

# Automatic Extraction of Aspectual Information from a Monolingual Corpus

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

## Abstract

This paper describes an approach to extract the aspectual information of Japanese verb phrases from a monolingual corpus. We classify verbs into six categories by means of the aspectual features which are defined on the basis of the possibility of co-occurrence with aspectual forms and adverbs. A unique category could be identified for 96% of the target verbs. To evaluate the result of the experiment, we examined the meaning of *-teiru* which is one of the most fundamental aspectual markers in Japanese, and obtained the correct recognition score of 71% for the 200 sentences.

## 1 Introduction

Aspect refers to the internal temporal structure of events and is distinguished from tense, which has a deictic element in it, of reference to a point of time anchored by the speaker's utterance. There is a voluminous literature on aspect within linguistics and philosophy. Recently, computational linguists also have joined in the act within the context of machine translation or text understanding etc. For example, consider the following Japanese sentences (quoted from (Gunji, 1992)).

- (a). Ken-wa ima tonarino heya-de kimono-wo ki-te-i-ru.  
Ken-TOP now next room-LOC kimono-ACC put-on-PRES  
'Ken is now putting on kimono in the next room.'
- (b). Ken-wa kesa-kara zutto ano kimono-wo ki-te-i-ru.  
Ken-TOP this morning-since always that kimono-ACC wear-PRES  
'Ken has been wearing that kimono since this morning.'
- (c). Ken-wa ano kimono-wo san-nen maeni ki-te-i-ru.  
Ken-TOP that kimono-ACC three-year before wear-PRES  
'Ken has the experience of wearing that kimono three years ago.'

Notice that English translations use separate lexical items (*put on* for (a) and *wear* for (b), (c))

and different aspectual configurations (the progressive for (a), the perfect progressive for (b), and another for (c)), while all Japanese sentences contain the same verbal form *ki-te-i-ru*. Thus, when the system tries to translate these sentences, it must be aware of the difference among them.

This paper describes an approach to extract the aspectual information of Japanese verb phrases from a monolingual corpus. In the next section, we will classify Japanese verbs into six categories by means of aspectual features following the framework of (Bennett et al., 1990). The aspectual forms<sup>1</sup> and adverbs are defined as the functions which operate on verbs' aspectual features and changes their values. By using the constraints of the applicability of the functions, we can identify a unique category for each verb automatically. If one can acquire aspectual properties of verbs properly and know how the other constituents in a sentence operate on them, then the aspectual meaning of the whole sentence will be determined monotonically. To evaluate the result of the experiment, we will examine the meaning of *-teiru* which is one of the most fundamental aspectual forms, since the classification itself is difficult to evaluate objectively.

## 2 Realization Process of Aspectual Meaning

We consider that the whole aspectual meaning of verb phrases is determined in the following order: **verbs** → **arguments** → **adverbs** → **aspectual forms**. Adverbs and aspectual forms are defined as indicators of such cognitive processes as "zooming" and "focusing" which operate on the time-line representation. They are similar to the notions "aspectual coercion" (Moens and Steedman, 1988) or

<sup>1</sup>The term "form" refers to grammatical morphemes which are defined in terms of derivation. In this paper, we refer to the aspectual morphemes which follow verbs as "aspectual forms", including compound verbs such as *-hazimeru* (*begin*), suffixes with epenthetic *-te* such as *-teiru*, and aspectual nominals such as *-bakari* (*just now*) etc.

“views” (Gunji, 1992). We explain each in turn.

## 2.1 Aspectual Categories of Verbs

A number of aspectually oriented lexical-semantic representations have been proposed. We adopt and extend the feature-based framework proposed by (Bennett et al., 1990) in the spirit of (Moens and Steedman, 1988). They uses three features:  $\pm$ dynamic,  $\pm$ telic, and  $\pm$ atomic. We add two more features:  $\pm$ process and  $\pm$ gradual.

The feature **dynamicity** distinguishes between states(-d) and events(+d), and **atomicity** distinguishes between point events(+a) and extended events(-a). The duration described by verbs is twofold: an ongoing process and a consequent state. The feature **process** concerns an ongoing process and distinguishes whether events described by verbs have the duration for which some actions unfold. The feature **telicity** distinguishes between culminative events(+t) and nonculminative events(-t). It presupposes a process. The feature **graduality** characterizes events in which some kind of change is included and the change gradually develops.

