

Charles University

Faculty of Arts

Institute of East Asian Studies

Bachelor Thesis

Mgr. Adam Nohejl

**Polysemy of Japanese V-V compound verbs:
a corpus account**

Polysémie japonských V-V složených sloves:
korpusová analýza

Prague 2018

Thesis supervisor: Mgr. Petra Kanasugi, PhD.

I would like to thank Petra Kansugi, who supervised my work on this thesis, for her insightful questions and comments. I am also very grateful to Dominika Džúriková for correcting my English and style.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 29. prosince 2017

Adam Nohejl

Abstract: The thesis analyses Japanese verb-verb compound verbs using a corpus in order to build a pedagogical word list of these verbs accounting for their polysemy. First, the typology and characteristics of Japanese compound verbs are discussed. The following review of pedagogical resources identifies the need for a list of compound verbs and their senses based on frequency criteria. A methodology for creating the word list and assessing its utility to learners is discussed with attention to the characteristics of the Japanese language. The resulting word list based on a corpus analysis (included in the appendix) consists of 37 compound verbs, out of which 32 are lexical, includes 45 senses of lexical compound verbs. It covers 17.95 % of the lexical compound verb occurrences, which is proportional to covering 85 % verbs overall. Finally, the quantitative characteristics of Japanese compound verbs and English phrasal verbs are compared. The comparison shows that the Japanese compound verbs are more frequent and diverse and therefore also likely to be an major stumbling block for language learners.

Keywords: Japanese language, compound verbs, polysemy, corpus analysis, vocabulary, language learning, pedagogical wordlist.

Abstrakt: Tato práce analyzuje japonská složená slovesa typu sloveso-sloveso pomocí korpusu s cílem vytvořit pedagogický seznam slov. Nejdříve jsou probrány typologie a vlastnosti japonských složených sloves. Následný přehled a zhodnocení výukových zdrojů identifikuje potřebu seznamu složených sloves a jejich významů sestaveného podle frekvenčních kritérií. Metodologie pro vytvoření seznamu slov a posouzení jeho užitečnosti pro studenty je rozebírána s ohledem na vlastnosti japonského jazyka. Výsledný seznam založený na korpusové analýze (příloha práce) se skládá ze 37 složených sloves, z nichž 32 je lexikálních, a zahrnuje 45 významů lexikálních složených sloves. Pokrývá 17.95 % výskytů lexikálních složených sloves, což pokrytí odpovídající celkovému pokrytí 85 % sloves. Nakonec jsou porovnány kvantitativní charakteristiky japonských složených sloves a anglických frázových sloves. Srovnání ukazuje, že japonská složená slovesa mají vyšší frekvenci a různorodost a mohou tedy také být zásadním úskalím pro studenty jazyka.

Klíčová slova: japonský jazyk, složená slovesa, polysémie, korpusová analýza, slovní zásoba, učení se jazyku, pedagogický seznam slov.

Contents

Introduction	6
1 Characteristics and typology of Japanese compound verbs	8
1.1 Japanese compound verbs in general	8
1.2 Lexical and syntactic compound verbs	12
2 Pedagogical resources on Japanese compound verbs	17
2.1 The Compound Verb Lexicon	18
2.2 “The structure and semantic usage of compound verbs”	19
2.3 “Japanese verbs in examples”	20
2.4 Conclusion	21
3 Research questions and methodology	22
3.1 Measuring vocabulary	23
3.2 The economics of vocabulary learning	24
3.3 Selection of items for the list	26
3.4 Information provided by the list	27
3.5 Japanese words and orthography	28
3.6 Compound verbs and morphological analysis	29
3.7 The corpus	31
3.8 Sampling	31
3.9 The analysis	32
4 Results	34
4.1 Discussion of the list	34
4.2 Comparing Japanese compound verbs and English phrasal verbs	36
Conclusion	38
Bibliography	40
Appendix: List of the most frequent compound verbs and their senses	i

Introduction

This thesis presents a corpus-based analysis of Japanese verb-verb compound verbs, and a pedagogical word list drawing from that analysis and accounting for polysemy of the verbs.

Verb-verb compound verbs (such as 思い出す *omoidasu*, “remember”, “recall”, or つきあう *tsukiau*, “associate with”, “keep company”) are an important part of Japanese vocabulary. Similarly to phrasal verbs in English (such as “go on” or “pick up”), they complement and build on the simple verbs and they are essential to achieving a certain level of proficiency in the language. They also may become a stumbling block for the learners. The difficulty does not stem only from their quantity, as would be the case with any sort of vocabulary acquisition, but also from the fact that their structure is often different from verbs or constructions with similar function in other languages. Furthermore, many of them possess multiple meanings, often not readily comprehensible from their structure and constituents alone.

Consequently textbooks almost universally deal with the following compound verb patterns: ～すぎる *sugiru*¹ at the beginner level, ～はじめる *hajimeru* and ～おわる *owaru* at the intermediate level², and a few others such as ～える *eru* and ～かねる *kaneru* at the advanced level. Unfortunately, aside from this, learners of Japanese can find very little guidance on acquiring compound verbs. The focus of the available resources is limited to dictionary-like extensive coverage and further linguistic analysis and systematization of the verb-verb compound verbs. From the learner’s point of view, extensive coverage is useful when looking up randomly encountered words, but it has little value as a feature of a vocabulary building resource. In addition, linguistic analysis or systematization is useful only as far as it helps acquisition (for instance by making common patterns more visible).

¹～すぎる *sugiru* stands for an adverbial form (“*i*-stem”, “*masu*-stem” in common text-book terms) of a verb on the first position, compounded with the verb すぎる *sugiru* on the second position, which is conjugated in the usual way. The other patterns share this form. It should be noted that in case of *sugiru*, not only verbs but also adjectives may be on the first position.

² For instance in the textbooks published by Japan Times, ～すぎる *sugiru* is introduced in the beginner text book (Banno et al., 2000), while ～はじめる *hajimeru* and ～おわる *owaru* are introduced in the intermediate one (Miura and McGloin, 2008), as the only compound verb patterns dealt with up to intermediate level by this series of textbooks.

What is lacking in current research, as well as in the textbook market, is a resource that would prioritize the verbs and their multiple meanings and systematize them in a useful way. So, which verbs are the most useful to the learner? What kind of systematization has the potential to accelerate learning? In order to answer this question, sensible quantitative methods have to be used in addition to theoretical linguistic research.

This thesis therefore surveys Japanese verb-verb compound verbs in order to bring results useful to Japanese language students. To this end, I employ corpus analysis and quantitative methods, while drawing on the previous linguistic research. The text is organized in the following chapters:

- Chapter 1 reviews characteristics and typology of Japanese compound verbs in the previous research. First, basic types based on parts of speech of compound verb constituents (e.g. noun+verb, verb+verb) are introduced. Then, the focus is shifted to the subject of this thesis: verb+verb compound verbs, and their detailed classification, especially the semantic-lexical dichotomy introduced by Kageyama (1993), is further explored.
- Chapter 2 reviews existing pedagogical resources on Japanese verb+verb compound verbs. We discuss their utility from a student's (or a teacher's) perspective and point out possible areas for improvement.
- Chapter 3 formulates several research questions related to the creation of a frequency-based list of Japanese verb+verb compound verbs. The research questions are followed by a methodology for creating such a list and also answering the questions. We discuss how items for the list will be selected, what information will be given for the items, what information sources will be used, and how information will be retrieved from the corpus.
- Chapter 4 discusses the main result of this thesis, the list of the most frequent compound verbs and their senses, and attempts to answer the research questions.
- The last chapter concludes the thesis and suggests possibilities for further research.

Additionally, the Appendix presents the final list.

Chapter 1

Characteristics and typology of Japanese compound verbs

In this chapter, we will review characteristics and typology of Japanese compound verbs in previous research. We will especially focus on the characteristics that will later be useful for the extraction of data from the corpus (e.g. word formation patterns, distinction from morphologically similar constructions), or which may be of interest to learners (e.g. irregular phonetic changes). In Section 1.2 we will introduce a distinction between lexical and syntactic compound verbs, which has recently been widely used, and may also be beneficial to learners to some extent.

1.1 Japanese compound verbs in general

Japanese compound verbs are verbs formed by compounding two or more components, the last of which is always a verb. At the highest level Japanese compound verbs may be classified according to their formation. As outlined by Himeno (1999, sec. 1.1) a Japanese compound verb may be formed in one of the four ways listed in Table 1.1.

The verb+verb pattern in particular has received attention of traditional Japanese linguistics (see Kageyama, 1993, p. 74–75 for an short overview, or Himeno, 1999, ch. 1–2 for a longer discussion). More recently, Kageyama has done widely cited research on both the verb+verb pattern and the noun+verb pattern from the point of view of generative linguistics. Kageyama calls the two patterns **N-V compound verbs** and **V-V compound verbs**, respectively¹.

¹The original Japanese terms are N-V 型複合動詞 *enubui gata fukugōdōshi* and V-V 型複合動詞 *buibui gata fukugōdōshi* (Kageyama, 1993). The English equivalents of the terms I use, are also used in *The Compound Verb Lexicon* (NINJAL, 2015), a project lead by Kageyama. In addition to these two types, the book also deals with *sahen* verbs considering them another type of compound verbs Kageyama (1993, p. 255). *Sahen* verbs are will be touched in item 2 of the following list of characteristics.

Pattern	Example			
noun + verb	目 <i>me</i>	+	さめる <i>sameru</i>	= 目ざめる <i>mezameru</i>
verb + verb	書く <i>kaku</i>	+	始める <i>hajimeru</i>	= 書き始める <i>kakihajimeru</i>
adjective + verb	近い <i>chikai</i>	+	寄る <i>yoru</i>	= 近寄る <i>chikayoru</i>
adverb + verb	ぶらぶら <i>burabura</i>	+	下がる <i>sagaru</i>	= ぶら下がる <i>burasagaru</i>

Table 1.1: Types of compound verbs according to their formation (Himeno, 1999, sec. 1.1). Approximate English translations: *mezameru* – “wake up”, *kakihajimeru* – “begin to write”, *chikayoru* – “come near”, *burasagaru* – “dangle”.

The following formal characteristics (which I tried to organize top-down from border of syntax and morphology to phonology) further describe Japanese compound verbs and distinguish them from other kinds of verbs:

1. A compound verb has all the features of a verb. It may be suffixed (e.g. by adding conjugated suffix for passive voice) and, more importantly, conjugated, forming the basis of a phrase. This is realized simply by suffixing and conjugating its final component, which is always a verb. (Table 1.1 lists the non-past forms.²) A compound verb has an argument structure (case relations), which may trivially correspond to the argument structure of its verb component (or components, if they share the same argument structure), but non-trivial combinations of constituent verbs (transitive+intransitive and vice versa) are possible and relatively frequent. Kageyama (1993) discusses the argument structure of compound verbs in detail.
2. Although the numerous *sahen* verbs³ can technically be analysed as noun+verb or adverb+verb compounds, they are usually considered to form a separate verbal category of their own (and consequently not treated by Himeno, 1999). The initial component of a *sahen* verb is often a verbal noun (e.g. 勉強する *benkyōsuru*, “to study”, 愛する *aisuru*, “to love”, 報ずる *hōzuru*, “to inform”), or an adverb (e.g. ぼんやりする *bon'yari suru*, “to be dim, indistinct”, ホカホカする *hokahokasuru*, “to be nice and warm” or “to be piping hot”).

²The non-past form is also called dictionary form because of its use in dictionary headwords, and is commonly used as a lemma for Japanese verbs. Verbs are in their non-past form throughout this thesis, unless otherwise stated.

³サ変 *sahen* is an abbreviation of サ変格活用 *sagyō henkaku katsuyō*, “irregular conjugation of (verbs with stem ending with) the *sa* column of Japanese syllabary”. *Sahen* verbs are the verb する *suru*, “to do”, and the compound verbs that end with する *suru* (or its variants ずる *zuru*, じる *jiru*).

3. Compound verbs should not be confused with verbs formed by derivation. Kageyama (1993, p. 1) gives *メモる memoru* (“to take notes”, from the English loanword noun *memo* and verb-forming suffix *ru*) as a clear example of a verb formed by derivation, and contrasts it with its synonym *メモする memosuru*, which falls in the *sahen* category of compound verbs (the item 2 above). There are, however, also less obvious cases of derivation, such as verbs formed from an adjective stem, a noun, or an adverb and the suffix *meku*, “to be like” (e.g. *皮肉めく hinikumeku*, “to sound ironic”), or verbs formed from an adjective stem and the suffix *garu*, “to feel” (e.g. *寂しがる sabishigaru*, “to feel lonely”). These derivational suffixes are conjugated like verbs, but cannot stand alone as a word, unlike the final constituents of compound verbs, such as *すぎる sugiru*, “too (much)” (e.g. *寂しすぎる sabishisugiru*, “to be too lonely”), which may also appear as stand-alone verbs.
4. The constituents of compound verbs may also be compounds. Namely in the case of the verb+verb pattern, the first constituent (but not the second constituent) may be a compound verb. In other words, compound verbs consisting of three verbs should be analysed as (verb₁ + verb₂) + verb₃ (Kageyama, 1993, p. 92–93). For instance, the initial top-level constituent of the verb *立ち上がりかける tachiarikakeru*, “begin to/be about to stand up” is the compound verb *tachiagaru*, “to stand up”. Longer chains do not seem to occur as compounds in actual usage.⁴
5. The initial constituents of a compound verbs may feature derivational suffixes, e.g. *子供っぽすぎる kodomoppoosugiru*, “to be too childish” (*kodomo*, “child” → *kodomoppoi*, “childish”).
6. The adjective component of a compound verb may be both an *i*-adjective (traditionally called 形容詞 *keiyōshi*) and a *na*-adjective (traditionally called 形容動詞 *keiyōdōshi*)⁵. There is a very productive compound verb pattern for both kinds of adjectives:

adjective+すぎる *sugiru*, e.g.

- 大きすぎる *ōkisugiru*, “to be too big” (*i*-adjective),
- 地味すぎる *jimisugiru*, “to be too plain” (*na*-adjective).

