

A Computational Treatment of V-V Compounds in Japanese

Chikara Hashimoto

Department of Informatics, Kyoto University

Francis Bond

Machine Translation Research Group,  
NTT Communication Science Laboratories

Proceedings of the HPSG05 Conference

Department of Informatics, University of Lisbon

Stefan Müller (Editor)

2005

CSLI Publications

<http://csli-publications.stanford.edu/>

## Abstract

We examine how a large-scale computational grammar can account for the complex nature of Japanese verbal compounds. Previous computational Japanese grammars have tried to avoid the problem by simple solutions such as enumerating as many verbal compounds in the lexicon as possible. In contrast, we develop the analysis that is linguistically adequate and computationally tractable and thus meets the requirement of a syntactically and semantically precise natural language processing of Japanese like Bond et al. (2005). Our analysis distinguishes between two kinds of verbal compounds: syntactic compounds, which are fully productive; and lexical compounds, which are of varying productivity.

## 1 Introduction

In this study, we examine how a large-scale computational grammar can account for the complex nature of Japanese verbal compounds ( $V_1$ - $V_2$  compounds, hereafter), such as *yomi-owaru* (read-finish) ‘finish to read’. It is necessary to develop a linguistically accurate and computationally tractable analysis for  $V_1$ - $V_2$  compounds, since they are common in written documents and spontaneous speech, and, despite their surface simplicity, they show various complexities. To date, several computational Japanese grammars have been developed, but little attention has been paid to  $V_1$ - $V_2$  compounds. In fact, their approaches are either enumerating all  $V_1$ - $V_2$ s in the lexicon as if they were single words without internal structures (the exhaustive listing approach) or simply concatenating the  $V_1$  and  $V_2$  of any kind of  $V_1$ - $V_2$  without taking into account the differences in their syntactic and semantic composition (the simple concatenation approach). The former suffers from undergeneration since some patterns are very productive and moreover a  $V_1$ - $V_2$  can embed another one.

- (1) Ken-ga musuko-o [[[nade-mawasi]-tuzuke]-sobire]-kake-ta  
Ken-NOM son-ACC [[[stroke-slua]-continue]-fail]-be.about.to-PAST  
‘Ken was about to fail to continue to caress his son.’

The latter approach leads to overgeneration since not all combinations of two verbs are allowed:

- (2) a. \*yu-ga waki-age-ta  
hot.water-NOM boil-raise-PAST  
‘Water reached a full boil.’  
b. yu-ga waki-aga-tta  
hot.water-NOM boil-go.up-PAST  
‘Water reached a full boil.’

---

<sup>†</sup>We appreciate many people for helping this research. We especially thank Takao Gunji, Melanie Siegel, Dan Flickinger, Sato Satoshi and the other members of the NTT Machine Translation Research Group.

We develop the analysis of  $V_1$ - $V_2$ s that is compatible with the linguistic analyses and observations made by Kageyama (1993) and Matsumoto (1996) while being computationally tractable. The analysis is implemented in JACY (Siegel and Bender, 2002) using the LKB system (Copestake, 2002) and evaluated with the Hinoki corpus (Bond et al., 2004) and the [incr tsdb()] system (Oepen and Carroll, 2000). A slightly different version of the grammar, along with the analysis tools, is available at the Deep Linguistic Processing with HPSG (DELPH-IN) website: <http://www.delph-in.net>.

## 2 Data

$V_1$ - $V_2$ s show differences in terms of how productive they are, how their transitivity and case-marking are determined, whether or not they are compositional, and what semantic composition they undergo if they are compositional. First, as for their productivity, some  $V_1$ - $V_2$ s are very productive and allow even a phrase in the  $V_1$  position. In (4), for example, the  $V_1$ - $V_2$  headed by *sobireru* (fail) allows the phrasal  $V_1$ , *nade-te age* (stroke-TE give), while the  $V_1$ - $V_2$  headed by *mawasu* (fondle) does not.

