# DETECTING THE ORGANIZATION OF SEMANTIC SUBCLASSES OF JAPANESE VERBS

Akira Oishi Yuji Matsumoto
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma,
Nara 630-01 Japan
{ryo-o,matsu}@is.aist-nara.ac.jp

**Summary**

This paper describes an approach to detect the organization of semantic subclasses of Japanese verbs. First, we classify verbs along two dimensions: thematic and aspectual. In the thematic dimension, we exploit the pattern of case marking particles which are attached to arguments of verbs. In the aspectual dimension, we exploit the classification of adverbs which modify verbs in a corpus. By combining the results of two classifications, we obtain an elaborate classification of verbs. We can incorporate the prototypicality of the members which constitute each semantic subclass by taking account of the frequency of case particles patterns and cooccurring adverbs. Moreover, the existence of close relationships among them enable us to detect the organization of these subclasses.

**Keywords:** Japanese verbs, semantic subclass, two-dimensional, case marking particles, adverbs, prototypicality

## 1. INTRODUCTION

A Chinese proverb says, '*it is born without a report and dies without a funeral.*' What does this mean? What is the referent of the subject *it*? The answer is a Chinese character. Since Chinese characters are ideogram and roughly correspond to morpheme, the true sense of this proverb amounts to language creativities.

Language creativities, i.e. the abilities to coin new words and to use some known words figuratively, bring about gaps in a lexicon and cause a bottleneck when one tries to scale up existing natural language processing (NLP) systems. This motivates a (semi-)automatic acquisition of lexical information for NLP systems from machine-readable resources such as dictionaries and corpora (Zernik, 1991). Furthermore, in order to construct a more complete lexicon, it is necessary to consider ways in which relevant lexical information can be identified and represented. The objective of this paper is to detect the organization of semantic classes of Japanese verbs on the basis of syntactic realizations of corpus data. By carefully classifying the verbs, we can determine many syntactic properties of a new verb, given its meaning (or semantic class membership) (Levin, 1993; Levin and Rappaport, 1995).

The next section briefly describes some of the issues about the syntactic-semantic interface and linguistic categories. The third section describes the classification method of verbs. The classification is carried out along two dimensions: thematic and aspectual. In the thematic dimension, we exploit the pattern of case marking particles which are attached to arguments of verbs. In the aspectual dimension, we exploit the classification of adverbs which modify verbs in a corpus. By combining the results of two classifications, we can obtain a more elaborate classification of verbs.

The second half of this paper presents the result of the classification. We adopt Lexical Conceptual Structure(LCS) developed by Jackendoff to represent the core meaning of each subclass of verbs (Jackendoff, 1990). Thematic roles are definable in terms of the position of the participants of the event in LCSs. Moreover, certain parts of LCSs are explicitly aspectual; the distinction between EVENT and STATE, for example. At the same time, we can detect the organization of semantic subclasses by considering the relationships among them. The obtained organization will be shown diagrammatically at the end of this paper.

## 2. THE SYNTACTIC-SEMANTIC INTERFACE

Work on argument structures in contemporary linguistics treats them as an interface between lexical semantics and sentential syntax. Syntactic argument structures are projections (via linking rules) of a verb's semantic structure. An event cannot be conceptualized without making some kind of mental reference to its participants. If these participants are absent altogether the meaning of the verb is incoherent. Thus, the arguments of a verb are salient facets of its semantic structure. In the last ten years, argument structure representations have become increasingly sophisticated. For instance, Grimshaw proposes the "two-dimensional" theory on argument realization (Grimshaw, 1990). She points out that an argument structure has internal organization, which is not stipulated for each predicate but is projected from lexical semantic representation along two separate dimensions: one in terms of thematic properties and one in terms of aspectual properties. This separation seems to correspond to the two of four levels of representation in Pustejovsky's Generative Lexicon framework (Pustejovsky, 1991), i.e. "Argument structure" and "Event structure".

In addition, there are regularities as to the combination of sentence constructions in which a verb can participate. For example, Pinker introduces an interesting hypothesis developed by Landau and Gleitman, although he doesn't fully accept this hypothesis (Pinker, 1989). He writes (pp.260-261):

> An even more specific form of syntax-guided learning of verb meaning has been sug-

> gested by Barbara Landau and Lila Gleitman. ... they suggest that children rely on a particularly rich form of verb meaning. This kind of inference, which they later call "Syntactic Bootstrapping", has the child examining an *entire set* of verb entries sharing the same stem, ... They give an example of how a child might learn the meaning of *see* using nothing but linguistic input. The stem *see*, unlike *touch*, takes sentential complements, suggesting that it is a cognitive/perceptual verb. It takes directional phrases (e.g., *to see from the room*), suggesting that it has an abstract motion component. But in comparison with *look*, it does not appear in the imperative, showing that it is stative. Nor does it appear in the pseudo-cleft **What John did was see Bill* or with purposive *to* adjuncts, which shows that it is nonagentive. Taken together, these inferences strongly constrain what the verb *see* could mean.

Thus, the syntactic formats for a verb provide crucial cues to its meanings and we can narrow down the possible meaning of a verb by observing its syntactic realizations.

On the other hand, cognitive linguists have pointed out that much in language is a matter of degree (Lakoff, 1987; Langacker, 1987; Langacker, 1991). Linguistic relationships are not invariably all-or-nothing affairs, nor are linguistic categories always sharply defined and never fuzzy around the edges. A simple plus/minus value or yes/no answer is not always sufficient in specifying whether a word has a certain property, belongs to a particular category, or participates in a given relationship. These conditions are often matter of degree, and we must devise some means of accommodating the complete range of possibilities.

As mentioned in the introduction, we will provide LCS representation for each subclass of verbs. It is an abstract characterization that is fully compatible with all the members of the subclass it defines (so membership is not a matter of degree). However, there exist prototypical members and marginal members of each subclass; there are degrees of membership based on their perceived resemblance to the prototype. Moreover, each subclass has fuzzy edge and might even merge into each other. Therefore, we must incorporate the prototypicality of the members which constitute each semantic subclass, and clarify the relationships among them.

## 3. CLASSIFICATION PROCESSES

In this section, we describe the classification processes of verbs along two dimensions: thematic and aspectual. We use the EDR Japanese Cooccurrence Dictionary (EDR, 1995) as material to extract syntactic clues in both processes. The EDR Cooccurrence Dictionary comprises surface cooccurring items from the EDR corpus. Each cooccurring item consists of a governing word, a dependent word, the relator between the words, and supplementary cooccurrence item information which consists of the frequency of the item in the corpus and examples. An example of this item is shown in Table 1.

### 3.1. Thematic dimension

There are two types of languages in the world. In languages such as English (so-called "fixed word-order" languages), the order of words in sentences is tightly constrained and the position of each noun phrase determines their roles in the event, like *A dog bites a man* and *A man bites a dog*. In many other languages such as Japanese (so-called "free word-order" languages), there are more options as to word order. Instead of phrase order, Japanese expresses who did what to whom by modifying nouns with case marking particles (*kakujoshi*, hereafter called "case particle"), like *Inu-ga(NOM) hito-wo(ACC) kamu "A dog bites a man"* and *Hito-ga(NOM) inu-wo(ACC) kamu "A man bites a dog"*. These case particles play an important role when we understand sentences. Moreover, their pattern in a sentence provides a crucial clue as to speakers' construal of an event. In other words, we have ability to construe a situation in alternative ways and this construal

Table 1: An example of EDR Japanese Co-occurrence Dictionary

| Governing word Information | Relator Information | Dependent word Information |
|---|---|---|
| *kake* verb | *wo* | *denwa* noun |
| Frequency:40 | | |
| Sentence#23/[*denwa*]*wo*[*kake*]*teita* | | |
| Sentence#50/[*denwa*]*wo*[*kake*]*nagara* | | |
| : | | |

determines the pattern of case particles, i.e. sentence construction.

For example, consider the locative, also known as the "spray/load" or "figure/ground" alternation, shown in example 1.

1. (a) Hanako-ga penki-wo kabe-ni nutta.
   Hanako-NOM paint-ACC wall-DAT smear-PRF.
   "Hanako smeared paint on the wall."

   (b) Hanako-ga kabe-wo penki-de nutta.
   Hanako-NOM wall-ACC paint-WITH smear-PRF.
   "Hanako smeared the wall with paint."

These two forms are expressing the same situation but are not synonymous. In (a), *penki(paint)* is accentuated and it means adhesion of the object to another object, while (b) takes up *kabe(the wall)* as an object undergoing changes. This difference gives the latter a holism effect "the whole wall is smeared", which is an epiphenomenon of the fact that the verb specifies a change of state: unless the wall is entirely smeared, there is no pragmatic sense in which it can be said to have changed state. In contrast, the former specifies adhesion of the object and does not imply such change of state.