⁴All occurrences I have found in a corpus (discussed in Section 3.7) are just sequences or alternations of two or more verbs, e.g. *見あげ見おろす miage miorosu*, “to look up and (look) down”, or reduplications, e.g. *考え抜き考え抜く kangaenuki kangaenuku*, “to think (very hard) and think very hard”.

⁵形容詞 *keiyōshi*, “adjectives”, and 形容動詞 *keiyōdōshi*, “adjectival verbs” have traditionally been considered separate parts of speech in Japanese. Some linguists prefer to call the latter, 形容動詞 *keiyōmeishi*, “adjectival nouns” (for instance Kageyama, 1993, p. 23) instead, which may be more appropriate. In line with the view that both classes are adjectives, and to avoid ambiguity, I use the terms “*i*-adjective” and “*na*-adjective”, based on the adnominal suffix that each class uses.

Apart from this pattern, instances of *i*-adjective+verb compound seem easier to think of, e.g. 長引く *nagabiku*, “to drag on”, 高鳴る *takanaru*, “to throb (about heart)”.⁶

7. As indicated by underline in the examples in Table 1.1, if the initial component is originally a conjugated part of speech, its form in the compound is an adverbial form.⁷ of a verb, a stem of an adjective, or a root of an adverb in the case of mimetic words consisting of a reduplicated root (Himeno, 1999, p. 3).
8. Euphonic changes are common on a boundary between two components of a compound verb:
 - a. The *rendaku* phenomenon⁸ may occur at the beginning of a non-initial component (Himeno, 1999, p. 3), e.g.
 - verb+verb: 引く *hiku* + 摺る *suru* = 引き摺る *hikizuru*,
引っ繰る *hikkuru* + 返す *kaesu* = 引っ繰り返す *hikkurigaesu* or *hikkurikaesu*,
 - noun+verb: 手 *te* + 掛ける *kakeru* = 手掛ける *tegakeru*,
 - adjective+verb: 長い *nagai* + 引く *hiku* = 長引く *nagabiku*.
 - b. The *sokuonbin* (or less frequently *hatsuonbin*)⁹ phenomenon may occur at the end of a non-final component if it is a verb, e.g.
 - 引く *hiku* + 掛ける *kakeru* = 引っ掛ける *hikkakeru* (*引き掛ける *hikikakeru*),
 - 酔う *you* + 払う *harau* = 酔っ払う *yopparau* (*酔い払う *yoiharau*),
 - 追う *ou* + 掛ける *kakeru* = 追っ掛ける *okkakeru* or 追い掛ける *oikakeru*,
 - 追う *ou* + 出る *deru* = 追っ出る *onderu* (*追い出る *oideru*),
 - 追う *ou* + 出す *dasu* = 追っ出す *ondasu* or 追い出す *oidasu*.

Note that(1) as *sokuonbin* requires the following consonant to be unvoiced, it is mutually exclusive with *rendaku*; (2) *hatsuonbin*, however, may occur together with *rendaku*, e.g. 踏む *fumu* + 張る *haru* = 踏ん張る *funbaru*; (3) the consonant *h* after a *sokuonbin* changes to the consonant *p* (see *yopparau* in the examples above); (4) both phenomena occur irregularly and in some cases both

⁶Himeno (1999, sec. 1.1) explicitly includes both kinds in adjectives, but gives an example only of the arguably more productive combination, *i*-adjective+verb (近寄る *chikayoru*).

⁷Adverbial form: 連用形 *ren'yōkei* in Japanese, or “*i*-stem”, “*masu*-stem” in common English textbook terms. Note that in the English terminology of Japanese linguistics, a distinction is usually made between “adverbial”, pertaining to an adverb, and “adverbial (form)”, a specific form of an inflected part of speech.

⁸連濁 *rendaku* is a change from an unvoiced consonant to a voiced consonant at the beginning of a non-initial morpheme of a word, which happens commonly in compound words.

⁹促音便 *sokuonbin* is a replacement of the final mora of a morpheme by 促音 *sokuon* (gemination of the following consonant, written as つ in hiragana). 撥音便 *Hatsuonbin* is a replacement of the final mora of a morpheme by 撥音 *hatsuon* (moraic *n*, written as ん in hiragana). *Sokuonbin* and *hatsuonbin* are prominent, along with *i-onbin*, as regular sound changes of the adverbial form of a verb if followed by the *te* (*de*) or *ta* (*da*) suffixes.

the variant with and the variant without it exists (see *hikkurigaesu*, *hikkurikaesu*; *okkakeru*, *oikakeru*; *ondasu*, *oidasu* in the examples above), which may or may not correspond to a semantic difference (see Himeno, 1999, p. 23–24 for details).

It is worth noting that there is a special case of compound verb formation that does not fit the above top-down organization very well, and is rather surprising in word formation: the initial constituent may be in passive voice, for instance 愛され続ける *aisaretsuzukeru*, “to continue to be loved”, composed of *aisuru*, “to love”, *reru*, conjugated suffix for passive voice, and *tsuzukeru*, “to continue”. Conjugated suffixes are considered to be syntactic morphemes and therefore are generally not supposed to occur within words, although exceptions exist (Kageyama, 1993, p. 10). We will deal with this special case in the following section.

In the chapters to follow I will focus on Japanese V-V compound verbs. From now on, I will therefore refer to them simply as **compound verbs**. This narrow use of the term is not uncommon, as witnessed by the title of a major work and a database, both concerning Japanese V-V compound verbs: 複合動詞の構造と意味用法 *Fukugōdōshi no kōzō to imiyōhō*, “The structure and semantic usage of compound verbs” (Himeno, 1999), and *The Compound Verb Lexicon* (NINJAL, 2015). I will also refer to the two constituent verbs of a compound verb as **V1** and **V2**, respectively, in the same way as Kageyama (1993) does.

1.2 Lexical and syntactic compound verbs

The current section will introduce further classification of compound verbs (that is V-V compound verbs). Himeno (1999, ch. 2) cites three ways to classify compound verbs proposed by different authors. While there are similarities in the resulting classifications, their criteria differ. In this thesis I make use of the most recent of them, introduced by Kageyama (1993), which applies formal criteria based on generative linguistics. Not only are the criteria clear-cut, but the classification has already been used as a basis for inclusion of compound verbs in pedagogical resources (Himeno, 1999; NINJAL, 2015; Kanasugi et al., 2015, all of which will be discussed in Chapter 2). Table 1.2 gives a short summary and examples of the two types introduced by Kageyama.

Kageyama (1993, p. 76–79) first verifies that compound verbs are in fact words, not just multi-word expressions. Then he proposes a distinction between compound verbs of “type A”, which “possess typical characteristics of a word – conventionalization of meaning and restriction of combinations with other words”, and compound verbs of “type B”, which even though they are words, function similarly to phrases:

Type: formation	Examples
Type A – lexical: formed in the realm of lexicon	飛び上がる <i>tobiagaru</i> , “to fly up”, “to take off” 押し開く <i>oshihiraku</i> , “to push open” 泣き叫ぶ <i>nakisakebu</i> , “to cry loudly” 売り払う <i>uriharau</i> , “to sell off” 受け継ぐ <i>uketsugu</i> , “to take over” 解き放す <i>tokihanasu</i> , “to set free” 飛び込む <i>tobikomu</i> , “to plunge”, “to burst in”
Type B – syntactic: formed in the realm of syntax	払い終える <i>haraioeru</i> , “to pay up” 話し終わる <i>hanashiowaru</i> , “to finish speaking” しゃべり続ける <i>shaberitsuzukeru</i> , “to talk on” 食べすぎる <i>tabesugiru</i> , “to eat too much” 食べそこなう <i>tabesokonau</i> , “to miss a meal” 助け合う <i>tasukeau</i> , “to help each other” 動き出す <i>ugokidasu</i> , “to start to move”

Table 1.2: The compound verb types with examples according to Kageyama (1993, p. 75). (I have abbreviated the lists of examples and added approximate English translations.)

The semantic relation of V1 and V2 in a compound verb of type B is entirely transparent and compositional, and can be analysed as a complement relation, as in 手紙を書き終える *tegami o kakioeru* = 手紙を書くことを終える *tegami o kakukoto o oeru*, 雨が降り始める *ame ga furihajimeru* = 雨が降ることが始まる *ame ga furukoto ga hajimaru* [...].¹⁰ Similar constructions with a complement could be hypothetically created even in the case of V2 that cannot stand alone, such as 出す *dasu* in 雨が降り出す *ame ga furidasu*, or 付け (ない) *tsuke(nai)* in 食べ付けない *tabetsukenai*.¹¹ Thus, the compositionality of inner relations of type B compound verb is parallel to the compositionality of semantic interpretation of an ordinary sentence.

The difference in semantic transparency of the two groups, corresponds on the whole with their difference in productivity. The productivity of type A compound verbs depends much on V2, [examples are given of V2s that combine only with a handful of V1s as well as of V2s that combine many V1s]. Even in the cases of high productiv-

¹⁰In the example sentences the compound verb (e.g. *o kakioeru*) is replaced with V2 (e.g. *oeru*) as a stand-alone verb and a nominalization of V1 (e.g. *kakukoto*) as a complement (e.g. an object) of V2: *kakukoto o oeru*, “start writing”. The meaning of the whole sentence, which also contains another complement of V1 (*tegami*, “a letter”) remains the same.

¹¹In contrast with the suffixes mentioned in item 3 on page 10, these V2 verbs can actually stand alone in their basic sense (e.g. *tsukeru* as “to attach”, “to acquire”, instead of the V2 meaning “to be used to”), but unlike the previously mentioned V2 verbs they cannot be used as stand-alone verbs in a complement+V2 construction of the same meaning as the compound verb V1+V2.

ity, though, there is a lexical restriction on the combinations, and it is necessary for each of them to be recorded in a dictionary. As a result, coming across an as yet unheard of compound verb would likely attract attention. Type B, on the other hand, is formed without lexical restrictions, just like sentences are made up freely.

(Kageyama, 1993, p. 78, footnotes mine)

As a final step in distinguishing the two types, Kageyama (1993, p. 80–92) indicates 5 criteria for verbs to be categorized as type B:

1. Pro-forms: V1 can be replaced by the pro-form そうする *sō suru*, “to do so”.
E.g. 調べ始める *shirabehajimeru* → そうし始める *sō shihajimeru*
 (“begin to investigate” → “begin to do so”)
2. Respectful honorific forms: V1 can be replaced by its respectful honorific form お〜になる *o... ni naru*.
E.g. 歌い始める *utaihajimeru* → お歌いになりはじめる *outai ni narihajimeru*
 (“start singing” → “start singing”, respectful)
3. Passive forms: V1 may be in the passive form.
E.g. 愛され続ける *aisaretsuzukeru*, “continue to be loved” (*aisuru*, “to love” + *reru*, conjugated suffix for passive voice + *tsuzukeru*, “continue”)
4. *Sahen* verbs: V1 may be a *sahen* verb.
E.g. 見物し続ける *kenbutsushitsuzukeru*, “continue watching” (*kenbutsusuru*, “to watch”)
5. Repetitive constructions: V1 can be replaced with the repetitive construction 〜に〜 ... *ni ...*:
E.g. 苦しみ抜く *kurushiminuku*, “suffer to the end” →
苦しみに苦しみ抜く *kurushimi ni kurushiminuku*, “suffer and suffer to the end”

All of these forms (the replacement forms in cases 1, 2, and 5) are syntactic constructions, and consequently, Kageyama argues, the whole compound verb of type B is formed in the realm of syntax. Kageyama notes that in cases 1, 2, and 5 above, the whole result of the replacement can no longer be considered a single word¹², and discusses the disparity between syntax and morphology in each case. He also indicates numerous semantic restrictions, which are outside the framework of morphology and syntax (for instance that the pro-form そうする

¹²The original compound is a single word nonetheless. After the replacement, only the last word of the resulting expression is a compound verb.

sō suru, “to do so” can replace only volitional verbs. Therefore, it can be said that none of the type A verbs can meet any of these criteria, and that some of the criteria may not apply to certain type B verbs due to semantic restrictions. In the rest of the thesis I will refer to verbs (either whole compound verbs, or their V2 verbs) of type A as **lexical verbs**, and to verbs of type B as **syntactic verbs**.

Based on these criteria, Kageyama has identified 30 V2 verbs used in syntactic verbs and categorized them according to meaning in the following way:¹³

- Inception: ～かける *-kakeru*, ～だす *-dasu*, ～始める *-hajimeru*,
 ～かかる *-kakaruru**
- Continuation: ～続ける *-tsuzukeru*, ～まくる *-makuru*
- Completion: ～終わる *-owaru*, ～終わる *-owaru*, ～尽くす *-tsukusu*,
 ～通す *-toosu*, ～抜く *-nuku*
- Incompletion: ～そこなう *-sokonau*, ～損ねる *-sokeneru**,
 ～損じる *-sonjiru*, ～そびれる *-sobireru*, ～かねる *-kaneru*,
 ～遅れる *-okureru*, ～忘れる *-wasureru*, ～残す *-nokosu*,
 ～誤る *-ayamaru*, ～あぐねる *-aguneru*
- Excessive action: ～過ぎる *-sugiru*
- Retrial: ～直す *-naosu*
- Habitual: ～つける *-tsukeru*, ～慣れる *-nareru*, ～飽きる *-akiru*,
 ～こなす *-konasu**
- Reciprocal action: ～合う *-au*
- Potential: ～得る *-eru/-uru*

The original list (Kageyama, 1993, p. 96) contained only the other 27 verbs. The three verbs marked with an asterisk (*) are considered syntactic V2 in *The Compound Verb Lexicon* (NINJAL, 2015). Out of these three verbs added later, the *kakaruru* and *sokeneru* have also been proposed as syntactic V2 based on Kageyama’s criteria by Himeno (1999).¹⁴

It should also be noted that some of the verbs listed above may be used in one sense as a syntactic V2, and in another sense as a lexical V2, sometimes even with the same V1. For instance the verb *dasu* occurs frequently both in syntactic verbs (e.g. 歩き出す *arukidasu*, “to start walking”, 飛び出す *tobidasu*, “to start flying, to take off” – the sense of inception) and lexical verbs (e.g. 引き出す *hikidasu*, “to pull

¹³The English names of the categories follow the terminology used in *The Compound Verb Lexicon* (NINJAL, 2015), which also adds the three compound verbs marked with an asterisk.