- (3) a. Ken-ga musuko-o nade-sobire-ta  
 Ken-NOM son-ACC stroke-fail-PAST  
 ‘Ken failed to stroke his son.’
- b. Ken-ga musuko-o nade-mawasi-ta  
 Ken-NOM son-ACC stroke-fondle-PAST  
 ‘Ken caressed his son.’
- (4) a. Ken-ga musuko-o nade-te age-sobire-ta  
 Ken-NOM son-ACC stroke-TE give-fail-PAST  
 ‘Ken failed to stroke his son.’
- b.\*Ken-ga musuko-o nade-te age-mawasi-ta  
 Ken-NOM son-ACC stroke-TE give-fondle-PAST  
 ‘Ken caressed his son.’

Second, some  $V_1$ - $V_2$ s inherit  $V_2$ 's transitivity and case-marking (5), while others are given those of  $V_1$  (6).

- (5) a. Ken-ga huku-o kiru  
 Ken-NOM clothes-ACC wear  
 ‘Ken wears clothes.’
- b. huku-ga kuzureru  
 clothes-NOM get.out.of.shape  
 ‘Clothes get out of the shape.’

c. huku-ga ki-kuzureru  
 clothes-NOM wear-get.out.of.shape  
 ‘Clothes get out of the shape by someone’s wearing.’

(6) a. Ken-ga siai-ni katu  
 Ken-NOM game-DAT win  
 ‘Ken wins games.’

b. Ken-ga siai-o tuzukeru  
 Ken-NOM game-ACC continue  
 ‘Ken continues games.’

c. Ken-ga siai-ni kati-tuzukeru  
 Ken-NOM game-DAT win-continue  
 ‘Ken continues to win games.’

Third, some  $V_1$ - $V_2$ s show semantic compositionality (7), but others are highly lexicalized (8).

(7) a. Ken-ga nikki-o kaki-hazime-ta  
 Ken-NOM diary-ACC write-begin-PAST  
 ‘Ken began to write a diary.’

b. Ken-ga naki-saken-da  
 Ken-NOM cry-shout-PAST  
 ‘Ken cried and shouted.’

(8) a. Ken-ga sono-ninmu-o uti-ki-tta  
 Ken-NOM that-mission-ACC hit-cut-PAST  
 ‘Ken aborted the mission.’

b. keisatu-ga hanzai-o tori-simaru  
 police-NOM crime-ACC take-fasten  
 ‘Police control crimes.’

Finally, compositional  $V_1$ - $V_2$ s are composed in diverse ways. (9a)–(9b) correspond to (7a)–(7b), respectively.

(9) a.  $\exists x \exists y \text{begin}(x, \text{write}(x, y))$

b.  $\exists x \text{and}(\text{cry}(x), \text{shout}(x))$

Table 1: Syntactic V<sub>1</sub>-V<sub>2</sub>s vs. Lexical V<sub>1</sub>-V<sub>2</sub>s

	Syntactic	Lexical
<b>Productivity</b>	Very productive; the V <sub>2</sub> s allow almost any V <sub>1</sub> .	Not so productive; the combination of V <sub>1</sub> and V <sub>2</sub> is more restricted.
<b>Transitivity</b>	The V <sub>1</sub> 's transitivity and case-marking are passed to the V <sub>1</sub> -V <sub>2</sub> .	Either V <sub>1</sub> or V <sub>2</sub> or both participate in the determination of transitivity and case-marking.
<b>Compositionality</b>	Compositional.	Some of them show varying degrees of compositionality, but others are highly lexicalized.
<b>Semantics</b>	The semantics of V <sub>2</sub> consistently embeds V <sub>1</sub> 's semantics.	There are various kinds of semantic composition.

### 3 Analysis

#### 3.1 Linguistic Analyses

Kageyama (1993)'s insightful analysis claims that different behaviors of different V<sub>1</sub>-V<sub>2</sub>s are mostly predictable from how they are composed. He distinguishes two major types: syntactic V<sub>1</sub>-V<sub>2</sub> compounds and lexical V<sub>1</sub>-V<sub>2</sub> compounds. The two component verbs of syntactic V<sub>1</sub>-V<sub>2</sub> compounds are combined in the syntax, while lexical V<sub>1</sub>-V<sub>2</sub> compounds are formed in the lexicon. Accordingly, syntactic V<sub>1</sub>-V<sub>2</sub>s are generally as productive and compositional as ordinary phrases, but lexical V<sub>1</sub>-V<sub>2</sub>s are often irregular and idiomatic. Table 1 summarizes the characteristics of the two types in more detail.