We can classify verbs by means of different combinations of the five features. Since there are dependences between features, only subsets of the combinatorially possible configurations of features are defined as shown in the Table 1.

In the Table 1, **1.stative verbs** are those that are not dynamic. **2.atomic verbs** are those that express an atomic event. **3.resultative verbs** express a punctual event followed by a new state which holds over some interval of time. **4.process+result verbs** are those that express a complex situation consisting of a process which culminates in a new state. **5.non-gradual process verbs** are those that express only processes and not changes of state. **6.gradual process verbs** are those that have graduality. Although the verbs of the categories 5 and 6 don't contain telicity, the arguments of the verbs or some kinds of adverbs can set up the endpoint of the process as discussed later. In Vendlerian classification, states correspond to 1, achievements to 2 and 3, accomplishments to 4 and 6, activities to 5, respectively (Vendler, 1957).

## 2.2 Arguments

Tenny points out that internal argument of a verb can be defined as that which temporally delimits or measures out the event (Tenny, 1994).

The direct internal argument can aspectually “measure out the event” to which the verb refers. To clarify what is meant by “measuring-out”, she gives examples of three kinds of measuring-out: incremental theme verbs (*eat an apple, build a house* etc.), change-of-state verbs (*ripen the fruit* etc.) and path objects of route verbs (*climbed the ladder, play a sonata* etc.).

On the other hand, the indirect internal argument can provide a temporal terminus for the event described by the verb. The terminus causes the event to be delimited as in *push the car to a gas station*. There is only one kind of internal argument, in terms of thematic roles, that does provide an event terminus, and that is a goal.

In terms of the current framework, both of them add the telicity to the verb which does not inherently contain the telicity. They play a role of framing the interval on which the focus should be brought.

## 2.3 Adverbs

In general, adverbs focus on the subpart of the event described by a verb and give a more detailed description. According to the discussion in (Moriyama, 1988), adverbs can be classified as follows in terms of the subpart on which they focus.

**Processes modifiers** modify verbs which have process (+p). This class includes reduplicative onomatopoeia such as *gasagasa, batabata, suisui, seseto, butubutu*, etc., which are expressing sound or manner of directed motion, and rate adverbs such as *yukkuri(slowly), tebayaku(quickly)*, etc., which express the speed of motions. They focus on the ongoing process of events described by verbs.

**Gradual change indicators** express the progress of change of state, such as *dandan(gradually), sukosizutu(little by little), jojoni(gradually), dondon(constantly), sidaini(by degrees)*, etc., which modify gradual process verbs (+g) and focus on the process.

**Continuous adverbs** are those that can modify both states verbs (-d) and process verbs (+p), such as *zutto(for a long time), itumademo(never)*, etc. They express a continuance of an event or a maintenance of a state.

Table 1: Aspectual categories of verbs

categories	features	examples
1.stative	-d]	<i>aru(be), sobieru(rise), sonzaisuru(exist)</i>
2.atomic	+d,+a]	<i>hirameku(flash), mikakeru(notice)</i>
3.resultative	+d,-a,-p]	<i>suwaru(sit down), tatu(stand up)</i>
4.process+result	+d,-a,+p,+t]	<i>korosu(kill), kiru(put on/wear), akeru(open)</i>
5.non-gradual process	+d,-a,+p,-t,-g]	<i>aruku(walk), iu(say), utau(sing)</i>
6.gradual process	+d,-a,+p,-t,+g]	<i>kusaru(turn sour), takamaru(become high)</i>

**Atomic adverbs** make any events instantaneous, such as *satto*, *ponto*, *gatatto*, *potarito*, *syunkan*, etc., which express instantaneous sound emission or an instant. When these adverbs co-occur with verbs, the events are understood as instantaneous. This doesn't necessarily imply that the verb itself is instantaneous.

**Quantity regulators** measure out events, such as *gokiro aruku* (walk 5km), *gojikan seizasita* (sit straight for 5 hours), etc. These include time, distance, and any quantity of contents.