¹⁴Interestingly, one more addition suggested by Himeno (1999), *hateru* (in the completion category), is considered lexical V2 of type “Verb + Subsidiary verb” in *The Compound Verb Lexicon* (NINJAL, 2015). Himeno also gives an example of its compound with a *sahen* V1 verb: 退屈し果てる *taikutsushihateru*, “to be utterly bored”. The compounds that end with *hateru* are, however, difficult to rephrase using a verb-complement construction. This suggests that although the criteria are formally clear, there may be cases, which fulfil some of the criteria but lack some of the expected characteristics.

out”, 飛び出す *tobidasu*, “to spring out” – the sense of outward movement).

Kageyama (2013) further analyses the inner structure of lexical compound verbs, and identifies four types. These types are also being used in *The Compound Verb Lexicon* NINJAL (2015), from where I quote the following English explanations and examples:

1. **VV (Verb + Verb):** In the VV type, each of the two component verbs has its own lexical meaning and semantic roles (case relations). [... T]he verb in V2 generally determines the argument relations (case relations) of a whole compound verb. E.g. *arukitsukareru*, “get tired from walking” = “walk and get tired”.
2. **Vs (Verb + Subsidiary verb):** In the Vs type, the verb in V2 has become a subsidiary verb due to loss of its literal meaning and (in most cases) its semantic roles (case relations), while the verb in V1 maintains its lexical meaning and semantic roles. [... T]he verbs in V2 in the compound verbs of the Vs type have changed to adverbial or auxiliary functions, and the case relations (semantic roles) of a whole compound verb are determined primarily by the verb in V1. E.g. (*ame ga*) *furi-shikiru*, “(rain) falls incessantly”.
3. **pV (prefix + V):** In this type, the verb in V1 functions as a prefix or a prefix-like verb because of the weakening of its lexical meaning and (in many cases) the accompanying phonetic erosion. Because the prefix or prefix-like verb has only the function of emphasizing the meaning of the second verb, the case relations (semantic roles) of a whole compound verb are determined wholly by the verb in V2. E.g. *buttobasu*, “punch forcefully”.
4. **V (one word):** In this type, the compound verb is regarded as being lexicalized as one word because it is no longer recognized as consisting of two members by contemporary speakers. E.g. *ochitsuku*, “settle down, be in a stable condition” (lit. “fall and arrive”).

(NINJAL, 2015, examples abbreviated to one per type)

In conclusion, Kageyama’s lexical-syntactic dichotomy promises to effectively set apart productive compound patterns, which produce verbs with meanings composed transparently from V1 and V2. The typology of lexical verbs may also be used to separate compound verbs with transparent meaning (the VV type) from the rest of lexical verbs, acquisition of which presumably requires more attention from the learners. Several of the formal criteria (e.g. use of *sahen* verbs as V1) may have pedagogical value as well. The lexical-syntactic distinction and lexical typology will be applied to the verbs and their meanings analysed in the Appendix.

Chapter 2

Pedagogical resources on Japanese compound verbs

In the Introduction we have already mentioned that the coverage of compound verbs in basic textbooks is quite limited. A learner seeking to fill this gap may attempt to use one of the specialized resources on compound verbs. I have been able to find the following three resources that are accessible to speakers of Japanese language, English language (presumably familiar to many Japanese language learners), and Czech language (my native language):

- *The Compound Verb Lexicon* (NINJAL, 2015): An online database “useful not only to specialists on linguistic analysis but also to learners of Japanese, researchers in language processing, and other fields”, comprising over 2,700 lexical verbs commonly used in contemporary Japanese. It features definitions and example sentences with English, Chinese and Korean translations for each meaning sense. Additionally, each sense is also tagged with a “case pattern” (argument structure) and “word structure” (the four types cited on page 16). The project is lead by prof. Kageyama.
- 複合動詞の構造と意味用法 *Fukugōdōshi no kōzō to imiyōhō* (Himeno, 1999), “The structure and semantic usage of compound verbs”: A monograph focusing mostly on lexical verbs that summarizes and updates the author’s long-standing research of compound verbs motivated by the difficulties experienced by learners of Japanese. Himeno first groups compound verbs by V2 verb, organizing them in more detail based on the meaning of the V2 verbs and various other criteria (in chapters 3–11), and then attempts to classify V1 verbs according to which V2 verbs they combine with (in chapter 12). In both cases, only several important V2 verbs are covered. The individual compound verbs are most often only listed as members of the different groups.

- *Japonská slovesa v příkladech* (Kanasugi, Kurihara, Labus, and Morita, 2015), “Japanese verbs in examples”: A “study material, which may be used by both beginner and advanced students”, resembling a Japanese-Czech dictionary of simple verbs, lexical compound verbs, and an overview of syntactic V2 verbs. It lists over “900 most frequent simple verbs and over 800 compound verbs”. It features short Czech and Japanese definitions, information about transitivity, argument structures, and perhaps most importantly, many examples with Czech translations for each meaning sense.

In the following three sections I will review these resources, focusing on how they could be helpful to learners or teachers in the vocabulary acquisition process. In each section I will address successively coverage, treatment of polysemy, definitions, examples, and possible uses cases for the given resource.

2.1 The Compound Verb Lexicon

The Compound Verb Lexicon (NINJAL, 2015) has clearly the largest coverage of the available resources. The web site of the database explains how the 2,700 lexical compound verbs were selected: “Of these, about 2,000 come from the data collected by Taro Kageyama from dictionaries and other sources, and the others were added by Kyoko Kanzaki from major works on the topic.” and “Only compound verbs that are used more or less commonly in contemporary Japanese are included.” None of the most frequent lexical verbs seems to be missing.

In principle, the entries should differentiate multiple meaning senses, but even some of the very frequent entries list only one. For instance 飛び出す *tobidasu* lists only the sense “To fly out forcefully.”¹ The coverage of senses is thus often in sharp contrast with the large number of verb entries. Verbs that have meanings falling both into the lexical and into the syntactic type, simply omit the latter. For instance 言い出す *iidasu* lists only the meaning (or rather two of them combined into one) “To say out loud. To propose.”²

¹The third of the reviewed resources (Kanasugi et al., 2015) lists four senses. I have identified three frequent senses: “spring out, rush out”, “appear unexpectedly” and “run away” in on page vii of the Appendix, of which the first one corresponds to the one listed in the database. Two large dictionaries, 大辞林 *Daijirin* (Sanseido Co., Ltd., 2013), 大辞泉 *Daijisen*, (Shogakukan Inc., 2010), list six senses.

²In this case, neither of the the two meanings is used in the English translation of the only listed example sentence: 夫は、なぜ妻が突然、離婚を言い出したのか、理解できなかった。 *Otto wa, naze tsuma ga totsuzen, rikon o iidashita no ka, rikai dekinakatta*. It is instead translated as if the verb was a syntactic compound with the meaning “start to say”: “The husband could not understand why his wife had suddenly started talking about divorce.”. In my opinion, this reflects ambiguity of actual usage, but it certainly also adds to the confusion of a learner who would not find a meaning sense that has been excluded on a theoretical basis, only to see it in an example.

The definitions often seem to focus not just on the overall meaning but also on the composition of the meanings of the two constituents. The English definitions are verbatim translations of the Japanese ones. Appropriate English equivalents of the verbs can usually be found in the translations of the example sentences. As an example I will cite the definitions and translations used in examples of several very frequent compound verbs:

- 繰り返す *kurikaesu*: “(From spinning again thread that one has already spun) to do the same thing many times.”, translated as “to keep repeating”.
- 思い出す *omoidasu*: “To revive in one’s memory forgotten things.”, translated as “to remember”.
- 見つける *mitsukeru*: “To find something for which one was looking.”, translated as “to find”.
- 酔っ払う *yopparau*: “To become completely intoxicated with alcohol.”, translated as “to get drunk”.

The example sentences and their translations are helpful but there is usually only one per sense. For instance, there is only one example sentence for each of the frequent polysemous verbs 飛び出す *tobidasu* and 言い出す *iidasu* mentioned above.

The database is freely accessible online and can be searched by a whole compound, V1, or V2. Alternatively, the complete data can be downloaded as a spreadsheet. The entries include case patterns and links to the Balanced Corpus of Contemporary Written Japanese.

The vast size of the database, compounded with lack of any organization, makes it useful to learners primarily when searching for a particular verb, a function that can be fulfilled by a general dictionary as well. As a very comprehensive database it is certainly useful to linguists, which are also cited as target audience before learners.

2.2 “The structure and semantic usage of compound verbs”

The exact coverage of the book by Himeno (1999) is unclear. It contains an appendix with a list of over 2,200 compound verbs, and although it is organized around the same V2 verbs as the preceding chapters, it also lists compound verbs that were not previously mentioned. Additionally the coverage is limited by the choice of V2 verbs the author covers. For instance the frequent lexical V2 verbs ~取る *toru*, ~返す *kaesu*, ~詰める *tsumeru* are not included. As there is no index, the monograph is impractical for looking up a particular verb .

The book is organized mainly around meanings of V2, but nuances in meaning of individual compound verbs are usually not discussed. When a V2 can also be used in a syntactic compound, it is discussed as a separate meaning of that V2. Rather than to facilitate understanding different meanings of one verb, the organization of the book makes it possible to find other verbs with related meanings.

The last chapter uses a different principle of organization, classifying V1 verbs according to which V2 verbs they combine with. Five such classifications of V1 verbs are made on different groups of V2.³ Given that Himeno, as she admits at the end of the chapter, was not able to use a computer for any automated analysis, the result is quite impressive. Unfortunately, in this chapter the author does not even list the actual compound verbs in most cases, giving only a V1 verb and a direction category (which involves up to six different V2 verbs), presumably enough for a native speaker to tell the resulting verb without hesitation.

Most of the covered compound verbs do not include examples or definitions, but the discussions of meanings and usage of selected verbs are very informative.

Supplemented with other resources, the book might be helpful to authors of textbooks, and perhaps teachers. Due to the lack of information provided on individual compounds and the absence of an index, its usefulness to learners is very limited.

2.3 “Japanese verbs in examples”

The handbook (Kanasugi et al., 2015) is specifically designed as a learning aid and its main focus are simple verbs (over 900 frequent verbs selected to cover the Japanese Language Proficiency Test Level 1). In addition to them, it contains over 800 compound verbs, but the selection criteria for these are not clear. Some very frequent verbs are missing (for instance 取り出す *toridasu*).

Meaning senses seem to be listed in a very comprehensive manner compared to *The Compound Verb Lexicon* database. Although purely syntactic compounds are not covered, if a verb has meanings falling both into the lexical and into the syntactic type, both types are included under the same headword. The number of meaning senses generally seems to strike a good balance, usually being higher than in *The Compound Verb Lexicon*, but also lower than in large Japanese dictionaries, which tend to cover common usage as well as rare and historical.

³For instance, the first and most extensive classification concerns combination with V2 verbs having directional meanings. The V2 verbs are first divided into four categories (upward, downward, outward, and inward direction). The V1 verbs are then classified based on their potential to combine with the four categories of V2. To give a few examples, the verb 浮く *uku*, “to float”, combines with (and only with) V2 verbs for upward and inward direction, the verb 走る *hashiru*, “to run”, combines with all four directional categories of V2 verbs, and so on.

Both Czech and Japanese definitions are succinct and easy to understand, although exact meaning is often clearer from the example sentences. Some very common meaning senses have up to four translated examples.

The lexical verbs are organized alphabetically and can be also looked up using an index. Information about transitivity and as argument structures are included too. The closing chapter lists all 30 syntactic V2 verbs, each with a definition and several example sentences.

Except for the short but helpful chapter on syntactic verbs, the book cannot be used alone as an aid for vocabulary acquisition due to its size and lack of organization other than alphabetical. It can, however, be more practical for looking up verbs than a general dictionary thanks to the example sentences, not to mention that a decent general Japanese-Czech dictionary is not yet available.

2.4 Conclusion

We have mentioned various useful features and use cases of the three resources, but we have also recognized how their use as pedagogical resources is limited.

The online database (NINJAL, 2015) and the Czech handbook (Kanasugi et al., 2015) are relatively accessible to learners, but the online database often fails to cover multiple meanings, and even intentionally omits the syntactic ones, and the handbook, while handling polysemy well, (unintentionally) omits several important verbs. Most importantly, none of the resources can be used to organize or prioritize vocabulary acquisition. The monograph by Himeno (1999) is organized very well, effectively grouping together verbs with similar meanings, but its coverage is limited to several V2 verbs, and it does not lists meanings and examples for many of the included verbs.

Chapter 3

Research questions and methodology

The main goal of this thesis is to build a pedagogical word list that would provide guidance to learners and teachers on which compound verbs and which of their meaning senses are the most frequent, and therefore the most useful. As we have argued in the previous chapters, this goal is not fulfilled by the currently available pedagogical resources.

We will also attempt to answer the following four quantitative questions that are instrumental in building the list and in assessing its utility to learners. Additionally, we will compare the results with quantitative characteristics of English phrasal verbs, which are also known for being difficult to acquire for learners.

1. What is the coverage of the resulting word list of compound verbs?
2. How frequent are simple verbs, lexical compound verbs, and syntactic compound verbs in Japanese?
3. How do the frequencies compare with the frequencies of simple and phrasal verbs in English?
4. How difficult is it to achieve a particular coverage of Japanese compound verbs compared to achieving the same coverage of English phrasal verbs?

The rest of this chapter will discuss the methodology used for creating a frequency list of Japanese compound verbs, as well as for answering the the above questions, proceeding from general methodology to technical matters. I am not aware of any previous attempt to create such a list of Japanese compound verbs, but a similar list of English phrasal verbs, dubbed *The PHaVE List*¹ by its authors (Garnier and Schmitt, 2015), has been published. The discussion of the extent to

¹The full title of the article is *The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses*.

which the methodology of *PHaVE* list can be applied when creating Japanese compound verb list will follow when appropriate. I will also cite a general vocabulary research manual (Schmitt, 2010).

