

## Verb Semantics and Argument Realization in Pre-Modern Japanese: A Preliminary Study of Compound Verbs in Old Japanese

BJARKE FRELLESVIG

*University of Oxford*

KERRI L RUSSELL

*University of Oxford*

STEPHEN W HORN

*University of Oxford*

PETER SELLS

*SOAS, University of London*

**Abstract:** This paper introduces the collaborative corpus-based research project *Verb semantics and argument realization in pre-modern Japanese*. As part of the project, we are developing a corpus of pre-modern Japanese texts which is encoded with grammatical, and in particular syntactic, information and we here present two pilot studies based on the corpus, concerning verb-verb (V1-V2) compounds in Old Japanese (8th century). We first focus on V2s, with a view to understanding what properties are characteristic of the loosely defined class of ‘auxiliary verbs’ (*bojodoshi* 補助動詞). We apply a number of tests to compounds, including for *lexical integrity* and *transitivity harmony*, and thereby identify a number of V2s that can take part in ‘non-lexical compounds’ (compounds relatively unconstrained by the semantics of their component verbs), as well as some distributional and combinatory patterns typical of non-lexical compounds. Second, we examine a single high-frequency verb, *omop-* ‘think, feel’, in order to examine its argument-taking properties when used as a predicate alone and when used as a V1 in a compound. We identify interesting differences, in particular finding that *omop*-V2 compounds are less likely to take clausal complements than when *omop-* is used as a predicate on its own.\*

**Key words:** verb semantics, argument realization, pre-modern Japanese, syntax, verb compounding

### 1. Introduction

Broadly defined, historical linguistics is a wide field which includes the attested and unattested past of all varieties of language. More narrowly defined, however, historical linguistics deals with the history of individual languages as reflected in the extant written sources, i.e., in the textual record. This includes synchronic descriptions of earlier stages of a language as well as diachronic descriptions of the

---

\*The research and research project reported in this paper are funded by a generous grant (AH/F018894/1) from the Arts and Humanities Research Council, UK. We would like to thank the anonymous reviewer who provided valuable comments and suggestions for improvement. We would also like to acknowledge the contributions of Janick Wrona to the research reported in this article.

changes that have taken place between those stages. In the case of Japanese, the textual record itself is complex and diverse, with many complicating issues regarding text tradition, the interpretation of script and orthography, and the relationship between the different varieties of language represented in the texts, combined with the fact that many texts are uneven and/or anachronistic in the language used within them. Even so, a long philological effort has given us reliable critical editions of most literary texts.<sup>1</sup>

Historical linguistics in this narrower sense has thus always been an extreme kind of *corpus linguistics*—long before that term became popular—with a closed text corpus as its data and no recourse to native speaker intuition or any kind of directed testing. This poses important challenges to studies of historical syntax, because the methodology used in current syntactic theories has been developed to investigate spoken languages and relies heavily on testing and introspection. On the other hand, advances in data processing technology over the past decades have made it possible to organize the text corpora used in historical linguistics electronically or digitally, and to annotate and encode them with linguistic or other information. The electronic text format, combined with the possibility of searching combinations of encoded information, make the texts available for analysis in ways that were unthinkable twenty or even ten years ago, ensuring a comprehensive and systematic overview over large sets of data. It therefore seems that one very important part of the future of Japanese historical linguistics will be the construction of extensive and reliable electronic corpora of the textual material, represented in a uniform and standardized way.

In the history of Japanese, most aspects of the phonology and morphology of older stages of Japanese as reflected in the available written sources have been studied in great detail, and on the whole, it is probably fair to say that most of the ‘big’ questions regarding earlier Japanese phonology and morphology have been answered, although there are issues which remain controversial. Within the traditional domains of grammar, it is in the field of the syntax of pre-modern Japanese that the most challenges and open questions may be found, and recently the study of pre-modern Japanese syntax has attracted interest from the point of view of various versions of current syntactic theory, with studies dealing with specific aspects of the syntax of OJ and MJ, such as complementation, negation, focus constructions, case marking, and word order. Valuable as such studies have been, the difficulties they confront also reflect the complexity and volume of the available material, as well as the general conditions of the study of historical syntax.

In this paper we first present (section 2) a large-scale collaborative research project hosted at the University of Oxford which uses text encoding technology to study the history of the Japanese language, in particular its syntax. The project is entitled *Verb semantics and argument realization in pre-modern Japanese: A compre-*

<sup>1</sup> More recently large amounts of first-hand material in the form *kunten* (訓点 scholastic text annotations) material has come to be studied and used to supplement other types of material.

*hensive study of the basic syntax of pre-modern Japanese* (abbreviated as ‘VSARPJ’) and is funded by a grant of almost £1 million from the Arts and Humanities Research Council in the UK. An important first phase of the project is the construction of an annotated and encoded corpus of texts. While the corpus is initially constructed specifically for the purpose of serving the VSARPJ project, we believe it will eventually become useful for the investigation of many other aspects of the syntax of pre-modern Japanese.

In the second half of the paper we present a preliminary study, based on our corpus, of some aspects of verb compounding in OJ. In section 4 we examine properties of the second verb in verb-verb compounds in order to understand what constitutes the class of ‘auxiliary verb’ (*hojodōshi* 補助動詞) in OJ. In section 5, we look at a single verb, *omop-* ‘think, feel’, examining its range of syntactic uses when used as a single predicate, and when used as the first verb in a verb-verb compound. The study will both in its own right contribute to the understanding of verb compounding in OJ, and the impact this has on syntax, as well as serving as an illustration of the kind of research which will become possible on the basis of the VSARPJ corpus and other comparable corpora.

## 2. The VSARPJ Project

The primary and immediate goal of the VSARPJ project is to investigate *argument realization* in detail for pre-modern Japanese. Argument realization is a fundamental aspect of the syntax of a language which concerns the ways in which verb meaning determines the number of arguments in a clause and their morpho-syntactic and semantic properties. In essence, the project will contribute to a comprehensive account of the basic syntax of each of the stages of the Japanese language, from the beginning of its recorded history in the 8th century until the end of the 16th century, and of the changes in basic syntax that have taken place between these stages.

In the words of Levin and Rappaport Hovav, argument realization is “the study of the possible syntactic expressions of the arguments of a verb” (2005: 1). The following are well-known examples from English where two verbs, *break* and *hit*, which at first glance seem to share syntactic properties, as in (1) and (2), on closer inspection differ in the kinds of arguments they can take, shown in (3)–(6).

- (1) The boy broke the window with a ball
- (2) The boy hit the window with a ball
- (3) The window broke
- (4) \*The window hit
- (5) I broke his leg / \*I broke him on the leg
- (6) I hit his leg / I hit him on the leg

Illustrative examples from modern Japanese (NJ, cf. Fukui, Miyagawa, Tenny 1985) include verbs such as *nur-*<sup>2</sup> ‘smear’ and *har-* ‘paste, stick’, which show similar

<sup>2</sup> We use the basic stem as the citation form for inflecting verbs, e.g., *nur-*, *tabe-*, etc.

patterns in (7)–(8), but different patterns in (9)–(10). Interestingly, if the auxiliary verb *-tsukus-* ‘exhaust, do fully’ is attached to *bar-*, it comes to behave similarly to *nur-*. This suggests that the constraints on the argument realization pattern of *bar-* do not simply reflect real-life constraints on the possible participants in the event that it denotes, but that they are genuine linguistic constraints.

- (7) kabe-ni      penki-o      nuru  
 wall-DAT    paint-ACC  
 ‘smear paint on the wall’
- (8) kabe-ni      posutaa-o      haru  
 wall-DAT    poster-ACC  
 ‘put posters up on the wall’
- (9) penki-de      kabe-o      nuru  
 paint-with    wall-ACC  
 ‘smear the wall with paint’
- (10) \*posutaa-de    kabe-o      haru  
 poster-with    wall-ACC
- (11) kabe-ni      posutaa-o      hari-tsukusu  
 wall-on        poster-ACC    stick-exhaust  
 ‘hang all the posters on the wall’
- (12) posutaa-de    kabe-o      hari-tsukusu  
 poster-with    wall-ACC    stick-exhaust  
 ‘hang the wall all over with posters’

In investigating this kind of phenomenon for pre-modern Japanese, the VSARPJ project has two parts: synchronic and diachronic. In the synchronic part, we investigate for the main stages of pre-modern Japanese the argument realization patterns of individual verbs and of verb classes. For each verb attested in the texts of pre-modern Japanese which we are using in the investigation, we will establish the syntactic frames in which it can occur and its basic argument realization pattern. We will attempt to determine which lexical semantic and/or syntactic features are relevant for argument realization in the classification of verbs, for example: semantic roles, lexical aspect, telicity, type of motion, causation, split intransitivity. We will also investigate differences and similarities in argument realization patterns for classes of verbs. An important part of this will be the determination of what counts as an argument, and to which extent a more finely graded range of categories between argument and adjunct is needed. We will also look at other grammatical phenomena than argument realization which may be explained by verb semantics, for example aspect, auxiliary selection, ellipsis, and case-drop.

The diachronic part of the project will build on the results of the synchronic part. In addition to charting changes affecting individual verbs, we will be able to establish an inventory of changes through the history of Japanese in argument realization both for individual verbs and for classes of verbs and thereby be able to investigate general patterns of change, including possible development pathways for verb meanings and argument realization.

Apart from the intrinsic value the results of the project will have to the description and understanding of Japanese grammar and its history, the project may also be expected to yield results of more general interest, as this will be the first detailed application of the type of framework employed here to a language such as Japanese, which frequently drops case markers, has extensive argument ellipsis (pro-drop), and has fairly free word order. It will also be the first large-scale investigation of this kind of a ‘dead’ language, which as noted above, poses particular challenges to research into syntax.

### 2.1. The VSARPJ corpus of pre-modern Japanese

In order to enable us to carry out the research described above, the initial phase of the VSARPJ project consists in building an electronic corpus of the texts which will be investigated (the texts initially included in the corpus are mentioned in section 2.2), encoded with information about a number of linguistic properties. The corpus has the following main features:

- It transcribes the texts phonemically (2.3)
- It uses XML markup following the internationally recognized standards of the Text Encoding Initiative (TEI) (2.4)
- It encodes the following types of linguistic information:
  - Orthography (2.4.1)
  - Words, morphemes, and part of speech (2.4.2)
  - Lexeme and morpheme ID; lexical information (2.4.3)
  - Syntax (2.4.4)

In the following sections we outline very briefly the main important features of the coding of linguistic information in the corpus. Details of coding, tagging and other conventions can be found at <http://vsarpj.orinst.ox.ac.uk/corpus.html>

### 2.2. Texts

The corpus will in the initial stage comprise a selection of texts from the three main periods of pre-modern Japanese: Old Japanese (‘OJ’, approximately 700–800), Early Middle Japanese (‘EMJ’, 800–1200), and Late Middle Japanese (‘LMJ’, 1200–1600). The corpus includes all main extant texts from the OJ period.<sup>3</sup> For EMJ, the corpus focuses on texts from the period 900–1100 which are thought to a large extent to reflect the (spoken) language of the time. From the LMJ period, where most of the textual material is written in ‘classical Japanese’ rather than in the contemporary language and is characterized by a high degree of fossilization, we use two texts produced by the Jesuit missionaries at the end of the 16th century, the *Esopo no fabulas* and the *Feiqe monogatari*, which both reflect the contemporary

<sup>3</sup> These texts are: *Kojiki kayō* (古事記歌謡, 712), *Nibon shoki kayō* (日本書紀歌謡, 720), *Fudoki kayō* (風土記歌謡, 730s), *Bussokuseki-uta* (仏足石歌, after 753), *Man'yōshū* (万葉集, after 759), *Shoku nihongi kayō* (続日本紀歌謡, 797), *Shoku nihongi Senmyō* (続日本紀宣命, 697–791), *Engishiki Norito* (延喜式祝詞).

language at the very end of the period, and also have the additional advantage of being written in alphabetic writing. We base the corpus on critical editions of the OJ and EMJ texts. For OJ we have, where possible, followed the critical editions of the *Nihon Koten Bungaku Taikei* edition.<sup>4</sup>

### 2.3. Phonemic transcription

All texts are romanized in a phonemic transcription. Concretely, this means that texts are romanized in accordance with the phonology of the time a text is thought to have been written, and reflecting the sound changes which had been completed by that time. Take for example the word which is often written by 恋, which in NJ has the shape *koi* and which may be glossed very roughly as ‘love’. In the historical *kana* spelling (歴史的仮名遣<sup>レ</sup>) this word is written 𪛗𪛘, regardless of the time from which the text dates. In a phonemic transcription, however, this word has the shape /kwopwi/ (𪛗<sub>甲</sub>𪛘<sub>乙</sub>) in OJ.<sup>5</sup> As a result of sound changes which took place since OJ, the shape of this word has changed as shown in (13) with approximate dating, and the corpus uses those shapes in accordance with the dates of the texts. Thus, in the *Tosa nikki* (from 935), this word is transcribed *kopi*, but in the *Genji monogatari* (from just after 1000) it is transcribed *kowi*. This is a very basic point, but one which is often ignored in the presentation of pre-modern Japanese texts.