**End state modifiers** express the consequent state of events, such as *mapputatuni* (into two exact halves), *konagonani* (into pieces), *pechankoni* (be flat), *barabarani* (come apart), etc. They focus on the resultant state.

So far we have described adverbs which concern a single event, but some adverbs regulate the multiple events which involves iteration of a single event. By iteration, the whole process of a collective event can be taken up regardless of the inherent features of verbs.

There are two kinds of **Repetition adverbs**: one regulates the whole quantity of the iteration of events such as *san-kai* (three times) or *nandomo* (many times) etc., and the other describes the habitual repetition of events such as *itumo* (always) or *syottyuu* (very often) etc. Both describe many events each of which involves one person's act.

Finally, we shall mention **Time in the past adverbs**. There are cases where the form *-teiru*, which marks the present tense, can co-occur with temporal adverbs describing the past. (See the example (1c) in the introduction.) It describes the experiential fact of an event. Such adverbs as *katute* (once), *mukasi* (in the past) and *izen* (before) determine the temporal structure of the event related with tense.

## 2.4 Aspectual Forms

The ability of aspectual forms to follow verbs is constrained by the inherent features of verbs. We briefly describe some of aspectual forms used in the experiment.

The forms *-you-to-suru* (be going to) and *-kakeru* (be about to) take up the occurrence of events. They can follow the verbs which are dynamic(+d).

The form *-tuzukeru* (continue) can follow the verbs which have duration(-a). It can take up either the ongoing process or the resultant state. The form *-hajimeru* (begin) can follow the verbs which have process(+p) and takes up the start time of the process. On the other hand, the forms *-owaru* (cease) and *-oeru* (finish) can follow the verbs which are telic(+t) and takes up the end point of the process. However, these constraints on the inherent features of verbs are only concerned with a single event. By iteration, the whole process of a collective event can be taken up regardless of the inherent features of verbs, as mentioned above.

The forms *-tutuaru* (be in progress), *-tekuru* (come into state) and *-teiku* (go into state) focus on the gradual process of change. *-Tutuaru* (be in progress) takes up it as a kind of state, *-tekuru* (come into state) views it from the end state of change while *-teiku* (go into state) from the initial state of change. Both of *-tekuru* and *-teiku* have usages other than aspect, as in *mot-tekuru* (bring) or *mot-teiku* (take).

## 3 Experiment

We carried out an experiment to classify Japanese verbs into six categories in the Table 1 by means of corpus data.

As shown in the Figure 1, each category is defined in terms of the ability to co-occur with aspectual forms. However, the discrimination of the categories needs negative evidence which we cannot use by definition. A corpus only provides positive evidence. Furthermore, some forms can be used regardless of the features and have usages other than aspect as discussed in the previous section. We must establish a method which takes into these facts into account.

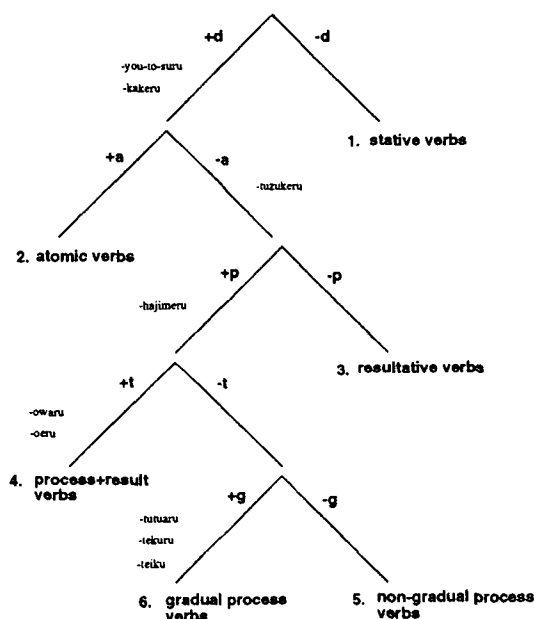


Figure 1: The relation between categories of verbs and features

### 3.1 Algorithm

We used the EDR Japanese Corpus and the EDR Japanese Co-occurrence Dictionary (EDR, 1995) as material to extract syntactic clues in the experiment. The corpus contains 220,000 sentences from various genres of text. The results of the parsing analysis of these sentences indicates that the constituents of the sentence have a dependency struc-

**STEP:1** Pick out the items of which the governing and dependent words are a verb and an adverb from the EDR Co-occurrence Dictionary and store them with the frequency in an array called PAIRS (cf. Table 2).