This phenomenon can be explained by referring to the core meaning of the verbs. The former has the core meaning "X moves Y into/onto Z", and the participants in the event, X, Y, and Z are mapped onto surface case particles *ga*, *wo*, and *ni*, respectively, following general linking rules that map semantic structures onto syntactic argument structures. The latter has the core meaning "X causes Y to change its state by means of moving Z onto Y", and X, Y, and Z are mapped onto surface case particles *ga*, *wo*, and *de*, respectively. Thus, we can define a class of possible verb meanings by running linking rules backward. That is, since there is some semantic commonality observable across an array of verbs that participate in the same sentence construction, we can schematize it by abstracting away from points of divergence. This schema is a type of event or relationship that lies at the core of the meaning of a class of possible verbs. It can then be used as a template for assembling new expressions on the same pattern. As for the relationship between case particles patterns and the core meaning of verbs, see (Oishi and Matsumoto, 1995).

However, there is an enormous reduction of information when going from the semantic structure of a verb to its syntactic argument structure, and several different semantic subclasses can be mapped onto the same argument structure. In order to resolve this problem, we explore regularities of the combination of case particles patterns which a verb takes in a sentence.

We extracted from the EDR Cooccurrence Dictionary the items of which the governing and dependent words are a verb and a noun. Each item has the sentence numbers in the EDR Corpus from which it is extracted. For example, if the sentence "230: *sensei ga jisyo wo hiku* (A teacher refers to a dictionary)" appears in the corpus, then the items "*hiku - ga - sensei*" and "*hiku - wo - jisyo*" are recorded in the Cooccurrence Dictionary with the sentence number "230" (cf. Table 1). By gathering the items which have an identical sentence number and contain the same verb, we reconstructed 236,157 surface argument structures. From these, we chose 858 verbs having more than 50 examples for processing (total of 158,888 structures and 210,242 arguments). In this stage, we excluded the items which include auxiliary verbs such as *(s)areru*(passivizer), *(s)aseru*(causativizer), *temorau*(receptivizer), etc. since these auxiliary verbs change the argument structure of the attached verbs. Each datum has the following form:

$$\begin{bmatrix} \text{verb} : kaisetusuru(establish) & \\ \text{args} : \begin{bmatrix} ga\text{−NOM} : [\text{noun} : kaisya(company)] \\ wo\text{−ACC} : [\text{noun} : jimusyo(office)] \\ ni\text{−DAT} : [\text{noun} : Osaka(Osaka)] \end{bmatrix} \end{bmatrix}$$

We extracted case particles patterns from the data, and for each datum we allotted a pattern label in the Table 2 by pattern matching. (The above example has the pattern "*ga,wo,ni*" and the label "h" is allotted.)

Table 2: Surface Case Patterns List

| particles | label | particles | label | particles | label | particles | label |
|---|---|---|---|---|---|---|---|
| *ga* | a | *ga,ni* | b | *ga,de* | b1 | *ga,ni-oite* | b2 |
| *ga,e* | b3 | *ga,ni-taisite* | b4 | *ga,ni-yotte* | b5 | *ga,kara* | c |
| *ga,to* | d | *ga,kara,ni* | e | *ga,kara,e* | e1 | *ga,kara,to* | e2 |
| *ga,to,ni* | f | *ga,to,ni-taisite* | f1 | *ga,wo,to* | g | *ga,wo,ni* | h |
| *ga,wo,de* | h1 | *ga,wo,ni-oite* | h2 | *ga,wo,e* | h3 | *ga,wo,ni-taisite* | h4 |
| *ga,wo,kara* | i | *ga,wo,kara,ni* | j | *ga,wo,kara,e* | j1 | *ga,wo,kara,to* | j2 |
| *ga,wo* | k | *ga,ni-tsuite* | k1 | | | | |

By summing up all of the data, we created a file with each verb as a key and registered the pattern labels and their frequency. A part of this file is given in Table 3.

Notice that pattern labels for each verb are arranged by the order of frequency. We eliminated labels which have less than 5% of the total number of examples for each verb, since these labels might have been possibly produced by erroneous analyses. Consequently, we obtained such data as shown in Table 4.

This table consists of 173 lines and each line contains an ordering of pattern labels and verbs which have this ordering of pattern labels. We call them as "thematic subclasses" of verbs and refer to each subclass by using the ordering of pattern labels.

### 3.2. Aspectual dimension

The properties that regulate the interface between syntax and lexical semantics other than thematic are aspectual ones. Constraints on the aspectual properties associated with direct internal arguments, indirect internal arguments, and external arguments in syntactic structure constrains the kinds of event participants that can occupy these positions (Tenny, 1994).

Table 3: A part of the case pattern file

| verb | pattern frequencies | total |
|---|---|---|
| *hasiru*(run) | k=169 a=139 b=110 c=16 b3=12 b2=12 e1=6 h=5 e=3 h3=3 h2=2 b5=2 | 480 |
| *tobu*(fly) | a=68 b=35 k=33 b3=20 c=16 b2=7 e=6 e1=4 b5=3 h=2 | 195 |
| *kankeisuru*(relate) | b=66 a=26 d=18 | 111 |
| *kanrensuru*(relate) | b=83 a=18 d=13 | 114 |
| *musubituku*(link) | b=58 d=33 a=16 | 107 |

Table 4: A part of the classification of verbs based on the ordering of case patterns

| patterns | verbs |
|---|---|
| a | *nakunaru(be lost),hakkirisuru(become clear),heru(get fewer), gensyousuru(decrease),jousyousuru(rise),fukamaru(get deeper), ...* |
| a b | *osamaru(be settled),sorou(be complete),tamaru(be accumulated), noboru(ascend),mieru(see),tikazuku(approach),arawareru(appear), ...* |
| a b b1 | *deru(go out),sodatu(grow),naku(cry),hiromaru(spread),furu(fall) gouisuru(agree),tomaru(stop),demawaru(appear on the market) ...* |
| a b b1 c | *umareru(be born)* |
| a b b1 k | *hikaru(shine),syoujiru(arise)* |
| a b b3 | *ugokidasu(start to move)* |
| a b b3 c | *nobiru(extend)* |
| a b c | *ukagaeru(can see),kuru(come),mieru(can see),agaru(rise),otiru(fall)* |

Perhaps the best-known taxonomy of verbs in terms of aspectual property was suggested by Vendler (Vendler, 1957). In his classification, verbs first divide into "states" and "processes", where a state is temporally homogeneous and static, and a process specifies something that changes over time. Processes in turn divide into three varieties: "activities" which are extended in time but have no clearly demarcated endpoints, "accomplishments" which are extended over time but are defined by the fact that they terminate in the attainment of some state, and "achievements" which are construed as referring to the instant at which a state is attained.

Pustejovsky collapses these classes to three distinct structural configurations: "states", "processes", and "transitions", where Vendler's "activities" is renamed as "processes" and "transitions" subsume both "accomplishments" and "achievements" (Pustejovsky, 1992). In his definitions, "state" is a single event, which is evaluated relative to no other event (e.g. *be sick, love, know*). "Process" is a sequence of events identifying the same semantic expression (e.g. *run, push, drag*). "Transition" is an event identifying a semantic expression, which is evaluated relative to its opposition (e.g. *give, open, build, destroy*). He gives a structural representation called "event structure" to each event type as in Figure 1 (where "e" is an event and "E" is a variable for any event type).

In order to obtain aspectual properties of verbs we exploit adverbs which modify verbs in a

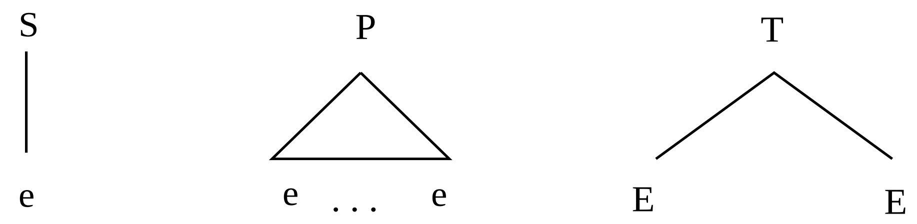


Figure 1: Event Structures

corpus, since many of adverbs are closely related with aspectual properties of verbs. For example, Moriyama has classified adverbs in terms of aspectual properties of their modifying verbs as follows (Moriyama, 1988). We connected them with event structures and allotted a label for each class in favor of the subsequent process.

**Processes modifiers** There are two kinds of process modifiers. One modifies "process" or the initial subevent of "transition". This class includes reduplicative onomatopoeia such as *gasagasa, batabata, suisui, sesseto, butubutu,* etc., which are expressing sound or manner of directed motion, and rate adverbs such as *yukkuri(slowly), tebayaku(quickly),* etc., which are expressing a speed of motions. We gave them a label "P(Process)". The other expresses the progress of change of state, such as *dandan(gradually), sukosizutu(little by little), jojoni(gradually), dondon(constantly), sidaini(by degrees),* etc., which modifies "transition" as a whole. We gave them a label "T(Transition)". They are illustrated in Figure 2 where the scope of each adverb is represented by "[adverb]".