3.1 Measuring vocabulary

In order to answer the first two research question, we must first adopt some method of measuring vocabulary size. In the field of vocabulary research, vocabulary may be measured in both absolute and relative terms. Let us initially assume that we can distinguish individual words. (In English, orthographic words, conveniently separated by spaces, usually overlap with morphological words. We will examine Japanese words in Section 3.5.)

The unit commonly used to measure vocabulary in absolute terms is a **word family** (Schmitt, 2010, p. 8). Each word family includes a base word (e.g. 話す *hanasu*, “to speak, to talk”), words regularly derived from it (e.g. 話 *hanashi*, “a talk”, 話し方 *hanashikata*, “a way of talking”) and their inflections (e.g. a past form 話した *hanashita*). Compounds (e.g. 話し言葉 *hanashikotoba*, “spoken language”) form separate word families. Schmitt (2010, p. 9) also notes that the members of a word family are not always transparently related and gives the English words “involved” and “involvedness” as an example. The same is true about Japanese (take for instance the various ways to form plural, or the adverbial forms of verbs which sometimes can be converted to nouns and sometimes not). Consequently acquiring a complete word family may not be completely straightforward, and its usage as a unit of vocabulary size is a necessary simplification rather than an accurate description of reality. Additionally, the concept of vocabulary consisting of word families ignores formulaic language altogether (see Schmitt, 2010, sec. 1.1.3, ch. 3 for further discussion).

Nonetheless, the concept of word family is very useful for creating a pedagogical list of Japanese compound verbs. Due to the presence of highly productive suffixes in Japanese, it is natural to work with whole word families of the compound verbs, for instance taking into account nouns derived using the suffix ～方 *-kata* (e.g. 受け取る *uketoru*, “to accept” → 受け取り方 *uketorikata* “a way to accept”).²

To measure the utility of a speaker’s vocabulary, the relative measure of **coverage** is used. A coverage of a vocabulary is the sum of the relative frequencies of its words. (Considering a vocabulary consisting of word families, it is the sum

²Additionally, conversion, which can be thought of as adding a zero suffix, is very common in Japanese, e.g. 思い込み *omoikomi*, “an impression”, a noun converted from the adverbial form of the verb 思い込む, “to be under an impression”, 思い切って, “resolutely”, an adverb converted from the *te*-form of the verb 思ひ切る *omoikiru*, “to resolve”.

of the frequencies of the words of its word families.) **Word frequency** is usually given as a percentage, indicating how many times a word occurs per hundred words. A frequency of a word can be calculated based on a corpus as a ratio of the word's tokens (occurrences) and the total number of tokens present in the corpus. Consequently, coverage of a vocabulary may theoretically range from 0 % (empty vocabulary) to 100 % (all words of the language discourse, or all words of a corpus, a level not achieved even by native speakers given a representative corpus).

In the research of vocabulary size of learners of English, target coverages of 95 % for being "conversant" or 98 % for being able read a novel or a newspaper have been suggested, and the requirements for Cambridge-ESOL Certificate of Proficiency in English or the highest level (C2) of the Common European Framework seem to fall within the range of these two values (see Schmitt, 2010, sec. 1.1.2). Of course, the percentages are arbitrary. In the case of reading, the 98 % figure simply means not knowing one in fifty words (more precisely tokens) and there is no particular reason that makes 98 % a more appropriate goal than 95 % or 99 %. Additionally, the corpus on which the coverage is measured is just an approximate representation of the language discourse, and even if it was a good representation overall, it would not represent well specific genres (e.g. legal or technical texts).

3.2 The economics of vocabulary learning

Vocabulary of our first language is acquired mostly naturally, but second language learners acquire large part of their vocabulary through some sort of organized learning and using learning resources such as textbooks. In that case, the learner (or to be more realistic, the teacher or the author of the textbooks) has to make a choice of which vocabulary to include at which stage of learning.

It is clear that the more vocabulary the better. Why should we give priority to particular words? Intuitive guesses may be made about which words are more basic or useful, but research has shown that only frequency determined from a corpus provides a reliable and measurable evidence of a word's utility (Schmitt, 2010, sec. 1.1.4). The second, perhaps obvious, reason for preferring some words over others, is acquiring vocabulary takes time and the time at hand of a second language learner is limited. In terms borrowed from economics, learning a word carries the opportunity cost of not being able to learn other words (given the limited time available to second language learners). Given this cost, a rational learner should maximize marginal utility by learning words (or word families) of higher frequency first. Although not expressed in these terms by its authors, this is also

the underlying rationale of the *The PHaVE List* (Garnier and Schmitt, 2015).

Of course other considerations, such as semantic relations between words, should be taken into account when designing teaching materials. Additionally, a distinction should be made between spoken and written discourse, and different genre-specific vocabulary may be important to each learner.

As this thesis is concerned with a very specific class of Japanese words, not just Japanese vocabulary in general, I believe it is vital to take this line of economic thinking one step further: It must be realized that learning a compound verb carries not only the opportunity cost of not being able to learn other compound verbs, but also the opportunity cost of not being able to learn some other simple verbs (or perhaps nouns). It would therefore be wrong to focus solely on frequencies and coverage of compound verbs. For instance, we cannot assume that in order to cover with one's vocabulary 98 % of a corpus, one also should cover 98 % of compound verbs. Telling by which proportions different parts of speech should be represented in one's vocabulary would require additional analysis, as each part of speech has an irreplaceable role in the language, but it seems reasonable to measure (or set as target) overall coverage of verbs, and then identify how many simple verbs and how many compound verbs, based on their respective frequencies, one should learn to reach this overall target for verb coverage.

To give a concrete example, let us consider the following frequencies within verbs:

- simple verbs and compound verbs have the total frequencies of 60 % and 40 %, respectively,
- the three most frequent simple verbs have the frequencies of 10 %, 10 %, and 5 %, respectively,
- the three most frequent compound verbs have the frequencies of 6 %, 4 %, and 3 %, respectively.

If our goal was to achieve 20 % coverage of verbs, it would be rational to limit ourselves only to the two most frequent simple verbs (covering $10\% + 10\% = 20\%$), not learning any compound verbs at all. To achieve 30 % coverage we would learn the three most frequent simple verbs and only one compound verb (reaching the coverage $10\% + 10\% + 6\% + 5\% = 31\%$), which would involve only 15 % ($= 6\% / 40\%$) coverage of compound verbs.

Although the example is only hypothetical, compound verbs indeed happen to be much less frequent than simple verbs (see Chapter 4). We will take this into consideration when calculating and comparing coverages (the second research question).

Similar considerations apply to learning the meaning senses of a polysemous word, as meaning senses too have their frequencies. It is useful to learn the dif-

ferent but related meanings of the same word (and words in its word family) and to study them together. At the same time, it is also reasonable to omit infrequent meaning senses. Therefore, we will not depart from using words and word families as units of vocabulary acquisition. We will, however, set a frequency threshold for the inclusion of a meaning sense. The threshold should be set relative to the frequency of the least frequent words we are willing to learn, consequently including less senses for less frequent words.³

3.3 Selection of items for the list

The list will have two levels of items: the compound verbs and their meaning senses. As the syntactic compound verbs are supposed to have meanings composed transparently from their V1 and V2 verb meanings (as discussed in Section 1.2, the list will focus on providing information on the lexical compound verbs. Nonetheless, the syntactic verbs will not be ignored. Their frequencies will be computed, samples from the corpus will be taken to verify the results of the automated extraction, and they will be included along with their frequency at the corresponding positions on the list, only their meanings will not be analysed. In case of a verb that has meanings falling both into the lexical and into the syntactic type, both will be included for clarity (like in the handbook by Kanasugi et al., 2015).

At the first level, compound verbs will be selected starting from the most frequent ones until reaching the desired level of coverage, as explained in the previous section.

At the second level, we will use a similar procedure for selecting meaning senses as Garnier and Schmitt (2015), that is including only meaning senses that are above a certain frequency threshold⁴. The procedure will differ, however, in how the threshold is set. Garnier and Schmitt set the threshold proportionally to the number of occurrences of each verb, so senses accounting for the selected proportion of the verb's occurrences are included regardless of the verb's frequency. As the most frequent verbs in their list are over ten times more frequent than the least frequent ones, it means that, in the extreme case, meanings of one verb may be omitted even if ten times more frequent than meanings of another verb, which are included. To maintain overall consistency, we will instead set the threshold relative to the frequency of the least frequent word on the list, as explained in the

³Setting it to the threshold equal to the frequency of the least frequent word would be equal to asserting that learning a new meaning of a known polysemous word requires the same effort as learning a completely new word of the same frequency.

⁴In addition, the authors also use an upper threshold, and the resulting selection procedure is rather complex Garnier and Schmitt (2015, p. 652–653).

previous section.

The procedure of determining the frequencies of the verbs and their meaning senses will be described in more details the subsequent sections of this chapter.

3.4 Information provided by the list

Each verb entry on the list will feature the following information:

1. *Headword*: The compound verb is written in a common Japanese orthography⁵ and transcribed to Latin alphabet. Multiple variants of the same verb may be listed.
2. *Frequency and coverage*: The total frequency of the verb and the coverage of the selected meanings are given as percentages of total tokens of lexical compound verbs.
3. *Formation and type*: The verbs V1 and V2 verb are written in the same orthography and transcription as the headword, with euphonic changes (see page 11) indicated. The type of the compound verb is given as “lexical”, “syntactic”, or “lexical and syntactic”. For the lexical verbs, the type of the word structure according to NINJAL (2015) (see page 16) is also given.
4. *Selected senses*: The selected meaning senses are ordered by decreasing frequency. Each of them features the following information:
 - a. *English equivalent*: An English equivalent or explanation, if necessary, is given.
 - b. *Frequency*: The frequency is given as a percentage of total tokens of the verb. Unlike the verb frequencies, the meaning frequencies have only informative character (see Section 3.8).
 - c. *Example sentences*: At least one example sentence from the corpus that illustrates the meaning well is given and translated to English. A reference to the work in which it originally appears is included.
5. *Syntactic sense*: A syntactic sense of the verb, if present, is listed separately. For verbs that have both lexical senses and a syntactic sense, example sentences for the syntactic sense are listed as well.

⁵We do not list multiple orthographic variants, although they may appear in the example sentences.

6. *Omitted senses*: To document the selection of meaning senses, all other meanings identified in the corpus sample are listed with their English equivalents and frequency information, but without examples.
7. *Notes*: Notes present in some of the entries document the preparation of the entry, or contain additional information about the verb.

The information given is based strictly on the corpus: word variants, meanings, or use as a syntactic verb not represented in the sample extracted from corpus is not indicated. The example sentences taken from the corpus sample are presented without changes, except for explicit abbreviation or addition of necessary context.

3.5 Japanese words and orthography

Japanese orthography has three characteristics that make corpus analysis relatively difficult. The first of them is the lack of word separation, which can be partially remedied by morphological analysis (discussed in Section 3.9). The second characteristic is the pervasive use of orthographic variants. Without delving too deep into the Japanese writing system or considering which variants are standard or appropriate, I will name the practical issues that can be encountered when extracting verbs from a corpus:

1. A simple verb may be written entirely in kana, or in a combination of kanji and okurigana⁶. In the latter case, the kanji used and the length of the okurigana may vary. As a result the verb *agaru* may be written as any of the following: あがる, 上がる, 上る, 挙がる, 挙る, etc., the verb *tomaru* may be written as とまる, 止まる, 止る, 停まる, 停る, etc. The various kanji may or may not distinguish homonyms or meaning senses. In some cases two kanji may be used in one simple verb (e.g. 出来る *dekiru*).
2. In a compound verb, the previous point applies both to V1 and V2. Commonly, but not regularly, the V1 verb is written using a kanji, while the V2 verb is written using only kana.
3. Additionally, if both the V1 and the V2 verbs are written using kanji and okurigana, the okurigana may be entirely left out from V1. For instance the verb *tachiagaru* (basic meaning: “to stand up”) is commonly written as 立ち上がる, 立ちあがる, 立上がる, or 起ち上がる.

⁶The Japanese writing system uses *kana*, a syllabary, and *kanji*, ideographic characters originating in China. If *kana* occurs at the end of the word after initial *kanji*, it is called *okurigana*.

4. In some cases, the whole compound verb may be written with a single kanji, obscuring the morpheme boundary, for instance 陥る *ochiiru*.

The third characteristic of Japanese orthography to be taken into account is the presence of homographs. For instance both the verb *agaru* and the verb *noboru* may be written as 上る, both the verb *tomaru* the verb *todomaru* may be written as 止まる (and also 止る, 停まる, 停る). This too applies to both constituents a compound verb. Fortunately, homography does not affect as many words as orthographic variation.

In order to effectively extract information from the corpus we will need to recognize the orthographic variants as belonging to a single lemma⁷, and also disambiguate homographs. Unlike word segmentation, these two tasks cannot be solved by tools for morphological analysis. We will return to them in Section 3.9.

3.6 Compound verbs and morphological analysis

As the first step in automated processing, a text must undergo morphological analysis, which consists of tokenization (segmenting the text into tokens) and tagging (adding linguistic information about the tokens). In the case of Japanese language the available tools segment the text into tokens, which do not have a precise linguistic definition, but are closer to morphemes than to words.

The following list summarizes the typical behaviour of morphological analysis with regards to compound verbs. (Hyphens in the examples indicate token boundaries.)

1. Tokens always follow character boundaries (i.e. individual characters are never segmented), e.g. 陥る *ochiiru* (kanji obscures morpheme boundary, no segmentation), 終わっ-ちゃう *owat-chau* (a contraction of 終わっ-て-しま-う *owat-te-shimau*).
2. A simple verb in non-past form, adverbial form, or any of the basic stems of Japanese conjugation is considered a single token, e.g. 読む *yomu*, 読み-ます *yomi-masu*, 読ん-だ *yon-da*, 読ま-ない *yoma-nai*.
3. A *sahen* verb is segmented as (at least) two tokens, e.g. 使用-する, デジ-タル-化-する *dejitaru-ka-suru*.
4. A potential form of a monograde⁸ verb is considered a single token, e.g. 読める *yomeru* (a different lemma from 読む *yomu*), 食べ-られる *tabe-rareru*.