**STEP:2** For each adverb in PAIRS, give an adverb class label (the initial letter of the class name) on the basis of the discussion in sec. 2.3 and store them in an array called ADVERBS (cf. Table 3 and Table 4).

**STEP:3** For each verb in PAIRS, add up the frequency of the co-occurrence with the adverbs contained in the array ADVERBS. If the sum is greater than 4, store the verb in a list called VERBS.

**STEP:4** For each sentence in the corpus, find a verb and if it is contained in VERBS, then:

**STEP:4-1** If the form following the verb is contained in the predefined list (Table 5), make an array FORMS[i, j] positive (where i is the position of the verb in the list VERBS and j is the position of the form in the Table 5, see Table 6), provided that the verb is not modified by repetition adverbs(R). When the form is *-tekuru* or *-teiku*, put it on record only if the verb is modified by gradual change indicators(G).

**STEP:4-2** If the verb is modified by the adverbs contained in the array ADVERBS, refer to the adverb class label and add 1 to an array MODIFIED[i, k] (where i is the position of the verb in the list VERBS and k is the position of the adverb class label in the Table 4. When the adverb is continuous one(C), distinguish the cases where the verb is followed by *-teiru*(C1) from the other cases(C2), see Table 7), provided that the verb is not followed by negative forms such as *-nai* or *-zu*, nor the forms which change the voice such as *-reru*(the passivizer) or *-seru*(the causativizer), since they affect the aspectual properties of the attached verb.

**STEP:5** For each verb in VERBS:

**STEP:5-1** Narrow down the candidates by means of the array FORMS (on the basis of possible categories shown in Table 5).

**STEP:5-2** In the case where the category of the verb cannot be uniquely identified in STEP:5-1, i.e., other than the category 6, determine it by means of the array MODIFIED as follows:

- the category 6 (if the verb is modified by gradual change indicators(G))
- the category 5 (if modified by process modifiers(P) and not by end state modifiers(E))
- the category 4 (if modified by both process modifiers(P) and end state modifiers(E))
- the category 3 (if modified by end state modifiers(E) and not by process modifiers(P))
- the category 2 (if modified by only atomic adverbs(A))
- the category 1 (if modified by continuous adverbs without being followed by *-teiru*(C2) and not modified by process modifiers(P) nor gradual change indicators(G) nor end state modifiers(E))
- ambiguous (otherwise)

ture. That is, the constituents have a governing-dependent relation. It is these constituents that form the head phrases of the Japanese Co-occurrence Dictionary which describes collocational information in the form of binary relations. Each item in the Japanese Co-occurrence Dictionary consists of a governing word, a dependent word, the relator between the words, and supplementary co-occurrence item information which is composed of the frequency of the co-occurrence relation and a portion of the actual example sentence from which the co-occurrence relation was taken.

The algorithm used for classifying verbs is shown in Figure 2.

Table 2: A part of the array PAIRS

adverb	verb	freq.
<i>aa</i> (like that)	<i>tu</i> (say)	1
<i>aa</i> (like that)	<i>suru</i> (do)	1
<i>ai</i> (mutually)	<i>au</i> (meet)	1
<i>aikawarazu</i> (as usual)	<i>iru</i> (be)	1
<i>aikawarazu</i> (as usual)	<i>otituku</i> (settle)	1
<i>aituide</i> (one after another)	<i>sannyuusuru</i> (join)	3
<i>aituide</i> (one after another)	<i>seturitusuru</i> (establish)	4

Table 3: A part of the array ADVERBS

adverb	label
<i>aikawarazu</i> (as usual)	C
<i>aegiaegi</i> (gasping)	P
<i>akaakato</i> (brightly)	P
<i>akuseku</i> (busily)	P
<i>atafuta</i> (in a hurry)	P
<i>atafutato</i> (in a hurry)	P
<i>attoiroma</i> (in an instance)	A
<i>ikiiki</i> (vividly)	P