(13)		800	950	1000	1100					
	OJ	<i>kwopwi</i>	> EMJ	<i>kwopi</i>	>	<i>kopi</i>	>	<i>kowi</i>	>	<i>koi</i>

### 2.4. Tagging

The corpus is encoded according to the guidelines of the *Text Encoding Initiative* (TEI), by means of markup in the coding language XML. Some of the coding has been semi-automated, but all of it is checked manually and much coding is entered manually; for the encoding we use the XML editor Oxygen.<sup>6</sup> Although not exclusively developed for encoding of linguistic features of texts, the TEI provides tools well suited for linguistic coding. The inventory of TEI coding is a small set of *tags* which are used to enclose portions of text; text enclosed by tags can further be characterized by various attributes, such as *type*, *subtype*, *function*, *inflection*, etc. The inventory of coding elements and conventions of the TEI are under constant development and improvement; they may be viewed at <http://www.tei-c.org/>. A major consideration for adopting TEI technology and guidelines for the corpus was that such standards ensure that the corpus we design will be long lasting, non-idiosyncratic, and updateable along with future changes in technology.

#### 2.4.1. Orthography

Japanese script comprises both logographic and phonographic writing. The interpretation of logographic writing relies on reading tradition and is in many respects

<sup>4</sup> Examples from (*Shoku nibongi*) *Senmyō* follow the readings in Kitagawa (1982).

<sup>5</sup> We use the Frellesvig & Whitman (2008) transcription system for OJ.

<sup>6</sup> See <http://www.oxygenxml.com/>

uncertain; this is sometimes reflected in the existence of significantly different reading traditions of some texts. If a text or crucial parts of it are written logographically, we can not strictly speaking be certain of which words, or inflected forms, are reflected in the text.

In the corpus we follow the reading tradition of the critical editions we base the texts on. However, in order to reflect the crucial distinction between logographic and phonographic writing and to represent information about how words and/or morphemes were written in the original script,<sup>7</sup> we have adopted the following policies. First, for OJ texts we preserve the original script together with the phonemically transcribed text. Thus, reference can be made to the original script. Second, we code for all strings of text whether they reflect original logographic or phonographic writing.<sup>8</sup> We do this by using a TEI ‘character tag’, <c/>, and a ‘type’ attribute. In this way we indicate three types of writing: phonographic writing versus logographic writing, and within the category of logographic writing, a subtype where an element has no overt representation in the text. In (14), 我屋戸乃 (from MYS 8.1606) is thought to represent *wa ga yadwo no* ‘of my hut’, exemplifying all three types of writing. The words *wa* and *yadwo* are written logographically, by 我 and 屋戸, respectively; the particle *ga* is not overtly represented in the original script; and the particle *no* is written phonographically, by 乃. In text examples cited in this paper, we show logographically written text portions in roman type, elements with no overt representation in the text with underline, and phonographically written text in italics.

(14)	我		屋戸		乃
	wa	<u>ga</u>	yadwo		<i>no</i>
	I	GEN	hut		GEN
	‘of my hut’				

#### 2.4.2. Words, morphemes, and part of speech

*Words* are enclosed in ‘word(-like)’ tags, <w/>, and information about part of speech is supplied by a ‘type’ attribute. Complex words can consist of more than one word, forming a compound word. And they can consist of one or more words followed or preceded by one or more *morphemes*. Morphemes are enclosed by ‘morpheme’ tags, <m/>, and categorized by ‘type’ attributes. Inflecting parts of speech are supplied with information about their inflectional forms, in an ‘inflection’ attribute.

<sup>7</sup> By ‘original script’ we mean the script employed in the critical edition upon which a text is based.

<sup>8</sup> For EMJ texts the issue of script is less complex. In most of the texts, by far the most writing is phonographic, in *kana*. However, also for these texts do we make a three-way distinction between text parts which are (a) logographically written, (b) have no overt representation in the original script, or (c) are phonographically written. The two LMJ texts included in the corpus are written phonographically in alphabet writing which also notes some allophonic variation; we change this to a strictly phonemic representation, but do not preserve the original script.

### 2.4.3. The lexicon

Each distinct item (word or morpheme) in the corpus is assigned a unique ID number. This has a number of advantages, in particular in making it possible to divorce searches in the corpus from actual strings of text.

- Searches for inflecting words or morphemes in the texts will not be linked to the actual inflected forms of an item. Thus, a search for the verb *sin-* ‘die’ will return all the inflected forms of that verb. However, searches can also be modified to give only a subset of forms, for example defined by specific inflected forms or combination with specific auxiliaries.
- Searches across time for items which have changed shape as a result of sound change will be straightforward. For example, as a result of sound change the verb OJ *kwopwi-* has a number of different shapes through time, as outlined above (13), and appears in texts from different periods in significantly different shapes (*kwopi-*, *kopi-*, *kowi-*, *koi-*). With unique ID numbering, it is not necessary to search for all of these shapes, but it is possible to search for all, or a specific set of, occurrences of this verb through the corpus, regardless of the actual shape of the verb at any particular stage.
- Searches are not contaminated by text strings which are identical to the intended target of a search. For example, the verb ‘request, ask’ OJ *kop-* has a number of forms which are segmentally identical with forms of ‘love’ from somewhere in the first half of the EMJ period (for example infinitive *kopi*, *kowi*, *koi*). With unique ID numbering, forms of one verb will not show up in searches for the other verb.

As the substantial research of the VSARPJ project progresses, *grammatical information* will also be entered into the lexicon. This will include information about the possible argument realization patterns of a verb, and for example information such as that preliminarily reported below (section 5) for the verb *omop-* will eventually be entered for all verbs in the corpus and linked to their instantiations in the corpus. In this way, the lexicon will also be an important tool for organizing the results of our research as they appear.

### 2.4.4. Syntax

Syntactic information is encoded by means of a minimal inventory of constituents, namely those of *clause* and *phrase*, and two relation types, *argument* (predicate-selected) and *adjunct* (free adverbial). Clauses are enclosed in clause tags, <cl/>, and noun phrases are enclosed in phrase tags, <phr/>.

Clauses can be embedded within other clauses, as subordinate clauses. Adnominal, or relative, clauses are embedded within phrases. Nominalized clauses are enclosed in phrase tags, outside of their clause tags. Predicate-selected clauses (including but not limited to complement clauses) are categorized by a ‘type’ attribute as *arguments*.

Phrases include adverbs and nominalized clauses, in addition to expressions

headed by a noun. Phrases are categorized by a ‘type’ attribute as *arguments* if they are clearly predicate-selected, and as *adjuncts* if they are clearly free adverbials. At this stage of markup, a large proportion of phrases are marked neither as argument nor as adjunct, because their status is not entirely clear. The investigation and clarification of such phrases, and other important issues such as the determination of whether categories may be needed which are intermediary between the poles of argument and adjunct, or whether argument-hood is a scalar property, are parts of the substantive research of the VSARPJ project. The corpus will eventually reflect the results of this research. The structure of both clauses and phrases is generally flat.<sup>9</sup> The words which can form predicates of clauses are verbs, adjectives, and copulas. Within a clause, the word or words which form its predicate are identifiable by not being enclosed in phrase tags.

This small inventory of syntactic elements and conventions for their use, combined with the material they can contain, will allow unique identification of at least all of these elements or properties in the corpus: *topics*, *right dislocated elements*, *focused elements*, *noun phrase heads*, *particle scope*, *clause predicates* (including analytic predicates), *zero marked arguments*, *topicalized arguments*, *relative order of case marked and zero marked arguments* (including ordering relative to focused elements), and *clause types* (main, subordinate, adnominal, nominalized). Furthermore, all such elements and properties, as well as combinations of them, and combinations with other items and properties coded in the corpus will become searchable and extractable from the corpus. For example, we will be able to use the corpus to extract all attested syntactic frames for individual verbs, within individual stages of the language as well as across different stages. All of this is highly relevant, not just to the VSARPJ research project, but also more generally and widely to investigation of most features of pre-modern Japanese syntax.<sup>10</sup>

### 3. A Preliminary Study of Verb Compounding in Old Japanese

The notion of ‘auxiliary verb’ (*hojodōshi* 補助動詞) is considered to be an important one in the description of Japanese grammar, and has been the focus of a considerable amount of research on NJ. Much less is known about auxiliary verbs in earlier stages of Japanese, and one of the outcomes of our VSARPJ project will be a much better understanding of what the relevant data are. Beyond that, we expect that research based on our corpus will show what kinds of criteria are relevant for defining the class of auxiliary verbs across different stages of Japanese, as well as showing which verbs belong to the class at a given stage. Naturally, we expect a further consequence to be evidence from which verbs have entered or left the group of auxiliary verbs in regard to what factors influenced or inhibited such changes.

For this paper, we have undertaken two pilot studies. The first involves a general

<sup>9</sup> Within phrases constituency is usually predictable from the sequence of constituents, but if not, constituency can be marked as necessary.

<sup>10</sup> Needless to say, these coding conventions easily lend themselves to the creation of equally powerful corpora of modern Japanese.

survey of verb–verb (V1-V2) compounds in OJ, focussing on V2s and particularly with a view to understanding what properties are characteristic of auxiliary verbs; this is the subject of section 4. In section 5, we look at the data specific to one verb, *omop-* ‘think, feel’, in order to examine its range of syntactic uses when used as a single predicate, and when used as a V1 in a V1-V2 compound. In this way, we can begin to investigate the effects on argument structure of compounding. Even in this pilot study, we have found some quite striking differences in the argument-taking properties of this verb when used alone and when part of a compound.

#### 4. Auxiliary Verbs in Old Japanese

To begin an inquiry into the properties of the class of auxiliary verbs in OJ, we made a survey of verbal compounds. The basic morphological segmentation and classification in the corpus make it possible to identify and examine all directly compounded verbs in the corpus. In this way we provisionally identify 2,550 tokens of verbal compounds in the *Man'yōshū*.<sup>11</sup> It is possible to analyze these compounds to determine, for example, which verb appears most frequently as a V2 in verbal compounds: in this case, the verb of motion *-ko-* ‘come’ with 267 tokens. We can also determine which auxiliaries are least frequently attached to V2s in verbal compounds in the *Man'yōshū*: causative *-(a)sime-* and passive *-(a)ye-*, neither of which is found after a V2.<sup>12</sup> It is interesting to note that both of these auxiliaries involve voice. Other more predictable restrictions on co-occurrence for common V2s are also verifiable: For example, in the *Man'yōshū* the potential V2 *-kane-* ‘fail’ never appears together with a negative auxiliary (in contrast to NJ), and the potential V2 *-e-* ‘be able’ is never followed by the potential V2 *-kate-* ‘prevail’.

The ability to inspect the full inventory of directly compounded verbs, and to examine each token in its context, makes it possible to identify a subset of V2s by referring to specific properties that they share. As a working hypothesis, we expect the class of auxiliary verbs to exhibit the following closely related properties:

- (15) (a) overall productivity  
 (b) semantic lightness (i.e., V2s whose contribution to the semantics of the compound is limited to or focused in categories such as spatial deixis, social deixis, aspect, polarity, potential, degree, etc.)

<sup>11</sup> Needless to say, correction of mistakes in the corpus, or changes in our coding practices, will affect that figure. For example, in the corpus, tokens of *tatematur-* ‘offer up (humble)’ and *noritamap-* ‘say (respect)’ are marked up as full lexical items rather than as synchronically productive compounds, and they are therefore not counted as compounds.

<sup>12</sup> We exclude from this count lexicalized passives appearing as V2s: The frequency for these V2s in the OJ corpus is as follows: *midaye-* ‘be confused’ (1 token); *sakaye-* ‘flourish’ (9 tokens); *sinaye-* ‘wither’ (3 tokens); *taye-* ‘be cut short’ (2 tokens). Note that even if such forms were included, the count would remain very low, and in most cases the passive morpheme could not be interpreted as affecting the voice of the host verb. The following lexicalized passives did not appear as V2s: *wakaye-* ‘become young’; *saye-* ‘cool off’; *kikoeye-* ‘be audible’; *moeye-* ‘burn’; *nipoye-* ‘take on a color’; *omopoye-* ‘come to mind’.