Kageyama further divides syntactic V<sub>1</sub>-V<sub>2</sub>s into three types: Raising (e.g. V<sub>1</sub>-*kakeru* (V<sub>1</sub>-be.about.to) 'be about to V<sub>1</sub>'), Control (e.g. V<sub>1</sub>-*sobireru* (V<sub>1</sub>-fail) 'fail to V<sub>1</sub>'), and  $\bar{V}$  complementation types (e.g. V<sub>1</sub>-*tukusu* (V<sub>1</sub>-exhaust) 'work out to V<sub>1</sub>'). This is supported by, among other things, a contrast in passivizability; Raising and Control types do not allow passivization of V<sub>1</sub>-V<sub>2</sub>, while the  $\bar{V}$  type does.

- (10) hon-ga Ken-ni yomi-{\*kake/\*sobire/tukus}-rare-ta  
 book-NOM Ken-DAT read-{\*be.about.to/\*fail/exhaust}-PASS-PAST

Also, the three kinds show differences in whether V<sub>2</sub>s thematically restrict their subjects and objects.

- (11) a. *ame-ga huku-o nurasi*-{*kake*/\**sobire*/\**tukusi*}-*ta*  
rain-NOM clothes-acc humidify-*{be.about.to*/\**fail*/\**exhaust*}-PAST  
‘The rain {was about/failed/worked out} to wet the clothes.’
- b. *Ken-ga atama-o hiyasi*-{*kake*/*sobire*/\**tukusi*}-*ta*  
Ken-NOM head-ACC cool-*{be.about.to*/\**fail*/\**exhaust*}-PAST  
‘Ken {was about/failed/worked out} to cool off.’

Since  $V_{2s}$  of Control (*-sobireru*) and  $\bar{V}$  (*-tukusu*) types put a thematic restriction on a subject, which the subject, *ame* (rain) in (11a), cannot satisfy, only the Raising type (*-kakeru*) is grammatical in the example. In (11b), only the  $\bar{V}$  type is ruled out since it restricts an object to something that can be exhausted, but the object, *atama*, which is a part of the idiom, *atama-o hiyasu* ‘cool off,’ cannot meet the restriction.

Matsumoto (1996) classifies lexical  $V_1$ - $V_{2s}$  into seven subtypes according to the semantic relations between  $V_1$  and  $V_2$ . Each subtype, its example and a tentative semantics of the example are depicted in (12).

- (12) a. Pair  $V_1$ - $V_{2s}$

ex) *naki-sakebu* (cry-shout)  $\cdots$  and(*shout*( $x$ ), *cry*( $x$ ))

*Ken-ga naki-saken-da*  
Ken-NOM cry-shout-PAST  
‘Ken cried and shouted.’

- b. Cause  $V_1$ - $V_{2s}$

ex) *yake-sinu* (burn-die)  $\cdots$  cause(*burn*( $x$ ), *die*( $x$ ))

*Ken-ga yake-sin-da*  
Ken-NOM burn-die-PAST  
‘Ken was burned to death.’

- c. Manner  $V_1$ - $V_{2s}$

ex) *kake-yoru* (run-come)  $\cdots$  in.manner.of(*come*( $x$ ), *run*( $x$ ))

*Ken-ga musuko-ni kake-yo-tta*  
Ken-NOM son-DAT run-come-PAST  
‘Ken ran up to his son.’

- d. Means  $V_1$ - $V_{2s}$

ex) *tataki-kowasu* (hit-break)  $\cdots$  by.means.of(*break*( $x$ ,  $y$ ), *hit*( $x$ ,  $y$ ))

Ken-ga sara-o tataki-kowasi-ta  
Ken-NOM dish-ACC hit-break-PAST  
'Ken battered down the dish.'

e.  $V_1$ - $V_2$ s with Deverbalized  $V_1$

ex) *sasi-semaru* (thrust-close) ··· emphasized.by(close( $x$ ), thrust)

Kiken-ga sasi-semaru  
danger-NOM thrust-close  
'Dangerous situation becomes imminent.'

f.  $V_1$ - $V_2$ s with Deverbalized  $V_2$

ex) *hare-wataru* (clear.up-cross) ··· modified.by(clear.up( $x$ ), cross)

sora-ga hare-wataru  
sky-NOM clear.up-cross  
'Skies are sunny.'