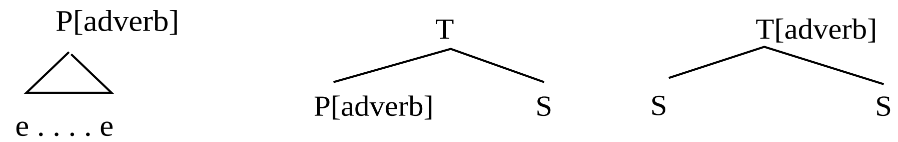


Figure 2: Process modifiers

**Continuance adverbs** There are adverbs which can modify both "state" and "process", such as *zutto(for a long time), itumademo(forever),* etc. They express a continuance of an event or a maintenance of a state, as illustrated in Figure 3.

We gave them a label "C(Continuance)".

We distinguish adverbs which express manner of existence and gave them a label "M(Manner of existence)". They can modify only "state".

**Instantaneous adverbs** There are three kinds of instantaneous adverbs. The first makes any events instantaneous, such as *satto, ponto, gatatto, potarito, syunkan,* etc. which are expressing instantaneous sound emission or an instant. When these adverbs cooccur with verbs, the events are grasped as instantaneous. This doesn't necessarily imply that the verb itself is instantaneous. We gave them a label "I(Instantaneous)". The second measures out events, such as *gokiro aruku(walk*


Figure 3: Continuance adverbs

*5km), gojikan seizasita(sit straight for 5 hours)*, etc. These include time, distance, and any quantity of contents. We gave them a label "Q(Quantity)". The third expresses the resultant state of events, such as *mapputatuni(into two exact halves), konagonani(into powder), pechankoni(be flat), barabarani(into peaces)*, etc. We gave them a label "R(Result)". They are illustrated in Figure 4.

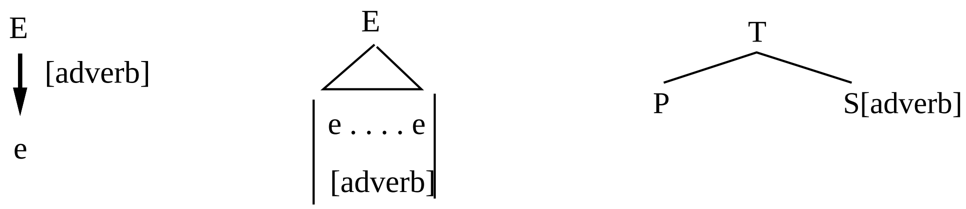


Figure 4: Instantaneous adverbs

This classification of adverbs does not directly correspond to event types of verbs. However, the tendency to cooccur with some type of adverbs provides a clue to aspectual properties of a verb.

We first extracted from EDR Cooccurrence Dictionary the items of which the governing and dependent words are a verb and an adverb. This time we didn't use the relator information but instead used the frequency information assigned at the items. Then, we created an index file with each verb as a key and registered the kinds of adverbs and their frequency. A part of the file is given in Table 5.

Table 5: Adverbs cooccurring with verbs

| verbs | adverbs |
|---|---|
| *intaisuru(retire)* | *sudeni(already)*=2, *totuzen(suddenly)*=1 |
| *hasyagu(frolic)* | *hitoitibaini(more than others)*=1, *henni(strangely)*=1 |
| *enjiru(perform)* | *sanzan(severely)*=1, *naze(why)*=1, *hajimete(first)*=1 |
| *furimuku(look back)* | *guito(with a jerk)*=1, *tittomo(at all)*=1 |
| *houjiru(report)* | *isseini(all together)*=2, *kou(like this)*=1, *sou(so)*=1, *tubusani(minutely)*=1, *motto(more)*=1, *issai(all)*=1, ... |

We then carried out the classification of adverbs by using morphological clues and corrected it by hand. The numbers and examples of each subclass are shown in the Table 6.

Table 6: Result of the classification of adverbs

| label | number | examples |
|---|---|---|
| P | 299 | *yukkuri(slowly), gasagasa, batabata, suisui, sesseto, butubutu,, ..* |
| T | 93 | *sidaini(gradually), masumasu(increasingly), jojoni(gradually), ..* |
| C | 65 | *sonomama(as it is), zutto(for a long time), itumademo(forever),..* |
| M | 79 | *jitto(still), akaakato(brightly), uttorito(absorbedly), ..* |
| I | 183 | *satto, ponto, gatatto, potarito, syunkan(instantaneously), ..* |
| Q | 69 | *180-do(180 degree), ippo(a step), ippai(a cup), ..* |
| R | 60 | *upputatuni(into two exact halves), konagonani(into powder), ..* |

By referring to this file, we created a file with each verb as a key and registered the labels and their frequency for each verb. (We summed up the frequency of adverbs which have the same label). By applying the process similar to the previous subsection, we obtained such data as shown in Table 7.

Table 7: A part of verb classification based on the frequency ordering of adverb classes

| labels | verbs |
|---|---|
| T Q | *matomaru(be settled), gensyousuru(decrease), zoudaisuru(become larger), hasseisuru(occur), toutatusuru(arrive at), syuuryousuru(finish), ..* |
| T Q I | *fueru(increase), akkasuru(get worse), kanseisuru(complete), ketteisuru(decide), kyoukyuusuru(supply), setuzokusuru(connect), ..* |
| T Q I C | *kaizensuru(improve), heru(decrease), jitugensuru(realize)* |
| T Q I C P | *kasaneru(pile up)* |
| T Q I C R P | *kuwaeru(add)* |
| T Q I P | *fukameru(deepen)* |
| T Q I P C | *agaru(rise)* |

This table consists of 328 lines and each line contains an ordering of adverb class labels and verbs which have this ordering of labels. We call them as "aspectual subclasses" of verbs and refer to each subclass by using the ordering of adverb class labels.

However, this classification is incomplete, because the perspective of speakers (writers) often changes the aspectual properties of verbs. For example, Japanese aspectual marker *-teiru* "durativizes" all types of verbs and changes their event type, e.g., "transition" into "state" or "process". We can recognize motion verbs as either "process" or change of location, i.e. "transition". Moreover, all verbs are not necessarily modified by adverbs (52 verbs out of 835 have an empty list of labels).

In order to solve these problems, in the next section, we combine this aspectual classification with the thematic one.

## 4. REPRESENTATION

In this section, we combine the classifications obtained by the above two processes, and represent core meaning of each subclass of verbs by Lexical Conceptual Structure(LCS) (Jackendoff, 1990).

Since each subclass is too small to capture their core meaning, we collected the verbs which have similar orderings of pattern labels and adverb class labels by hand. However, the manual analysis was an easy task, since the verbs were sorted out in advance.

**Subject change-of-state verbs** For a simplest example, consider the subclass members which belong to the thematic <a> and aspectual <T> ("a" indicates particles pattern $< ga >$ and "T" is the representative marker of such subclasses as consisting of subclasses <T>, <T,I>, <T,I,C>, <T,Q>, <T,Q,I>, etc. Below we use the most frequent adverb class as the label of aspectual classes for simplicity.)

This subclass contains:

> *nakunaru(become lost), heru(get less), kireru(be disconnected), tuyomaru(become strong), fukamaru(get deeper), taeru(cease), otoroeru(get weak), totonou(be ready), usureru(fade),* etc.

These verbs express "change of state" and many of them are "de-adjectival" intransitive verbs such as *nai(lost)* $\rightarrow$ *nakunaru(become lost)*, *fukai(deep)* $\rightarrow$ *fukamaru(get deeper)*, *usui(pale)* $\rightarrow$ *usureru(fade)*, etc. The adjectives from which these verbs derived express the results of change of state. Thus, the LCS of this subclass can be defined as in (1):

$$[_{\text{Event}}\text{GO}_{\text{Ident}}([_{\text{Thing}}\alpha]_{\text{A}}, \left[_{\text{Path}} \begin{array}{l} \text{FROM}([_{\text{State}}\text{BE}([\alpha],[\text{NOT AT}([\beta])])]) \\ \text{TO}([_{\text{State}}\text{BE}([\alpha],[\text{AT}([\beta])])]) \end{array}\right])] \tag{1}$$

Here, the square brackets identify "conceptual constituents". Each constituent is labeled as belonging to a "major conceptual category" — one of the kinds of entities the world is conceptualized as containing, for example, Thing, Event, State and Path. Within the brackets, the expressions in capital letters denote the conceptual content. GO, FROM, TO, BE, and AT are functions. GO is a two-place function that maps a Thing and a Path into an Event denoting the Thing traversing along the Path. The subscript on this function is a "semantic field" in which the Event or State is defined. Ident means Identificational field and $\text{GO}_{\text{Ident}}$ denotes change of state. FROM and TO are one-place functions that map a State into an abstract Path that starts from or terminates at that State. BE is a two-place function that maps a Thing and a Place into a State denoting the Thing existing at the Place. Semantic field also makes this Place abstract. AT is a one-place function that maps a State into an abstract Place. $\beta$ indicates this State and is contained in the verb's meaning.