## (c) free combination with V1s of all transitivity types

After scanning the inventory of V2s in the *Man'yōshū* for tokens that exhibit these properties, we applied these conditions to the whole OJ corpus. We find that the 20 most productive of such V2s are as follows, with the number of tokens in parentheses:

- |  |                                      |
|--|--------------------------------------|
| (16) <i>-tamap-</i> 'give (respect)' (650) | <i>-kane-</i> 'fail' (77)            |
| <i>-matur-</i> 'offer up (humble)' (403)   | <i>-ide-</i> 'go out, put out' (69)  |
| <i>-ko-</i> 'come' (314)                   | <i>-tug-</i> 'continue, convey' (47) |
| <i>-myes-</i> 'see (respect)' (179)        | <i>-sugwi-</i> 'go beyond' (42)      |
| <i>-tamape-</i> 'receive (humble)' (159)   | <i>-kose-</i> 'do for me' (38)       |
| <i>-imas-</i> 'exist (respect)' (182)      | <i>-kate-</i> 'prevail' (38)         |
| <i>-watar-</i> 'go across' (134)           | <i>-sake-</i> 'separate' (33)        |
| <i>-yuk-</i> 'go' (115)                    | <i>-ok-</i> 'put, land' (30)         |
| <i>-mi-</i> 'see' (93)                     | <i>-watas-</i> 'put across' (25)     |
| <i>-tat-</i> 'stand' (85)                  | <i>-sik-</i> 'spread' (25)           |

We augment this list with a number of semantically light and freely combining V2s. With regard to semantic lightness, judgments were made on a case by case basis. Examples include: *-yuk-*, *-tug-*, and *-watar-* used to denote continuative aspect; *-ide-* and *-some-* used to denote inceptive aspect, *-kate-*, and *-kane-* used to denote potential, etc.

A given V2 that is attested in compounds with transitive V1s, unergative intransitive V1s, and unaccusative V1s is categorized as being freely combining, for example:

- (17) *-ape-*, *-e-*, *-ide-*, *-imas-*, *-kane-*, *-kapyer-*, *-ko-*, *-kose-*, *-matur-*, *-mi-*, *-ok-*, *-sik-*, *-some-*, *-sugwi-*, *-topor-*, *-tug-*, *-tuk-*, *-watar-*, *-yuk-*

In addition to looking for V2s having the three properties in (15), we extended the list in the following way: For any V2 on the list that is a member of a transitive/intransitive pair (e.g., *watas-* 'put across' and *watar-* 'go across'), we add the other member of the pair to the full list. The pairs are as follows:

- (18) *-idas-/ide-*; *-ire-/ir-*; *-kapyes-/kapyer-*; *-kwos-/kwoye-*; *-sugus-/sugwi-*; *-tate-/tat-*; *-topos-/topor-*; *-tuge-/tug-*; *-tuke-/tuk-*; *-watas-/watar-*; *-yame-/yam-*

In this way we provisionally identify 44 semantically light or productive or freely combining V2s as candidates for inclusion in the class of auxiliary verbs.<sup>13</sup>

<sup>13</sup> A common definition of auxiliary verbs includes their ability to appear as independent verbs, as opposed to auxiliaries (*jodōshi*) which are not used as independent verbs. On this definition, we might eliminate *-kate-*, *-kane-*, *-some-*, and *-kose-* (which do not appear in isolation) from the list, but they are all productive, semantically light, and freely combining. Of these, only *-kane-* passes a significant number of the syntactic tests we established as indications of membership in the class of auxiliary verbs (see sections 4.1 and 4.3).

- (19)
- |                                       |                                     |
|---------------------------------------|-------------------------------------|
| - <i>ape-</i> 'endure'                | - <i>pate-</i> 'end'                |
| - <i>e-</i> 'be able'                 | - <i>sake-</i> 'separate'           |
| - <i>idas-</i> 'remove'               | - <i>sik-</i> 'spread'              |
| - <i>ide-</i> 'exit, remove'          | - <i>some-</i> 'begin'              |
| - <i>imas-</i> 'exist (respect)'      | - <i>sugus-</i> 'put beyond'        |
| - <i>ir-</i> 'go in'                  | - <i>sugwi-</i> 'go beyond'         |
| - <i>ire-</i> 'put in'                | - <i>tamap-</i> 'give (respect)'    |
| - <i>itadak-</i> 'receive'            | - <i>tamape-</i> 'receive (humble)' |
| - <i>kake-</i> 'hang'                 | - <i>tar-</i> 'suffice'             |
| - <i>kane-</i> 'fail'                 | - <i>tat-</i> 'stand'               |
| - <i>kapyer</i> 'come back'           | - <i>tate-</i> 'make stand'         |
| - <i>kapyes-</i> 'put back'           | - <i>topor-</i> 'go through'        |
| - <i>kate-</i> 'prevail'              | - <i>topos-</i> 'put through'       |
| - <i>ko-</i> 'come'                   | - <i>tug-</i> 'continue, convey'    |
| - <i>kose-</i> 'do for me'            | - <i>tuge-</i> 'continue, convey'   |
| - <i>kwos-</i> 'put over'             | - <i>tuk-</i> 'stick'               |
| - <i>kwoye-</i> 'go over'             | - <i>tuke-</i> 'affix'              |
| - <i>masar-</i> 'exceed'              | - <i>watar-</i> 'go across'         |
| - <i>matur-</i> 'serve (humble)'      | - <i>watas-</i> 'put across'        |
| - <i>mi-</i> 'see'                    | - <i>yam-</i> 'stop'                |
| - <i>myes-</i> 'see (respect)'        | - <i>yame-</i> 'make stop'          |
| - <i>ok-</i> 'put' - <i>yuk-</i> 'go' |                                     |

The main factors by which we identified the 44 V2s (viz. productivity, semantic lightness, and combinatory freedom) are interdependent to some extent. Productivity is dependent both on semantic lightness and on combinatory freedom. Combinatory freedom is arguably dependent on semantic lightness to some extent. In compounds of OJ verbs carrying their full semantic weight, combinatory freedom is (as will be shown below) constrained by the argument structure that a given V2 exhibits when it appears in isolation (an important component of that verb's semantics). We used the basic morphological segmentation and classification of the VSARPJ corpus to find independent factors characterizing OJ compounds with V2s from the list in (19). We now report on the results of applying different syntactic tests to OJ compounds with V2s from the list in (19), and the correspondences we found, particularly between the factors of argument structure and *lexical integrity*. OJ verb compounds with V2s from the list in (19) are relatively unconstrained by the semantics of their component verbs, and we will use the provisional term *non-lexical compound* to refer to them.<sup>14</sup>

Interestingly, of these four, only *-kane-* survives as a productive form in NJ.

Some verbs often treated as auxiliary verbs are not on the list, for example: *agar-* 'rise', *age-* 'raise', *kikos-* 'hear (repect)', *marwir-* 'go (humble)', *marwos-* 'speak (humble)', *tamapar-* 'accept (humble)', *wos-* 'rule', *yor-* 'approach'. We expect that further research may confirm the status of these verbs as auxiliary verbs.

<sup>14</sup> At this point we make no claims about the extent to which the class of non-lexical

#### 4.1. Lexical integrity

One factor in differentiating between types of OJ verbal compounds could be the varying degrees to which they exhibit what Kageyama (1999: 298) calls ‘lexical integrity’: A word’s resistance to having any part of it separated, moved, or deleted by rules of syntax. We tested the degree of lexical integrity of OJ non-lexical compounds in two ways: (a) by the ability of V1 and V2 to appear in the pattern [V1+particle+V2] where the particle is a non-case particle;<sup>15</sup> and (b) by the ability of V1 and V2 to appear in the pattern [V1-auxiliary-V2].<sup>16</sup> If one of these patterns is possible, that indicates a low degree of lexical integrity for the V1-V2 compound, and suggests that the V2 in question forms compounds of low lexical integrity in general.

Regarding the first test for lexical integrity, ideally we would like to restrict our judgements of the pattern [V1+particle+V2] to cases where there is overt case-marking by V1, such as in the examples below.<sup>17</sup>

- (20) *yono*naka    *wo*    *somuki*    *si*    *e-neba*  
 this.world    ACC    turn.away    EMPH    can-NEG.PROV  
 ‘because (I) can’t turn away from this world’ (MYS 2.210)
- (21) *wagimokwo*    *ni*    *kwopwi*    *pa*    *masaredo*  
 my.beloved    DAT    yearn    TOP    exceed.CONC  
 ‘although my yearning for my beloved increases’ (MYS 11.2597)

Only in such cases can we be sure that these tokens of *somuki* and *kwopwi* are in fact not nouns, strictly speaking. However, in a case-drop and pro-drop language such as Old Japanese, such unambiguous evidence is difficult to come by. Nevertheless, but with this *caveat* in mind, we find non-case particles (specifically, *ka*, *koso*, *mo*, *pa*, *si*, *so*, *ya*) appearing in the pattern [V1+particle+V2] with 13 V2s from the list in (19):

- (22) *-ape-*, *-e-*, *-imas-*, *-kane-*, *-ko-*, *-masar-*, *-mi-*, *-sugwi-*, *-tamap-*, *-tug-*, *-tuge-*,  
*-watar-*, *-yuk-*<sup>18</sup>

compounds in OJ share properties with the class of syntactic compounds in NJ, where the latter is defined as involving some kind of complement structure (Matsumoto 1996: 170). Here the term “non-lexical compound” simply refers to compounds in which V2s have one or more of the properties described in (15).

<sup>15</sup> A reviewer notes that the pattern [V1+particle+V2] is not possible for syntactic compounds in Modern Japanese (see also Matsumoto 1996: 171).

<sup>16</sup> Kageyama (1999: 302, #6) proposes that the pattern [V1-passive-V2] is a diagnostic for syntactic compounds in NJ.

<sup>17</sup> The segmentation, analysis and glossing of OJ examples follow Frellesvig 2010. A list of abbreviations is given at the end of this paper.

<sup>18</sup> The pattern [V1+particle+V2] is not by itself sufficient to establish whether a given compound is non-lexical: lexical compounds forming the same kind of patterns are also attested in OJ, as can be seen in section 5. In particular, it seems that lexical compounds form this pattern in the case of a repetitive construction. For example, in the pattern [V<sub>i</sub>1+*nomwi*+V<sub>i</sub>2], the particle *nomwi* expresses exhaustive extent and potential when mark-

Regarding the second test for lexical integrity, we find the following V2s from the list in (19) occurring in the pattern [V1-auxiliary-V2]:

**Passive -(a)ye<sup>-19</sup>**

(23) **-kapyer-** ‘return’ (1 token)

ime *ni si* **mi-ye-kapyeru** *ramu*  
 dream DAT EMPH see-PASS-return.CONCL PCONJ.ADN  
 ‘it seems she comes back to appear in my dreams’ (MYS 12.2890)

(24) **-imas-** ‘exist (respect)’ (1 token)

pwi *wo* umi-tamapite *mi-poto wo* **yaka-ye-masi-ki**  
 fire ACC bear-give.GER HON-sex ACC burn-PASS-exist-SPST  
 ‘giving birth to fire, her sex was burned’ (EN 12)

(25) **-yuk-** ‘go’ (1 token)

momoyo *pete* **sinwopa-ye-yuka-mu**  
 hundred.age transpire.GER admire-PASS-go-CONJ.CONCL  
 ‘(it) will go on being admired for one hundred ages’ (MYS 6.1065)

(26) **-ko-** ‘come’ (6 tokens)

ime ni **mi-ye-ko** to  
 dream DAT see-PASS-come.IMP COMP  
 ‘saying “come appear to me in my dreams!”’ (MYS 12.3128)

(27) **-kane-** ‘fail’ (1 token)

waka-*kye*-mu *kwo-ra ni*  
 young-ACOP-CONJ.ADN child-PLUR DAT  
**nora-ye-kane-me** *ya*  
 tease-PASS-fail-CONJ.EXCL Q  
 ‘would you fail to be teased by the children who’ll be young then?’  
 (MYS 16.3793)

ing a repeated V<sub>i</sub>.

- (i) akipagwi *wo* *wori* *nomwi* *worite*  
 autumn.bush.clover ACC break.INF only break.GER  
 ‘snapping off all the autumn bush clover that you can’ (MYS 10.2099)

Another example of a lexical compound of the pattern [V<sub>i</sub>1+particle+V<sub>i</sub>2] with repetition is found in constructions stating a condition with a compound verb and repeating the compound in an immediately following command. The repeated V<sub>i</sub> is marked by a topic particle in the command:

- (ii) kwopwi-sinaba *kwopwi mo sine to ya*  
 yearn.INF-die.COND yearn.INF ETOP die.IMP COMP Q  
 ‘is it to say, “if you would die of yearning, then, even of yearning, die!”?’ (MYS 11.2370)

<sup>19</sup> We exclude lexicalized passives appearing as V1s from our count. The frequency for these V1s in the OJ corpus is as follows: *kikoye-* ‘be audible’ (2 tokens); *nipoye-* ‘take on a color’ (1 token); *sakaye-* ‘flourish’ (5 tokens); *sinaye-* ‘wither’ (2 tokens). The following lexicalized passives did not appear as V1s: *midaye-* ‘be confused’; *omopoye-* ‘come to mind’; *wakaye-* ‘become young’.

