo Z-ni tabe-sase-rare-taku-nai.
- X-NOM Y-ACC Z-DAT eat-CAUS-PASS-WANT-NOT.
- "X does not want to be forced to eat Y by Z"

Since three auxiliary verb forms CAUS, PASS and WANT appear by this ordering in e.g. 2-3 c, a simple, natural solution to correctly recognize the scope of complex modality features is to recursively apply the permutations of surface case set as is described in the following section.

3. Combining Surface and Deep Frames

3-1. Basic Representation for Japanese Verb Subcategorization Frame

Empirical studies as we observed in the previous section have suggested that combining syntactic and semantic frames could lead to an optimum efficiency of lexicon descriptions. Thus we created the basic description framework in the verb lexicon as is shown in fig.1.

Slot name:	WITH FROM BY/IN NOM2		
	NOM	ACC	DAT
Surface case frame	GA/deha(pl) <animate>	WO	NI/he <time>
Deep case frame	AGenT	PAT-ient	GOAL

Fig.1 Subcategorization Frame for "ageru"

The example content in Fig.1 is that of the verb "ageru (give)" used in e.g. 2-1 and e.g. 2-2. A natural language analyzer in an MT system is supposed to convert the case elements in e.g. 2-1 a,...,f, X, Y and Z to AGenT, PATient and GOAL, respectively, as shown in e.g. 3-1.

- e.g. 3-1 a. ok "X-ga Y-wo Z-ni age-ru"
 b. ok "Y-wo X-ga Z-ni age-ru"

X = John -> AGenT ; Y = the book -> PATient ;
 Z = Mary -> GOAL

The analyzer looks up the slots in Surface case frame and find the match of the case postposition; for "-ga", GA in the NOM case slot matches and the deep case that is stored in the NOM slot is taken out from the subcategorization frame. The analyzer checks if the semantic restriction 'animate' matches the case element X (John). If it fails, the analyzer looks for other slots, the other subcategorization frame of the same verb, and then the frames of other verbs that appear in the different place of the sentence.

Fig. 1 shows a fixed frame with seven case slots, and this is exactly what the record format of our Japanese lexicon is. Why is it not necessary to have more slots though we know there are definitely more than seven case postpositions in Japanese? One of the reasons⁴ is that other postpositions that can be mapped into a thematic role are supposed to fall into either of the seven slot and take the position as the alternative case markers. For example, "deha" in Fig.1 could only be used with animate plural nouns such as "kotira ('our side')," but it certainly could mark the nominative case.

⁴ The other reason is that there are case postpositions that are not mapped into thematic role. They constitutes not argument structure but just adjuncts (free elements) as is explained in modern Linguistics.

The alternative case postposition "deha" also complies with the Unique Case Principle⁵ that prohibits other case elements from filling the same slot as NOM that is already filled by "X-deha". This is why this use of case postposition "deha," with a different semantic restriction, is supposed to occupy the same slot NOM with the major case postposition "-ga". Another slot DAT in Fig.1 shows that "-he" could replace the major case postposition "-ni" and be assigned the thematic role GOAL. Again, the following ungrammatical example e.g. 3-2 that violates the Unique Case Principle shows that "-ni" and "-he" for the verb "ageru" have to share the same slot in the subcategorization frame.

e.g. 3-2 *X-ga Y-wo Z-ni Z'-he age-ta.

X-NOM Y-ACC Z-DAT Z'-DAT give-PERFective
 "X gave Y to Z'."

For a simple Japanese analyzer that tries to fill as many slots as possible for a verb, the Unique Case Principle is virtually embedded in the subcategorization frame of our architecture for the computational lexicon.

The slot name NOM2 represent the typical case with two-term adjectival predicates⁶ that require two nominative cases.

e.g.3-3 "X-ga Y-ga sukina-no-ha.."
 X-NOM
 "That X likes Y is..."

3-2. Generating Permutational Subcategorization Frame Triggered by AUX Verbs

We have generalized the notion of voice conversion for Japanese auxiliary verbs and equivalents by abstracting 14 codes of case frame permutation. These codes, the contents of which are to be elaborated in the fig. 2 & 3, are assigned on the extended category of auxiliary verbs "dekiru (can)," "rareru (be pp./ can)," "kotoga-dekiru (be able to)," "tai (want to)," "seru (make/let)," "garu (feel complement)". Below is the description of the procedure by which the Japanese analyzer performs the permutation of the verb subcategorization frame.

When the morphological analyzer detects an auxiliary verb or an equivalent while checking the information contained in the predicate phrase, the analyzer develops the verb subcategorization frame from the code in the verb's lexicon and read from the

⁵ The Unique Case Principle in Case Grammar and empirical studies is formulated and explained by the Lexicalist Hypothesis about thematic roles and the X-bar Theory in the school of Universal Grammar (Chomsky88).

