

# Adjective-to-Verb Paraphrasing in Japanese Based on Lexical Constraints of Verbs

Atsushi Fujita<sup>†</sup> Naruaki Masuno<sup>††</sup> Satoshi Sato<sup>†</sup> Takehito Utsuro<sup>‡</sup>

<sup>†</sup>Graduate School of Engineering, Nagoya University

<sup>††</sup>IBM Engineering & Technology Services, IBM Japan, Ltd.

<sup>‡</sup>Graduate School of Systems and Information Engineering, University of Tsukuba

## Abstract

This paper describes adjective-to-verb paraphrasing in Japanese. In this paraphrasing, generated verbs require additional suffixes according to their difference in meaning. To determine proper suffixes for a given adjective-verb pair, we have examined the verbal features involved in the theory of Lexical Conceptual Structure.

## 1 Introduction

Textual expressions that (roughly) convey the same meaning are called paraphrases. Since generating and recognizing paraphrases has a potential to contribute to a broad range of natural language applications, such as MT, IE, and QA, many researchers have done a lot of practices on automatic paraphrasing in the last decade.

Most previous studies have addressed paraphrase phenomena where the syntactic category is not changed: e.g., noun-to-noun (“document”  $\Leftrightarrow$  “article”), verb-to-verb (“raise”  $\Leftrightarrow$  “bring up”). In these *inner-categorial paraphrasing*, only limited types of problems arise when replacing words or phrases with their synonymous expressions. On the other hand, this paper focuses on *inter-categorial paraphrasing*, such as adjective-to-verb (“attractive”  $\Leftrightarrow$  “attract”) that leads to novel type of problems due to the prominent differences in meaning and usage. In other words, calculating those differences is more crucial to determine how they can or cannot be paraphrased.

The aim of this study is to clarify what lexical knowledge is required for capturing those differences, and to explore where such a knowledge can be obtained from. Recent work in lexical semantics has shown that syntactic behaviors and semantic properties of words provide useful information to explain the mechanisms of several classes of paraphrases. More specifically, lexical properties involved in the theory of *Lexical Conceptual Structure* (LCS) (Jackendoff, 1990) have seemed

to be beneficial because each verb does not function idiosyncratically. However, in the literature, there have been less studies for other syntactic categories than verbs. To the best of our knowledge, the Meaning-Text Theory (MTT) (Mel’čuk and Polguère, 1987) is one of the very few frameworks. In MTT, lexical properties and inter-categorial paraphrasing are realized with a unique semantic representation irrespective of syntactic categories and what are called lexical functions, e.g.,  $S_0(\textit{receive}) = \textit{reception}$ .

To make out how the recent advances in lexical semantics for verbs can be extended to other syntactic categories, we assess LCS for inter-categorial paraphrasing. We choose adjectives as a counterpart of paraphrasing because they behave relatively similar to verbs compared with other categories: both adjectives and verbs have inflection and function as predicates, adnominal elements, etc. Yet, we speculate that their difference in meaning and usage reveal intriguing generation problems. To put it briefly, adjective-to-verb paraphrasing in Japanese requires verbal suffixes such as “*ta* (past / attributive)” in example (1)<sup>1</sup>:

- (1) s. *furui otera-no jushoku-o tazune-ta.*  
be old temple-GEN priest-ACC to visit-PAST  
I visited a priest in the old temple.  
t. *furubi-ta otera-no jushoku-o tazune-ta.*  
to olden-ATTR temple-GEN priest-ACC to visit-PAST  
I visited a priest in the olden(ed) temple.

## 2 Preliminary investigation

To make an investigation into the variation and distribution of required verbal suffixes, we collected a set of paraphrase examples through the following semi-automatic procedure:

**Step 1.** We handcrafted adjective-verb pairs based on JCore (Sato, 2004), which classifies Japanese words into five-levels of readability. Our 128 pairs (for 85 adjectives) contain only those sharing first few phonemes (reading)

<sup>1</sup>For each example, “s” and “t” denote an original sentence and its paraphrase, respectively.