⁷I use the term lemma the way it is commonly used in corpus linguistics or morphological tagging: a word representing the set of inflected word forms of the same word.

⁸一段 *ichidan*, “monograde”, is one of the conjugation classes of Japanese verbs.

Consequently no difference is made between 切れる *kireru*, “to break”, and 切れる *kireru*, the potential of 切る *kiru*.

5. Suffixes (including conjugated suffixes) are segmented as a separate tokens, e.g. 読ま-さ-れ-な-なかつ-た *yo-ma-sa-re-na-kat-ta*, 使用-さ-れ-始める *shiyō-sa-re-hajimeru* (a syntactic compound verb).
6. Some compound verbs are considered a single token, some are parsed according to the above points as a sequence of tokens, e.g. 見回す *mimawasu*, 眺め-回す *nagame-mawasu* (both lexical), 有り得る *ariuru*, 勉強-し-始める *benkyō-shi-hajimeru* (both syntactic).⁹

The tags assigned by morphological analysis to each token give information about the token’s reading, lemma, part of speech, conjugation class (e.g. monograde), and conjugation form (e.g. adverbial).

Consequently, compound verbs can be extracted by searching for the following patterns of part of speech and form of conjugation:

1. verb (any form), if its lemma can be identified as a compound verb,
2. verb (adverbial form) + verb (any form),¹⁰
3. verb (irrealis form)¹¹ + passive suffix *-reru/-rareru* (adverbial form) + verb (any form).

Note that the patterns correspond to the formation of compound verbs as discussed in Chapter 1, and therefore the first verb in pattern 2 and pattern 3 may be either a simple verb or a compound verb corresponding to one of these three patterns. The identification of compound verbs left unsegmented by morphological analysis (pattern 1) requires additional data not available from the morphological analysis.

The occurrences extracted in this way will include not only compound verbs but also parts of more complex compounds or derived words. This may or may not correspond to actual derivation or compounding that involves a compound verb, e.g. であり得る *deariuru* (tokenized as *de-ari-uru*) contains a pattern corresponding to the compound verb あり得る *ariuru*, but it is actually a formed by compounding である *dearu* and 得る *uru*. We will therefore have to verify if the word actually occurring in the text is indeed a compound verb or a word from its family.

⁹Especially the last point, whether compounds are recognized as a single tokens, depends on both the lexicon and the statistical model used by the tool morphological analysis. The same compound word may be tokenized in a different way based on context and orthography.

¹⁰There are, however, several common grammatical patterns, in which adverbial forms of a verb are followed by another verb (e.g. ~なさい *...nasai*, お~ください *o...kudasai*) or a contracted form tagged as a verb (e.g. ~ちゃう *...chau*). Such patterns have to be excluded.

¹¹Irrealis form: 未然形 *mizenkei* in Japanese, or “*a*-stem”, “*nai*-stem” in common English textbook terms.

3.7 The corpus

I have first considered using the largest corpus of modern Japanese language, the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (NINJAL, 2009). The BCCWJ is tokenized and tagged using the morphological analysis tool Mecab, so extracting occurrences of compound verbs using the patterns from the previous section would be possible. The corpus would certainly be suitable for the task, but none of its freely available interfaces allows for search by such patterns, let alone automated extraction of their occurrences for further preprocessing. Access to the full data is paid only.

Instead, I have opted to create a smaller corpus from texts, mostly literary works, freely available from the 青空文庫 *Aozora bunko* online library. I have used only works written using new orthography (新仮名遣い *shinkanazukai*), whose first publication date is from 1945 onwards.¹² The resulting corpus consists of 12,566,822 tokens (compared to 104,911,464 tokens in the BCCWJ, about eight times as many), which is sufficient, as we will focus on the most frequent words. As it consists predominantly of literary works, it cannot be considered balanced. However, out of the possible/available genre-specific vocabularies, the one of literary works is relatively acceptable. Further considerations about the corpus suitability for this task, as well as the technical details about the corpus creation and its morphological analysis can be found in my research report about the corpus (Nohejl, 2016)¹³.

3.8 Sampling

Even in our small corpus the most frequent compound verbs have hundreds of occurrences. In order to review them for errors, homographs, and to analyse the polysemy, we extract random stratified samples from the occurrences.¹⁴

The sample size is calculated for each compound verb to meet the following two conditions:

1. *Representativeness of the verb*: The sample must be large enough so that the

¹²These 1,185 works comprise about 9 % of *Aozora bunko*, which concentrates mostly on older works. This is because it contains only works that are out of copyright (i.e. mostly with expired copyright). Thankfully, in Japan, copyright expires after 50 years since the death of the author.

¹³The report, titled *Gendai nihongo no kōpasu: fukugōdōshi no kenkyū ni mukete*, “A corpus of contemporary Japanese: towards a study of compound verbs”, is also available online: <http://nohejl.name/files/2016/vv-20160705-report-corpus-fukugoudoushi.pdf>

¹⁴The stratification is done according to decimal classification of genres available in *Aozora bunko*. Of course, it cannot make corpus more balanced, but it ensures that genres in the sample are distributed in the same way as in all the occurrences, which is important given a relatively small sample size and that the corpus is biased towards literary works.

proportions (e.g. frequencies of meaning senses relative to the number of occurrences of the verb, which will be included in the list) computed from the samples are within $\pm 20\%$ of the actual proportions with 95 % confidence.¹⁵

2. *Representativeness of the meaning senses included in the list:* As we have decided to set threshold for inclusion of a meaning relative to the frequency of the least frequent words in the list (see Section 3.2), the sample of each verbs should be large enough to include (at least one instance of) meaning senses with frequency just above this threshold with 95 % confidence.¹⁶

3.9 The analysis

The morphologically analysed data are further processed and analysed in the following steps:

1. *Extraction of compound verbs:* Compound verbs are extracted by searching for patterns according to Section 3.6. We use data from *The Compound Verb Lexicon* (NINJAL, 2015) to identify and segment compound verbs that were left unsegmented by the morphological analysis (pattern 1 in Section 3.6).
2. *Assignment of a common lemma to unambiguous orthographic variants:* Possible ambiguities are determined and one of the variants is assigned as a lemma using data from Jim Breen's EDICT dictionary¹⁷.
3. *Computation of frequencies and sample sizes:* For each verb, its frequency and the size of the sample to extract according to (Section 3.8) are computed.
4. *Extraction of samples:* For each verb, a sample of contexts in which it occurs is extracted according to (Section 3.8).
5. *Manual verification:* We manually verify that each item in the sample contains the supposed compound verb or a word from its word family. Otherwise it is excluded from the sample, namely in the following cases:

¹⁵There are two reasons why we can afford to use the relatively high 20 % margin of error: First, our primary goal is to prioritize verbs, not meaning senses, therefore the exact percentage is only informative. Second, the chance of an important meaning senses being omitted even if above threshold because of the error in frequency is still very low. It would be less than 5 % for meaning senses with frequency over 20 % + the threshold.

¹⁶The first condition will result in sample sizes that differ only very little based on the number of occurrences of the verb, as is usual in representative sample size calculation. In contrast, the effect of the second condition is to enlarge sample sizes of very frequent verbs.

¹⁷The EDICT dictionary files and documentation are available online: http://edrdg.org/jmdict/edict_doc.html

- a. *Errors*: results of incorrect segmentation or tagging (e.g. 陥す *otosu*, orthographic variant of 落す *otosu*, segmented as 陥-す *ochii-su*, where 陥 *ochii* is tagged as an irregular form of the compound verb 陥る *ochiiru*)
- b. *Verb mismatches*: homographs (see Section 3.5), potential forms of different verbs (see Section 3.6),
- c. *Word family mismatches*: more complex words that contain a compound verb pattern as their part (see Section 3.6), but do not belong to word family of that compound verb.

To compensate for any exclusions, larger samples are extracted. In case of a verb mismatch, frequencies of the two respective verbs are also adjusted.

6. *Analysis of polysemy*: Meaning senses in the samples are identified, their frequencies computed, and suitable example sentences selected. Possible distinctions of meaning senses are first examined in existing compound verb verb lexicons (NINJAL, 2015; Kanasugi et al., 2015) and major Japanese dictionaries, namely the electronic editions of 広辞苑 *Kōjien* (Iwanami shoten, 2011), 大辞林 *Daijirin* (Sanseido Co., Ltd., 2013), 大辞泉 *Daijisen* (Shogakukan Inc., 2010), and 明鏡国語辞典 *Meikyō kokugo jiten* (Kitahara and Taishukan, 2011).

Chapter 4

Results

In this chapter we will discuss the main result of this thesis, the list of the most frequent compound verbs and their senses, which can be found in the Appendix, and answer the four research questions from the previous chapter, namely:

In Section 4.1, we will discuss the list and compare it with the currently available pedagogical list, answering the first two research questions.

In Section 4.2, we will compare the quantitative characteristics of Japanese compound verbs and English phrasal verbs, answering the third and fourth research question.

4.1 Discussion of the list

The methodology adopted to create the list was discussed in the previous chapter. To proceed with the preparation of the list, its extent also had to be decided. As follows from our methodology, the size of the list is directly linked with its coverage, which is a more useful parameter to the learner. We have also argued that it is more sensible to consider the overall coverage of verbs than only the coverage of compound verbs. (In the following discussion of the coverage, we will take into account only lexical compound verbs, as syntactic compound verbs do not need to be acquired by learning the individual compounds, see also Section 3.3.)

To find the desired level of coverage, we will compare the estimated numbers of simple and compound verbs necessary for levels of coverage from 75 % to almost 100 % of all verb occurrences using the data extracted from our corpus.¹

¹The data for the estimates were obtained by the following procedure: The first two steps of the procedure in Section 3.9 were applied to both simple and compound verbs and frequencies of the individual verbs were computed. The verbs featuring one of the syntactic V2 and at the same time not occurring in *The Compound Verb Lexicon* (NINJAL, 2015) were considered likely to be syntactic and therefore omitted. *Sahen* verbs, which cover about 12 % of all verbs occurrences are considered one verb due to being analysed so by the morphological analysis. Samples were not manually analysed.

All verbs		Compound verbs	
coverage	verbs	verbs	coverage
99.73 %	6,004	* 2,759	96.08 %
98.00 %	2,799	896	80.11 %
95.00 %	1,566	317	57.12 %
92.00 %	† 1,048	133	39.54 %
85.00 %	517	27	35.04 %
75.00 %	222	4	5.43 %

* = *The Compound Verb Lexicon* entries

† ≈ Kanasugi et al. entries (over 900 simple)

Table 4.1: The second column indicates the number of verbs (both compound and simple) necessary to achieve the coverage given in the first column. The third column indicates how many of these verbs are compound verbs. The fourth column indicates the resulting coverage of lexical compound verbs. (Syntactic compounds are not considered and *sahen* verbs are counted as one verb, *suru* for these estimates.)

The results are shown in Table 4.1. Two of the rows are included in reference to two of the compound verb resources discussed in Chapter 2. The 2,759 most frequent (lexical) compound verbs, the same number as included in *The Compound Verb Lexicon* (NINJAL, 2015), give us impressive 96.08 % coverage of all compound verb occurrences. To aim for learning that many compound verbs would make sense as part of a hypothetical goal to cover 99.73 % of all Japanese verb occurrences. If one were to set a more realistic goal to learn the over 900 simple verbs featured in the handbook by Kanasugi et al. (2015), the proportional amount of compound verbs would be a little over one hundred.²

A list of about 100 to 300 entries seems appropriate for achieving a good coverage while keeping the list reasonably long for learners. For the purpose of this thesis we have set a lower goal, 85 % coverage, which corresponds to 27 compound verbs. The number of verbs finally included in the list is actually higher as we have compensated for the omitted meaning senses by including more verbs.

The last general decision to be made the list while preparing was to set the threshold for omitting a meaning sense. The threshold was finally set to 0.1 % of all compound verb occurrences, which corresponds to 43 % of the occurrences of the last verb on the list. Consequently, even the last verb on the list has the chance to have two meanings included on the list if both are frequent enough.

The final list consists of 37 entries, contains 32 lexical compound verbs (in-

²The publication does not claim that the 800 compound verbs it contains in its second part are proportional to the simple verbs in the first part. It states, however, that the 900 simple verbs correspond to Japanese Language Proficiency Test Level 1, which could make learning them, and a corresponding amount of compound verbs, an attractive goal for many learners.

cluding verbs with both lexical and syntactic meanings) and 5 purely syntactic compound verbs. 45 lexical senses have passed the threshold for being included on the list, another 2 syntactic senses are included for completeness and clarity, while another 29 lexical meanings have been identified but omitted. 6 supplementary verbs (in addition to the initial estimate of 27 lexical compound verbs) were added to compensate for the omission of these senses and reach the goal of covering 85 % verb occurrences, two of the verbs in the initial estimate were manually joined as orthographic variants of one verb.

We have therefore succeeded in reducing the number of senses by almost one sense per verb on average (from 2.31 to 1.41) with a very small effect on the overall coverage (compensated by adding 6 more verbs to the list). Overall, the average included number of lexical senses is similar to that of *The Compound Verb Lexicon* for the same verbs (1.56), but it has been done based on a clear standard. Additionally, syntactic senses are included as well.

The list covers 17.95 % of the lexical compound verb occurrences, which is proportional to covering 85 % verbs overall. The resulting coverage shows that the list, despite being short, can cover an important part of the vocabulary effectively thanks to being firmly grounded in corpus analysis.³ It can therefore be useful as a basic pedagogical resource on compound verbs.

4.2 Comparing Japanese compound verbs and English phrasal verbs

To view Japanese verbs and compound verbs from a broader perspective, their frequencies are compared in Table 4.2. The data for Japanese were extracted and computed as we have already described. The frequencies of English verbs are based on data from the British National Corpus (BNC)⁴ and the number of phrasal verb occurrences in BNC indicated by Garnier and Schmitt (2015).