⁶ There is only one verb "komaru (be in trouble)", the active voice of which shows two nominative cases ("GA"s).

auxiliary verbs' lexicon what we call the Surface Case Permutation Frame code (SCPF code). The analyzer generates the subcategorization frame for the entire predicate by applying the permutation commands developed from the SCPF code for one auxiliary verb at a time. The first permutation is performed for the first auxiliary verb next to the main verb, and the focus moves on from the main verb to the first auxiliary verb. The second permutation is performed for the second auxiliary verb next to the first auxiliary verb, and the focus moves on to the second auxiliary verb. And so on: the N-th permutation is performed for the N-th auxiliary verb next to the (N-1)-th auxiliary verb, and the focus moves on from the (N-1)-th auxiliary verb to the N-th auxiliary verb. The maximum number for N is actually set to three in our MT system, reflecting the numbers of auxiliary verbs in real utterances and written sentences.

- e.g.3-3 a. X-ga Y-wo taberu.
 X-NOM Y-ACC eat.
 "X eats Y "
- b. Z-ga Y-wo X-ni tabe-saseru.
 Z-NOM Y-ACC Z-DAT eat-CAUS.
 "Z makes X to eat Y"
- c. X-ga Y-wo Z-ni tabe-sase-rareru.
 X-NOM Y-ACC Z-DAT eat-CAUS-PASS.
 "X was made to eat Y by Z"

A correct process would generate the subcategorization frames represented in the example sentences from e.g. 3-3a via e.g. 3-3b to e.g. 3-3c, where all case elements X, Y and Z are consistent in these three sentences. The SCPF code in the causative auxiliary verb "saseru" has two ambiguities of the set of permutation commands as is shown in Fig.2.

	permutation commands		permutation commands
Caus.	NULL > GA [CAUser]	Caus.	NULL > GA[CAUser]
A	GA > NI/NIYORI	B	GA > WO

Fig.2 SCPF Codes and Permutation Commands for "saseru"

The set of original postpositions is described in the left term (in the source direction of an arc) in the permutation commands. 'NULL' term unconditionally matches and adds an extra case slot with a new deep case described within the bracket on the right term ([CAUser]). If all the left term condition for matching the case markers meet, the permutation frame is valid and the number of subcategorization frames for a predicated is sometimes increases. In the above example, however, the permutation commands of Causative B results in generating two identical case markers "WO" violating the Unique Surface Case Principle as is shown in e.g.3-4, so the permutation is blocked.

- e.g.3-4 a. X-ga Y-wo taberu.
 X-NOM Y-ACC eat.
 "X eats Y "
- b. *Z-ga Y-wo X-wo tabe-saseru.
 Z-NOM Y-ACC Z-ACC eat-CAUS .

The second auxiliary verb "rareru" has even more ambiguities in meaning, each of which corresponds to six different Permutation Frames as is shown in Fig.3.

Direct Passive	GA > NI/ NIYORI WO > GA	Possibility	GA > NI WO > GA
Indirect Passive	NULL >GA[EXP] GA >NI/NIYORI	Honorific	N/A
Dative Passive	GA >NI/NIYORI/KARA NI >GA	Autonomous	GA > NI WO > GA

Fig.3 SCPF Codes and Permutation Commands "rareru"

The analyzer follows the permutation procedure described above for the second auxiliary verb. All the permutation commands in fig.3 actually can match the original surface case frame of e.g. 3-3b. It is a set of independent semantic heuristic rules that drops the 'Autonomous' reading of "rareru" and almost drops the 'Honorific' reading of "rareru". All the other subcategorization frames for 'Direct Passive,' 'Indirect Passive,' 'Dative Passive,' and 'Possibility' are generated with slightly different variations of alternative surface case markers. The sentence in e.g. 3-3c can mean any of these but 'Indirect Passive,' the whole sentence of which is shown in e.g. 3-3d.

- e.g.3-3 d. E-ga X-ni Y-wo Z-ni tabe-sase-rareru.
 E-NOM X-DAT Y-ACC Z-DAT eat-CAUS-PASS.
 "E experienced that X was made to eat Y by Z"

It is still grammatical, but is much more difficult to get the meaning of it because it has four arguments for the single verb. This factor alone can be used by the analyzer to restrain the application of the generated subcategorization frame for the 'Indirect Passive' interpretation.

3-3. Two Cases of Multiple Deep Cases in a Single Slot

There are two kinds of description by which multiple deep cases are described in a deep case slot of a subcategorization frame (Fig.1). One is selectional, and the other is overlapping. The selectional one is the use of alternative deep cases and meets the needs of economical description of the lexicon and also the manageability of it.