Table 1: Distribution of verbal suffixes used.

Verbal suffix	$C_{ad_c}$	$C_{pr_1}$	$C_{pr_2}$
<i>ru</i>	9	16	0
<i>tei-ru</i>	5	42	0
<i>re-ru</i>	14	8	0
<i>re-tei-ru</i>	2	5	0
<i>ta</i>	57	0	7
<i>tei-ta</i>	2	0	2
<i>re-ta</i>	6	0	1
<i>re-tei-ta</i>	0	0	1
both <i>ta</i> and <i>tei-ru</i>	4	0	0
both <i>ta</i> and <i>ru</i>	1	0	0
<i>tea-ru</i>	0	2	0
Total	100	73	11

where  
*ru*: base form  
*tei*: progressive / perfective  
*re*: passive / potential  
*ta*: past / attributive  
*tea*: perfective

and characters (*kanji*), and either of adjective or verb falls into the easiest three levels.

**Step 2.** Candidate paraphrases for a given sentence collection are automatically generated by replacing adjectives with their corresponding verbs. Multiple candidates are generated for adjectives that correspond to multiple verbs.

**Step 3.** The correctness of each candidate paraphrase is judged by two human annotators. The basic criterion for judgement is that *two sentences are regarded as paraphrases if and only if they share at least one interpretation*. In this step, the annotators are allowed to revise candidates: (i) append verbal suffixes, (ii) change of case markers, and (iii) insert adverbs. Finally, candidates that both annotators judge correct qualify as paraphrases.

Assuming that the variation and distribution of verbal suffixes vary according to the usage of adjectives, we separately collected paraphrase examples for adnominal and predicative usages.

**Adnominal usages:** For 960 sentences randomly extracted from a one-year newspaper corpus, Mainichi 1995, we obtained 165 examples for 142 source sentences. We then divided them into two portions: 12 adjectives that appeared only once and at least one examples for the other adjectives were kept unseen ( $C_{ad_o}$ ), while the remaining examples ( $C_{ad_c}$ ) were used for our investigation.

**Predicative usages:** For 157 example sentences within IPAL adjective dictionary (IPA, 1990), we generated candidate paraphrases. 84 candidates for 70 sentences qualified as paraphrases. They are then divided into two portions according to the tense of adjectives:  $C_{pr_1}$  consists of examples where adjectives appear in base form and  $C_{pr_2}$  is for “*ta*” form (past tense).

Table 1 shows the distribution of verbal suffixes used for given *adjective-verb pairs* in each portion of example collections. We confirmed that their distribution was fairly different. In the remaining sections, we focus on adnominal usages because examples of predicative usages have displayed a

degree of compositionality. Which of “*ru*” or “*ta*” must be used is given by the input: if a given adjective accompanies past tense, the resultant verbal suffix is necessarily that for present tense followed by “*ta*.”

### 3 Determining verbal suffixes

The task we address here is to determine verbal suffixes for a given input, a pair of an adnominal usage of adjective in a certain context and a candidate verb given by our adjective-verb list.

From the viewpoint of language generation, this task can be thought of as generating verbal expressions where options are already given in Table 1. A straightforward way for determining verbal suffixes is to make use of lexical properties of verbs as constraints on generation. To manifest them, in particular aspectual properties involved in LCS, we first designed seven types of linguistic tests shown in Table 2. They are derived from a classical analysis of verb semantics in Japanese (Kageyama, 1996) and some ongoing projects on constructing LCS dictionaries (Kato et al., 2005; Takeuchi et al., 2006). We then manually examined 128 verbs in Section 2 under those tests. To determine the word sense in which the derivative relationship hold good, example sentences in IPAL verb dictionary (IPA, 1987) for each verb were used. For a verb which was out of the dictionary, we manually gave a sample sentence.