The apparent discrepancy between the frequency of verbs in the two languages may be partially due to the fact that tokens resulting from the morphological analysis of Japanese are closer to morphemes than to words (see Section 3.6). There is, however, no such technical explanation for the difference between the frequencies of compound verbs in Japanese and phrasal verbs in English (both relative to verbs in general). We can therefore say that compound verbs (and even lexical compound verbs) are more frequent in Japanese than phrasal verbs are in Eng-

³Incidentally, the list includes three compound verbs (取り出す *toridasu*, 酔っ払う *yopparau*, 見上げる *miageru*) that are not included in the much more extensive handbook by Kanasugi et al. (2015).

⁴Available from <https://kilgarriff.co.uk/bnc-readme.html>.

	Japanese (CV)	English (PV)
Frequency of verbs	12.12 %	17.83 %
Frequency of CV (PV) rel. to verbs	7.19 %	2.87 %
Frequency of lexical CV rel. to CV	85.98 %	–

Table 4.2: Overall frequencies of compound verbs (CV) in Japanese compared with phrasal verbs (PV) in English. Frequencies in each row are to be interpreted as a proportions of the frequencies in the previous rows.

lish.

To further extend this comparison, let us consider the parameters of the list of the most frequent 150 English phrasal verbs. The list covers 62.95 % occurrences of phrasal verbs (Garnier and Schmitt, 2015). To cover the same proportion of the occurrences of Japanese compound verbs, one would need a list of 413 compound verbs.⁵ It should also be noted that the average number of frequent senses per English phrasal verbs indicated by Garnier and Schmitt (2015), 1.92, is not markedly different from the number we have derived for Japanese compound verbs (1.42, which is smaller only by factor of 1.36).⁶

⁵The computation has been done in the same way as the computations in Table 4.1.

⁶Out of the comparisons that been made, this one should be considered the most imprecise. It would be necessary to compare with a larger set of Japanese compound verbs and verify that the methodologies are similar enough to afford direct comparison of the results.

Conclusion

In this thesis, we have analysed Japanese verb-verb compound verbs using a corpus and we have successfully built a pedagogical word list of the compound verbs and their meanings based on frequency criteria. We have also assessed the resulting lists utility.

In Chapter 1, we have reviewed the characteristics and typology of Japanese compound verbs in the previous research and noted that Kageyama's lexical-syntactic could be useful for set apart the productive compound patterns, which produce verbs with meanings composed transparently from its constituents, an observation we have later used to focus our list on lexical compound verbs.

In Chapter 2, we have reviewed existing pedagogical resources on Japanese verb+verb compound verbs. Having discussed their utility from a student's (or a teacher's) perspective, we have identified a need for a list of compound verbs and their senses based on frequency criteria.

In Chapter 3, we have formulated several research questions related to the creation of a frequency-based list of compound verbs. We have discussed a methodology for creating such a list and assessing its utility to learners, while paying attention to the characteristics of the Japanese language. We have described how items for the list will be selected, what information will be given for the items, what information sources will be used, and how information will be retrieved from the corpus.

In chapter Chapter 4, we have discussed the resulting list and attempted to answer our research questions. We have shown the list consisting of 37 items covers 17.95 % of the lexical compound verb occurrences, which is proportional to covering 85 % verbs overall. We have shown how frequent are simple verbs, lexical compound verbs, and syntactic compound verbs in Japanese and compared their frequencies with the frequencies of simple and phrasal verbs in English. The Japanese compound verbs are more frequent and also more diverse (we need more of them to reach the same coverage), while the analysis of polysemy made in this thesis suggests that they are only marginally less polysemous. Therefore, they are also likely to be a major stumbling block for language learners.

I hope the resulting list could find at least some use in practise. Despite being relatively limited, it provides effective information grounded in corpus analysis, which can help acquire compound verbs and their senses that account for 17.95 % of the occurrences. Such coverage is proportional to 85 % of verbs overall, which seems like a reasonable goal for an intermediate learner.

Clearly, the present list also has also shortcomings other than its relatively limited extent. A more balanced corpus consisting of more recent texts should be used to confirm the frequencies as well as the analysis of meanings, whether by extending the current corpus or by using a well established one, such as the BCCWJ. A separate analysis of spoken discourse would also be valuable.

The analysis itself could be made more precise by taking larger samples, and perhaps more importantly, recruiting several independent reviewers. I have consulted the assignment of senses to the sentences from a corpus sample of several verbs with a well-educated native speaker, only to find that the task was difficult for her as well, because senses that are defined as separate often overlap in practical use.

Finally, given how many compound verbs are necessary for coverage of 90 % and beyond, it is clear that a for a more comprehensive learning resource it would be necessary to employ organizational principles other than only ordering by frequency or omitting less common senses. The more detailed categorization of lexical compound verbs proposed by Kageyama (2013) or the organization based on possible combination with specific V2 verbs attempted by Himeno (1999) are promising, as they could be used to find productive patterns with relatively transparent composition of meanings outside syntactic compound verbs.

Bibliography

- Eri Banno, Yutaka Ohno, Yoko Sakane, Chikako Shinagawa, and Kyoko Tokashiki. *Genki : an integrated course in elementary Japanese*. Japan Times, 1st edition, 2000.
- Mélodie Garnier and Norbert Schmitt. The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19(6):645–666, 2015.
- Masako Himeno. *Fukugōdōshi no kōzō to imiyōhō*. Hitsuji Shobō, 1999.
- Iwanami shoten. *Kōjien [Software]*. 6th edition, distributed with a Casio electronic dictionary, 2011.
- Tarō Kageyama. *Bunpō to Gokeisei*. Hitsuji Shobō, 1993.
- Tarō Kageyama. Goitekifukugōdōshi noshintaikei. In Tarō Kageyama, editor, *Fukugōdōshikenkjū no saisentan : nazo no kaimei ni mukete*, chapter 1, pages 3–46. Hitsuji Shobō, 2013.
- Petra Kanasugi, Sachiko Kurihara, David Labus, and Mamoru Morita. *Japonská slovesa v příkladech*. Karolinum, 2015.
- Yasuo Kitahara and Taishukan. *Meikyō kokugo jiten [Software]*. Taishūkan shoten, 2nd edition, distributed with a Casio electronic dictionary, 2011.
- Akira Miura and Naomi Hanaoka McGloin. *An integrated approach to intermediate Japanese*. Japan Times, revised edition, 2008.
- NINJAL (National Institute for Japanese Language and Linguistics). *Gendai nihongo kakikotoba kinkō kōpasu (BCCWJ)*, 2009. URL http://pj.ninjal.ac.jp/corpus_center/bccwj/.
- NINJAL (National Institute for Japanese Language and Linguistics). *The Compound Verb Lexicon [Software]*, 2015. URL <http://vlexicon.ninjal.ac.jp>.
- Adam Nohejl. Gendai nihongo no kōpasu : fukugōdōshi no kenkyū ni mukete [research report]. In *2015–2016 nendo nihongo nihonbunka kenshūsei repōto shū*, pages 283–300. Nagoya daigaku kokusai gengo sentā, 2016.

Sanseido Co., Ltd. *Sūpā daijirin / Super Daijirin Japanese Dictionary [Software]*. Distributed with Apple Mac OS X, 2013.

Norbert Schmitt. *Researching Vocabulary: A Vocabulary Research Manual*. Research and Practice in Applied Linguistics. Palgrave Macmillan UK, 2010.

Shogakukan Inc. *Daijisen [Software]*. Distributed with Apple Mac OS X, 2010.

Electronic dictionaries and databases that do not state their authors, are listed with the publisher instead of author. (In the case of *The Compound Verb Lexicon*, whose authors are Taro Kageyama, Kyoko Kanzaki and Shiro Akasegawa, such citation format is requested by its terms of use.)

Both Japanese and European names are listed as given names followed by surnames. I have used the Revised Hepburn transcription for transcribing Japanese characters, or respected any transcription that was already used in the publications.

Appendix

List of the most frequent compound verbs and their senses

The following list is organized according to Section 3.4. Senses are numbered sequentially within each entry. If a sense has a frequency equal to that of the preceding sense, its number is marked with a prime, e.g. senses 3 and 4 in the following list: **1** (30 %) **2** (20 %) **3'** (20 %) **4'** (20 %) **5** (10 %).

Where senses are specific to a certain form of a compound verb, the sense starts with one of following abbreviations followed by the respective form or phrase:

adv. A form of the compound verb used as (converted to) an adverb.

attr. A form of the compound verb used as an attribute.

phr. A phrase idiom containing a form of the compound verb.

n. A form of the compound verb used as (converted to) a noun.

Each of the example sentence is marked with a number in parentheses referring to its source work in Aozora bunko. The list of references (author, work, first publication year, and URL) follows on page xvi of the Appendix.

When example sentences are abbreviated or otherwise edited, the changes are marked in square brackets.

1. 思い出す omoidasu

Frequency: 1.64 % **Formation:** 思う *omou* + 出す *dasu*

Coverage: 1.64 % **Type:** lexical (Vs)

Selected senses:

1 remember, recall (100 %)

- 私は何となく高等学校を思い出した。(2)
For some reason, I **remembered** high school.
(Somehow it **reminded** me of high school.)
- それで思い出したのが、ロンドンで毎日食べていた、フランス風のサラダである。(55)
That **reminded** me of something, the French-style salads we used to eat every day in London.
- [...] 三十年も昔に言われた先生の言葉を、しみじみと思い出している [...] (56)
Even though thirty years have already passed, I still have a keen **memory** of the words the teacher said to me.

2. 出掛ける dekakeru

Frequency: 1.23 % **Formation:** 出る *deru* + 掛ける *kakeru*

Coverage: 1.23 % **Type:** lexical (Vs)

Selected senses:

1 go out (100 %)

- 「恐竜も散歩に行くんですかい」「散歩じゃない。朝になれば食物をさがしに出かけるだろう」(8)
“Dinosaurs also go out for a stroll?” “Not for a stroll. I guess that in the morning they **go out** to search for food.”
- 帆村もメモをしまつて、出掛ける用意をした。(10)
Homura too put the notes away and prepared to **go out**.
- こんな日ならば気軽に出かける気持ちになるであろう、出かけさえすればあとは何とかなるであろう、と思ったのである。(48)
On a day like this, he could feel like **going out** leisurely, and if we **go out** together, something could happen, I thought.

3. 立ち上がる tachiagaru

Frequency: 1.00 % **Formation:** 立つ *tatsu* + 上がる *agaru*

Coverage: 0.91 % **Type:** lexical (Vs)

Selected senses:

1 stand up (74 %)

- 幹男はその灰皿に煙草を捨てると、急にむっくり立ち上って、出て行こうとした。(5)
Mikio dropped his cigarette into the ashtray, suddenly **stood up** and was going to leave.
- 立ち上ると、じゃ明日また……と、雨の中へ風のように出て行った。(6)
She **got up**, said “See you tomorrow...”, and dashed into the rain.

2 recover (11 %)

- [...] 災害の中から「立ち上ろうとする大阪」 [...] (4)
“Ōsaka trying to recover” from the disaster

3 start up, boot (about a computer) (7 %)

- [...] パソピア 16 は、MS-DOS マシンとして立ち上がり、GWベーシックもフロッピーディスクから読み込む形をとった。(7)
Pasopia16 was a type of computer that booted as an MS-DOS machine, and loaded GW Basic from a floppy disk.

Omitted senses:

4 stand, rise (about a building) (4 %)

5' take action, rebel (4 %)

Notes: In the corpus, this verb was most often spelled as 立ち上る or 起ち上る (note the different okurigana), which ought to be the standard spelling of a different verb, *tachinoboru*. Such occurrences were consequently tagged as *tachinoboru* during the analysis. The results were recomputed manually based on the analysis of the samples.

4. 繰り返す *kurikaesu*

Frequency: 0.89 % **Formation:** 繰る *kuru* + 返す *kaesu*

Coverage: 0.89 % **Type:** lexical (V)

Selected senses:

1 repeat (64 %)

- [...] 京都弁そのものが変化に乏しく、奥行きが浅く、ただ紋切型をくりかえしているだけにすぎないのではあるまいか。(3)
Isn't it that the Tokyo dialect itself, poor in variation and shallow, amounts only to **repeating** stereotyped phrases?

2 *adv.* 繰り返して *kurikaeshite*, again, several times (20 %)

- 京野等志は、この詩のようなものを、三度繰り返して読んだ。(39)
Kiyōno Hitoshi read this thing resembling poetry three times **in a row**.

3 *adv.* 繰り返し繰り返し *kurikaeshikurikaeshi*, again and again (16 %)

- そして宗匠から伝えられてくる手前を繰り返し繰り返し復習してから、控帳へ書き留めをする。(28)
Then, after reviewing the rituals passed on from masters **over and over**, he made a record to his notebook.

5. 見詰める *mitsumeru*

Frequency: 0.87 % **Formation:** 見る *miru* + 詰める *tsumeru*

Coverage: 0.80 % **Type:** lexical (Vs)

Selected senses:

1 stare, gaze (92 %)

- 豹吉は、カッと拳銃の先を見つめながら、「射て！」と、言った。(6)
Staring at the muzzle of the gun with his eyes wide open, Hyōkichi said “Fire!”
- 彼女は私の顔を見つめた。(21)
She gazed at my face.
- みんなが私の口を開くのを待って、じっと私をみつめている。(27)
Everyone was waiting for me to open my mouth, staring at me fixedly.

Omitted senses:

2 examine (8 %)

6. 見付ける mitsukeru

Frequency: 0.76 % Formation: 見る *miru* + 付ける *tsukeru*

Coverage: 0.76 % Type: lexical (V)

Selected senses:

1 find, spot (100 %)

- 首スジまで赤くなるタチであった。少年は目ざとくそれを見つけ[た。] (32)
He tended to flush to the nape of his neck. The boy was quick to spot this.
- 彼等が山でマイタケを見つけた [...] (53)
They found *maitake* mushrooms in the mountains.
- 私もきっかけを見つけて話に加わった。(58)
I spotted an opportunity, and joined the conversation too.