- (28) **-tug-** ‘continue, convey’ (1 token)  
 so no ime ni dani **mi-ye-tugu** ya  
 this GEN dream DAT even see-PASS-continue.CONCL Q  
 ‘will you at least continue to appear to me in my dreams?’ (MYS 12.2849)
- (29) **-kose-** ‘do for me’ (6 tokens)  
 ime ni **mi-ye-koso**  
 dream DAT see-PASS-do.IMP  
 ‘appear for me in my dreams!’ (MYS 11.2501)

**Respect -(a)s-:**

- (30) **-itadak-** ‘receive (humble)’ (1 token)  
 ko no opo-kimi ni **narapa-si-itadaki**  
 this GEN HON-lord DAT imitate-RESP-receive.INF  
 mota-simete  
 hold-CAUS.GER  
 ‘have him learn it from this lord and have him own it’ (SM 9)
- (31) **-matur-** ‘offer’ (20 tokens, all with *yos-* as V1)  
 sumyegamwi-tati *no* **yosa-si-matura-mu** oki  
 ruling.deity-PLUR GEN bestow-RESP-offer-CONJ.ADN behind  
*tu* mi-tosi  
 GEN HON-year  
 ‘the late-ripening grain that the ruling deities will bestow’ (EN 1)
- (32) **-myes-** ‘see (respect)’ (127 tokens, all with either *siras-*, *omopos-*, or *kikos-*)  
 ame no sita **sira-si-myesi-turu** sumyeramikoto  
 heaven GEN under know-RESP-see-PERF.ADN emperor  
 ‘the emperor who rules below heaven’ (SM 3)
- (33) **-imas-** ‘exist (respect)’ (20 tokens, 19 with either *siras-* or *omopos-*)  
 opo-mi-ne **naka-si-masu**  
 HON-HON-noise cry-RESP-exist.CONCL  
 ‘he cries’ (SM 51)
- (34) **-tamap-** ‘give (respect)’ (4 tokens)  
*mi-tedukara* **oka-si-tamapite**  
 HON-hand.by put-RESP-give(RES).GER  
 ‘placed with his own hand’ (MYS 5.813)

**Causative -(a)sime-:**

- (35) **-imas-** ‘be (respect)’ (1 token)  
 yasurake-ku **tukape-matura-sime-masu** *ni* yorite  
 peaceful-ACOP.INF serve-offer-CAUS-be.ADN DAT depend.GER  
 ‘owing to your allowing us to serve you peacefully’ (EN 8)
- (36) **-tamap-** ‘give (respect)’ (3 tokens)  
 ko no opo-kimi *wo* **tukape-matura-sime-tamapeba**  
 this GEN HON-lord ACC serve-offer-CAUS-give.PROV  
 ‘because he allows this lord to serve’ (SM 10)

As may be seen, examples of this kind are very limited in number and in most cases what appears to be V1-auxiliary may in fact be a single lexicalized verb. Out of 17 tokens of the form [V1-passive-V2], 11 tokens had *miye-*. Out of 172 tokens of the form [V1-respect-V2], 166 tokens had either *yosas-*, *siras-*, *omopos-*, or *kikos-*. These may all be lexicalizations.

In total, 18 of the 44 V2s listed in (19) appeared in compounds with low lexical integrity: *-ape-*, *-e-*, *-imas-*, *-itadak-*, *-kane-*, *-kapyer-*, *-ko-*, *-kose-*, *-masar-*, *-matur-*, *-mi-*, *-myes-*, *-sugwi-*, *-tamap-*, *-tug-*, *-tuge-*, *-watar-*, *-yuk-*. Of these, the following 6 appeared in both pattern [V1+particle+V2] and pattern [V1-auxiliary-V2]: *-imas-*, *-kane-*, *-ko-*, *-tamap-*, *-tug-*, *-yuk-*.

#### 4.2. Transitivity Harmony in OJ lexical compounds

Another notion that is useful in delimiting the class of auxiliary verbs is an adaptation of Kageyama's (1993, 1999) 'Transitivity Harmony Principle'. For NJ, Kageyama proposes the following generalization for lexical compounds.

(37) Given the three argument structures below, indicating external and internal arguments, lexical compounds are built by combining two verbs either both with external arguments or both without:

- a. transitive verbs: (x <y>)
- b. unergative intransitive verbs: (x<>)
- c. unaccusative intransitive verbs: <y> (Kageyama 1999: 309, #15)

According to Kageyama (1999: 309), subtypes (a) and (b) can combine by virtue of having VP external arguments, a characteristic which subtype (c) doesn't share. Applying this reasoning to all possible combinations of the three types, the combinatory possibilities for lexical compounds in NJ are predicted to be as follows:

- (38) transitive+transitive  
 unergative+unergative  
 unaccusative+unaccusative  
 unergative+transitive  
 transitive+unergative  
 \*transitive+unaccusative  
 \*unaccusative+transitive  
 \*unergative+unaccusative  
 \*unaccusative+unergative

Testing the principle against the data for OJ, as predicted, unaccusatives combine fairly freely with other unaccusatives: *tabure-matwop-* 'go crazy-get lost', *sinaye-urabure-* 'wither-faint', *tiri-midare-* 'scatter-be confused', etc.<sup>20</sup> One

<sup>20</sup> Some criteria for determining unaccusativity: (1) combination with perfective auxiliary *-(i)n-*, (2) lack of agency, transitivity, volition, telicity, (3) affected theme, (4) ability to co-occur with resultative phrases. Relatively free compounding with other unaccusatives is another characteristic, but for our purposes, this can not be considered an independent test.

common compound of this type is *puki-midare-* ‘blow-be confused’. As an independent verb, *puk-* ‘blow’ shows the expected properties of an unaccusative: It generally does not appear with agent arguments, it frequently takes the perfective auxiliary *-(i)n-*, and it never takes perfective auxiliary *-(i)te-*.

However, for OJ, we find patterns suggesting that while there are restrictions on combinatory possibilities between lexical verbs, they are not identical to those proposed by Kageyama for NJ.<sup>21</sup> For example, while a transitive+unaccusative lexical compound ostensibly is ruled out for NJ, in OJ we find unambiguous transitive+unaccusative lexical compounds in *mi-matwop-* ‘see-get lost’ and *kiki-matwop-* ‘hear-get lost’:

- (39) ata mi-taru twora ka poyuru to  
 enemy see-STAT.ADN tiger Q howl.ADN COMP  
 moropito no **kiki-matwopu** made-ni  
 everyone GEN hear-get lost.ADN ALL-COP.INF  
 ‘until everyone misheard it to be the howling of a tiger that sees an enemy’  
 (MYS 2.199)

- (40) *i-maki-wataru* to moropito no  
 prefix-bundle-go.across.CONCL COMP everyone GEN  
**mi-matwopu** made-ni  
 see-get lost.ADN ALL-COP.INF  
 ‘until everyone mistook it to be a whirlwind rolling across’ (MYS 2.199)

It is possible to find more compounds that are counterevidence to the idea that unaccusatives only form lexical compounds with other unaccusatives: Many lexical compounds with V1 *omop-* ‘think’ (discussed in detail in section 5) have unaccusative V2, for example *omopi-midare-* ‘think-be confused’, *omopi-tawam-* ‘think-falter’, *omopi-matwop-* ‘think-get lost’, *omopi-urabure-* ‘think-faint’, *omopi-yase-* ‘think-grow thin’. Other transitive+unaccusative lexical compounds include *pumi-nadum-* ‘step on-have difficulty’, *mi-nagwi-* ‘see-become calm’, *osi-ter-* ‘push-shine’, *tarapasi-ter-* ‘fill-shine’, etc.<sup>22</sup>

As an example of an unaccusative+transitive lexical compound, we find *puk-* compounding with *midar-* ‘make confused’ (the transitive counterpart of unaccusative *midare-*). Interestingly, as a V2 *midar-* seems to allow the verb *puk-* (which is not usually transitive) to take an accusative object:

- (41) yanagwi no ito wo **puki-midaru** kaze  
 willow GEN thread ACC blow-make.confused.ADN wind  
 ‘wind that blows into confusion the threads of the willow’ (MYS 10.1856)

<sup>21</sup> Counterexamples to Kageyama’s Transitivity Harmony Principle as it applies to NJ have been noted previously (e.g., Matsumoto 1996, pp. 202–204, 228–231). A full account of the constraints on lexical compounding in OJ is beyond the scope of this paper.

<sup>22</sup> We do not include examples where the V2 of a given compound is included in our list in (19); we classify such compounds as non-lexical rather than lexical.

The ability of a transitive V2 to add to the valence of a lexical compound can be seen in the compounds *puki-kapyes-* ‘blow-put back’ and *puki-panat-* ‘blow-release’ as well.

While unaccusative+transitive and transitive+unaccusative lexical compounds are predicted to not occur insofar as their components do not both select VP *external* arguments, we find they are not uncommon in OJ. But the verbs combined in these compounds are actually similar insofar as their components both select VP *internal* arguments. Considering VP internal arguments as a second factor in argument structure similarity, we get a cline of Transitivity Harmony: Like combines best with like, verbs that share some structure but not all structure are less likely to combine, and verbs that share no structure are ruled out at the bottom of the cline:

- (42) transitive+transitive  
 unergative+unergative  
 unaccusative+unaccusative  
 ?transitive+unaccusative  
 ?unaccusative+transitive  
 ?transitive+unergative  
 ?unergative+transitive  
 \*unergative+unaccusative  
 \*unaccusative+unergative

This new set of predictions conforms better with the data for OJ. The Transitivity Harmony cline predicts unergative+transitive compounds to be less likely to occur. In fact, this type is very rare in the *Man'yōshū*. When we examine the few tokens that are attested, it is difficult to see whether the transitive verb retains its argument structure within the compound. For example, while we do find *ki-yose-* ‘come-push near’, *ki-ywob-* ‘come-call’, *ki-naki-toyomos-* ‘come-cry-make resound’, and *ki-sik-* ‘come-spread’ in the *Man'yōshū*, none of the tokens takes an overt patient argument. Another common unergative+transitive lexical compound in the *Man'yōshū* is *kaduki-tor-* ‘dive-take’, but again, none of the tokens takes an overtly marked patient argument. A cursory search of the *Man'yōshū* only yields three more examples: *yuki-wakare-* ‘go-separate’, *ide-tati-naras-* ‘go out-stand-flatten’, and *tati-naras-* ‘stand-flatten’.

Transitive+unergative compounds are slightly easier to find, and overtly expressed patient arguments can be found co-occurring with such compounds. A cursory survey of the *Man'yōshū* yields the following compounds: *kogi-kakur-* ‘row-hide’, *maki-ne-* ‘bundle-sleep’, *mati-wi-* ‘wait for-sit’, *marwosi-wakare-* ‘speak-separate’, *piki-nobor-* ‘pull-climb’, *seme-yori-ko-* ‘attack-approach-come’, *wake-nak-* ‘divide-cry’.<sup>23</sup>

<sup>23</sup> Again, we do not include examples such as *pumi-kwoye-* ‘step on-go over’, *pumi-watar-* ‘step on-go across’, *moti-kapyer-* ‘hold-come back’, because we classify as non-lexical compounds having a V2 from our list in (19).

Looking at the combinations ruled out in the Transitivity Harmony cline, the prediction based on the idea that V1+V2 compatibility is affected by the ability to take VP internal arguments seems to be borne out. Combining the two types of intransitive verbs, as in the last two cases in (42), is very rare. Unergative+unaccusative compounds are very hard to find in the *Man'yōshū*: The two which we have identified are *yuki-tukare-* 'go-get tired' and *tati-nure-* 'stand-get wet'. The second of these is arguably not an unergative+unaccusative compound, if *tati-* is a prefix or a syntactic element here:

- (43) akatokituyu ni ware tati-nure-si  
 dawn.dew DAT I stand-get.wet-SPST.ADN  
 'I stood and got wet in the dawn dew' (MYS 2.105)

We have not been able to find examples of unaccusative+unergative lexical compounds in the *Man'yōshū*. A closer look at the rest of the OJ corpus may yield compounds combining an unaccusative V1 with an unergative V2 that is not on our list in (19), but it seems that they are rare.<sup>24</sup>

### 4.3. Transitivity Harmony in OJ non-lexical compounds

Now that we have a better understanding of how Transitivity Harmony works in OJ, we can contrast lexical compounds and non-lexical ones in a meaningful way by referring to the argument structures that V2s on our list exhibit when they appear in isolation: Any V2 in a compound which would be of low acceptability as part of a lexical compound due to the Transitivity Harmony cline is more likely to be part of a non-lexical compound. By extension, compounds formed with such V2s should in general be more likely to exhibit a low degree of lexical integrity according to other syntactic tests.