Since our aim is to explain why a certain verbal suffix is used for a given input, we have not feverishly applied a machine learning algorithm to the task. Instead, we have manually created a rule-based model shown in Table 3 using  $C_{ad_c}$ , where each *if-then rule* assigns either of verbal suffixes in Table 1 to a given input based on verbal features in Table 2 and some other features below:

- *D*: affix pair of the adjective and the candidate verb: e.g., “A\_shii-V\_mu” for “*kuyashii* (be regretful)”  $\Leftrightarrow$  “*kuyamu* (to regret)”
- *N*: disjunction of semantic classes in a thesaurus (The Natural Institute for Japanese Language, 2004) for the modified noun
- *C*: whether the adjective is head of clause

### 4 Experiment and discussion

By conducting an empirical experiment with  $C_{ad_c}$  and  $C_{ad_o}$ , we evaluate how our model (**RULE**) properly determines verbal suffixes. A comparison with a simple baseline model (**BL**) is also done. **BL** selects the most frequently used suffix (in this experiment “*ta*”) for any given input.

Table 2: Linguistic tests for verbs derived from Lexical Conceptual Structure (Kageyama, 1996).

Label	Description
$V_a$	whether the verb allows accusative case
$V_b$	whether the verb can co-occur with a temporal adverb “ <i>ichi-jikan</i> (for one hour)” or its variant
$V_c$	whether the verb can co-occur with a temporal adverb “ <i>ichi-jikan-de</i> (in one hour)” or its variant
$V_d$	whether the verb can be followed by “ <i>tearu</i> (perfective)” when its accusative case is moved to nominative
$V_e$	interpretation of the verb followed by “ <i>tei-ru</i> (progressive / perfective)”
$V_f$	when followed by “ <i>ta</i> ,” whether the verb can have the perfective interpretation or just past tense
$V_g$	whether the verb can co-occur with a sort of adverb which indicates intention of the action: e.g. “ <i>wazato</i> (purposely)” and “ <i>iyaiya</i> (reluctantly)”

Table 3: The rule-set for determining verbal suffixes, where “(non)” indicates non-paraphrasable.

Order	Condition (conjunction of “feature label = value”)	Verbal suffix
1	$V_a = \text{“yes”} \wedge V_b = \text{“yes”} \wedge V_f = \text{“no”} \wedge N = \text{“except Human (1.10)”} \wedge D = \text{“A}_{ui}\text{-V}_{bumu} \vee \text{“A}_{j}\text{-V}_{mu} \vee \text{“A}_{asii}\text{-V}_{u}$	<i>re-ru</i>
2	$V_a = \text{“yes”} \wedge V_b = \text{“yes”} \wedge V_f = \text{“no”} \wedge V_d = \text{“no”} \wedge N = \text{“Mind: mind, attitude (1.303)”}$	<i>ta</i>
3	$V_a = \text{“no”} \wedge V_g = \text{“yes”}$	<i>ta</i>
4	$V_a = \text{“no”} \wedge V_f = \text{“yes”} \wedge D = \text{“A}_{j}\text{-V}_{migakaru}$	<i>ta / tei-ru</i>
5	$C = \text{“clause”} \wedge D = \text{“A}_{i}\text{-V}_{maru}$	<i>ru</i>
6	$V_a = \text{“no”} \wedge V_f = \text{“yes”}$	<i>ta</i>
7	$V_a = \text{“no”} \wedge V_b = \text{“yes”} \wedge V_f = \text{“no”}$	<i>ta</i>
8	$V_b = \text{“yes”} \wedge V_f = \text{“no”} \wedge V_c = \text{“yes”} \wedge V_d = \text{“yes”} \wedge V_e = \text{“progressive”} \wedge N = \text{“Subject (1.2)”}$	<i>tei-ru</i>
9	*	(non)

Table 4 shows the experimental results, where recall and precision are calculated with regard to input adjective-verb pairs. Among rules in Table 3, rules 1 (for “*re-ru*”), 3, 6, and 7 (for “*ta*” where  $V_a = \text{“no”}$ ) performed much better than the other rules. This indicates that these rules and features in their conditions properly reflect our linguistic intuition. For instance, rule 6 reflects that a change-of-state intransitive verb expresses resultative meaning as adjectives when it modifies *Theme* of the event via “*ta*” (Kageyama, 1996) as shown in (1), and rule 2 does that a psychological verb modifies a nouns with “*re-ru*” when the noun arouses the specific emotion, such as regretting mistakes (e.g., “*kuyashii* (be regretful)”  $\Leftrightarrow$  “*kuyama-re-ru* (be regretted)”). The aspectual property captured by the tests in Table 2 is used to classify verbs into these semantic classes.