- e.g.3-5
 a. "The typhoon[REASON] has broken a part the city block."
 b. "A monster [AGENT] has broken a part the city block."

In these examples, not the verb "break", but the semantics of the subject decides what deep case the subject should be allocated. So, instead of assigning only one deep case onto the deep case slot and create bunch of whole subcategorization frames, we introduced an ambiguity marker such as AIRM (AGenT/INStrument/REAsOn/MEANs) to be assigned on the case slot. In this case, the analyzer does not have to decide the deep case until when necessary at whatever point in the phases of MT⁷.

The other kind of description is 'overlapping' of deep cases.

Slot name:	NOM	ACC	DAT	WITH
Deep case frame:	AGenT &[SOuRce]	PAT- ient	GOAL	n/a

Fig. 4 Deep Case Overlap

Fig.4 shows almost the same deep case frame as in Fig.1 that shows the subcategorization frame of verb "ageru (*give*)". The only difference is the deep case [SOuRce] added to the AGenT. Other kind of transitive verbs such as "taberu (*eat*)" may not let the [SOuRce] be added because the AGenT here is not the SOuRce position of the PATient in the event/action. "Taberu (*eat*)" may let [GOAL] be added to the AGenT. This distinction may work in the later knowledge-based inference phases of the MT system.

3-4. Lexicon Structure in Relation to Word Senses

There are cases in which one word sense corresponds to multiple subcategorization frames, other cases in which one word sense corresponds to one subcategorization frame each, and the other cases in which multiple word senses correspond to less number of subcategorization frames. Since our approach here is rather empirical so any guidelines that help the lexicon to be uniform in quality, we take advantage of other literature that aimed at some exhaustive listing of interesting cases. The example sentences in e.g. 3-6 are cited from Fillmore68, and e.g. 3-7, from Levin93.

- e.g.3-6 a. John opened the door with the key.
 b. The key opened the door.
 c. The door opened.
 d. John ate the meal. e. John ate.

- e.g.3-7 a. John pounded the metal [flat].
 b. Metal pounded flat.
 c. *Metal pounded.

⁷ The decision point could be even delayed into the generation phase of MT.

As is briefly mentioned in the previous sections, the entry in our lexicon is composed of three blocks: M (Morphology)-Block, S (Syntax)-Block and C (Concept)-Block. M-Block contains the very surface information and can in general be linked to multiple S-Blocks. A whole subcategorization frame is described and stored in a S-Block coupled with corresponding other syntactic features such as aspect features. A C-Block linked from an S-Block or more represents an independent word sense, and, ideally, is linked to by other S-blocks that are linked to by other M-Blocks, in effect, other words of the same or the different language.

The basic principle of the lexicon requires one to one correspondences between a subcategorization frame and an S-Block. So, any sentence in e.g. 3-6 or e.g. 3-7⁸ corresponds to a different S-Block from the others (except for e.g. 3-7c that does not exist). Any two S-Blocks can share the same word sense (C-Block) as long as the deep case (thematic role) frame is consistent. That is, all the case roles that appear in e.g.3-6 a, b and c are assigned different deep cases from one another: {John = AGenT, door = PATient, key = INStrument}. So are all the case roles in e.g. 3-6 a and b, and all the case roles in e.g. 3.7. Thus, our system of the lexicon could guarantee that these subcategorization frame s of intuitively the same word sense share the same C-Block.

There are other cases in which the above criteria require to separate the intuitively single word sense as is shown in e.g. 3-8 e.g.3-8

- a. John smeared the *window*[PATient] with the *paint*.
 b. John smeared the *paint*[PATient] on the *window*.

If a lexicographer is asked to fill in the deep cases as usual in the S-Blocks of e.g.3-8 a and b, he or she will assign PATient on window in e.g.3-8 a, and on *paint* in e.g.3-8 b. This inconsistency in the case assignment does not allow the lexicon to allocate the same C-Block to both e.g.3-8 a and b. NJ&B94 gives a solution to this kind of case by introducing some *deeper* conceptual primitives than our deep cases.

4. The Development Results of a Computational Japanese Lexicon for MT

We have developed a computational Japanese lexicon with more than 80 thousand words, 30 thousand of which are verbs and their derivations. A key part of the development was to establish word

⁸ Lexical Entry for the verb 'pound': M-block:{pound} S-block1:{Sub=AGT, Dobj=OBJ, [Comp=TARstat]} S-block2 {Subj=OBJ, Comp=TARstat } C-block: CP{POUND} /*Strike repeatedly & forcefully, so the form of object be altered to meet the agent's purposes */

senses by means of comparing synonymous vocabulary sets of Japanese and English [Nomura89]. Each verb subcategorization frame is coded in what we call S-Block that is placed between M-Block that contains the very surface level information and C-Block that is supposed to contain language independent, purely conceptual information.