As combinatory freedom is one of the criteria in our original assumptions about the constitution of the class of auxiliary verbs, it is no surprise that most of the V2s in our list in (19) are not subject to combinatory restrictions involving the transitivity they exhibit when used in isolation. All of the V2s in (19) combine with transitive verbs, except for *-tat-* (85 tokens), which combines with unergatives and unaccusatives, and *-ir-* (6 tokens) which combines with unergatives only. The V2s that combine freely with V1s of all types (transitive, unergative intransitive, and unaccusative verbs) are as follows:

- (44) *-ape-*, *-e-*, *-ide-*, *-imas-*, *-kane-*, *-kapyer-*, *-ko-*, *-kose-*, *-mas-*, *-matur-*, *-mi-*,  
*-ok-*, *-sik-*, *-some-*, *-sugwi-*, *-topor-*, *-tug-*, *-tuk-*, *-watar-*, *-wi-*, *-yuk-*

A total of 10 of the V2s in (19) behave as unergatives in isolation: *-imas-*, *-ir-*, *-ko-*, *-kapyer-*, *-kwoye-*, *-sugwi-*, *-tat-*, *-topor-*, *-watar-*, *-yuk-*. Of these, 8 combine as V2 with preceding unaccusative verbs to form compounds whose mem-

<sup>24</sup> These appear to be rare in NJ as well, but not entirely unattested. Matsumoto (1996: 211, #22a, b) cites *ukare-aruk-* 'make merry, walk' and *yopparai-aruk-* 'get drunk, walk' as examples.

bers have maximally different argument structures. Below are some examples of unaccusative+unergative non-lexical compounds for these unergative V2s:

- (45) a. *woti-mas-* ('become young-be'), *sidumari-mas-* ('become quiet-be'), *sakaye-mas-* ('flourish-be')
- b. *sini-kapyer-* ('die-come back'), *woti-kapyer-* ('become young-come back')
- c. *miti-ko-* ('fill up-come'), *ake-ko-* ('dawn-come'), *nagare-ko-* ('flow-come'), *puki-ko-* ('blow-come'), *kikoye-ko-* ('be audible-come'), *tiri-ko-* ('scatter-come')
- d. *tiri-sugwi-* ('scatter-go beyond')
- e. *urabure-tat-* ('faint-stand'), *opwi-tat-* ('grow-stand'), *sigeri-tat-* ('be abundant-stand')
- f. *nure-topor-* ('get wet-go through'), *simi-topor-* ('seep-go through')
- g. *miti-watar-* ('fill up-go across'), *nagarape-watar-* ('flow-go across'), *nadusapi-watar-* ('be bouyed-go across'), *kopori-watar-* ('freeze-go across')
- h. *uturopi-yuk-* ('change-go'), *usure-yuk-* ('thin out-go'), *sakaye-yuk-* ('flourish-go'), *tiri-yuk-* ('scatter-go'), *kikoye-yuk-* ('be audible-go'), *nadusapi-yuk-* ('be bouyed-go'), *puke-yuk-* ('grow late-go'), *kure-yuk-* ('grow dark-go'), etc.

Of these unergative V2s that combine with unaccusatives, all show a positive result on at least one of our syntactic tests for low lexical integrity, with the exception of *-tat-* and *-topor-*. It therefore seems that there is a correlation between low lexical integrity and combinatory freedom (i.e., not respecting the Transitivity Harmony cline).

In cases where a V2 combines with both a transitive V1 and an unaccusative V1, we can check whether the transitivity of the resulting compounds varies according to the choice of perfective marker. We found 26 V2s from the list in (19) that combine with both transitive V1s and unaccusative V1s, but only six of these appeared both in compounds taking the perfective auxiliary *-(i)n-*, and in compounds taking the perfective auxiliary *-(i)te-*:

- (46) *-tamap-*, *-ko-*, *-mas-*, *-some-*, *-tug-*, *-yuk-*

Variation in the type of perfective auxiliary that a V2 takes is a clear indication that the argument structure of the compound is determined by V1. Of the six V2s that exhibit this pattern, all but *-some-* appear with intervening particles and all but *-some-* appear with an auxiliary attached to V1. Furthermore, all but *-tamap-* combine with V1s of any transitivity.<sup>25</sup> Similar to what we saw in verbs that fail to respect the Transitivity Harmony cline, this result too suggests a correspondence between low relevance of the argument structure of V2 in compounds and low lexical integrity.

We further examined the characteristics of V2s in (19) to find out whether the combinatory possibilities for V2s forming non-lexical compounds are constrained

<sup>25</sup> It appears that OJ respect auxiliaries do not combine with unaccusative verbs with the exception of *-imas-*. This needs to be confirmed by a more exhaustive study.

in any unexpected ways. Those that form transitive+intransitive pairs (viz. *-idas/-ide-*; *-ire/-ir-*; *-kapyes/-kapyer-*; *-kwos/-kwoye-*; *-sugus/-sugwi-*; *-tate/-tat-*; *-topos/-topor-*; *-tuge/-tug-*; *-tuke/-tuk-*; *-watas/-watar-*; *-yame/-yam-*) show an interesting pattern: While the intransitive member of each pair is able to combine with verbs of a variety of transitivities (including, in many cases, unaccusatives), in almost every pair the transitive member only combines with transitive verbs (the only exceptions being *-kapyes-* and *-kwos-*, which also combine with unaccusative V1s).

#### 4.4. Summary

Our examination of 44 semantically light, productive, or freely combining V2s produced the following results for our various tests. The full data for all the V2s in (19) are shown in Appendix A.

- V2s for which choice of perfective auxiliary is seen to vary according to the properties of V1 are as follows: *-ko-*, *-mas-*, *-some-*, *-tamap-*, *-tug-*, *-yuk-*. All of these V2s appear in the pattern [V1+particle+V2] and in the pattern [V1-auxiliary-V2], with the exception of *-some-*.
- Of the 14 V2s which appear in the pattern [V1+particle+V2], 12 combine freely with V1s of any transitivity, the exceptions being *-masar-* and *-tamap-*.
- The V2s that show free combination, particles, auxiliaries, and perfective variation are: *-kane-*, *-ko-*, *-tug-*, *-yuk-*.<sup>26</sup> They are also among the 13 most productive V2s examined in this study.
- V2s which appear after a V1 with a passive auxiliary *-(a)ye-* (i.e., in the environment [V1-passive-V2]) are a subset of the V2s that combine freely with V1s of any transitivity. More importantly, with the exception of the V2 *-kapyer-*, they are also a subset of V2s that appear after an intervening non-case particle.
- On the other hand, out of the six V2s which appear either after a V1 with a respect auxiliary *-(a)s-* or after a V1 with a causative auxiliary *-(a)sime-*, only two also appeared in the pattern [V1+particle+V2]: *-imas-* and *-tamap-*. Furthermore, these two are also the only V2s out the six that also show variation in perfective auxiliary selection.
- Finally and most importantly, we observed that almost all of the V2s that combine with V1s of maximally different argument structure (specifically, unergative V2s that follow unaccusative V1s) appear in compounds that have low lexical integrity (i.e., compounds that allow either non-case particles or auxiliaries to attach to V1).

The patterns we have identified among non-lexical compounds contribute to

<sup>26</sup> No V2s pass all of the syntactic tests for low lexical integrity. No V2 that appears in the construction [V1-passive-V2] also appears in either the construction [V1-causative-V2] or the construction [V1-respect-V2]. The set of verbs that appear either the construction [V1-causative-V2] or the construction [V1-respect-V2] is wholly comprised of verbs of social deixis: *-imas-*, *-itadak-*, *-matur-*, *-myes-*, *-tamape-*, *-tamap-*.

an understanding of the constitution of the class of auxiliary verbs in OJ, particularly in terms of the correspondence between combinatory freedom and lexical integrity. We expect that further research based on the VSARPJ corpus will help us find still other syntactic properties of compounds relevant to the definition of the class of auxiliary verbs, based on an even larger inventory of V2s.

One thing that became clear during the course of this preliminary investigation is that in order to understand the contribution of one member of a compound to the semantics of the whole, it is necessary to examine full predications of the tokens of that compound as well as the behaviours of each member in isolation. One question that we would ultimately like to investigate in a comprehensive way for all compound verbs is how each member of a given compound contributes to the argument structure of the whole. The next section presents a preliminary study of how one of the most commonly attested verbs in the corpus behaves in isolation and in compounds, focussing on argument structure.

### 5. The Relationship of Verb Compounding and Argument Structure: OJ *omop-*

In our investigation of verbal compounds, we noticed that the verb *ip-* 'to say' is less likely to have clausal arguments marked with the complementizer *to* when it is the V1 in a verbal compound than when used in isolation, where it often takes clausal arguments marked with the complementizer *to* (so often, in fact, that the fused forms *topu* and *tipu* are quite common). However, there are not enough tokens of *ip-* as a V1 in a compound to make any generalizations.<sup>27</sup>

In order to determine what effect, if any, the V2 has on the argument structure of the V1 we examined other verbs which can select clausal complements. For the purpose of the present study, we focus on the verb *omop-* 'think, feel' which is well attested in OJ.<sup>28</sup> We include here occurrences of *omop-* and its abbreviated form *mop-* followed by another verb or in a [V1+particle+V2] pattern, where the particle is a non-case particle.<sup>29</sup> Fifty unique V2s were identified in a verbal compound with *omop-*, and seven unique V2s were identified in [V1+particle+V2] patterns.<sup>30</sup>

Information about the argument structure was then extracted. When an argument was present, we considered whether it was marked with a case particle or not, and, if the argument was marked with a case particle, we noted whether the case

<sup>27</sup> The compounds with *ip-* as the V1 are: *ipi-some-* 'speak-begin; begin to speak', which has two tokens with one argument marked by dative *ni*; *ipi-yar-* 'say-send; send word' which has one token with one argument marked by accusative *wo*; *ipi-ko-* 'say-come; come to say' which has three tokens, with one argument marked by genitive *no*, and one with complementizer *to*; *ipi-tug-* 'say-convey; transmit by word of mouth' has nine tokens where four out of nine times it selects an argument marked with complementizer *to*. The high frequency of complementizer clauses with *ipi-tug-* may be because *tug-* also specifies for this kind of clausal complement.

<sup>28</sup> For this experiment, the corpus consists of all OJ texts except *Norito*.

<sup>29</sup> Passive and causative forms were excluded as such constructions are likely to have different argument realization patterns and should be treated in another study.

<sup>30</sup> This does not include *omopi-V2-V3* sequences.

particle was written logographically or phonographically.

We present our findings by first discussing the argument realization patterns for *omop-* used in isolation. Then, we discuss the arguments which can occur with verbal compounds with *omop-* as the V1 in terms of their case marking. Finally, we discuss verbal compounds of the pattern [*omopi*+particle+V2].<sup>31</sup>

### 5.1. The verb *omop-* in isolation

Our examination of *omop-* in isolation revealed that it selects both nominal<sup>32</sup> and clausal arguments. The nominal arguments can be: zero-marked arguments (see 47a); zero-marked arguments followed by a non-case particle (47b); arguments with genitive *ga* (47c); arguments with genitive *no* (47d); arguments with accusative *wo* (47e); arguments with dative *ni* (47f).<sup>33</sup> Zero-marked arguments and zero-marked arguments with a non-case particle can be either internal arguments (objects) or external arguments (subject); zero-marked arguments without a particle are more frequently internal arguments than external ones. The selected arguments of *omop-* which are marked with *ga* and *no* are always external arguments, while those marked with *wo* and *ni* are always internal. Clausal complements can be divided into two categories: complementizer clauses with the complementizer *to* (48a); and infinitive clauses with the copula *ni* (48b), copula *to* (48c), adjective-*mi* (48d), and adjective-*ku* (48e). Examples are shown below with the argument being exemplified in bold.

(47) Nominal arguments:

(a) Zero-marked arguments

(i) **mono** *mopa-zu*

**thing** think-NEG.CONCL

'not thinking about it' (MYS 15.3760)

(ii) *miyabwi-taru* *pana to* **are** *mopu*

be elegant-STAT.ADN flower COMP I think.CONCL

'I think about the elegant flowers' (MYS 5.852)

(b) Arguments marked with a non-case particle

(i) *siraga* *opuru* **koto pa** *omopa-zu*

white.hair grow.ADN thing TOP think-NEG.INF

'not thinking about the fact that my hair grows white'

(ii) *kokoro* **pa** *mopedo*

heart TOP think.CONC

'although my heart feels' (MYS 14.3367)

<sup>31</sup> The particle in these cases is not a case particle, but a topic, focus, or restrictive particle.