7. 受け取る uketoru

Frequency: 0.73 % Formation: 受ける *ukeru* + 取る *toru*

Coverage: 0.73 % Type: lexical (VV)

Selected senses:

1 receive (a thing, information) (79 %)

- 私達祖国を愛する者は、この戦争の結果を悲しい心で受取った。(17)
We, who love our homeland, received the outcome of the war with a heavy heart.
- 照子はハンカチを差出しました。木原はそれを受取って、ポケットに納めました。(26)
Teruko held out a handkerchief. Kihara took it from her and put it into his pocket.

2 interpret (21 %)

- [...] さてどんなかたちをして見せれば、火星人们はそれを敬礼だと受取ってくれるだろうかと思いなやんだ。(9)
Now, how should we do it so that the Martians take it as a salute, I worried.
- この教のみが真実の教である。[...] しかもこの絶対的真理の開示は我々において歴史的なものとして受取られなければならぬ。(15)
Only this sutra is a sutra of truth. Moreover, we must interpret the revelation of these absolute truths as a historical fact.

8. 言い出す *iidasu*

Frequency: 0.60 % **Formation:** 言う *iu* + 出す *dasu*
Coverage: 0.50 % **Type:** lexical (Vs) and syntactic

Selected senses:

1 utter, say (out loud), say unexpectedly, be first to say, bring up (52 %)

- 痔の方はどうやらなおったが、しばらくすると、お静はまた気分が悪いと言い出した。(58)
The haemorrhoids were cured somehow, but soon Oshizuka **said** she doesn't feel well again.

2 suggest, propose (30 %)

- もっとも、碁を打とうといいだしたのは升田だった。(38)
But it was Masuda, who suggested to play *go*.

Syntactic sense:

3 begin to say (17 %)

- 「[...] 本当のねうちがあるって云えやしないでしょう?……」雑誌によせていったが、それをいい出す伸子の心のうちでは、自分の書く小説のことであ[った]。(16)
"But it's difficult to say it has a genuine value, right?..." She was contributing to a magazine, but Nobuko, who **began to say that**, felt in her heart it was about the novel she was writing.

9. 有り得る

Frequency: 0.58 % **Formation:** 有る + 得る
Coverage: 0.00 % **Type:** syntactic

Syntactic sense:

1 possible, likely (often in negative: impossible, cannot be true) (100 %)

10. 取り上げる *toriageru*

Frequency: 0.57 % **Formation:** 取る *toru* + 上げる *ageru*
Coverage: 0.52 % **Type:** lexical (VV)

Selected senses:

1 treat, deal with (a topic) (70 %)

- ベルリンには、ヘーゲル哲学の進歩的な面をとりあげて、その弁証法的方法を発展させようとする若い哲学者の一群があった。(18)
In Berlin, there is a group of young philosophers who want to **explore** the progressive aspect of Hegel's philosophy and develop its dialectical method.

2 pick up; take (from someone) (0993, 1112) (22 %)

- チョット書く気が起きても、さてペンを取りあげるとどう書いてよいかわからなくな [つた]。(44)

He felt like writing a bit, but when he **picked up** a pen, he suddenly didn't know what he should write.

Omitted senses:

3 collect (taxes), confiscate (4 %)

4' accept, adopt (a proposal, ...) (4 %)

11. ～し得る

Frequency: 0.57 % **Formation:** ～する *suru* + 得る *eru*

Coverage: 0.00 % **Type:** syntactic

Syntactic sense:

1 can do ...; be able to do ... (100 %)

12. 引っ張る *hipparu*

Frequency: 0.54 % **Formation:** 引く *hiku* + 張る *haru* (*sokuonbin*)

Coverage: 0.35 % **Type:** lexical (VV)

Selected senses:

1 take, haul, cajole (someone); lead (people, a field, ...) (52 %)

- 君と逢ったらすぐに、ものも言わずに、その吉祥寺のスタンドに引っばって行くつもりでいたのだが、しかし、君の汽車は、ずいぶん遅れた。(1)

I was going to **take** you to the bar in *Kichijōji* straight away after meeting with you, without telling you anything, but your train came really late.

- 戦争中のことだから、生きのいいのは大てい兵隊とか工場に引っばられている。(61)

It was during the war, so those who were in good health were usually **taken to (forced to go)** to the army or to factories.

2 prolong, drag out; prolong a sound, drawl (22 %)

- [...] 明治以来の保守的な日本の支配権力は、この委員会の仕事を、蝸牛の這うようなテンポで引っばった。(17)

The ruling powers of Japan from the Meiji period on, **dragged out** the work of this committee, so it proceeded at a snail's pace.

- [ウグイスの] 鳴きはじめの「ホー」のひっぱり方にも、「ケキョ」の早さにも、いろいろ特徴のあることなど教えて貰った。(41)

I've been told about many peculiarities of the bush warbler's call, about the way it **prolongs** the initial *hō* sound, or about the speed of the *kekyo* sound.

Omitted senses:

3 pull, stretch (13 %)

4' drag (13 %)

13. 飛び出す *tobidasu*

Frequency: 0.53 % **Formation:** 飛ぶ *tobu* + 出す *dasu*

Coverage: 0.48 % **Type:** lexical (Vs)

Selected senses:

1 spring out; rush out (48 %)

- ドロボー君は飯も食わずに大急ぎでとびだした。(31)
Without even eating his meal, the burglar **rushed out** in great hurry.

2 appear unexpectedly (22 %)

- それか時に言葉の調子の中にも飛び出して来る。(46)
At times **it was apparent** even from the tone of his speech.

3' run away (22 %)

- ふん、飛び出して、どこへ行って、どうすんの？(47)
Ha, and once you **run away**, where' you gonna go, whatcha' gonna do?

Omitted senses:

4 start to fly, take off (4 %)

5' protrude (4 %)

14. 取り出す *toridasu*

Frequency: 0.52 % **Formation:** 取る *toru* + 出す *dasu*

Coverage: 0.47 % **Type:** lexical (VV)

Selected senses:

1 take out (91 %)

- そして、男はべつの新しい煙草を取り出して、火をつけた。(6)
Then, the man **took out** another cigarette and lit it up.
- 菊千代は笑って、戸棚からウイスキーの瓶を取り出しました。(24)
Kikuchiyo laughed and **took** a whisky bottle **out** of a cupboard.

Omitted senses:

2 pick out, select (9 %)

15. 見出す *miidasu*

Frequency: 0.49 % **Formation:** 見る *miru* + 出す *idasu*

Coverage: 0.49 % **Type:** lexical (VV)

Selected senses:

1 find, discover (usually abstract sense) (100 %)

- あんなに落ちつきのない遊びに私はタノシミを見出すことはできないのである。(38)
I cannot **find** pleasure in such a restless pastime.
- 「真実を表現するためには、真実を見出すことが必要ではないでしょうか」(50)
In order to express the truth, isn't it necessary to **find** the truth?
- その意味で、日本語の中にアイヌ語の要素を見いだすことになんのふしぎもないわけである。(52)
In that sense, it is no wonder we **find** elements of Ainu in Japanese.

16. ～し始める

Frequency: 0.47 % Formation: ～する *suru* + 始める

Coverage: 0.00 % Type: syntactic

Syntactic sense:

1 begin to do ..., start to do ... (100 %)

17. 飛び込む *tobikomu*

Frequency: 0.46 % Formation: 飛ぶ *tobu* + 込む *komu*

Coverage: 0.38 % Type: lexical (VV)

Selected senses:

1 fly, dive, jump into; throw oneself (48 %)

- 「おや。へんなものがあるぞ」「あっ、そうだ。窓から飛びこんできたんだ (13)
“Look! There’s something weird.” “Oh, you’re right. It came **flying in** through the window.”
- [彼らは] すぐ、あそこに見える横穴にとびこむんだ。(45)
At once, they **jumped** into the cave over there.

2 rush in, burst in (35 %)

- 何かいざこざが起ったりすると、[...] 床屋だの銭湯に飛び込んだ。(20)
If you run into any trouble, just **rush** into a barbershop or a public bath.

Omitted senses:

3 suddenly appear, come (news, information), catch one’s eye (13 %)

4 plunge into, commit oneself to (4 %)

18. 落ち着く *ochitsuku*

Frequency: 0.43 % Formation: 落ちる *ochiru* + 着く *tsuku*

Coverage: 0.30 % Type: lexical (V)

Selected senses:

- 1 calm down (about conditions, feelings), settle down (about people) (70 %)
 - ヨーコとの不安定な関係も、結婚という手続きを踏むことで落ちつくのではないか。(7)
Maybe, the unstable relationship with Yōko too will **settle down** after going through the formalities of marriage.
 - そして火星人が少しおちついたところを見計って、外交交渉を始めるんだね。(9)
After that, we will wait until the Martians **calm down** a little, and start diplomatic negotiations.

Omitted senses:

- 2 *attr.* 落ち着いた *ochitsuita*, calm (atmosphere), sober, tasteful (appearance) (13 %)
- 3' settle in, make one's home (13 %)
- 4 come to a conclusion (4 %)

19. 酔っ払う *yopparau*

Frequency: 0.43 % **Formation:** 酔う *you* + 払う *harau* (sokuonbin)
Coverage: 0.36 % **Type:** lexical (Vs)

Selected senses:

- 1 get drunk (84 %)
 - 酔っ払えば [人間は] 本音をはく。(37)
When people **get drunk**, they say what they really think.
 - 「君はずいぶん酔っばらってるね」(60)
“You’re awfully **drunk**, aren’t you?”

Omitted senses:

- 2 *n.* 酔っ払い *yopparai* a drunk (16 %)

20. 見上げる *miageru*

Frequency: 0.41 % **Formation:** 見る *miru* + 上げる *ageru*
Coverage: 0.36 % **Type:** lexical (Vs)

Selected senses:

- 1 look up (87 %)
 - 「そうですか………」とおだやかな調子で眼を伏せてしばらく何か考えていたが、今度はジロリと見上げて「まさか、お宅にいるんじゃないでしょうね？」(44)
“Is that so...?” she said in a quite voice, lowered her eyes, thinking about something for a while, and then she **looked up** sharply and said: “Don’t tell me she is staying at your home!”

Omitted senses:

- 2 admirable (“見上げた”) (13 %)

21. 見回す mimawasu

Frequency: 0.40 % Formation: 見る *miru* + 回す *mawasu*

Coverage: 0.40 % Type: lexical (Vs)

Selected senses:

1 look around [a place], take a glance around (100 %)

- 彼女はそっとあたりを見廻すと、素早く^{カン}罐の中へ手を突っ込んだ。(6)
She **threw a stealthy glance around** and quickly thrust her hand into the can.
- 三百六十度、どこを見まわしても海と空と積乱雲の群像ばかりで、船影はおろか、島影一つ見えない。(8)
He **looked 360 degrees around**, but wherever he looked at, it was just the sea, the sky and the thunderclouds; an outline of an island, let alone one of a boat, was nowhere to be seen.

22. 歩き出す

Frequency: 0.36 % Formation: 歩く + 出す

Coverage: 0.00 % Type: syntactic

Syntactic sense:

1 start walking (100 %)

23. 逃げ出す nigedasu

Frequency: 0.36 % Formation: 逃げる *nigeru* + 出す *dasu*

Coverage: 0.14 % Type: lexical (Vs) and syntactic

Selected senses:

1 run away, escape (39 %)

- 「すると、その窓を明けて、誰か外へ逃げだしたんだな」(13)
“In that case, someone probably opened that window to **escape** outside.”
- あのように厳しく、そこから逃げ出せば法律で以て罰せられ[る]。(19)
You will be punished this severely by the law, if you **run away** from there.

Syntactic sense:

2 start to run (61 %)

24. 振り返る furikaeru

Frequency: 0.33 % Formation: 振る *furu* + 返る *kaeru*

Coverage: 0.33 % Type: lexical (V)

Selected senses:

1 turn around, look back (65 %)

- そして、その間にも、ときどきうしろをふりかえって、このガラス廊下の入り口の方を気にしていた。(12)

Then, he sometimes **looked back** again, anxiously watching the glass entrance to the corridor.

2 look back (on past); reflect (on) (35 %)

- ここでしばらくこの時から十年あまり前のことをふりかえってみる。(51)

Now I will try to **reflect** for a while on what happened a little over ten years ago.

25. 立ち止まる tachidomaru

Frequency: 0.33 % **Formation:** 立つ *tatsu* + 止まる *domaru*

Coverage: 0.33 % **Type:** lexical (VV)

Selected senses:

1 stop (walking); stand still (100 %)

- 彼はその路地の入口に立ち止って、彼女へ野菜の袋を渡しました。(25)

He **stopped** at the entrance to the alley and handed her a bag with vegetables.

26. 立ち去る tachisaru

Frequency: 0.33 % **Formation:** 立つ *tatsu* + 去る *saru*

Coverage: 0.33 % **Type:** lexical (VV)

Selected senses:

1 leave, go away (100 %)

- そして、ふりむいて立去った。(29)

Then she turned around and **left**.

27. 打ち明ける uchiakeru

Frequency: 0.32 % **Formation:** 打つ *utsu* + 開ける *akeru*

Coverage: 0.32 % **Type:** lexical (pV)

Selected senses:

1 confess (100 %)

- 「怖しくて誰にも打ちあける勇気がありませんでしたが、はじめてあなたに打ち開け[た]。」(36)

I was afraid and didn't have the courage to **confess** to anyone, for the first time I **confessed** to you.

28. 思い切る omoikiru

Frequency: 0.31 % Formation: 思う *omou* + 切る *kiru*

Coverage: 0.24 % Type: lexical (Vs)

Selected senses:

1 *adv.* 思い切って *omoikitte*, dare to ...; pluck up one's courage and ... (77 %)

- もう、思いきって、言ってしまうほかに手がなくなったんだ。(34)
Now you have no other option left than to **pluck up your courage and** say it.