Figure 2: The algorithm for classifying verbs

Table 4: Results of the classification of adverbs

adverb class(label)	total	examples
process modifiers (P)	470	yukkuri(slowly), gasagasa, batabata, suisui, sesseto, butubutu,...
gradual change indicators (G)	52	sidaini(gradually), masumasa(increasingly), jojoni(gradually)...
continuous adverbs (C)	78	sonomama(as it is), zutto(for a long time), itumademo(never)...
atomic adverbs (A)	294	satto, ponto, gatatto, potarito, syunkan(instantaneously)...
quantity regulators (Q)	12	180-do(180 degree), ippo(a step), ippai(a cup)...
end state modifiers (E)	86	mapputatuni(into two exact halves), konagonani(into powder)...
repetition adverbs (R)	122	nandomo(many times), itumo(always), syottyuu(very often)...
time in the past adverbs (T)	11	katute(once), mukasi(in the past), izen(before)...

Table 5: The aspectual forms used in the experiment

forms	followable verb categories
-you-to-suru(be going to), -kakeru(be about to)	2, 3, 4, 5, 6
-tuzukeru(continue)	3, 4, 5, 6
-hajimeru(begin)	4, 5, 6
-owaru(end) and -oeru(finish)	5, 6
-tutuaru(be in progress), -tekuru(come into state), -teiku(go into state)	6

Table 6: A part of the array FORMS

verb	forms				
	-kakeru	-tuzukeru	-hajimeru	-owaru	-tutuaru
akkasuru(become worse)	-	-	+	-	+
nigiru(catch)	+	+	-	-	-
anteisuru(become stable)	-	-	+	-	+
isikisuru(become conscious)	-	-	+	-	+
kotonaru(differ)	-	-	-	-	-
idousuru(move)	-	-	+	-	-
ijisuru(maintain)	-	+	-	-	-
tigau(differ)	-	-	-	-	-
sodatu(grow)	-	-	-	-	+
sodateru(bring up)	-	-	-	-	+
ittisuru(agree)	-	-	+	-	-

Table 7: A part of the array MODIFIED

verb	adverb class labels						
	P	G	C1	C2	A	Q	E
akkasuru(become worse)	0	5	0	0	1	0	0
nigiru(catch)	0	1	0	1	0	0	1
anteisuru(become stable)	0	1	1	1	0	0	1
isikisuru(become conscious)	0	1	0	1	0	0	0
kotonaru(differ)	0	1	0	0	0	0	1
idousuru(move)	1	1	0	1	1	0	0
ijisuru(maintain)	0	0	0	4	0	0	0
tigau(differ)	0	1	0	0	1	0	0
sodatu(grow)	5	3	0	0	0	1	1
sodateru(bring up)	3	1	0	1	0	0	0
ittisuru(agree)	0	0	0	0	3	0	2

The steps 1, 2 and 3 are the processes to determine the target verbs. There are 431 verbs modified by the classified adverbs more than 4 times. In step 2, we classify adverbs on the basis of the discussion in the previous section. Although the classification has been done by hand, it is much easier than that of verbs, since adverbs are fewer than verbs in number (2,563 vs. 12,766 in the corpus) and have higher “iconicity” — the isomorphism between form and meaning — than verbs. This classification of adverbs is used not only for determining the aspectual categories of verbs but also for examining the meaning of *-teiru* as mentioned later.

The step 4 is a process to register the co-occurring forms and adverbs for each verb. By using these data, we identify the aspectual categories of verbs in the step 5. Since the categories cannot be uniquely identified by aspectual forms only, we use adverbs which can modify the only restricted set of verbs as shown in Table 8.

Table 8: Adverb classes and their modifiable verb categories

adverb class(label)	verb categories
process modifiers (P)	4, 5, 6
gradual change indicators (G)	6
continuous adverbs (C)	1, 3, 4, 5, 6
atomic adverbs (A)	2, 3, 4, 5, 6
quantity regulators (Q)	1, 3, 4, 5, 6
end state modifiers (E)	3, 4, 6

### 3.2 Evaluation and Discussion

Out of 431 target verbs, we could uniquely identify categories for 375 verbs. As for the rest 56 verbs, 37 verbs were identified in the step 5-2 as the category which was not included in the set of categories outputted by the step 5-1. This seems to be due to the failure to detect the expression of repetition, therefore, we chose the category determined in the step 5-2. Table 9 shows the results.