<sup>32</sup> This includes both nouns and nominalized clauses.

<sup>33</sup> Some noun phrases marked with dative *ni* which do not express what someone is thinking about (e.g., *kokoro ni* 'to my heart') and nouns marked with ablative *yu* (e.g., *kokoro yu* 'from my heart') may be arguments or adjuncts depending on whether *omop-* can select source or goal arguments; we set this aside for further research. Only the ablative *yu*, and not the other forms of the ablative (namely *ywori*, *ywo*, and *yuri*) occurs with *omop-*.

- (c) *ga*-marked arguments  
**a ga mop-yeru**  
 I GEN think-STAT.ADN  
 'I am thinking' (MYS 15.3627)
- (d) *no*-marked arguments  
**tuma no omopu tori tatu**  
 spouse GEN think.ADN bird rise.CONCL  
 'the bird, which my husband loved, has flown' (MYS 2.153)
- (e) *wo*-marked arguments  
**kimi wo omopi**  
 you ACC think.INF  
 'thinking about you' (MYS 15.3683)
- (f) *ni*-marked arguments  
**soko ni omopi**  
 there DAT think.INF  
 'thinking of that place' / 'thinking about that' (NSK 43)
- (48) Clausal complements:  
Complementizer clauses:
- (a) the complementizer *to*  
**mi-mu to omopi-si**  
 see-CONJ.CONCL COMP think-SPST.ADN  
 'I thought that I would see you.' (MYS 15.3681)
- Infinitive clauses:
- (b) the copula *ni*  
**iki no wo ni omop-yeru ware wo**  
 breath GEN string COP.INF think-STAT.ADN I INTJ  
 'I, who think of you as the string of life' (MYS 7.1360)
- (c) the copula *to*  
**masurawo to omop-yeru ware wo**  
 strong.man COP.INF think-STAT.ADN I ACC  
 'I, who am thought of as a strong man'
- (d) adjective-*mi*  
**urupasi-mi omopu**  
 lovely-ACOP.INF think.CONCL  
 'I think of you as lovely' (KK 46)
- (e) adjective-*ku*  
**kwoposi-ku omopeba**  
 love-ACOP.INF think.PROV  
 'think of you as dear' (MYS 12.3140)

Table 1 shows the argument realization patterns for *omop-* when used in isolation.

Table 1. Argument realization patterns for *omop-*

raising	clausal complements	non-case particle (1)*	non-case particle (2)	zero-marked (1)	<i>wo</i> -marked	<i>ga</i> -marked	<i>no</i> -marked	<i>ni</i> -marked	non-case particle (3)	zero-marked (2)
				NP		NP <sub>ga</sub>				NP
				NP	NP <sub>wo</sub>					
		NP <i>mo</i>								
		NP <i>pa</i>								
		NP <i>pa</i>								NP
		NP <i>pa</i>	NP <i>ka</i>							
		NP <i>pa</i>			NP <i>woba</i>					
		NP <i>si</i>								
		NP <i>si</i>			NP <i>wo</i>					
		NP <i>si zo</i>								
		NP <i>ya</i>								NP
		NP <i>ya</i>								
						NP <sub>ga</sub>				
							NP <i>no</i>			
								NP <i>ni</i>		
					NP <i>wo</i>					
					NP <i>wo</i>	NP <sub>ga</sub>				
					NP <i>woba</i> **					
					NP <i>woba</i>				NP <i>pa</i>	
					NP <i>wo dami ka</i>					
					NP <i>wo ka</i>					
					NP <i>wo koso</i>					
					NP <i>wo so</i>					
					NP <i>wo si</i>					
					NP <i>wo si so</i>					
					NP <i>wo si so</i>					NP

*omop-*

Table 1. (Continued)

raising	clausal complements	non-case particle (1)*	non-case particle (2)	zero-marked (1)	<i>wo</i> -marked	<i>ga</i> -marked	<i>no</i> -marked	<i>ni</i> -marked	non-case particle (3)	zero-marked (2)
					NP <i>wo ya</i>					NP
	adjective- <i>ku</i>									
	adjective- <i>ku</i>				NP <i>ga</i>					
	adjective- <i>mi</i>									
	adjective- <i>mi</i>				NP <i>ga</i>					
	adjective- <i>mi</i>				NP <i>ga</i>					NP
	adjective- <i>mi ka</i>									
	N <i>to</i>									
	N <i>to</i>				NP <i>ga</i>					
	clause <i>to</i>									
	clause <i>to</i>									NP
	clause <i>to</i>									
	clause <i>to</i>				NP <i>ga</i>					
	clause <i>to</i>				NP <i>ga</i>					NP
	clause <i>to</i>									
	clause <i>to</i>				NP <i>wo</i>					
	clause <i>to</i>				NP <i>wo</i>					
	clause <i>to</i>									
	clause <i>to</i>									
	clause <i>to so</i>									
	clause <i>to pa</i>									
	clause <i>ka to</i>									
	clause <i>ka to so</i>									
NP <i>wo</i>	clause <i>to</i>									
NP <i>wo</i>	N <i>ni</i>									
NP	adjective- <i>mi</i>									
NP <i>pa</i>	adjective- <i>ku</i>									

*note*: NP = noun phrase; N = noun. \*The number in parentheses indicates that there is more than one possible position for this type of argument. In cases where the position is ambiguous the argument is listed in the position closest to the verb. \*\*The particle *wo/ka* is a combination of the accusative particle *wo* and topic particle *pa*.

## 5.2. The verb *omop-* as the first member of a V1-V2 compound

Next, we looked at compounds with *omop-* to learn if the V2 has the ability to affect the argument structure of *omop-*. There are fifty unique V2s that compound with *omop-*, for a total of 137 tokens of *omopi*-V2 compounds. We first present information about argument structure in terms of the marking of arguments and then present argument realization patterns for these compounds.

### 5.2.1. Verbal compounds that lack an overt argument

Of the fifty unique verbal compounds, twenty never occur with an overt argument. This number does not include compounds that lack an overt argument in some cases but have them in others; it only counts occurrences which never select an argument:

<i>omopi</i> -V2	<i>gloss</i>	total number of tokens
<i>omopi-amar-</i>	'think-excess; be at a loss'	1
<i>omopi-ko-</i>	'think-come; agonize over'	1
<i>omopi-kwopwi-</i>	'think-love; think lovingly'	3
<i>omopi-matupar-</i>	'think-be wrapped; think deeply'	1
<i>omopi-matwop-</i>	'think-get lost; be confused'	2
<i>omopi-mi-</i>	'think-see; try to think'	1
<i>omopi-musubore-</i>	'think-be contrite; be depressed'	1
<i>omopi-nagek-</i>	'think-sigh; be depressed'	1
<i>omopi-sin-</i>	'think-die; mope, be depressed'	1
<i>omopi-tamape-</i>	'think-[respect]'	1
<i>omopi-tawam-</i>	'think-falter; brood, mope'	1
<i>omopi-taye-</i>	'think-be cut short; resign oneself'	2
<i>omopi-tuk-</i>	'think-attach; think of, come up with'	1
<i>omopi-wabwi-</i>	'think-be bewildered; feel embarrassed'	1
<i>omopi-wabure-</i>	'think-be discouraged; be discouraged'	1
<i>omopi-wadurap-</i>	'think-suffer; be depressed'	1
<i>omopi-wasure-</i>	'think-forget; forget'	1
<i>omopi-yam-</i>	'think-quit; forget'	2
<i>omopi-yar-</i>	'think-send away; set at ease'	7
<i>omopi-yase-</i>	'think-get thin; think about less'	1
<i>omopi-yor-</i>	'think-approach; recall'	1

The lack of argument selection should not be interpreted as anything significant, as most of these compounds are attested only once. Since the data are lacking, it is not possible to make any claims about the argument realization patterns for any of these compounds.

### 5.2.2. Zero-marked arguments

There are seven compounds which occur with a zero-marked argument:

<i>omopi</i> -V2	gloss	total number of tokens	number of zero-marked arguments
<i>omopi-ide-</i>	'think-go out; remember, recall'	18	1
<i>omopi-imase-</i>	'think [respect]'	1	1
<i>omopi-kane-</i>	'think-fail; be unable to think about'	11	6
<i>omopi-masar-</i>	'think-excess; obsess'	1	1
<i>omopi-midare-</i>	'think-be confused; be confused by, be unsettled, be distracted by'	17	1
<i>omopi-sugwi-</i>	'think-pass; be forgotten'	6	1
<i>omopi-urabure-</i>	'think-faint; be depressed'	2	2
<i>omopi-wak-</i>	'think-divide; distinguish'	1	1

Three of the six examples of *omopi-kane-* and the only example of *omopi-ide-* occur with a zero-marked argument that is not followed by a non-case particle. The examples with *omopi-kane-* all involve internal arguments while the argument with *omopi-ide-* is external. The other compounds occur with a zero-marked argument followed by a non-case particle and only *omopi-urabure-* and two of the remaining tokens of *omopi-kane-* are found with internal arguments; the other examples are all external arguments (49).

- (49) *tuma omopi-kanete*  
 spouse think-fail.GER  
 'I am unable to think about my wife' (MYS 15.3678)

### 5.2.3. Arguments marked by genitive *ga*

There are two verbal compounds with *ga*-marked arguments:

<i>omopi</i> -V2	gloss	total number of tokens	number of <i>ga</i> -marked arguments
<i>omopi-kui-</i>	'think-regret; feel regretful'	1	1
<i>omopi-sinaye-</i>	'think-wither; be depressed'	3	2

In all three cases the *ga*-marked arguments are external (50).

- (50) *wa ga tuma no kwo ga* natukusa  
 I GEN spouse COP child GEN summer grass  
*no omopi-sinayete*  
 COP.INF think-wither.GER  
 'my wife forgets me like the summer grass' (MYS 2.138)

### 5.2.4. Arguments marked by genitive *no*

There are five compounds with a *no*-marked argument:

<i>omopi</i> -V2	gloss	total number of tokens	number of <i>no</i> -marked arguments
<i>omopi-ip-</i>	'think-say; consider'	1	1
<i>omopi-takyebwi-</i>	'think-get angry; be fierce'	1	1
<i>omopi-tanom-</i>	'think-rely; think as reliable'	12	1
<i>omopi-wabrwi-</i>	'think-be bewildered; feel embarrassed'	1	1
<i>omopi-yasum-</i>	'think-rest; forget'	1	1

As might be expected, the *no*-marked arguments are external (51).

- (51) **masurawo** *no* *omopi*-midarete  
 strong.man GEN think-confuse.GER  
 'the strong man is unsettled (by thoughts of his secret wife)' (MYS 11.2354)

### 5.2.5. Arguments marked by dative *ni*

There are five compounds which select an argument case marked with dative *ni*:

<i>omopi</i> -V2	gloss	total number of tokens	number of <i>ni</i> -marked arguments
<i>omopi-ape-</i>	'think-join; come up with, think of'	3	1
<i>omopi-ide-</i>	'think-go out; remember, recall'	18	4
<i>omopi-motome-</i>	'think-seek; seek out'	1	1
<i>omopi-tarapas-</i>	'think-fill; fill with thoughts'	2	2
<i>omopi-watar-</i>	'think-cross; dwell on, contemplate, brood over'	4	1

In all cases, the argument marked with *ni* is an internal argument (52).

- (52) *kasumi* *tatu* *naga-ki* *parupi wo*  
 mist rise.ADN long-ACOP.ADN spring.day ACC  
**ametuti** *ni* *omopi-tarapa-si*  
 heaven.earth DAT think.fill.RESP.INF  
 'filling heaven and earth with thoughts about the mist rising on a long  
 spring day' (MYS 13.3258)

### 5.2.6. Arguments marked by accusative *wo*

There are eight compounds which mark an argument with *wo*:

<i>omopi</i> -V2	gloss	total number of tokens	number of tokens with pattern
<i>omopi-ide-</i>	'think-go out; remember, recall'	18	9
<i>omopi-kane-</i>	'think-fail; be unable to think about'	11	3
<i>omopi-kuras-</i>	'think-pass time; spend time thinking about'	2	2

<i>omopi-V2</i>	gloss	total number of tokens	number of tokens with pattern
<i>omopi-nobe-</i>	'think-stretch; be carefree'	2	1
<i>omopi-pakar-</i>	'think-measure; reflect on'	1	1
<i>omopi-parakas-</i>	'think-flutter; be scattered'	1	1
<i>omopi-sugus-</i>	'think-pass; forget'	2	1
<i>omopi-tarapas-</i>	'think-fill; fill with thoughts'	2	2

In all cases, *wo*-marked arguments are internal arguments (53).