$\alpha$ indicates a Thing which changes its state and is marked with the "linking subscript" $A$. An argument so marked must be expressed by a syntactic argument; by virtue of a general principle of linking, this syntactic argument will appear as the subject and be marked with case particle *ga*. Thus, the entire Event in (1) can be read roughly as "A Thing $\alpha$ changes its state from not being $\beta$ to being $\beta$. We refer to this class of verbs as "subject change-of-state verbs".

Some verbs which were borrowed from Chinese can be used as both intransitive and transitive. They belong to the thematic <a,k> ("k" indicates particles pattern $< ga, wo >$ ) and aspectual <T>:

> *kaifukusuru(recover), koujousuru(improve), gensyousuru(decrease) kanryousuru(complete/be completed), syuuryousuru(end),* etc.

When they are used as transitive, i.e. with the case particles pattern of $< ga, wo >$, the things which change their states are not subjects (marked with *ga*) but objects (marked with *wo*). In this case the LCS is the one whose main Event is the function CAUSE with (1) as its second argument. It will be described later under the title of "object change-of-state verbs".

On the other hand, when "subject change-of-state verbs" take the case particle *de* that indicates "cause of change", they appear with the particles pattern "b1" ($< ga, de >$). Since "cause of change" is not inherent in verb meaning, this *de* case is optional and the verbs belong to thematic subclass <a,b1>.

This subclass contains:

> *fueru(increase), nobiru(get longer), matomaru(be settled), kyuuzousuru(increase rapidly), hyoumenkasuru(come into the open), katamaru(get hard), sibousuru(die), sinu(die), nakunaru(die),* etc.

The LCS of this subclass can be defined as:

$$\left[_{\text{Event}} \begin{array}{l} \text{GO}_{\text{Ident}}([_{\text{Thing}}]_{\text{A}}, \left[_{\text{Path}} \begin{array}{l}\text{FROM}([_{\text{State}}])\\ \text{TO}([_{\text{State}}])\end{array}\right]) \\ [\text{BY}[_{\text{Event}}]]_{(\text{A})} \end{array}\right] \tag{2}$$

BY is a one-place function that takes an Event as its argument and indicates the cause of the main event $\text{GO}_{\text{Ident}}$. The fact that this function is subordinated represents an adjunct and its argument Event noun is marked with *de* case. The parentheses of the linking subscript indicate the fact that it is an optional case. The arguments of the function Path are abbreviated.

Verbs like *sinkousuru(advance)*, *zoukasuru(increase)*, *honkakukasuru(make/become real)*, *gekikasuru(become violent)*, etc. can be used as both intransitive and transitive. So they belong to thematic subclasses <a,b1,k> or <a,k,b1>. They have the same LCS as (2) or the one which has the function CAUSE with (2) as its second argument.

The thematic subclass <b,a,b1> ("b" indicates particles pattern $< ga, ni >$) contains:

> *nayamu(be troubled with), kurusimu(suffer from), odoroku(be surprised at)*

These verbs can take either *ni*-case or *de*-case to mark the cause of change-of-state. Although they are continuous with verbs that belong to the thematic subclass <a,b1>, *ni*-case marking a causing event characterize this class of verbs. They are also continuous with "relational verbs" that are mentioned later. When the causing event is marked with *ni*-case, the event itself, rather than the indication of the cause, is focused on. Therefore, the LCS is:

$$\left[_{\text{Event}} \begin{array}{l} \text{GO}_{\text{Ident}}([_{\text{Thing}}]_{\text{A}}, \left[_{\text{Path}} \begin{array}{l}\text{FROM}([\text{BE}_{\text{Psychological}}])\\ \text{TO}([\text{BE}_{\text{Psychological}}])\end{array}\right]) \\ [\text{BY}[_{\text{Event}}]_{\text{A}}] \end{array}\right] \tag{3}$$

where linking subscript A marks the Event which is an argument of BY rather than BY itself. Moreover, since all verbs in this subclass express the mental change-of-state, the semantic field of the states is "Psychological".

Although most members of these subclasses appear in the aspectual subclass <T> or <I>, a few of them do not. However, we can infer aspectual properties by virtue of thematic properties, even when verbs have no membership in the aspectual classification. For example, *ippensuru(change completely)* has no modifying adverb, but we can infer that the event type of this verb is "transition" because the thematic subclass <a,k,b1> contains nothing but change of state verbs. Thus, aspectual properties of verbs have a close relation with thematic properties of them.

**Subject movement/existence verbs** As for the subclasses of "subject change-of-state", we can infer that the event type of these verbs is "transition". Conversely, we can divide a thematic subclass by virtue of the aspectual classification. For example, the thematic subclass <a,b> ("b" indicates particles pattern $< ga, ni >$) contains both motion verbs and existence verbs. This is because Japanese *ni*-case is used to mark both "place" and "goal". When motion verbs take *ni*-case, it means "goal", while the *ni*-case with existence verbs means "place". The distinction between them can be obtained by virtue of the aspectual properties of the verbs. Motion verbs are contained in the aspectual subclasses which include the aspectual label "P(Process)" or "T(Transition)", while existence verbs are contained in the aspectual subclass which include labels "M(Manner of existence)" and "C(Continuance)". Thus, we can obtain the two subclasses for one thematic subclass <a,b> by virtue of aspectual properties of verbs. the LCSs of these subclasses can be defined as in (4) and (5) respectively:

$$[_{\text{Event}}\text{GO}([_{\text{Thing}}]_{\text{A}}, [_{\text{Path}}\text{TO}(\left[\left\{\begin{array}{c}\text{PLACE}\\ \text{THING}\end{array}\right\}\right]_{(\text{A})})])] \tag{4}$$

$$[_{\text{State}}\text{BE}([_{\text{Thing}}]_{\text{A}}, [_{\text{Place}}]_{(\text{A})})] \tag{5}$$

The former contains:

> *noboru(ascend), tikazuku(approach), otikome(fall in), tassuru(reach),* etc.

as its members and the latter contains:

> *osamaru(be settled), sorou(be complete), tamaru(be accumulated), hirogaru(spread), sonzaisuru(exist), sadamaru(be decided),* etc.

The former also contains "appearance verbs" such as *arawareru(appear), syutugensuru(appear), toujousuru(come onstage)* etc. The LCS of these verbs is:

$$[_{\text{Event}}\text{GO}_{\text{Ident}}([_{\text{Thing}}\alpha]_{\text{A}}, [_{\text{State}}\text{BE}([\alpha], [_{\text{Place}}]_{(\text{A})})])] \tag{6}$$

This LCS shows both the relationship between (4) and (5) by encoding the latter into the former and the relationship between (4) and (1) by substituting GO for $\text{GO}_{\text{Ident}}$. Each subclass has fuzzy edge and there is no distinct border between them. These verbs lie halfway between them.

Moreover, if the focus is brought into the place rather than the existing thing, the thematic subclass amounts to <b,a>. This subclass contains:

> *itisuru(take one's position), komoru(be confined), tamaru(accumulate), tatu(be build), otiiru(fall into), suwaru(sit down), sumu(live in),* etc.

In this subclass <b,a>, many verbs which have the relatively high ratio of "b" express "participation" or "admission":

> *kuwawaru(join), kanyuusuru(become a member of), hairu(enter), syuuninsuru(assume office), sankasuru(participate),* etc.

Some of these verbs belong to the aspectual subclasses which include the aspectual label "P" or "T", while others are contained in the aspectual subclass which include labels "M" and "C". The LCS of these verbs is determined by speakers who can change focus variously. If the speaker focuses on the movement of the participant, the LCS of the verb amounts to (4), while if the resultant state of the participation is taken as a focal point, it amounts to (5). Thus, movement and existence are continuous categories; if something moves to some place then it exist that place. The perspective taken by the speaker determines their categorization. Verbs that normally express existence can sometimes receive a movement construal, and vice versa. When an existence verb is construed as an event rater than a state, the LCS of the verb may be defined as:

$$[_{\text{Event}}\text{STAY}([_{\text{Thing}}]_{\text{A}}, [_{\text{Place}}]_{(\text{A})})] \tag{7}$$

The thematic subclasses <a,b,b1> and <a,b1,b> also contain existence verbs. These verbs can take either *ni*-case or *de*-case to mark the place where the subject exists. They express not only existence but also manner of existence and are often modified by adverbs of "M" class:

> *hikaru(shine), uku(float), saku(bloom), nemuru(sleep), yureru(shake), hataraku(work), seikatusuru(live), sodatu(grow), medatu(stand out),* etc.

Similarly, the thematic subclasses <b,a,b1>, <b,b1,a> and <b1,a,b> contain:

> *yokotawaru(lie down), neru(lie down), kurasu(live), asobu(play), hagemu(strive for),* etc.

Verbs like *syoujiru(come out/produce), hasseisuru(occur/produce),* etc. can be used as both intransitive and transitive and verbs like *ikiru(live)* can take a cognate object as in *jinsei wo ikiru (live a life).* They belong to the thematic subclasses <a,b,b1,k><a,b1,k,b> and aspectual <M>.