Table 9: The verb classification obtained by the experiment

verb category	no.	examples
1.stative	30	<i>mitumeru(stare) ijisuru(maintain) sumu(live) sonzaisuru(exist) nagameru(view) damaru(be silent) kurikaesu(repeat) tukaeru(can be used) ...</i>
2.atomic	19	<i>nageru(throw) haneagaru(leap up) kizuku(notice) mikakeru(happen to see) gousuru(arrive at an agreement) kireru(snap) fumikiru(launch out) ...</i>
3.resultative	29	<i>nureru(become wet) tumaru(become packed) tunagaru(make a connection) au(meet) suwaru(sit down) tatamu(fold) kureru(get dark) atehamaru(fit) ...</i>
4.process+result	30	<i>tateru(build) nobasu(lengthen) matomeru(put together) narabu(form a line) tutumu(wrap) majiwaru(associate) tiru(fall) torikakomu(surround) ...</i>
5.non-gradual process	94	<i>nomu(drink) hakobu(carry) ianosimu(enjoy) kansatusuru(observe) furueru(shake) hibiku(ring) tobimawaru(fly about) taberu(eat) sugosu(spend) ...</i>
6.gradual process	210	<i>akkasuru(get worse) tuyomaru(get strong) takamaru(become raised) sinkoukasuru(get more acute) seiyousuru(grow up) kappatukasuru(make active) ...</i>
ambiguous	19	<i>kuwawaru(join) tutomeru(be employed) tomonau(accompany) tazuneru(visit) rainitisuru(come to Japan) uwamawaru(be more than) hokoru(boast) ...</i>

We confirmed that more than 80% of verbs are correctly classified. However, this is a subjective judgement. To evaluate the results of the classification more objectively, we focus on one evaluation metric; namely the automatic examination of the meaning of *-teiru* which can represent several distinct senses as described in the introduction.

The form *-teiru* indicates “zoom in” operation: it is a function that takes an event as its input and returns a type of states, which refers to unbounded regions i.e., a part of the time-line with no distinct boundaries. Figure 3 shows the time-line representation for each aspectual category of verbs. Aspectual distinctions correspond to how parts of the time-line are delineated.

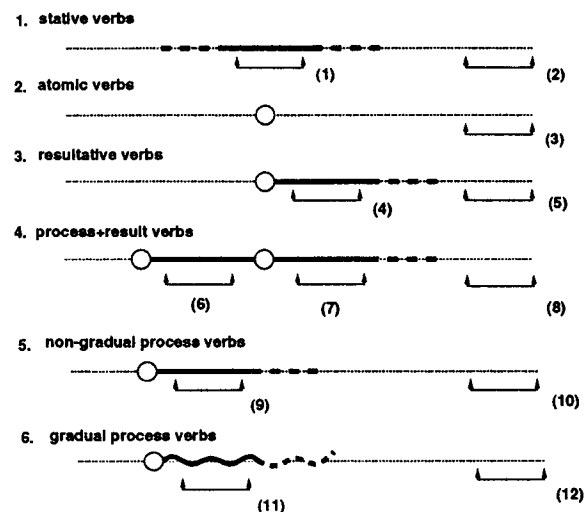


Figure 3: The time-line representation for aspectual categories of verbs

In Figure 3, thick line segments signify regions, dashed line segments signify unbounded ends of regions and large open dots signify points in time (boundaries or punctate events).

Since *-teiru* cannot include a time instant at which a state is drastically changed, it must denote one of the intervals depicted below the lines. The interval (1) in Figure 3 designates a state which is a part of the state described by a lexical stative verb. It means a state holding before a speaker's eyes.

It has been stated from (Kindaichi, 1976) that the form *-teiru* has three distinct senses: 'a simple state', 'a progressive state' and 'a consequent state'. (1) corresponds to a simple state, (4) and (7) to a consequent state, (6), (9) and (11) to a progressive state, respectively. Though not represented in Figure 3, a consequent state can be taken up with the verbs of categories 5 and 6 if the endpoints of the processes are set up by explicit expressions.