Among 34 deep cases we defined for the Interlingua, which is fewer than those in previous (e.g. Nagao80), 16 are currently used in the deep frame of our Japanese verb frame system.(Fig.5)

PATient, EXPeriencer, AGenT, INStrument, MEAns, REASon, SOuRce, GOAL, LOCation, BENeficiary, TARGet, PaRTicipant, CAPacity, FoCuS, MATerial, ELMenT
--

Fig.5 The 16 Deep Cases Described in the Subcategorization Frames for Japanese Verbs and Predivative Adjectives

The result of coding subcategorization frames for 30 thousand verbs have listed up 18 case postpositions and standard word order used for verbs, and six, used for predicative adjectives. These figures are much smaller than the number of simple combinations: $7! = 823,543$. The numbers of voice conversion types that affects the surface case pattern was 14^9 . The non-weighted mean number of the case slots for each lexicon is counted to 1.6 ; 30% of verbs are listed up to take multiple case patterns. This figure seems appropriate, considering the fact that most of Japanese transitive verbs and intransitive verbs take separate word forms.

The numbers of the variations of subcategorization frames in the lexicon was about 250 for ordinary verbs and adjectives, and we have 150 more for idiomatic ones. These are the figures after disregarding, of course, the variation of selectional restrictions. The sum of these figures are also much smaller than that of simple combinations: $(16C7) * (7!) = 57657600$.

Exhaustive listing of 400 combinatorial subcategorization frames has contributed much to improve the accuracy the contents in the lexicon. The lexicon specification by the proposed verb subcategorization codes and SCPF codes has improved uniformity in quality and the speed of lexicon development as well.

5. Conclusion

We proposed a knowledge representation framework for verb subcategorizations with combinatorial codes for the verb's surface case frame and deep case (thematic role) frame. The Japanese lexicon developed by this design has comprehensive

⁹ There are four other special ones that only replaces deep case labels.

information on the mapping between the surface case frame and the deep case frame, and yet is free of potential combinatorial explosion due to an exhaustive empirical research and development of Japanese verbs and auxiliary verbs. The reduction of the number of verb subcategorization codes was made possible by carefully identifying superficially different case frames with the idea of *alternative* case markers and semantic roles. The benefits include more manageable, repeatable lexicon realized by reducing the underlying redundancy of information in some distributed architecture of the computational lexicon. The future tasks should include further explorations of providing the concept dictionary with more syntactic test conditions and extensions to more than two languages other than English and Japanese.

References:

- Chomsky88, Chomsky, N., Language and the Problems of Knowledge: the Managua Lectures, MIT Press
 EDR94, Electronic Dictionary Research Institute Ltd., EDR Electronic Dictionary User's Guide Ver. 2.0, 1994
 Fillmore68, Fillmore, C., J., The Case for Case, in : Bach and Harms (eds.), Universals in Linguistic Theory (Holt, Rinehart and Winston, New York, 1968) 1-90
 FM&T85: Fukui, N., S. Miyagawa, and C. Tenny. (1985). Verb Classes in English and Japanese: A Case Study in the Interaction of Syntax, Morphology and Semantics. , Lexicon Project Working Papers #3, MIT.
 Grishman94: Grishman, R. et al. , Complex Syntax : Building a Computational Lexicon section 2.1 Subcategorization
 Hornby75, Hornby, A., S., Guide to Patterns and Usage in English, second edition, Oxford University Press, 1975
 Jackendoff90, Jackendoff, R., Semantic Structures, Cambridge, MA: MIT Press, 1990
 Jackendoff93, Jackendoff, R. (1993). On the Role of Conceptual Structure in Argument Selection: A Reply to Emonds. Natural Language and Linguistic Theory, 11.
 Levin93: Levin, Beth, English Verb Classes and Alternations, English Verb Classes and Alternations ~ A Preliminary Investigation, The University of Chicago Press.
 Martin75, S.E. (1975). A Reference Grammar of Japanese, Yale University Press.
 Muraki87, K.Muraki "PIVOT: A Two Phase Machine Translation System" , MT Summit, pp.81-83, Japan, 1987
 Nagao80: Nagao, M., Tsujii, J., Mitamura, K., Hirakawa, H. and Kume, M. , A Machine Translation System from Japanese into English," COLING-80, pp414-423
 NJ&B94: Nomura, N., Jones, D., & Berwick, R., An Architecture for a Universal Lexicon, in the Proceedings of COLING94
 Nomura89: Nomura, N., Muraki, K. : 'Case frame model of Machine Translation system PIVOT', Proceedings of 38th IPSJ Conference, 1989
 Pustejovsky91: Pustejovsky, J., The Generative Lexicon. Computational Linguistics, 17.4.