Matsumoto notes how the semantic relation determines the transitivity and the semantic composition of  $V_1$ - $V_2$  and posits a semantic analysis to deal with the phenomena. Although Matsumoto presents a precise and comprehensive analysis, it assumes fine-grained semantic notions and a complicating mapping theory. To implement this, the grammar would have to recognize which semantic relation holds between the two component verbs. But this depends heavily on world knowledge and pragmatic inference, and hence is not currently computationally tractable.

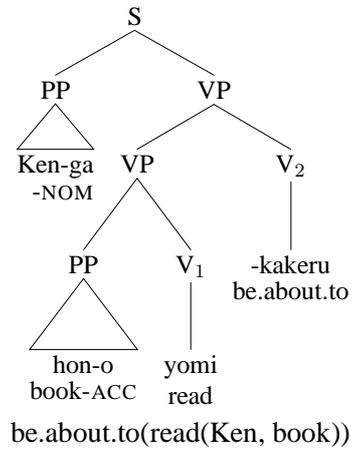
In sum, Kageyama (1993) and Matsumoto (1996) present useful analyses, but these must be revised to make them computationally tractable.

### 3.2 Computational Analysis — Proposal

Our analysis of syntactic  $V_1$ - $V_2$ s is mostly compatible with Kageyama (1993) but, as an HPSG analysis, assumes neither PRO nor government. (13) illustrates the analysis. (the V-embedding type corresponds to Kageyama's  $\bar{V}$  complementation type.)

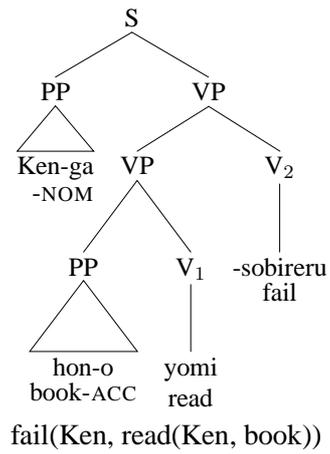
(13) a. **Raising**

'Ken is about to read a book.'



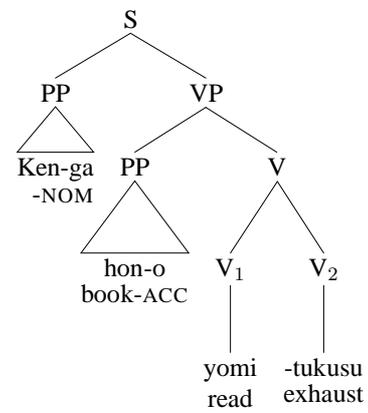
b. **Control**

'Ken fails to read a book.'



c. **V-embedding**

'Ken reads a book thoroughly.'



exhaust(Ken, book, read(Ken, book))

The Raising and Control structures are almost the same as those of Sag et al. (2003); the subject of Raising type  $V_2$  is “raised” from the  $V_1$ , and the subject of Control type  $V_2$  controls that of the  $V_1$ . The V-embedding type has a structure where the subject and object of the  $V_2$  control the subject and object of the  $V_1$ , respectively. These characteristics of the three are reflected in their semantic representations in (13). That is, the Raising type  $V_2$ , *kakeru* (be.about.to) in (13a), does not thematically restrict its subject, *Ken*, and object, *hon* (book), while the Control type  $V_2$ , *sobireru* (fail), puts a thematic restriction on its subject, *Ken*. The V-embedding type  $V_2$  assigns thematic roles to both the subject and object. Clearly, these differences account for (11). Note, in addition, that the Raising and Control types have a VP embedding structure, while the V-embedding type does not. The contrast in (10) is accounted for by this difference; only the object of the V-embedding type is selected by both the  $V_1$  and  $V_2$ , thus only this structure allows the passivization of  $V_1$ - $V_2$  as a whole. Other things to notice are that it is the  $V_1$  that determines the  $V_1$ - $V_2$ 's transitivity and, in most cases, case-marking, and that their semantic structures are consistently embedding structures.