Kudo (Kudo, 1982) has pointed out that there are inherent meaning and derivative meaning for both progressive and consequent states and has sorted out them as follows.

- (i) inherent meaning of 'a progressive state': an ongoing process
- (ii) derivative meaning of 'a progressive state': an iteration
- (iii) inherent meaning of 'a consequent state': a resultative state
- (iv) derivative meaning of 'a consequent state': an experiential state
- (v) otherwise: a simple state

(ii) is the above-mentioned process of a collective event; "a line as a set of points", so to speak. (iv) is a state where a speaker has an experience of the event described by a verb and corresponds to the intervals (2), (3), (5), (8), (10), (12) in Figure 3. These derivative meanings are conditioned syntactically or contextually, that is, they are stipulated as derivative by explicit linguistic expressions such as adverbials etc., while not concerned with the inherent features of verbs — they can appear with most of verbs regardless of their aspectual categories.

We carried out an experiment to examine the meaning of *-teiru* automatically by means of the classifications of verbs and adverbs obtained in the previous experiment. Table 10 shows the determination process of the meaning of *-teiru*. We checked the cases in Table 10 downward from the top.

Table 11 shows the results obtained from running the process of Table 10 on 200 sentences containing *-teiru* which are randomly selected from the EDR Japanese Corpus.

The precision on the whole is 71%. Note that the sense (i) 'an ongoing process' has high recall but low precision, while (iii) 'a resultative state' and (iv) 'an experiential state' show the opposite. This is due to the fact that the test sentences contain many "speech-act" verbs such as *syuchousuru* (*insist*), *setumeisuru* (*explain*), *hyoumeisuru* (*declare*) etc. They are classified as 5.non-gradual process verbs, and by

Table 10: The determination process of the meaning of *-teiru*

case	output
(1).the verb is modified by repetition adverbs(R)	(ii) an iteration
(2).the verb is modified by time in the past adverbs(P) or its category is 2. atomic verbs	(iv) an experiential state
(3).the category of the verb is 1. stative verbs	(v) a simple state
(4).the category of the verb is 3. resultative verbs	(iii) a resultative state
(5).the verb is modified by process modifiers(P) or gradual change indicators (G)	(i) an ongoing process
(6).the endpoint of the process is explicitly set up (the verb is modified by end state modifiers(E) or quantity regulators(Q) or it takes a goal argument i.e., <i>ni(to)</i> -case etc.	(iii) a resultative state
(7).the process cannot be taken up (the verb is modified by atomic adverbs(A) or <i>sudeni(already)</i> , etc.)	(iii) a resultative state
(8).the category of the verb is 5. non-gradual process or 6. gradual process verbs	(i) an ongoing process
(9).the category of the verb is 4. process+result verbs	ambiguous: (i) or (iii)

the case 8 in Table 10, the senses of *-teiru* following them are determined as (i) 'an ongoing process'. However, they takes a quotative *to*-case that marks the content of the statement and this measures out the event described by verbs. Therefore the resultative or experiential readings are preferred.

The other errors are caused by polysemous verbs such as *kakaru* (*hang/lie/fall...*) or *ataru* (*hit/strike/be exposed/shine...*). Their aspectual properties are changed by the complements they take. The analysis of how complements influence the aspectual properties of their governing verbs is beyond the scope of this paper. It seems to be a matter of pragmatic world knowledge rather than sense-semantic (but see (Verkuyl, 1993) for English).

## 4 Related Work

The approach proposed here is similar to that of Dorr's (Dorr, 1992; Dorr, 1993), but different from it in scale and determinability of the categories. She adopts the four-way classification system following Vendler (Vendler, 1957) and utilizes Dowty's test (Dowty, 1991) for determining aspectual categories of English verbs. She reports the results obtained from running the program on 219 sentences of the LOB corpus. Although we cannot know how many verbs she tested because she has shown only a subset of the verbs, the program was not able to pare down the aspectual category to one in 18 cases out of 27 verbs.