- (53) **na wo** omopi-kanete  
 you ACC think-fail.GER  
 'I cannot think about you' (MYS 11.2425)

### 5.2.7. Clausal complements

#### 5.2.7.1. Complementizer clauses marked by *to*

There are eight compounds which are attested with a clausal complement marked with complementizer *to* (54):

<i>omopi-V2</i>	gloss	total number of tokens	number of tokens with pattern
<i>omopi-ip-</i>	'think-say; consider'	1	1
<i>omopi-midare-</i>	'think-be confused; be confused by, be unsettled, be distracted by'	17	1
<i>omopi-pokor-</i>	'think-boast; be boastful'	1	1
<i>omopi-samatag-</i>	'think-prevent; stop thinking (about)'	1	1
<i>omopi-some-</i>	'think-begin; begin to think'	4	2
<i>omopi-tanom-</i>	'think-rely; think as reliable'	12	6
<i>omopi-tutusim-</i>	'think-show caution; be reverent'	1	1
<i>omopi-yasum-</i>	'think-rest; forget'	1	1

- (54) **odapi ni ari to tune pito no**  
 calm COP exist.CONCL COMP usual people GEN  
 omopi-ip-yeru koto *ni* ari  
 think-say-STAT.ADN thing COP exist.CONCL  
 'it is a thing which the common people consider to be calm' (SM 31)

#### 5.2.7.2. Infinitive clausal complements

The only type of infinitive clause found in the verbal compound data is adjective-*ku* clauses (55) which occur with only two compounds:

<i>omopi-V2</i>	gloss	total number of tokens	number of tokens with pattern
<i>omopi-kane-</i>	'think-fail; be unable to think about'	11	1
<i>omopi-mas-</i>	'think-[respect]'	4	1

- (55) **itapasi-ku** *na-omopi-masi-so*  
 dear.INF PROH-think-RESP-PROH  
 'don't think of me as dear' (SM 58)

### 5.2.8. Summary of argument realization patterns for compound verbs

As with *omop-* used in isolation, the argument realization patterns for the compound verbs were extracted from the data, as shown in Table 2 below.

Table 2. *omopi-V2* arguments realization patterns

<i>omopi-V2</i>	number of tokens	argument realization pattern	where attested
<i>omopi-amar-</i> 'think-excess; be at a loss'	1	n/a <sup>34</sup>	MYS 7.1335
<i>omopi-ape-</i> 'think-join; come up with, think of'	3	NP <i>ni</i>	MYS 4.671; MYS 6.962; MYS 20.4389
<i>omopi-(i)de-</i> 'think-go out; remember, recall'	18	NP NP <i>ni</i> NP <i>wo</i>	KK 51 (4 tokens); NSK 43 (2 tokens); MYS 11.2521; MYS 12.3133, 2885, 3133; MYS.17.3944, 3969 (2 tokens)
<i>omopi-imase-</i> 'think [respect]'	1	NP	MYS 3.443
<i>omopi-ip-</i> 'think-say; consider'	1	clause <i>to</i> NP <i>no</i>	SM 31
<i>omopi-kane-</i> 'think-fail; be unable to think about'	11	NP NP <i>mo</i> NP <i>mo</i> NP <i>pa</i> NP <i>wo</i> adjective- <i>ku</i>	MYS 4.503; MYS 11.2425, 2499, 2664; MYS 12.3019, 3096; MYS 14.3475, 3528; MYS 15.3678, 3696; MYS 20.4479
<i>omopi-ko-</i> 'think-come; agonize over'	1	n/a	MYS 10.2089
<i>omopi-kuras-</i> 'think-pass time; spend time thinking about'	2	NP <i>wo</i>	MYS 10.1934, 1936
<i>omopi-kui-</i> 'think-regret; feel regretful'	1	NP <i>ga</i>	MYS 11.2528
<i>omopi-kwopwi-</i> 'think-love; think lovingly'	3	n/a	MYS 2.217; MYS 17.4011; MYS 19.4214

<sup>34</sup> There are no examples with an argument for these compounds.

<i>omopi-mas-</i> 'think-[respect]'	4	adjective- <i>ku</i>	MYS 4.595; MYS 12.2882; MYS 14.3557; SM 58
<i>omopi-masar-</i> 'think-excess; obsess'	1	NP <i>pa</i>	MYS 13.3306
<i>omopi-matupar-</i> 'think-be wrapped; think deeply'	1	n/a	MYS 13.3248
<i>omopi-matwop-</i> 'think-get lost; be confused'	2	n/a	MYS 9.1804; MYS 13.3344
<i>omopi-mi-</i> 'think-see; try to think'	1	n/a	MYS 12.2986
<i>omopi-midare-</i> 'think-be confused; be confused by, be unsettled, be distracted by'	17	NP <i>no</i> clause <i>to</i> NP <i>no</i>	MYS 4.679, 724; MYS 7.1280; MYS 9.1804; MYS 10.2092; MYS 11.2354, 2365, 2620, 2764, 2765; MYS 12.2969, 3065, 3182, 3204; MYS 13.3272; MYS 15.3753; MYS 17.3973
<i>omopi-motome-</i> 'think-seek; seek out'	1	NP <i>ni</i>	SM 33
<i>omopi-musubore-</i> 'think-be contrite; be depressed'	1	n/a	MYS 18.4116
<i>omopi-nagek-</i> 'think-sigh; be depressed'	1	n/a	MYS 17.3969
<i>omopi-nobe-</i> 'think-stretch; be carefree'	2	NP <i>wo</i>	MYS 19.4154, 4177
<i>omopi-pakar-</i> 'think-measure; reflect on'	1	NP <i>wo</i>	SM 41
<i>omopi-pararakas-</i> 'think-flutter; be scattered'	1	NP <i>wo</i>	MYS 13.3326
<i>omopi-pokor-</i> 'think-boast; be boastful'	1	clause <i>to</i>	MYS 17.4011
<i>omopi-samatag-</i> 'think-prevent; stop thinking (about)'	1	clause <i>to</i>	SM 31
<i>omopi-sin-</i> 'think-die; mope, be depressed'	1	n/a	MYS 4.683
<i>omopi-sinaye-</i> 'think-wither; be depressed'	3	NP NP <i>ga</i>	MYS 2.131, 138; MYS 2.196
<i>omopi-some-</i> 'think-begin; begin to think'	4	clause <i>to</i>	MYS 11.2430, 2488; MYS 18.4070, 4087
<i>omopi-sugus-</i> 'think-pass; forget'	2	NP <i>wo</i>	MYS 14.3564; MYS 17.4003
<i>omopi-sugwi-</i> 'think-pass; be forgotten'	6	NP	MYS 3.325, 422, 668; MYS 10.2024; MYS 13.3228; MYS 17.4000
<i>omopi-takyebwi-</i> 'think-get angry; be fierce'	1	NP <i>no</i>	MYS 11.2354
<i>omopi-tamape-</i> 'think-[respect]'	1	n/a	SM 15

<i>omopi-tanom-</i> 'think-rely; think as reliable'	12	NP <i>no</i> clause <i>to</i>	MYS 2.167, 207; MYS 3.423; MYS 4.550; MYS 5.904; MYS 10.2089; MYS 13.3251, 3281, 3288, 3302, 3324, 3344
<i>omopi-tarapas-</i> 'think-fill; fill with thoughts'	2	NP <i>wo</i> NP <i>ni</i>	MYS 13.3258, 3276
<i>omopi-tarwam-</i> 'think-falter; brood, mope'	1	n/a	MYS 6.935
<i>omopi-taye-</i> 'think-be cut short; resign oneself'	2	n/a	MYS 4.750; MYS 15.3686
<i>omopi-tuk-</i> 'think-attach; think of, come up with'	1	n/a	MYS 13.3248
<i>omopi-tutusim-</i> 'think-show caution; be reverent'	1	clause <i>to</i>	SM 59
<i>omopi-urabure-</i> 'think-faint; be depressed'	2	NP <i>sape</i>	MYS 11.2465; MYS 17.3978
<i>omopi-wabure-</i> 'think-be discouraged; be discouraged'	1	n/a	MYS 15.3759
<i>omopi-wabwi-</i> 'think-be bewildered; feel embarrassed'	1	NP <i>no</i>	MYS 4.750; MYS 15.3727, 3686
<i>omopi-wadurap-</i> 'think-suffer; be depressed'	1	n/a	MYS 5.897
<i>omopi-wak-</i> 'think-divide; distinguish'	1	NP <i>pa</i>	SM 35
<i>omopi-wasure-</i> 'think-forget; forget'	1	n/a	MYS 6.914
<i>omopi-watar-</i> 'think-cross; dwell on, contemplate, brood over'	4	NP <i>ni</i>	MYS 4.714; MYS 12.2869, 3045, 3060
<i>omopi-yam-</i> 'think-quit; forget'	2	n/a	MYS 2.149; MYS 16.3811
<i>omopi-yar-</i> 'think-send away; set at ease'	7	n/a	MYS 1.5; MYS 4.707; MYS 9.1792; MYS 12.2892, 2941; MYS 13.3261; MYS 17.4008
<i>omopi-yase-</i> 'think-grow thin; think about less'	1	n/a	MYS 2.122
<i>omopi-yasum-</i> 'think-rest; forget'	1	clause <i>to</i> NP <i>no</i>	MYS 6.928
<i>omopi-yor-</i> 'think-approach; recall'	1	n/a	MYS 11.2404
<b>Total</b>	<b>137</b>		
<b>[V1+particle+V2]</b>			
<i>omopi-kane-</i> 'think-fail; be unable to think about'	1	n/a	MYS 11.2802
<i>omopi-masar-</i> 'think-excess; obsess'	1	n/a	MYS 18.4094
<i>omopi-sugwi-</i> 'think-pass; be forgotten'	2	n/a	MYS 9.1773; MYS 10.2269

<i>omopi-tuk-</i> 'think-attach; think of, come up with'	1	n/a	MYS 2.199
<i>omopi-yake-</i> 'think-bake; yearn for, burn with desire for'	1	n/a	MYS 1.5
<i>omopi-yam-</i> 'think-quit; forget'	3	n/a	MYS 11.2670; MYS 19.4185, 4186
<i>omopi-yuk-</i> 'think-go; come to think'	1	n/a	MYS 4.536
<b>TOTAL</b>	<b>10</b>		

### 5.3. The verb *omop-* as the first member of the [V1+particle+V2] pattern.

The following compounds can be found in the [V1+particle+V2] pattern:

<i>omopi-V2</i>	gloss	[ <i>omopi</i> +particle+V2]	number of tokens
<i>omopi-kane-</i>	'think-fail; be unable to think about'	<i>omopi mo kane-</i>	1
<i>omopi-masar-</i>	'think-excess; obsess'	<i>omopi si masar-</i>	1
<i>omopi-sugwi-</i>	'think-pass; be forgotten'	<i>omopi mo sugwi-</i> <i>omopi pa sugwi-</i>	2
<i>omopi-tuk-</i>	'think-attach; think of, come up with'	<i>omopi mo imada tuk-</i> <sup>35</sup>	1
<i>omopi-yake-</i>	'think-bake; yearn for, burn with desire for'	<i>omopi so yake-</i>	1
<i>omopi-yam-</i>	'think-quit; forget'	<i>omopi pa yam-</i>	3
<i>omopi-yuk-</i>	'think-go; come to think'	<i>omopi ya yuk-</i>	1

There are no overt arguments with these examples.

### 5.4. Summary

In comparing the use of *omop-* in isolation with *omop-* as a V1 we made several observations. First, we noticed fewer arguments marked with genitive *ga* and more marked with genitive *no* in the verbal compound data. In the case of verbal compounds only three arguments are marked with *ga*, while *ga*-marked arguments are fairly frequent when *omop-* is used in isolation. In all cases, *ga* is used to mark a selected external argument. More interestingly, we find six different verbal compounds which mark an argument with genitive *no*, and this is attested seven times. With *omop-* in isolation, however, *no*-marked arguments occur only twice. Thus, there are proportionally more *no*-marked arguments and fewer *ga*-marked arguments in the verbal compound data than when *omop-* is used in isolation.

Similarly, we noticed a significant decrease in the selection of a complement clause marked with complementizer *to*; most of the compounds do not take a

<sup>35</sup> Note that in this example the adverb *imada* 'not yet' is between the particle and verb.

clausal complement marked with *to*, and many of those that do take a complement include a V2 which can also take a clausal complement when used on its own. It should be noted that the verbs *tanom-* ‘trust, rely on’, *ip-* ‘say’, and *pokor-* ‘boast’ all take clausal complements with *to*, and it is not surprising that when two verbs that can select clausal complements are compounded, there is no effect on the ability to select such complements. As for the other V2s with clausal complements, one is the only attestation of the verb *samatag-* ‘plug, prevent’ in OJ (in SM 31), so it is not possible to determine its argument structure in isolation. The verb *tutusim-* ‘show caution’ has one other attestation in our corpus, which, like *omopi-tutusim-*, is also from Senmyō and in that attestation an argument is not present, so once again its argument structure in isolation is unknown.<sup>36</sup> Finally, the verbs *midare-* ‘be confused’, *some-* ‘begin’, and *yasum-* ‘rest’ appear to have no effect on the argument structure of *omop-* and do not take complement clauses when used in isolation.