One of the divergences from Kageyama (1993) involves the  $V_1$  passivization. Kageyama (1993) always accepts the  $V_1$  passivization of Control type but necessarily rules out that of his  $\bar{V}$  complementation type, based on the difference in their syntactic configurations: the VP complement vs. the  $\bar{V}$  complement. But this is incorrect as shown in (14).

- (14) a. \*hon-ga yom-are-sobireru  
book-NOM read-PASS-fail  
'A book fails to be read.'
- b. Ken-ga nagur-are-tukusu  
Ken-NOM punch-PASS-exhaust  
'Ken endures the successive punches.'

We basically allow all  $V_1$  passivizations but semantically restrict them. In (14a), for example, the subject, *hon* (book), cannot be construed as FAILER. In (14b), on the other hand, *Ken* can be interpreted as the one who exhausts himself by being punched a lot.

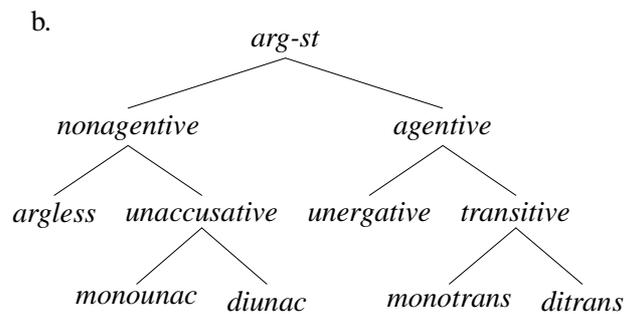
As for lexical  $V_1$ - $V_2$ s, we classify them into five subtypes roughly following Matsumoto (1996).

- (15) a. **Right-headed  $V_1$ - $V_2$ s**
- b. **Argument mixing  $V_1$ - $V_2$ s**
- c.  **$V_1$ - $V_2$ s with deverbialized  $V_1$**
- d.  **$V_1$ - $V_2$ s with deverbialized  $V_2$**

e. **Non-compositional V<sub>1</sub>-V<sub>2</sub>s**

The Right-headed and Argument mixing types jointly cover most of Matsumoto's Pair, Cause, Manner and Means compounds. The Non-compositional type is introduced to distinguish compositional and non-compositional V<sub>1</sub>-V<sub>2</sub>s. Unlike the finer grained semantic analysis of Matsumoto, our analysis leaves the exact semantic relationship under-specified. The constraints on composition come from an extended ARG-ST. As illustrated in (16), the ARG-ST consists of one EXTERNAL argument and two INTERNAL arguments and is classified into six types, following Imaizumi and Gunji (2000).

(16) a. 
$$\begin{bmatrix} \textit{arg-st} \\ \text{EXT} & \textit{index} \\ \text{INT1} & \textit{index} \\ \text{INT2} & \textit{index} \end{bmatrix}$$



c.

	EXT	INT1	INT2
<i>argless</i>	×	×	×
<i>monounac</i>	×	○	×
<i>diunac</i>	×	○	○
<i>unergative</i>	○	×	×
<i>monotrans</i>	○	○	×
<i>ditrans</i>	○	○	○

First, the Right-headed V<sub>1</sub>-V<sub>2</sub> obeys the Shared Participant Condition proposed by Matsumoto (1996), which requires that the two component verbs share at least one argument that is co-indexed with an argument of the other component verb. Any two arguments can be co-indexed between V<sub>1</sub> and V<sub>2</sub> if the arguments agree in the EXT/INT distinction. The transitivity and case-marking of the V<sub>1</sub>-V<sub>2</sub> are inherited from the V<sub>2</sub> (hence Right-headed). The semantics is totally compositional; the two semantic representations of the V<sub>1</sub> and V<sub>2</sub> are predicated by an underspecified semantic relation, which can be specified as Pair, Cause, Manner or Means by a component outside the grammar. For example, the semantic representations of the first two V<sub>1</sub>-V<sub>2</sub>s in (12) can be glossed as *unspec\_rel(shout(x),cry(x))*

and *unspec\_rel(burn(x),die(x))*. The semantic relation cannot be fully specified in a purely syntactic account since it is affected by contexts, pragmatics, and world knowledge, as these become available, the relation can be constrained further. Research on specifying the semantic relation typically uses information about verb selectional restrictions and noun semantic classes that is not available in our grammar (Uchiyama et al., 2005).

