Subjective frequency norms for 100 Japanese verb-verb compounds: The first step toward constructing a behavioral database for Japanese compound words

Yijia Yao（Graduate school of Languages and Linguistics Sophia University）

Abstract

Lexical frequency has been commonly used in previous studies to investigate how compound words are processed in the mental lexicon. Since the work of Balota et al., (2007), a growing number of databases have been constructed using normative and behavioral data of compound words (for English, see Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007; for French, see Ferrand, New, Brysbaert, Keuleers, Bonin, Méot, Augustinova & Pallier, 2010; for Chinese, see Sun, Hendrix, Ma, & Baayen, 2018; for Dutch, see Keuleers, Diependaele, & Brysbaert, 2010), which make it possible to investigate and compare the mechanism of compound processing across languages. However, despite the fact that compound words are prevalently used in Japanese, to the best of our knowledge, such database have not previously been constructed for Japanese compound words. This study reported two subjective ratings and six corpus-based objective frequency estimates for 100 Japanese verb-verb compounds: familiarity, age-of-acquisition (AoA), whole-word surface frequency, whole-word lemma frequency, first-constituent surface frequency, first-constituent lemma frequency, second-constituent surface frequency, second-constituent lemma frequency. The relationship among these frequency measures was examined by using correlational and hierarchical regression analyses. Negative correlation was obtained between the familiarity and AoA ratings, indicating that more familiar compounds tended to be rated as being of earlier-acquired words by native speakers of Japanese. Moreover, the whole-word lemma frequency was the best single predictor of both familiarity and AoA ratings, revealing that when native speakers of Japanese are asked to rate a Japanese verb-verb compound on a particular variable, the verb-verb compound is processed as a whole unit. These normative data should be beneficial to researchers who are interested in selecting stimuli for psycholinguistic experiments, and it will also help us to gradually construct a normative and behavioral database for Japanese compound words.

1. Introduction

Lexical processing is tuned to lexical frequency. Generally, high-frequency words are processed more quickly and accurately than low-frequency words. With respect to compound words, which are composed of two or more monomorphemic words (e.g., *watermelon*), the past four decades have seen a simmering debate as to whether compound words are processed via a holistic route or via an analytic route in the mental lexicon (for review, see Libben, Gagné, & Dressler, 2020), and this debate has centered largely around the question of which frequency variable is the best predictor for compound processing. Specifically, it is assumed that if the processing efficiency of a compound word varies according to the frequency of the whole-word form but not according to the frequencies of its corresponding constituents, then the compound word might be processed as a whole unit in the mental lexical (Full-listing models, Butterworth, 1983; Bybee, 1995). On the other hand, if the processing efficiency of a compound word is modulated by the constituent frequencies but not by the whole-word frequency, then the compound word might be processed as separate components in the mental lexicon

(Full-parsing models, Libben, Derwing, & de Almeida, 1999; Taft & Forster, 1976). In addition, jointly influences of the whole-word and constituent frequencies are interpreted as results of dual-route processing (for Sub-lexical model, see Rastle, Davis, & New, 2004; for Supra-lexical model, see Giraudo & Grainger, 2000; for Parallel dual-route model, see Schreuder & Baayen, 1995).

Because of its critical importance as a diagnostic factor in investigating the mechanism of compound processing, the problem of how to select the most appropriate indicator of lexical frequency for compound words have increasingly attracted the attention of psycholinguistic researchers. Previous studies have showed that different types of frequency measures may tap into different underlying constructs (for French, see Bonin, Laroche, & Méot, 2021; for English, see Juhasz, Lai, & Woodcock, 2011; for Chinese, see Su, Li, & Li, 2023). This means that using different frequency measures may yield somewhat different experimental results, and without solid validation of the measurements that are used to gather the frequency information, comparisons across studies and languages might be misleading. Therefore, to attempt to better understand the essential frequency effects on compound processing, this paper reported normative data for 100 Japanese verb-verb compounds and statistically examined the relationships among the different frequency measures. To the best of our knowledge, this work is the first normative study on Japanese verb-verb compounds.

2. Previous studies

2.1 Corpus-based measures of objective frequency

In psycholinguistic studies, the most widely used measure to estimate lexical frequency is the corpus-based frequency measure. Generally, two types of frequency estimates can be derived from a corpus: Surface frequency and lemma frequency. Surface frequency refers to the frequency of the particular word form. In the case of compound words, such as *watermelon*, three types of surface frequency can be distinguished: Whole-word surface frequency (the surface frequency of *watermelon*), first-constituent surface frequency (the surface frequency of *water*), and second-constituent surface frequency (the surface frequency of *melon*). On the other hand, lemma frequency refers to the summed frequency of a word form and its inflectional variant. For example, for a compound verb, such as *double-check*, three types of lemma frequency can be derived: Whole-word lemma frequency (the summed frequency of *double-check*, *double-checked*, *double-checking*, *double-checks*), first-constituent lemma frequency (the summed frequency of *double*, *doubled*, *doubling*, *doubles*), and second-constituent lemma frequency (the summed frequency of *check*, *checked*, *checking*, *checks*).