The LCS of these subclasses is:

$$\left[_{\text{Event}} \begin{array}{l} \text{MOVE}([_{\text{Thing}}]_{\text{A}}) \\ [\text{AT}[_{\text{Place}}]_{(\text{A})}] \end{array} \right] \tag{8}$$

MOVE is a one-place function which denotes an event without a path. AT is a one-place function that takes an Place as its argument where the main event MOVE takes place. The fact that this function is subordinated represents adjunct and its argument is marked with *de* case. When this place is marked with *ni* case, it can be construed as goal and the whole LCS amounts to (4) or (6).

As in the case of the thematic subclass <a,b>, the thematic subclass <b,a> also contains the verbs which express the movement. They belong to aspectual subclass <T> or <P> and are distinguishable from existence verbs. The fact that "b" (the particles pattern $< ga, ni >$) is at the top of the ordering indicates that the goal of the movement is a focal point in the event. Hence, these verbs express "entering" or "arriving" without mentioning the source of the movement. The thematic subclass <b,a,b3> ("b3" denotes particles pattern $< ga, e >$) also contains this kind of verbs, since *e*-case marks the direction of the goal and is often substitutable for *ni*-case without changing the primary meaning of the whole sentence. These subclasses contain:

> *toutyakusuru(arrive at), toutatusuru(arrive at), noru(ride), noridasu(launch into), totunyuusuru(rush into), sinsyutusuru(advance into),* etc.

The LCS of these subclasses is the same as (4) and these subclasses are continuous with "participation" or "admission" class.

Subject movement verbs are also contained in the thematic subclasses which include such patterns as "b3" ($< ga, e >$), "c"($< ga, kara >$), "e" ($< ga, kara, ni >$), and "e1" ($< ga, kara, e >$) in addition to "a" ($< ga >$) and "b" ($< ga, ni >$). Since the case particles *e* and *kara* mark the goal and the source respectively, these verbs express the transfer of something along a path. If they have the pattern "k"($< ga, wo >$), the *wo*-case marks the path. The LCS of this type of verbs is:

$$[_{\text{Event}}\text{GO}([_{\text{Thing}}]_{\text{A}}, \left[_{\text{Path}} \left\{ \begin{array}{l} \text{FROM} \\ \text{TO} \\ \text{TOWARD} \\ \text{VIA} \end{array} \right\} \left( \left[ \left\{ \begin{array}{l} \text{THING} \\ \text{PLACE} \end{array} \right\} \right]_{\text{A}} \right) \right] )] \tag{9}$$

FROM, TO, TOWARD, and VIA are one-place functions and correspond to *kara, ni, e* and *wo* respectively. The ordering of patterns varies according as which part of the path is focused on. The aforementioned LCS (4) is a special case in which the function TO is focused on. When the goal is mentioned, the aspectual subclass tends to be <T>; otherwise <I> or <P>.

This subclass contains:

> *nigeru(run away), tobu(fly), idousuru(move), nagareru(flow), tobidasu(fly out), agaru(go up), sutaatosuru(start), kuru(come),* etc.

and many other verbs. It may be possible to subdivide these verbs into such smaller subclasses as "starting" or "coming" etc. by virtue of the ordering of the particle patterns, but the LCS is the same as (9) except for the lack of some path functions.

**Relation verbs** The verbs that have only the particles pattern "b" ($< ga, ni >$), i.e. belong to the thematic subclass <b>, and that belong to the aspectual subclass <C> or <R> expless "binomial relation". This subclass contains:

> *taisuru(confront), mensuru(face), sou(be parallel to), motozuku(be based), tugu(be next to), sakidatu(be previous to), kansuru(be concerned with),* etc.

Since these verbs express the static relation between two things, both things must be realized as syntactic arguments; one marked with *ga*-case and the other marked with *ni*-case. However, when the context imply the object which should normally be marked with *ni*-case, it can be omitted. In this case, the thematic subclass amounts to <b,a> which contains:

> *taiousuru(correspond to), kanyosuru(participate in), zokusuru(belong to), tunagaru(be connected with), tyokumensuru(face), soutousuru(be suitable for),* etc.

The LCS is:

$$[_{\text{State}}\text{BE}_{\text{Relational}}([_{\text{Thing}}]_{\text{A}}, [_{\text{Place}}\text{AT}([_{\text{Thing}}]_{\text{A}})])] \tag{10}$$

This LCS is similar to (5) except that the semantic field is Relational and that the second linking subscript marks the thing (which is the argument of the function AT) rather than the place. We adopt the function AT to represent the relation between two things, although it should be elaborate for each verb.

Verbs expressing an attitude like *sanseisuru(approve of)*, *douisuru(consent to)*, *hantaisuru(be opposed to)* are contained in this subclass, since determining one's attitude toward something is equal to having some relation with it. These verbs relate to the thematic subclass <b,a,b4> ("b4" indicates the particle pattern $< ga, nitaisite >$) which contains *hanpatusuru(be offended)*.

On the other hand, the thematic subclasses <a,d> ("d" indicates the particle pattern $< ga, to >$) and <a,d,b1> contains

> *kotonaru(differ from), tigau(differ from), ittisuru(agree with), kasanaru(overlap with)*

These verbs also express the relation between two things; one marked with *ga*-case and the other marked with *to*-case. Since the particle *to* represents mutuality or accompaniment, the relation is mutual rather than static. They belong to aspectual subclass <C> or <R>. Therefore, the LCS can be defined as:

$$\left[_{\text{State}} \begin{array}{l} \text{BE}_{\text{Relational}}([\alpha]_{\text{A}}, [\text{AT}([\beta])]) \\ [\text{WITH}[\text{BE}_{\text{Relational}}([\beta]_{\text{A}}, [\text{AT}([\alpha])])]] \end{array} \right] \tag{11}$$

WITH is a one-place function which takes a State as its argument and indicates the coexistence of the two States $\text{BE}_{\text{Relational}}$.

The thematic subclasses <b,a,d> and <a,b,d> contains:

> *kankeisuru(be related to), kanrensuru(be related to), niru(resemble), au(fit), kakawaru(be concerned with), narabu(be parallel with),* etc.

and the thematic subclasses <b,d> contains

> *kuraberu(compare), sessuru(contact with), butukaru(hit against), musubituku(be related to), deau(come across), au(meet with),* etc.

These verbs lie halfway between (10) and (11). When they appear in the particles pattern "b", i.e. with case particle *ni*, the noun phrase marked with *ni*-case has a static interpretation and the relation implies one-way directionality. On the other hand, when they appear in the particles pattern "d", i.e. with case particle *to*, they express equal relations between two things. Moreover, verbs which have the relatively high ratio of "d" express "mutual action" rather than "mutual relation". The aspectual subclasses also illustrate this difference; from <C>, <R> to <I>, <T>. The verbs which belong to thematic subclass <d,a> and aspectual <T> are:

> *kekkonsuru(marry), kaidansuru(confer with), syoutotusuru(collide), tatakau(fight with), kyougisuru(consult with), sessyokusuru(contact with),* etc.

The LCS of this class is:

$$\left[_{\text{Event}} \begin{array}{l} \text{ACT}([\alpha], [\beta]) \\ [\text{WITH}[\text{ACT}([\beta], [\alpha]) \end{array} \right] \tag{12}$$

ACT is a two-place function which indicates that the first argument acts on the second argument. In order to deal with Actor-Patient relations, Jackendoff introduced an "action tier" apart from a "thematic tier" dealing with motion and location. Although it is useful in dealing with varieties of causation, we collapse two tiers for readability. (ACT was used in (Pinker, 1989).) The LCS (12) represents the fact that $\alpha$ acts on $\beta$ with $\beta$ acting on $\alpha$.

**Mental action verbs** The *to*-case can also mark the resultant state of mental change or the quotation of the cognitive/perceptual contents. Therefore, the thematic subclasses <a,b,d> contains:

> *wakaru(understand), omoeru(seem to)*

and the thematic subclasses <a,d>, and <d,a> contain:

> *ieru(can be said), yobareru(be called)*

These verbs are sometimes called as "spontaneous verbs". They express the fact that a noun phrase marked with *ga*-case is judged to be identical to a noun phrase marked with *to*-case. The agent of the judgement is not explicitly represented. The LCS is:

$$[^{\text{Epistemic}}_{\text{Event}} \text{GO}_{\text{Ident}}([\text{THING}]_{\text{A}}, [_{\text{Path}} \text{TO}([\text{THING}]_{(\text{A})})])] \tag{13}$$

or:

$$[^{\text{Epistemic}}_{\text{State}}\text{BE}_{\text{Ident}}([\text{THING}]_{\text{A}}, [_{\text{Place}}\text{AT}([\text{THING}]_{(\text{A})})])] \tag{14}$$

The distinction can be made by referring to the aspectual subclasses. That is, *wakaru(understand)* belongs to the aspectual subclass <T> while the others belong to <M> or <C>. "Epistemic" is an abstract semantic field where things or events or states can be interpreted as ideas which are changed or contained in sentient minds.