Further, the underspecification greatly simplifies the implementation. The Right-headed  $V_1$ - $V_2$ , formulated in this way, covers most of the lexical  $V_1$ - $V_2$ s (Pair, Cause, Manner and Means of Matsumoto's) without making the grammar complicated.

Second, the Argument mixing  $V_1$ - $V_2$  has a peculiarity; it is ambiguous in that they can take arguments from either the  $V_1$  or  $V_2$ . *nomi-aruku* (drink-walk), for example, can take as the object either something to drink ( $V_1$ 's argument) or a place to walk ( $V_2$ 's argument), according to Matsumoto (1996). To account for this, we underspecify the transitivity and case-marking of the  $V_1$ - $V_2$  such that they can be inherited from either the  $V_1$  or  $V_2$ . Another peculiarity involves the fact that the  $V_2$  is restricted to a *monotrans* verb that expresses a spatial motion,<sup>1</sup> while the  $V_1$  is *transitive* and must not be a spatial motion verb. As for the semantics, it is the same as that of the Right-headed  $V_1$ - $V_2$  except that the semantic relation is always construed as Manner.

Third, the  $V_1$ - $V_2$  with deverbalized  $V_1$  includes a  $V_1$  that is deverbalized and only emphasizes the content of  $V_2$  in some way (Kageyama, 1993; Matsumoto, 1996). For instance, *sasi-semaru* (thrust-close), in our analysis, represents something like *emphasize(close(x))*. In the sense that the  $V_1$  is deverbalized, the  $V_1$ - $V_2$  is considered not fully compositional. Naturally, as the  $V_1$  is deverbalized, it is the  $V_2$  that determines the transitivity and case-marking of the  $V_1$ - $V_2$ . As Kageyama (1993) notes, there is no restriction on the possible combinations of the  $V_1$  and  $V_2$  in terms of ARG-ST.

Fourth, the  $V_1$ - $V_2$  with deverbalized  $V_2$ , as the name implies, includes a  $V_2$  that loses its original verbal meaning and takes on an adverbial meaning that modifies the  $V_1$  (Kageyama, 1993; Matsumoto, 1996). *hare-wataru* (clear.up-cross), for instance, can be glossed as *cross(clear.up(x))* in our analysis. Similarly to the  $V_1$ - $V_2$  mentioned in the last paragraph, this type of  $V_1$ - $V_2$  is also considered not fully compositional, since the  $V_2$  has lost its original verbal meaning. Regarding the transitivity and case-marking of the  $V_1$ - $V_2$ , the  $V_1$  determines them since the  $V_2$  is deverbalized. In addition, according to Kageyama (1993), the  $V_1$  and  $V_2$  of this type must agree in agentivity, unlike the  $V_1$ - $V_2$  with semantically deverbalized  $V_1$ .

The two types with a deverbalized component verb lexically encode an embedding semantic structure, similarly to the lexical treatment of the 'biclausal' nature of Japanese causatives proposed by Manning et al. (1996).

---

<sup>1</sup>In the JACY framework, a locative accusative argument is considered an object.

As for productivity, the first two types are more productive than the last two. Actually, we can freely coin a  $V_1$ - $V_2$  that belongs to the first one, the Right-headed  $V_1$ - $V_2$ , as long as it is semantically and pragmatically plausible. On the other hand, the Non-compositional  $V_1$ - $V_2$  is absolutely not productive and literally non-compositional; the  $V_1$ - $V_2$  is totally lexicalized and should be analyzed as a single word.

All in all, even though our analysis might be coarser than Kageyama (1993) and Matsumoto (1996), it is sufficient to account for  $V_1$ - $V_2$ 's complex characteristics summarized in §2 and Table 1. Where there is insufficient information to decide the semantics we under-specify, which makes the analysis both correct and tractable.