Omitted senses:

2 *attr.* 思い切った *omoikitta*, radical, bold (23 %)

29. ~し切る

Frequency: 0.31 % Formation: ~する *suru* + 切る *kiru*

Coverage: 0.00 % Type: syntactic

Syntactic sense:

1 do ... completely (100 %)

30. 振り向く furimuku

Frequency: 0.31 % Formation: 振る *furu* + 向く *muku*

Coverage: 0.27 % Type: lexical (VV)

Selected senses:

1 turn one's face, to look over one's shoulder (86 %)

- 私は眉をしかめて振り向いた。(22)
I **turned my face** to him with a frown.
- 座敷の入り口から、クルリとふりむいて、お風呂へ行ってしまった。(33)
He **did an about-turn** from the sitting room's entrance and went to the bathroom.

Omitted senses:

2 *phr.* 振り向きもしない *furimuki mo shinai* not pay attention (14 %)

31. 陥る ochiiru

Frequency: 0.30 % Formation: 落ちる *ochiru* + 入る *iru*

Coverage: 0.29 % Type: lexical (V)

Selected senses:

1 fall, lapse into a bad or difficult condition (95 %)

- 私の書くものは次第にマンネリズムに陥って、精彩を欠くようになった。(21)
Whatever I tried to write, it gradually **fell** into mannerism and ended up poorly.
- [...] 潮田さんが昏睡に陥って居るんです。(57)
Mr Ushioda has **lapsed** into a coma.

Omitted senses:

2 fall (to enemy, into enemy's hands), become captured (5 %)

32. 押し付ける oshitsukeru

Frequency: 0.30 % **Formation:** 押す *osu* + 付ける *tsukeru*

Coverage: 0.30 % **Type:** lexical (VV)

Selected senses:

1 press (one thing to another) (55 %)

- 五郎は内ポケットから金を取り出した。百円玉を少年に渡した。「あそこに茶店があるだろう。ジュースを二本買って来て呉れ、咽喉が乾いた」少年はちょっとためらったが、五郎は無理に掌てに押しつけた。(60)
Gorō took money from his inner pocket. He gave the boy a hundred yen banknote.
“There’s a tea shop over there. I’m thirsty, buy me two bottles of juice.” The boy hesitated a little, but Gorō **pressed** the banknote into his palm, forcing him to take it.

2 force, compel (to accept, take responsibility, buy something) (45 %)

- 勝手に責任を押しつけられてはお巡りさんも堪らない [...] (35)
Even policemen hate when you force them to take responsibility as it suits you.

33. 見掛ける mikakeru

Frequency: 0.30 % **Formation:** 見る *miru* + 掛ける *kakeru*

Coverage: 0.30 % **Type:** lexical (Vs)

Selected senses:

1 (happen to) see; catch sight of (100 %)

- しかしその四方木田鶴子さんの姿を今年になってから突然見掛けたのでびっくりしていました。(11)
I was surprised when I suddenly **saw** that Ms Yomogita Zuko early this year.
- 雪子さんの事件以来、二度と姿を見かけないわね」(13)
After the incident with Yukiko, no one **saw** him again.

34. 引き摺る hikizuru

Frequency: 0.29 % **Formation:** 引く *hiku* + 刷る *suru* (rendaku)

Coverage: 0.15 % **Type:** lexical (VV)

Selected senses:

1 drag (rubbing on the ground), trail (50 %)

- 椅子をひきずってきて腰かけ [た。] (42)
She **dragged** a chair over, and sat on it.
- ふたりは、重たい足を引き摺って、しばらく歩いていた [...]。(59)
The two of them walked some more, **dragging** their heavy feet.

Omitted senses:

- 2 influence, control (often in passive) (27 %)
- 3 be unable to forget; be haunted (by one's past); cannot get over; carry a heritage (14 %)
- 4 take someone by force (5 %)
- 5' prolong, drag out (5 %)

35. 持ち出す *mochidasu*

Frequency: 0.29 % **Formation:** 持つ *motsu* + 出す *dasu*
Coverage: 0.29 % **Type:** lexical (VV)

Selected senses:

- 1 get out; carry out, take out, sneak out; embezzle; rescue (from fire) (55 %)
 - 「外へ持ちだして焼け残ったものを、盗まれたのではないのでしょうか」 (30)
“Don't tell me someone has stolen what we **carried out** from fire!”
- 2 bring up (a subject, sometimes personal, not belonging to the discussion) (45 %)
 - 子供のころ、まるで理由なしになぐられたり、どなられたりした話を、いくつでも持ち出して、反駁するばかりであった。 (48)
I **brought up** several stories about how, when I was a child, I was being beaten and yelled at without any reason, only to hear denial.

36. 追い掛ける・追っ掛ける *oikakeru/okkakeru*

Frequency: 0.29 % **Formation:** 追う *ou* + 掛ける *kakeru* (opt. *sokuonbin*)
Coverage: 0.24 % **Type:** lexical (Vs)

Selected senses:

- 1 run after, chase; pursue (a goal, idea) (82 %)
 - 君を追いかけて表門の方へ行った連中もいるから、すぐに帰ると、また、あぶない。 (43)
There's another group **running after** you, which went to the front gate; they're going to be back soon, and then it's going to be dangerous again.
 - [...] あの問題を絶えず追いかけている [...] (49)
I am incessantly **pursuing** that problem.

Omitted senses:

- 2 *adv.* 追い掛けて *oikakete*, 追い掛けるように *oikakeruyōni*, one after another (18 %)

Notes: Both variants (*oikakeru* and *okkakeru*) were equally common (50 %) in the sample. *Okkakeru* is more informal.

37. 思い付く omoitsuku

Frequency: 0.29 % **Formation:** 思う *omou* + 付く *tsuku*

Coverage: 0.29 % **Type:** lexical (Vs)

Selected senses:

1 think of, hit upon (100 %)

- [...] サムとぼくとか、すばらしい計画を思いついたからだ。(14)
It's because Sam and I **hit upon** a terrific plan.
- ほんの思いついた例にすぎません。(40)
It is just an example I **thought of**.

List of source references

- (1) 太宰治: 未帰還の友に (1946)
<http://www.aozora.gr.jp/cards/000035/card297.html>
- (2) 織田作之助: 髪 (1945)
<http://www.aozora.gr.jp/cards/000040/card46301.html>
- (3) 織田作之助: 大阪の可能性 (1947)
<http://www.aozora.gr.jp/cards/000040/card46351.html>
- (4) 織田作之助: 起ち上る大阪: 戦災余話 (1945)
<http://www.aozora.gr.jp/cards/000040/card46353.html>
- (5) 織田作之助: それでも私は行く (1946)
<http://www.aozora.gr.jp/cards/000040/card47287.html>
- (6) 織田作之助: 夜光虫 (1947)
<http://www.aozora.gr.jp/cards/000040/card47288.html>
- (7) 富田倫生: パソコン創世記 (1985)
<http://www.aozora.gr.jp/cards/000055/card365.html>
- (8) 海野十三: 恐竜島 (1948)
<http://www.aozora.gr.jp/cards/000160/card2718.html>
- (9) 海野十三: 火星探険 (1945)
<http://www.aozora.gr.jp/cards/000160/card3049.html>
- (10) 海野十三: 地獄の使者 (1947)
<http://www.aozora.gr.jp/cards/000160/card3052.html>
- (11) 海野十三: 千早館の迷路 (1947)
<http://www.aozora.gr.jp/cards/000160/card3053.html>
- (12) 海野十三: 時計屋敷の秘密 (1948)
<http://www.aozora.gr.jp/cards/000160/card3054.html>
- (13) 海野十三: 四次元漂流 (1946)
<http://www.aozora.gr.jp/cards/000160/card3358.html>
- (14) 海野十三: 恐竜艇の冒険 (1948)
<http://www.aozora.gr.jp/cards/000160/card874.html>
- (15) 三木清: 親鸞 (1946)
<http://www.aozora.gr.jp/cards/000218/card46946.html>
- (16) 宮本百合子: 二つの庭 (1947)
<http://www.aozora.gr.jp/cards/000311/card2011.html>
- (17) 宮本百合子: 私たちの建設 (1946)
<http://www.aozora.gr.jp/cards/000311/card3254.html>
- (18) 宮本百合子: カール・マルクスとその夫人 (1947)
<http://www.aozora.gr.jp/cards/000311/card3265.html>
- (19) 宮本百合子: 青年の生きる道 (1946)
<http://www.aozora.gr.jp/cards/000311/card3438.html>
- (20) 宮本百合子: 栄蔵の死 (1981)
<http://www.aozora.gr.jp/cards/000311/card7908.html>
- (21) 豊島与志雄: 憑きもの (1949)
<http://www.aozora.gr.jp/cards/000906/card42649.html>
- (22) 豊島与志雄: 悲しい誤解 (1949)
<http://www.aozora.gr.jp/cards/000906/card42656.html>

- (23) 豊島与志雄: 窓にさす影 (1952)
<http://www.aozora.gr.jp/cards/000906/card42679.html>
- (24) 豊島与志雄: 高尾ざんげ: 近代説話 (1946)
<http://www.aozora.gr.jp/cards/000906/card42728.html>
- (25) 豊島与志雄: 白蛾: 近代説話 (1946)
<http://www.aozora.gr.jp/cards/000906/card42730.html>
- (26) 豊島与志雄: 道標: 近代説話 (1947)
<http://www.aozora.gr.jp/cards/000906/card42734.html>
- (27) 永井隆: 長崎の鐘 (1949)
<http://www.aozora.gr.jp/cards/000924/card50659.html>
- (28) 蒲原有明: 夢は呼び交す: 黙子覚書 (1946)
<http://www.aozora.gr.jp/cards/001055/card43739.html>
- (29) 坂口安吾: 夜長姫と耳男 (1952)
<http://www.aozora.gr.jp/cards/001095/card42614.html>
- (30) 坂口安吾: アンゴウ (1948)
<http://www.aozora.gr.jp/cards/001095/card42837.html>
- (31) 坂口安吾: 吝嗇神の宿: 人生オペラ 第二回 (1953)
<http://www.aozora.gr.jp/cards/001095/card42953.html>
- (32) 坂口安吾: 発掘した美女 (1953)
<http://www.aozora.gr.jp/cards/001095/card42958.html>
- (33) 坂口安吾: 呉清源 (1948)
<http://www.aozora.gr.jp/cards/001095/card43144.html>
- (34) 坂口安吾: 行雲流水 (1949)
<http://www.aozora.gr.jp/cards/001095/card43167.html>
- (35) 坂口安吾: 安吾巷談: 06 東京ジャングル探検 (1950)
<http://www.aozora.gr.jp/cards/001095/card43177.html>
- (36) 坂口安吾: 明治開化 安吾捕物: 11 その十 冷笑鬼 (1951)
<http://www.aozora.gr.jp/cards/001095/card43213.html>
- (37) 坂口安吾: 安吾人生案内: 03 その三 精神病診断書 (1951)
<http://www.aozora.gr.jp/cards/001095/card45918.html>
- (38) 坂口安吾: 明日は天気になれ (1953)
<http://www.aozora.gr.jp/cards/001095/card45935.html>
- (39) 岸田国土: 光は影を (1950)
<http://www.aozora.gr.jp/cards/001154/card43859.html>
- (40) 岸田国土: あるニュー・フェイスへの手紙 (1951)
<http://www.aozora.gr.jp/cards/001154/card44833.html>
- (41) 宮城道雄: 春雨 (1949)
<http://www.aozora.gr.jp/cards/001288/card47115.html>
- (42) 徳永直: 白い道 (1948)
<http://www.aozora.gr.jp/cards/001308/card49290.html>
- (43) 三好十郎: その人を知らず (1948)
<http://www.aozora.gr.jp/cards/001311/card49746.html>
- (44) 三好十郎: 肌の匂い (1949)
<http://www.aozora.gr.jp/cards/001311/card49747.html>
- (45) 三好十郎: 樹氷 (1955)
<http://www.aozora.gr.jp/cards/001311/card49855.html>
- (46) 三好十郎: 夜の道づれ (1950)

- <http://www.aozora.gr.jp/cards/001311/card50051.html>
(47) 三好十郎: 胎内 (1949)
- <http://www.aozora.gr.jp/cards/001311/card50080.html>
(48) 和辻哲郎: 漱石の人物 (1950)
- <http://www.aozora.gr.jp/cards/001395/card49898.html>
(49) 和辻哲郎: 露伴先生の思い出 (1947)
- <http://www.aozora.gr.jp/cards/001395/card49914.html>
(50) 北大路魯山人: 梅にうぐいす (1953)
- <http://www.aozora.gr.jp/cards/001403/card49960.html>
(51) 津田左右吉: 学究生活五十年 (1951)
- <http://www.aozora.gr.jp/cards/001535/card53725.html>
(52) 知里真志保: 日本語とアイヌ語の関係: マタギという言葉の存在について (1961)
- <http://www.aozora.gr.jp/cards/001540/card53894.html>
(53) 知里真志保: アイヌ宗教成立の史的背景 (1955)
- <http://www.aozora.gr.jp/cards/001540/card53899.html>
(54) 柳田国男: 海上の道 (1961)
- <http://www.aozora.gr.jp/cards/001566/card54331.html>
(55) 中谷宇吉郎: サラダの謎 (1960)
- <http://www.aozora.gr.jp/cards/001569/card53204.html>
(56) 中谷宇吉郎: 線香の火 (1955)
- <http://www.aozora.gr.jp/cards/001569/card53222.html>
(57) 野村胡堂: 笑う悪魔 (1948)
- <http://www.aozora.gr.jp/cards/001670/card56721.html>
(58) 井上貞治郎 日本経済新聞社: 私の履歴書: 放浪の末、段ボールを思いつく (1959)
- <http://www.aozora.gr.jp/cards/001685/card54902.html>
(59) 山之口貊: 野宿 (1950)
- <http://www.aozora.gr.jp/cards/001693/card55180.html>
(60) 梅崎春生: 幻化 (1965)
- <http://www.aozora.gr.jp/cards/001798/card56634.html>
(61) 梅崎春生: 魚の餌 (1953)
- <http://www.aozora.gr.jp/cards/001798/card57172.html>