The thematic subclasses that include the particles pattern "k"$(< ga, wo >)$ or "k1"$(< ga, nituite >)$ in addition to the pattern "a"$(< ga >)$ and "d"$(< ga, to >)$, i.e. <a,k,d> or <k,a,d> or <k,a,k1> contain:

> *omou(think), handansuru(judge), kangaeru(consider), sinjiru(believe), kentousuru(examine), noberu(state),* etc.

These verbs express mental actions and the *to*-case denotes the quotation and marks the content of ideas. When the content of ideas is represented as a noun phrase, it can also be marked with *wo*-case and the particles pattern amounts to "k". The noun phrase marked with *ga*-case expresses the agent of the mental action. The aspectual subclass of these verbs is <P>. Thus, the LCS of them is:

$$\left[^{\text{Epistemic}}_{\text{Event}}\ \text{ACT}([_{\text{Thing}}]_{\text{A}}, \left[\left\{\begin{matrix}\text{THING}\\ \text{EVENT}\end{matrix}\right\}\right]_{\text{A}})\ \right] \tag{15}$$

The second argument of ACT indicates the content of ideas by virtue of the semantic field "Epistemic". Although thinking is not a concrete action on a physical object, we use the function ACT in order to represent the agentive properties of thinker, believer, etc.

The thematic subclasses <a,k,d,b> and <a,k,k1,b> contain:

> *iu(say), hanasu(talk to), kataru(tell)*

They express the transmission of information and *ni*-case in the pattern "b" marks a recipient of the information. The LCS is:

$$[^{\text{Epistemic}}_{\text{Event}}\text{CAUSE}([_{\text{Thing}}]_{\text{A}}, [_{\text{Event}}\text{GO}(\left[\left\{\begin{matrix}\text{THING}\\ \text{EVENT}\end{matrix}\right\}\right]_{\text{A}}, [_{\text{Path}}\text{TO}([\text{THING}]_{(\text{A})})])])] \tag{16}$$

The thematic subclasses that include the pattern "g"$(< ga, wo, to >)$ in addition to "d" and "k" contains *minasu(regard)* (<d,g,k>) and *yobu(call)* (<d,k,a,g>). They express the treatment of something. The LCS of them is the one with the function CAUSE whose second argument is (13). It is similar to (16) except for the difference between the function GO and $\text{GO}_{\text{Ident}}$.

**Object adhesion verbs** The thematic subclasses which include the pattern "h" $(< ga, wo, ni >$ in addition to "b" and "k", such as <b,h,k> or <h,b,k> or <h,k,b>, contains:

> *ateru(apply), todokeru(forward), kuwaeru(add), oikomu(drive into), mukeru(direct), okuru(send), ataeru(give), hakensuru(send),* etc.

They express an event where someone causes something to move to somewhere. Since the head of the ordering is the pattern "b" or "h" which contain *ni*-case, rather than the pattern "k" which doesn't, they focus on the goal of the movement. They belong to the aspectual subclass <T> or <I>. The LCS of them is the one with the function CAUSE whose second argument is (4), i.e.:

$$[_{\text{Event}}\text{CAUSE}([_{\text{Thing}}]_{\text{A}}, [_{\text{Event}}\text{GO}([_{\text{Thing}}]_{\text{A}}, [_{\text{Path}}\text{TO}(\left[\left\{\begin{matrix}\text{PLACE}\\ \text{THING}\end{matrix}\right\}\right]_{\text{A}})])])] \tag{17}$$

When the moved object is an abstract noun, the semantic field becomes “Psychological”. In this case, the meaning of verbs is an event where someone imposes a burden on someone else, i.e. “dependence” or “reliance”. They include:

*iraisuru(request), izonsuru(rely on), makaseru(entrust to), tanomu(beg),* etc.

In these subclasses, some verbs like *henkansuru(transform)*, *tenkansuru(convert)* express object change-of-state rather than object movement. This shows the relationship between two meanings. Its semantic field is “Identificational” and the *ni*-case marks resultant state of the change.

**Object change-of-state/Action verbs** Many verbs in the thematic subclasses <k> or <k,a> and the aspectual subclasses <T> or <Q> express change-of-state of the object:

*takameru(raise), tuyomeru(make stronger), kyoukasuru(strengthen), fukameru(deepen), fuyasu(increase), herasu(decrease),* etc.

They are the transitives corresponding to the intransitives which belong to the thematic subclass <a>, i.e. subject change-of-state verbs. Therefore, the LCS of them is:

$$[_{\text{Event}}\text{CAUSE}([_{\text{Thing}}]_{\text{A}}, [_{\text{Event}}\text{GO}_{\text{Ident}}([_{\text{Thing}}]_{\text{A}}, \begin{bmatrix}_{\text{Path}} & \text{FROM}([_{\text{State}}]) \\ & \text{TO}([_{\text{State}}])\end{bmatrix})])] \tag{18}$$

There are various kinds of change-of-states. For example, some express disappearance of the object such as *kesu(get rid of)*, *korosu(kill)*, *nakusu(lose)*, etc., some express production of object such as *seisansuru(produce)*, *umidasu(give rise to)*, *seiseisuru(create)*, *tukuridasu(make)*, etc., some express a change in quality such as *kaizensuru(improve)*, *tuyomeru(make strong)*, etc., others express a momentary change of posture or stance:

*niramu(stare), mitumeru(gaze), nigiru(grasp), suu(breathe), nerau(aim at), motu(hold), sasu(point to), kakaeru(hold),* etc.

There are still more verbs in these subclasses. We may be able to classify them into subtle subclasses by virtue of the meaning of the object noun phrases. However, since it must be subjective and complicated by idiosyncrasies, we don’t specify each of them but put them together as verbs acting on something and give some examples. The LCS of them can be defined as:

$$\left[_{\text{Event}}\ \text{ACT}([_{\text{Thing}}]_{\text{A}}, \left[\left\{\begin{matrix}\text{THING}\\ \text{EVENT}\end{matrix}\right\}\right]_{\text{A}})\ \right] \tag{19}$$

Verbs acting on a Thing express “treatment” or “management” of the object:

*hakaru(measure), syorisuru(manage), tasukeru(help), sapootosuru(support), seigyosuru(control), atukau(treat), sijisuru(support), toriatukau(treat), sousasuru(operate),* etc.

Verbs acting on an Event i.e., whose direct object is an event noun, express intellectual or purposeful action:

*aisuru(love), isikisuru(be conscious of), imisuru(mean), tasikameru(make sure of), kakuninsuru(ascertain), kibousuru(hope), ninsikisuru(perceive), ketteisuru(decide), kiteisuru(regulate),* etc.

Of course, there is no clear distinction between them.

The verbs take *ni*-case and belong to the thematic subclasses <k,a,b>, <k,a,h>, <k,a,b,h> or <k,a,h,b> express “employment” or “object change-of-state”. The former takes an event noun marked with *ni*-case and this means a purpose of the action:

*katuyousuru(utilize), tukau(use), siyousuru(employ), youisuru(prepare), motiiru(make use of), ouyousuru(apply), tekiyousuru(apply),* etc.

The LCS of them is the one in which (19) subordinates $[\mathrm{FOR}[_{\mathrm{Event}}]_{\mathrm{A}}]$ as an adjunct. They belong to the aspectual subclass <C> or <Q> since the employment of something tends to be continuous. On the other hand, the latter belong to the aspectual subclasses <T> or <I> and *ni*-case marks the resultant state of the change:

*kaeru(change), kakutyousuru(extend), hikiageru(increase), aratameru(revise), henkousuru(alter), naosu(amend), waru(divide),* etc.

The LCS of them is the same as (18) of which the argument of the path function TO is marked with a linking subscript A, i.e.:

$$[_{\mathrm{Event}}\mathrm{CAUSE}([_{\mathrm{Thing}}]_{\mathrm{A}}, [_{\mathrm{Event}}\mathrm{GO}_{\mathrm{Ident}}([_{\mathrm{Thing}}\alpha]_{\mathrm{A}}, \left[_{\mathrm{Path}} \begin{array}{l} \mathrm{FROM}([_{\mathrm{State}}]) \\ \mathrm{TO}([_{\mathrm{State}}]_{(\mathrm{A})}) \end{array}\right])])] \qquad (20)$$

**Passage/Starting verbs** Some verbs belonging to the thematic subclass <k,a> and the aspectual <C> or <P> express movement and the *wo*-case marks the course:

*koeru(go across), norikoeru(climb over), tuukasuru(pass), kudaru(go down),* etc.