Perhaps even more significantly, only two of the *omopi*-V2 compounds selected infinitive clausal complements, in the form of adjective-*ku* clauses. This indicates that the V2 inhibits *omop-* from selecting infinitive clausal arguments in compounds. In addition, raising constructions do not occur in the compound data, though they do occur with *omop-* in isolation.

Thus, we observe that the argument realization patterns differ when the verb *omop-* is used in isolation and when it is the first member of a verb compound. We have found out that there are far fewer patterns with the verbal compounds than with the verb *omop-* in isolation. Yet there are also four patterns which are only found in the compound data: NP-*mo* NP-*pa* verb; NP-*sape* verb; NP-*wo* NP-*ni* verb; and clause-*to* NP-*no* verb. The results of this study comparing the argument realization patterns of *omop-* in isolation and as a V1 are summarized in Table 3.

Table 3. Comparison of argument realization patterns

pattern	<i>omop-</i>	<i>omopi</i> -V2
NP	✓	✓
NP NP <i>ga</i>	✓	
NP NP <i>wo</i>	✓	
NP <i>mo</i>	✓	✓
NP <i>mo</i> NP <i>pa</i>		✓
NP <i>pa</i>	✓	✓
NP <i>pa</i> NP	✓	
NP <i>pa</i> NP <i>ka</i>	✓	
NP <i>pa</i> NP <i>woba</i>	✓	
NP <i>sape</i>		✓
NP <i>si</i>	✓	
NP <i>si</i> NP <i>wo</i>	✓	
NP <i>si zo</i>	✓	

<sup>36</sup> However, a related verb, *tutusimar-*, is found in Senmyō with a clausal complement marked with *to*.

<b>pattern</b>	<b><i>omop-</i></b>	<b><i>omopi-V2</i></b>
NP <i>ya</i>	✓	
NP <i>ya</i> NP	✓	
NP <i>ga</i>	✓	✓
NP <i>no</i>	✓	✓
NP <i>ni</i>	✓	✓
NP <i>wo</i>	✓	✓
NP <i>wo</i> NP <i>ga</i>	✓	
NP <i>wo</i> NP <i>ni</i>		✓
NP <i>woba</i>	✓	
NP <i>woba</i> NP <i>pa</i>	✓	
NP <i>wo dani ka</i>	✓	
NP <i>wo ka</i>	✓	
NP <i>wo koso</i>	✓	
NP <i>wo so</i>	✓	
NP <i>wo si</i>	✓	
NP <i>wo si so</i>	✓	
NP <i>wo si so</i> NP	✓	
NP <i>wo ya</i> NP	✓	
adjective- <i>ku</i>	✓	✓
adjective- <i>ku</i> NP <i>ga</i>	✓	
adjective- <i>mi</i>	✓	
adjective- <i>mi</i> NP <i>ga</i>	✓	
adjective- <i>mi</i> NP <i>ga</i> NP	✓	
adjective- <i>mi ka</i>	✓	
N <i>ni</i>	✓	
N <i>ni</i> NP <i>wo si</i>	✓	
N <i>ni</i> NP <i>ga</i>	✓	
N <i>ni si</i>	✓	
N <i>to</i>	✓	
N <i>to</i> NP <i>ga</i>	✓	
clause <i>to</i>	✓	✓
clause <i>to</i> NP	✓	
clause <i>to</i> NP <i>pa</i>	✓	✓
clause <i>to</i> NP <i>wo</i>	✓	
clause <i>to</i> NP <i>ya</i>	✓	
clause <i>to</i> NP <i>ga</i>	✓	
clause <i>to</i> NP <i>no</i>		✓
clause <i>to</i> NP <i>ga</i> NP	✓	
clause <i>to</i> NP <i>wo</i> NP <i>ga</i>	✓	
clause <i>to</i> NP <i>ni pa</i>	✓	✓
clause <i>to so</i>	✓	
clause <i>to pa</i> NP <i>pa</i>	✓	

<b>pattern</b>	<i>omop-</i>	<i>omopi-V2</i>
clause <i>ka to</i>	✓	
clause <i>ka to so</i>	✓	
NP <i>wo</i> clause <i>to</i>	✓	
NP <i>wo</i> N <i>ni</i>	✓	
NP adjective- <i>mi</i>	✓	
NP <i>pa</i> adjective- <i>ku</i>	✓	

## Abbreviations

### General

- TEI Text Encoding Initiative  
 VSARPJ Verb semantics and argument realization in pre-modern Japanese

### Grammatical Terms

- ACC accusative  
 ACOP adjectival copula<sup>37</sup>  
 ADN adnominal  
 ALL allative  
 CAUS causative  
 COMP complementizer  
 CONC concessive  
 CONCL conclusive  
 COND conditional  
 CONJ conjunctural  
 COP copula  
 DAT dative  
 EMPH emphatic  
 ETOP emphatic topic  
 GER gerund  
 HON honorific  
 IMP imperative  
 MPST modal past  
 NEG negative  
 NMNL nominal  
 OPT optative  
 PASS passive  
 PCONJ present conjunctural  
 PLUR plural  
 PROV provisional  
 Q interrogative  
 RESP respect

<sup>37</sup> The ‘adjectival copula’ (Frellesvig 2010: 80-1) is the morpheme which usually follows adjective stems, with forms like conclusive *-si*, adnominal *-ki*, and infinitive *-ku*.

SPST simple past  
TOP topic

### Languages

EMJ Early Middle Japanese  
LMJ Late Middle Japanese  
MJ Middle Japanese  
NJ Modern Japanese  
OJ Old Japanese

### Texts

EN (*Engishiki*) *Norito*  
MYS *Man'yōshū*  
SM *Senmyō*  
KK *Kojiki kayō*  
NSK *Nihon shoki kayō*

### **References**

- Frellesvig, Bjarke (2010) *A history of the Japanese language*. Cambridge: Cambridge University Press.
- Frellesvig, Bjarke and John Whitman (eds.) (2008) *Proto-Japanese: Issues and prospects*. Amsterdam: John Benjamins.
- Fukui, Naoki, Shigeru Miyagawa and Carol Tenny (1985) Verb classes in English and Japanese: A case study in the interaction of syntax, morphology and semantics. *Lexicon Project Working Papers* 7: 87–101. MIT: Center for Cognitive Science.
- Kageyama, Taro (1999) Word formation. In: Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*, 297–325. Malden, MA: Blackwell Publishers.
- Kageyama, Taro (1993) *Bunpō to gokeisei* [Grammar and word formation]. Tokyo: Hituzi Shobō.
- Kitagawa, Kazuhide (1982) *Shoku nihongi senmyō: Kōhon, sōsakuin*. Tokyo: Yoshikawa Kōbunkan.
- Levin, Beth and Malka Rappaport Hovav (2005) *Argument realization*. Cambridge: Cambridge University Press.
- Matsumoto, Yo (1996) *Complex predicates in Japanese: A syntactic and semantic study of the notion 'word'*. Stanford, CA: CSLI Publications.

*Author's contact information*

Research Centre for Japanese Language and Linguistics  
Oriental Institute  
Pusey Lane, OX1 2LE  
United Kingdom  
bjarke.frellesvig@orinst.ox.ac.uk

[Received 6 May 2010;  
Accepted 25 June 2010]

## 【要 旨】

日本語史における動詞の意味構造と項の具現化  
 ——上代日本語の複合動詞の分析——

BJARKE FRELLESVIG

(ビャーケ・フレレスビグ)

オックスフォード大学

KERRI L RUSSELL

(ケリー・ラッセル)

オックスフォード大学

STEPHEN W HORN

(スティーブン・ホーン)

オックスフォード大学

PETER SELLS

(ピーター・セルズ)

ロンドン大学東洋アフリカ研究学院

本稿では「日本語史における動詞の意味構造と項の具現化」という共同研究プロジェクトの概要とそのプロジェクトから得られた研究の成果を紹介する。このプロジェクトの一環として、上代から室町時代までの日本語文献に依拠した言語コーパスを開発中である。形態素から統語のレベルまでの情報をコード化したこのコーパスを利用して行った、上代日本語の複合動詞（V1 + V2 複合語）に関する分析の結果を紹介する。まず補助動詞の特徴を明らかにするため、V1とV2との意味関係が薄い複合動詞を「非語彙的複合動詞」(non-lexical compound)と定め、そういった複合動詞の単語完結性と他動性の調和を調べた。この結果、非語彙的複合動詞のV2の結合と分布に幾つかの特徴的パターンを見出した。次に、出現頻度の高い動詞「思ふ」が孤立した動詞の形でどのような項を取るかを調べ、V1を「思ふ」とする複合動詞が取る項と比較した。この分析により、[思ふ + V2]が文節を項として取ることが著しく少ないという興味深い結果が得られた。

Appendix A: Summary of syntactic tests for V2s in non-lexical compounds

V2	combination with a transitive V1	combination with an unergative V1	combination with an unaccusative V1	[V1+ particle+ V2]	occurrence with the -(i)n-perfective	occurrence with the -(i)te-perfective	[V1-passive V2]	[V1-respect V2]	[V1-causative V2]	tokens in MYS	tokens in OJ corpus
<i>ape-</i>	✓	✓	✓	<i>mo</i>						8	10
<i>e-</i>	✓	✓	✓	<i>mo, si</i>						11	11
<i>idas-</i>	✓									1	1
<i>ide-</i>	✓	✓	✓		✓					55	69
<i>imas-</i>	✓	✓	✓	<i>si</i>	✓	✓		✓		170	182
<i>ir-</i>		✓								1	6
<i>ire-</i>	✓									4	4
<i>itadak-</i>	✓							✓		0	2
<i>kake-</i>	✓									10	19
<i>kane-</i>	✓	✓	✓	<i>mo, so, ya</i>	✓	✓				77	77
<i>kapyer-</i>	✓	✓	✓							17	17
<i>kapyes-</i>	✓	✓	✓							9	9
<i>kate-</i>	✓	✓	✓							37	38
<i>ko-</i>	✓	✓	✓	<i>mo, si, so</i>	✓	✓		✓		267	314
<i>kose-</i>	✓	✓	✓							37	38
<i>kuos-</i>	✓	✓	✓			✓				5	5
<i>kuoye-</i>	✓	✓	✓							10	10
<i>masar-</i>	✓	✓	✓	<i>kosu, pa, si</i>	✓					9	9
<i>matur-</i>	✓	✓	✓			✓		✓		29	403
<i>mi-</i>	✓	✓	✓	<i>mo</i>						86	93
<i>myes-</i>	✓	✓	✓					✓		21	179
<i>ok-</i>	✓	✓	✓							20	30
<i>pate-</i>	✓	✓	✓							2	3
<i>sake-</i>	✓	✓	✓							31	33
<i>sik-</i>	✓	✓	✓		✓					22	25

V2	combination with a transitive V1	combination with an unergative V1	combination with an unaccusative V1	[V1+ particle+ V2]	occurrence with the -(i)n-perfective	occurrence with the -(i)te-perfective	[V1-passive-V2]	[V1-respect-V2]	[V1-causative-V2]	tokens in MYS	tokens in OJ corpus
<i>some-</i>	✓	✓	✓		✓	✓				22	22
<i>sugus-</i>	✓									2	2
<i>sugwi-</i>	✓	✓	✓	<i>ka, mo, pa</i>	✓					42	42
<i>tamap-</i>	✓	✓		<i>ya</i>	✓	✓		✓	✓	81	650
<i>tamape-</i>	✓	✓							✓	0	159
<i>tar-</i>	✓		✓							12	12
<i>tat-</i>	✓	✓	✓							71	85
<i>tate-</i>	✓									7	19
<i>topor-</i>	✓	✓	✓							5	8
<i>topos-</i>	✓									1	1
<i>tug-</i>	✓	✓	✓	<i>si</i>	✓	✓	✓			45	47
<i>tuge-</i>	✓	✓		<i>mo</i>						3	3
<i>tuk-</i>	✓	✓	✓							7	9
<i>tuke-</i>	✓									7	9
<i>watar-</i>	✓	✓	✓	<i>ka, si, ya</i>	✓					130	134
<i>watas-</i>	✓									23	25
<i>yam-</i>	✓		✓	<i>pa</i>						12	13
<i>yame-</i>	✓	✓								0	2
<i>yuk-</i>	✓	✓	✓	<i>ka, si, so</i>	✓	✓	✓			107	115