On the other hand, the rules for the other types are immature due to lack of examples: we cannot find out even necessary conditions to be “*ru*,” “*tei-ru*,” etc. What is required to induce proper conditions for these suffixes is a larger example collection and discovering another semantic property and a set of linguistic tests for capturing it.

## 5 Conclusion and future work

In this paper, we focused on inter-categorical paraphrasing and reported on our study on an issue in adjective-to-verb paraphrasing. Two general-purpose resources and a task-specific rule-set have been handcrafted to generate proper verbal suffixes. Although the rule-based model has achieved better performance than a simple baseline model, there is a plenty of room for improvement.

Table 4: Recall and precision of determining verbal suffix for given adjective-verb pairs.

Verbal suffix	$C_{adc}$		$C_{adp}$	
	Recall	Precision	Recall	Precision
<i>ta</i> ( $V_a = \text{“yes”}$ )	3/13	3/3	1/6	1/1
<i>ta</i> ( $V_a = \text{“no”}$ )	42/44	42/63	18/18	18/29
<i>re-ru</i>	12/14	12/19	7/13	7/11
<i>ru</i>	3/9	3/6	0/2	0/5
<i>tei-ru</i>	1/5	1/7	2/8	2/6
<i>ta / tei-ru</i>	2/4	2/2	1/2	1/1
No rule	for 11 inputs		for 7 inputs	
Total (RULE)	63/100 (63%)	63/100 (63%)	29/56 (52%)	29/53 (55%)
<b>BL</b>	57/100 (57%)	57/148 (39%)	24/56 (43%)	24/83 (29%)

Future work includes (i) to enlarge our two resources as in (Dorr, 1997; Habash and Dorr, 2003) evolving an effective construction method, (ii) intrinsic evaluation of those resources, and, of course, (iii) to enhance the paraphrasing models through further experiments with a larger test-set.

## References

- B. J. Dorr. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12(4):271–322.
- N. Habash and B. J. Dorr. 2003. A categorial variation database for English. In *Proceedings of the 2003 Human Language Technology Conference and the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pages 17–23.
- IPA. 1987. *IPA Lexicon of the Japanese language for computers (Basic Verbs)*. Information-technology Promotion Agency. (in Japanese).
- IPA. 1990. *IPA Lexicon of the Japanese language for computers (Basic Adjectives)*. Information-technology Promotion Agency. (in Japanese).
- R. Jackendoff. 1990. *Semantic structures*. The MIT Press.
- T. Kageyama. 1996. *Verb semantics*. Kurosio Publishers. (in Japanese).
- T. Kato, S. Hatakeyama, H. Sakamoto, and T. Ito. 2005. Constructing Lexical Conceptual Structure dictionary for verbs of Japanese origin. In *Proceedings of the 11th Annual Meeting of the Association for Natural Language Processing*, pages 871–874. (in Japanese).
- I. Mel’čuk and A. Polguère. 1987. A formal lexicon in meaning-text theory (or how to do lexica with words). *Computational Linguistics*, 13(3-4):261–275.
- S. Sato. 2004. Identifying spelling variations of Japanese words. In *Information Processing Society of Japan SIG Notes, NL-161-14*, pages 97–104. (in Japanese).
- K. Takeuchi, K. Inui, and A. Fujita. 2006. Construction of compositional lexical database based on Lexical Conceptual Structure for Japanese verbs. In T. Kageyama, editor, *Lexicon Forum No.2*. Hitsuji Shobo. (in Japanese).
- The Natural Institute for Japanese Language. 2004. *Word list by semantic principles, revised and enlarged edition*. Dainippon Tosho. (in Japanese).