Since they take a place noun marked with *wo*-case which cannot be substituted for *wotootte(across)*, they contains the passage of the course as their own meaning. Therefore, the LCS of them is:

$$[_{\mathrm{Event}}\mathrm{GO}([_{\mathrm{Thing}}]_{\mathrm{A}}, [_{\mathrm{Path}}\mathrm{VIA}([_{\mathrm{Place}}])]_{\mathrm{A}})] \qquad (21)$$

In this LCS, differing from the ordinary movement verbs (9), the linking subscript A marks a Path containing the function VIA rather than the Place which is an argument of VIA. These verbs means passage without mentioning the source and goal of the movement. Other general movement verbs such as *hasiru(run)* and *aruku(walk)* can take *ni*-case marking the goal or *kara*-case marking the source of the movement and therefore belong to the thematic subclass <k,a,b> *or* <k,a,c>.

Some *wo*-case can be substituted for *kara*-case. The verbs that take this *wo*-case focus on the source of the movement and belong to the thematic subclasses which contain the pattern "c" ($< ga, kara >$) such as <k,c,a>, <k,c,b> or <k,a,c,b>:

*saru(leave), syuppatusuru(start), dassuru(get out of), hanareru(depart from), oriru(get off), nogareru(get away off),* etc.

The LCS of them is:

$$[_{\mathrm{Event}}\mathrm{GO}([_{\mathrm{Thing}}]_{\mathrm{A}}, [_{\mathrm{Path}}\mathrm{FROM}([_{\mathrm{Place}}]_{(\mathrm{A})})]_{(\mathrm{A})})] \qquad (22)$$

The linking subscript A marks either a Path containing the function FROM or the Place which is an argument of FROM. The former is realized as *wo*-case and the latter *kara*-case. This explains the fact that *wo*-case marking an explicit place noun *naka(inside)* is anomalous, since *naka(inside)* cannot be a path but a source place from which someone leaves.

Heya {kara/wo} deru.
"go out of a room."
Heya no naka {kara/?wo} deru.
"go out of the inside of a room."

**Object movement verbs** As opposed to passage/starting verbs, the verbs which belong to the thematic subclasses that include such particles patterns as "i" ($< ga, wo, kara >$), "h" ($< ga, wo, ni >$), "h3" ($< ga, wo, e >$) or "j" ($< ga, wo, kara, e >$), i.e., the patterns including both *wo*-case and *ni*-case or *e*-case or *kara*-case, express an event where someone causes something to move to somewhere. Among them, the subclasses whose head of the ordering of the particles pattern is "k" not "b" or "h" contain:

> *kounyuusuru(purchase), syukkasuru(forward), teikyousuru(offer), harau(pay), yusyutusuru(export), hakobu(carry), yunyuusuru(import), osu(push), hiku(pull), hanatu(release), uru(sell),* etc.

They are the transitives corresponding to the intransitive subject movement verbs. Like them, the ordering of particles patterns varies according as which part of the path is focused on. The aforementioned object adhesion verbs (17) are special cases where the function TO is focused on. On the other hand, when FROM is focused on, the ratio of the pattern "i" ($< ga, wo, kara >$) becomes relatively high. They express "receiving" or "bringing out":

> *uketoru(receive), yomitoru(read one's thought), morau(be given), motidasu(carry out), hikidasu(pull out), ubau(rob), torinozoku(remove),* etc.

When the goal is mentioned, the aspectual subclass tends to be <T>; otherwise <I> or <P>. We can distinguish perceiving verbs that also have abstract motion components such as *miru(see), nagameru(look at)*, etc., since they belong to the aspectual subclass <M>.

The LCS of them is the one with CAUSE whose second argument is (9):

$$\left[_{\text{Event}}\text{CAUSE}\left([_{\text{Thing}}]_{\text{A}}, \left[_{\text{Event}}\text{GO}\left([_{\text{Thing}}]_{\text{A}}, \left[_{\text{Path}}\begin{Bmatrix}\text{FROM}\\ \text{TO}\\ \text{TOWARD}\end{Bmatrix}\left(\left[\begin{Bmatrix}\text{THING}\\ \text{PLACE}\end{Bmatrix}\right]_{\text{A}}\right)\right]\right)\right]\right)\right] \tag{23}$$

However, as opposed to (9), this LCS does not include the path function VIA. This is because when someone moves something *wo*-case marks the moved object rather than the path. In other words, the moved object itself forms the path. If the function GO is substituted for $\text{GO}_{\text{Ident}}$ then the LCS amounts to the one defining object change-of-state verbs mentioned above. Moreover, when the moved object is an abstract noun, i.e. some information, they express "transmitting" or "hearing":

> *sijisuru(instruct), setumeisuru(explain), miseru(show), simesu(indicate), houkokusuru(report), youkyuusuru(demand), siraseru(inform), tutaeru(transmit), kiku(hear), yakusokusuru(promise),* etc.

The LCS is the same as (16). However, since the head of the ordering of particles patterns is "k" ($< ga, wo >$), they focus on the content of the information rather than the recipient.

**Setting verbs** Among object movement verbs, the verbs belonging to the thematic subclasses <k,a,b,h>, <k,a,h>, <k,a,h,b>, <k,b,a,h>, <k,h,a,b>, etc. express "setting" or "building". They belong to the aspectual subclass <R> or <I> or <T>. They have the same LCS as (17) except that the object, rather than the place where it is set, is focused on:

> *kaku(write), kizuku(build), tateru(set up), sinsetusuru(found), seturitusuru(establish), tukuru(make), tousaisuru(load), kazaru(decorate), haru(stick), umeru(bury), kensetusuru(construct), nyuuryokusuru(input), noseru(put on),* etc.

When they contain change-of-state components in their meaning, they participate in the locative alternation as mentioned in the previous section.

**Object exchange/connection verbs** The verbs belong to the thematic subclasses whose head of the ordering of the particles pattern is "k" and also include the pattern "d" ($< ga, to >$) or "g" ($< ga, wo, to >$) express an exchange or connection of the object:

> *ketugousuru(join), koukansuru(exchange), tyouseisuru(coordinate), kumu(unite), kumi-awaseru(combine), naraberu(arrange), setuzokusuru(connect), tougousuru(unify), kyougisuru(talk with), arasou(compete with),* etc.

They are the transitives corresponding to the intransitive mutual relation/action verbs. The LCSs are therefore:

$$\left[_{\text{Event}}\text{CAUSE}\left([_{\text{Thing}}\alpha]_{\text{A}}, \left[_{\text{State}} \begin{array}{l} \text{BE}_{\text{Relational}}([\beta]_{\text{A}}, [\text{AT}([\gamma])]) \\ [\text{WITH}[\text{BE}_{\text{Relational}}([\gamma]_{\text{A}}, [\text{AT}([\beta])])]] \end{array}\right]\right)\right] \quad (24)$$

or:

$$\left[_{\text{Event}} \begin{array}{l} \text{ACT}([\alpha]_{\text{A}}, [\gamma]_{\text{A}}) \\ [\text{WITH}[\text{ACT}([\beta]_{\text{A}}, [\gamma]) \end{array}\right] \quad (25)$$

The former (24) indicates that the subject $\alpha$ causes $\beta$ and $\gamma$ to have some relation, while the latter (25) indicates that both $\alpha$ and $\beta$ act on the object $\gamma$.

**Other verbs** In the thematic subclasses <k,a,h1,b1> and <k,a,h,b>, there are verbs expressing "finding" such as *hakkensuru(discover)*, *mitukeru(find)* and *miidasu(find out)*. They belong to the aspectual subclass <T>. The LCS of them can be defined as the one with the function CAUSE whose second argument is "appearance" (6) and its semantic field is Epistemic.

The thematic subclasses <k,i>, <k,i,a> and <k,a,i> contains "defending" verbs such as *sukuu(rescue)*, *mamoru(guard)* and *hogosuru(protect)*. They have the same particles patterns as "bringing out" verbs. This can be explained by "force dynamics" theory of Talmy (Talmy, 1988). In the force-dynamic model, objects are conceived as bodies with inherent tendencies toward motion or rest. Talmy introduces the roles of "agonist": an entity under focus which has a tendency either for motion or rest, and "antagonist": an entity which opposes the agonist. In the case of bringing out, a stronger antagonist affecting a weaker agonist with a tendency towards rest, results in moving. While in the case of defending, a weaker antagonist affecting a stronger agonist with a tendency towards rest, results in resting. That is, the former corresponds to the occurrence of an event with an effect involving a moving, while the latter corresponds to an event whose effect is the cessation or failure of occurrence of a moving. Thus, in defending verbs, *kara*-case marks the source of an implicit movement force. The LCS is:

$$\left[_{\text{Event}}\text{CAUSE}\left([_{\text{Thing}}\alpha]_{\text{A}}, \left[_{\text{State}}\text{BE}\left([_{\text{Thing}}\beta]_{\text{A}}, \left[\text{NOT AT}([\text{ACT}([\left\{ \begin{array}{c} \text{THING} \\ \text{EVENT} \end{array} \right\} \gamma]_{\text{A}}, [_{\text{Thing}}\beta])])\right]\right)\right]\right)\right] \quad (26)$$

The entire Event in (26) can be read roughly as "A Thing $\alpha$ causes a Thing $\beta$ not to be acted on by an Event or a Thing $\gamma$.