Brent (Brent, 1991) discusses an implemented program that automatically classifies verbs into two groups, stative vs. non-stative, on the basis of their syntactic contexts. He uses the progressive and rate-

Table 11: The results of the evaluation experiment

the sense of -teiru	judgement by human(a)	output of program(b)	number of agreements(c)	recall(%) c/a×100	precision(%) c/b×100
(i) an ongoing process	95	137	88	93	64
(ii) an iteration	4	2	2	50	100
(iii) a resultative state	29	15	14	48	93
(iv) an experiential state	39	15	14	36	93
(v) a simple state	19	19	15	79	79
ambiguous	14	12	9	64	75
total	200	200	142	71	71

adverbs constructions in combination with some sort of statistical smoothing technique. He identified eleven verbs as purely stative, of the 204 distinct verbs occurring at least 100 times in the LOB corpus.

We think that the extraction of aspectual information must be based on principles that are well-grounded in linguistic theory. However, some sort of noise reduction technique such as the confidence intervals used by Brent may be needed to detect the cue more accurately.

## 5 Conclusion

In this paper, we have proposed a method for classifying Japanese verbs on the basis of surface evidence from a monolingual corpus, and examined the meaning of the form *-teiru* by means of the classifications of verbs and adverbs.

The aspect of verb phrases provides not only the temporal configuration within a single event but also the information needed for processing temporal relation between multiple events (Dowty, 1986; Passonneau, 1988; Webber, 1988).

Furthermore, the lexical aspect of verbs is closely related with their deep complement structures which may not be directly reflected on the surface argument structures. Therefore, by combining the aspectual categories of verbs and those that are defined in terms of their surface argument structures, we can obtain an elaborate classification based on semantic types of verbs. (Preliminary experiments on this issue can be seen in (Oishi and Matsumoto, 1996).)

Thus, the information obtained here can be used for various applications.

## References

S. W. Bennett, T. Herlick, K. Hoyt, J. Liro, and A. Santisteban. 1990. A computational model of aspect and verb semantics. *Machine Translation*, 4(4):247-280.

M. R. Brent. 1991. Automatic semantic classification of verbs from their syntactic contexts: An implemented classifier for stativity. In *Proceedings of the 5th ACL European Chapter*, pages 222-226.

B. J. Dorr. 1992. A parameterized approach to integrating aspect with lexical-semantic for machine translation. In *Proceedings of the 30th Annual Meeting of ACL*, pages 257-264.

B. J. Dorr. 1993. *Machine Translation — A View from the Lexicon*. The MIT Press.

D. R. Dowty. 1986. The effects of aspectual class on the temporal structure of discourse. *Linguistics and Philosophy*, 9(1):37-61.

D. R. Dowty. 1991. *Word Meaning and Montague Grammar : The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*, volume 7 of *Studies in Linguistics and Philosophy(SLAP)*. Kluwer Academic Publishers.

Japan Electronic Dictionary Research Institute Ltd. EDR. 1995. *the EDR Electronic Dictionary Technical Guide*. (in Japanese).

T. Gunji. 1992. A proto-lexical analysis of temporal properties of Japanese verbs. In *Linguistics Studies on Natural Language*, Kyung Hee Language Institute Monograph One, pages 197-217. Hanshin Publishing.

H. Kindaichi. 1976. *Nihongo Dousi-no Asupekuto ('Aspect of Japanese Verbs')*. Mugi Shobo. (in Japanese).

M. Kudo. 1982. Siteiru-kei-no imi-kijutu ('the description of the meaning of the form *-teiru*'). *Musashi University Jinbun Gakkai Zasshi*, 13(4).

M. Moens and M. Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics*, 14(2):15-28.

T. Moriyama. 1988. *Nihongo Doushi Jutsugobun no Kenkyuu ('A Study of Japanese Verb-pradicate Sentences')*. Meiji Shoin. (in Japanese).

A. Oishi and Y. Matsumoto. 1996. Detecting the organization of semantic subclasses of Japanese verbs. Technical Report NAIST-IS-TR96019, Nara Institute of Science and Technology.

R. J. Passonneau. 1988. A computational model of the semantics of tense and aspect. *Computational Linguistics*, 14(2):44-60.

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

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

H. Verkuyl. 1993. *A Theory of Aspectuality*. Cambridge University Press.

B. L. Webber. 1988. Tense as discourse anaphor. *Computational Linguistics*, 14(2):61-73.