## 4 Evaluation

To see if our implementation works well in practice, we conducted a corpus-based evaluation and examined its coverage, the amount of ambiguity, and efficiency. First, we extracted a small evaluation corpus from the Hinoki corpus (Bond et al., 2004). The evaluation corpus consists of 219 sentences, where each sentence contains at least one  $V_1$ - $V_2$ . In addition, we prepared two versions of JACY: JACY-plain and JACY-vv. JACY-plain is given no  $V_1$ - $V_2$  implementation but contains 1,325 lexical entries in the lexicon, which were added by the developers over the course of its development. In contrast, JACY-vv is equipped with all the  $V_1$ - $V_2$  implementations but without any compositional  $V_1$ - $V_2$  entries in the lexicon. Table 2 shows the results of the experiment. We find that JACY-vv gains

Table 2: Experimental results

	<b>JACY-plain</b>	<b>JACY-vv</b>
<b>Coverage (%)</b>	52.1	63.5
<b>Ambiguity (<math>\phi</math>)</b>	53.41	50.78
<b>time (<math>\phi</math>)</b>	4.85	6.43
<b>space (<math>\phi</math>)</b>	816779	995681

more coverage and less ambiguity than JACY-plain. The increased coverage is due to the remarkable productivity of the Right headed type. The reduction in ambiguity involves the more restricted nature of our approach to the free word order of Japanese. The table also shows the two versions' processing efficiency: **time** and **space**.<sup>2</sup> Adding the rules and lexical types for  $V_1$ - $V_2$ s slightly degrades JACY-vv's efficiency. However, JACY-vv still works fast enough for practical NLP applications.

<sup>2</sup>**time** shows how long the grammar needs to parse one sentence, and **space** shows how much memory the grammar consumes to parse one sentence.

## 5 Conclusion

We have provided and implemented an analysis for Japanese verbal compounds that captures their syntactic and semantic properties. We follow Kageyama (1993) in dividing them into syntactic verbal compounds and lexical verbal compounds.

Syntactic compounds are fully compositional. There are three types: raising, control and  $\bar{V}$  complementation.

Lexical compounds are of varying compositionality. We further divided them into five subtypes depending on how their argument structures combine: right-headed, argument mixing, deverbilized  $V_1$ , deverbilized  $V_2$ , and fully lexicalized non-compositional compounds. These types make use of an extended argument structure to constrain the classes of verbs that can appear in each type.

We implemented the analyses in the JACY grammar. We then tested them against corpus data to confirm their correctness.

## References

- Bond, Francis, Copestake, Ann, Flickinger, Dan, Oepen, Stephen and Siegel, Melanie. 2005. Open Source Machine Translation with DELPH-IN. In *Proceedings of The 10th Machine Translation Summit*.
- Bond, Francis, Fujita, Sanae, Hashimoto, Chikara, Nariyama, Shigeko, Nichols, Eric, Ohtani, Akira, Tanaka, Takaaki and Amano, Shigeaki. 2004. The Hinoki Treebank — A Treebank for Text Understanding. In *Proceedings of the First International Joint Conference of Natural Language Processing*, pages 554–559.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications.
- Imaizumi, Shinako and Gunji, Takao. 2000. Complex Events in Lexical Compounds. In Tanaka Itou and Shuichi Yatabe (eds.), *Lexicon and Syntax* (in Japanese), pages 33–59, Hitsuji Shobou.
- Kageyama, Taro. 1993. *Grammar and Word Formation* (in Japanese). Hitsuji Shobou.
- Manning, Christopher D., Sag, Ivan A. and Iida, Masayo. 1996. The Lexical Integrity of Japanese Causatives. In Takao Gunji (ed.), *Studies in the Universality of Constraint-Based Structure Grammars*, pages 9–37, Osaka University.
- Matsumoto, Yo. 1996. *Complex Predicates in Japanese: A Syntactic and Semantic Study of the Notion ‘Word’*. CSLI Publications.

- Oepen, Stephen and Carroll, John. 2000. Performance profiling for grammar engineering. *Natural Language Engineering* pages 81–97.
- Sag, Ivan A., Wasow, Thomas and Bender, Emily M. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications, second edition.
- Siegel, Melanie and Bender, Emily M. 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization*, Taipei, Taiwan.
- Uchiyama, Kiyoko, Baldwin, Timothy and Ishizaki, Shun. 2005. Disambiguating Japanese Compound Verbs. *Computer Speech and Language, Special Issue on Multiword Expressions* 19, Issue 4, 497–512.