## 5. ORGANIZATION OF SUBCLASSES

We have classified about 80% of 835 verbs semi-automatically. One group of verbs we could not classify properly are the verbs which take optional and semantically ambiguous case markers such as *de*. These verbs have got mixed in with the wrong subclass due to these optional cases. The other group are polysemous verbs such as "light verbs": *suru(do)*, *naru(become)*, etc., which may be included in several distinct subclasses or in superclass of these subclasses. We need to structure the subclasses hierarchically and to clarify the relations among them. These issues remain to be studied in the future.

As mentioned in the section 2, there can be graded membership in a subclass, i.e. there exist prototypical members and marginal members of each subclass. A prototype is a typical instance of a subclass, and other members are assimilated to the subclass on the basis of their perceived resemblance to the prototype; there are degrees of membership based on degrees of similarity. We can incorporate the prototypicality of the members which constitute each semantic subclass by taking account of the frequency of case particle patterns and cooccurring adverbs.

For example, the thematic subclass <a,b> consists of verbs which have various ratios of "a" to "b". We can arrange them in the order of percentage of "a". If the percentage of "a" becomes lower than 50%, then the verbs belong to the thematic subclass <b,a>. The subclass <b,a> also consists of verbs with various ratios of "b" to "a". If the percentage of "b" approaches to 100 %, then the verbs have similar properties with the verbs in the thematic subclass <b>. Moreover, as oft-mentioned in previous section, each subclass has close relationship with each other. Thus, we can detect the organization of the semantic subclasses of all verbs. It is illustrated in the Figure 5. (where each circle expresses a subclass and consists of a mnemonic name and the number of the LCS in the text. We represent them by separate circles and connected them with lines for expository purpose, but of course there is no clear demarcation between them.)

## 6. CONCLUSION

In this paper, we proposed a two-dimensional semantic classification of Japanese verbs. One dimension is thematic, and we exploited patterns of case particles which can be obtained by gathering syntactic argument structures of verbs in a corpus. The other dimension is aspectual, and adverbs which modify verbs of various event types played an important role. By combining these two classifications, we obtained an elaborate classification of verbs and gave a LCS representation for each subclass. Then, we detected the organization of these subclasses by considering interrelationship between them.

Although we consider this classification and the organization of subclasses as rather adequate, it is difficult to evaluate their accuracy. In order to evaluate them, we are implementing the linking rules which map LCSs into the HPSG-style lexical entries and we intend to parse sentences in a unparsed corpus by using this lexicon. This will be described in a subsequent paper.

We used a parsed corpus to extract the dependency relations between words, and this approach depends on the syntactic pre-processing of texts. However, parsed corpora can be achieved by many natural NLP systems or manual analyses with low cost, while the classification and the organization of verbs are hard to obtain with high quality and tend to be subjective unless we use such method as described here.

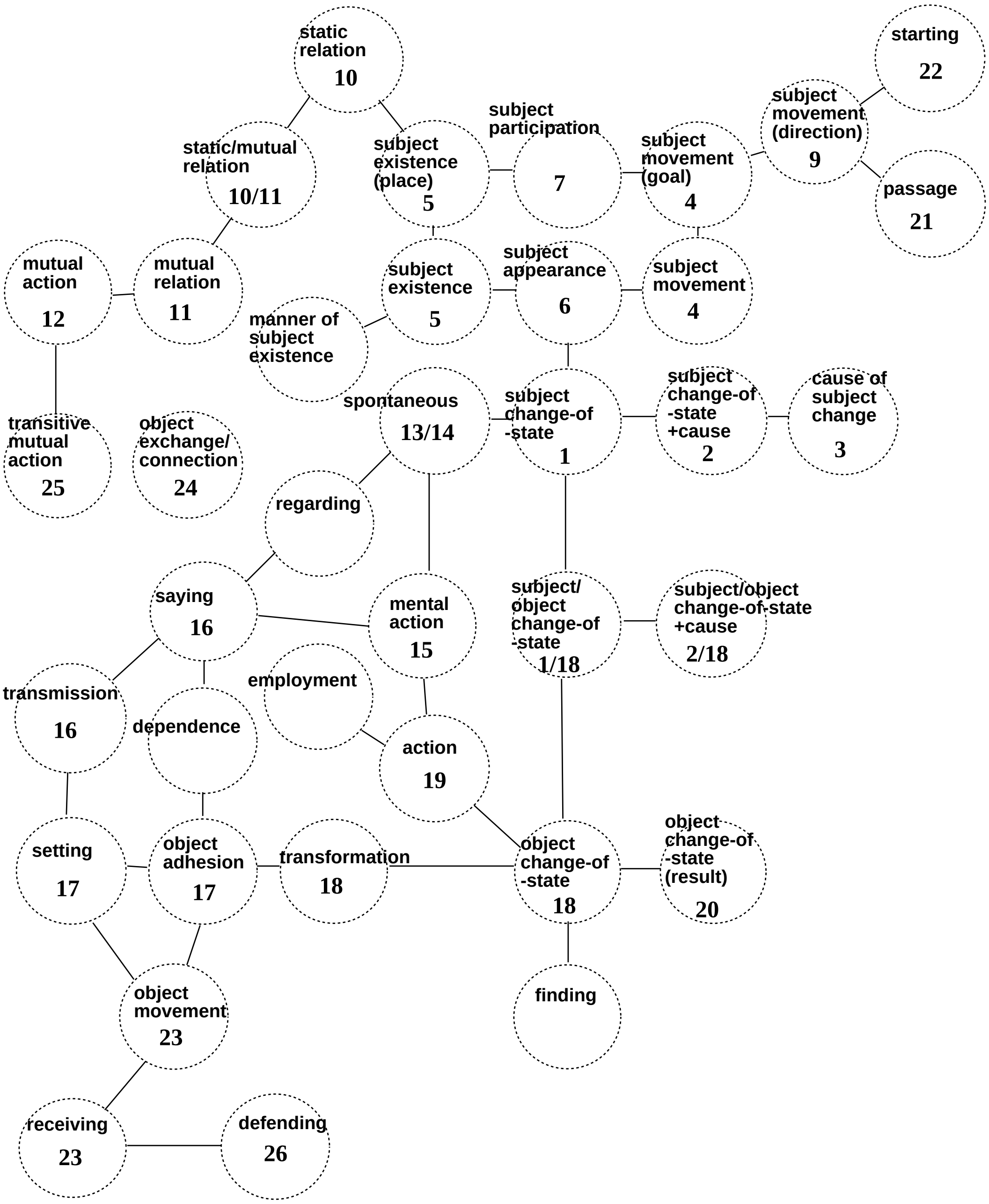


Figure 5: The network of semantic subclasses of verbs

# REFERENCES


Japan Electronic Dictionary Research Institute Ltd. EDR. 1995. *Japanese Co-Occurrence Dictionaries.* (in Japanese).

J. Grimshaw. 1990. *Argument Structure*, volume 18 of *Linguistic Inquiry Monograph.* MIT Press.

R. Jackendoff. 1990. *Semantic Structure.* MIT Press.

G. Lakoff. 1987. *Women Fire,and Dangerous Things : What Categories Reveal about the Mind.* The University of Chicago press.

R.W. Langacker. 1987. *Foundations of Cognitive Grammar, Vol.1 Theoretical Prerequisites.* Stanford University Press.

R.W. Langacker. 1991. *Foundations of Cognitive Grammar, Vol.2 Descriptive Application.* Stanford University Press.

B. Levin and H. M. Rappaport. 1995. *Unaccusativity : At the Syntax-Lexical Semantics Interface*, volume 26 of *Linguistic Inquiry Monograph.* MIT Press.

B. Levin. 1993. *English Verb Classes and Alternations.* The University of Chicago Press.

T. Moriyama. 1988. *Nihongo Doushi Jutsugobun no Kenkyuu (A Study of Japanese verb-pradicate sentences).* Meiji Shoin. (in Japanese).

A. Oishi and Y. Matsumoto. 1995. A method for deep case acquisition based on surface case pattern analysis. *Proceedings of the 3rd Natural Language Processing Pacific Rim Symposium*, 2:678–684.

S. Pinker. 1989. *Learnability and Cognition: The Acquisition of Argument Structure.* MIT Press.

J. Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4):409–441.

J. Pustejovsky. 1992. The syntax of event structure. In B. Levin and S. Pinker, editors, *Lexical and Conceptual Semantics*, chapter 3, pages 47–81. Blackwell.

L. Talmy. 1988. Force dynamics in language and cognition. *Cognitive Science*, 12:49–100.

C. L. Tenny. 1994. *Aspectual Roles and the Syntax-Semantics Interface*, volume 52 of *Studies in Linguistics and Philosophy(SLAP).* Kluwer Academic.

Z. Vendler. 1957. Verbs and times. *Philosophical Review*, 66:143–160.

U. Zernik. 1991. Introduction. In U. Zernik, editor, *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pages 1–26. Lawrence Erlbaum Associates.