These corpus-based estimates serve as important indicators of compound processing in psycholinguistic studies. For example, most studies showed that the visual recognition of compound words involves access to both whole-word and constituent representations, as evidenced by an interactive effect of whole-word and constituent frequencies on visual lexical decision times (for English, see Schmidtke, Gagné, Kuperman, Spalding & Tucker, 2018; for German, see Günther, Marelli, & Bolte, 2020; for Spanish, see Güemes, Gattei, & Wainselboim, 2019). However, this is not always the case when data from Asian languages is used. It was shown that direct access to the whole-word representation of compound words is more preferable than morpheme-based access in Asian languages, such as Chinese (Myers, Huang, & Wang, 2006; Tse, Yap, Chan, Sze, Shaoul, & Lin, 2017), Japanese (姚 2020), and Vietnamese (Pham & Baayen, 2015), as evidenced by the predominant effect of whole-word frequency. In other words, there seems to exist a typological distinction in compound processing. However, before we delve into this linguistic typological puzzle, the first thing we want to know is whether surface and lemma frequency measures orient to the same aspects of mental representations of compound words. This is important because differences in measurements might

make previous results incomparable with each other.

Although consensus has not been reached yet, previous studies dealing with this ongoing problem reported two important findings: Firstly, the predictive power of surface and lemma frequency measures on lexical decision times is different. Surface frequency is demonstrated to be less powerful to predict the effect of whole-word frequency on lexical decision times, as compared to lemma frequency (Ji, Gagné, & Spalding, 2011; Juhasz, Lai, & Woodcock, 2015; Keuleers, Brysbaert, & New, 2010). For example, by collecting both constituent-based surface and lemma frequency information on the same stimuli, Ji, Gagné, & Spalding (2011) found that whereas lemma frequency is the most important variable to predict lexical decision times, no influence of surface frequency on lexical decision times was observed. In addition, Keuleers, Brysbaert, & New (2010) reported that adding compound-based lemma frequency to the stepwise multiple regression model increases the explained variance in lexical decision times by 10% points for compound words. Secondly, there is a nonlinear relationship between surface frequency and lexical decision times. The predictive power of surface frequency for words in the high (i.e., above 100 per million) and low (i.e., below 1 per million) frequency bands were weaker than those in the medium frequency band (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Kuperman & Van Dyke, 2013). Moreover, surface frequency tends to be less reliable in predicting of proficient readers' (i.e., readers have more reading experience or larger vocabulary size) task performance (Falkauskas & Kuperman, 2015; Kuperman & Van Dyke, 2013). For example, Kuperman & Van Dyke (2013) reported that proficient readers are equally faster in processing both high- and low-frequency words. These results indicate that although both surface and lemma frequencies are well-accepted diagnostic factors in previous studies, they may represent some related but distinct constructs, and they might not be reliable indicators of individual differences in language experience.

2.2 Rating-based measures of subjective frequency

In order to better understanding the effects of experience-driven individual differences on compound processing, two rating-based frequency measures are increasingly used: Familiarity and AoA. Familiarity refers to the degree of the mental impression with a word. This estimate is typically obtained by asking participants to evaluate how often they encounter or produce a given word in daily communication, by using a Likert scale. For example, Tagalakis & Keane (2006) asked their participants to judge whether they had encountered the compound word before and how frequently using a 7-point Likert scale (1 = very unfamiliar, 7 = very familiar). Momenian, Cham, Amini, Radman, & Weekes (2021) asked their participants to rate the compound word and its corresponding constituents by using a 7-point Likert scale familiarity questionnaire (1 = very unfamiliar, 7 = very familiar). AoA refers to the age at which people learn a particular word and are often obtained by asking subjects to indicate their own age at which they learned the words. For instance, Juhasz, Lai, & Woodcock (2015) used a 7-point scale with 2-year age bands to collect the age at which compound words were acquired (7 = the word was learned at age 13 or order). Although it is known that rated AoA is less reliable for the words which are learned before the age of 4 and after the age 15 (Brysbaert, 2017), compound words that were judged to be earlier-acquired were shown to be recognized and produced faster and more accurately than those judged to be later-acquired (Juhasz, 2018; Song & Li, 2021).

By collecting these two rating-based frequency estimates on the same compound word dataset, previous studies reported several interesting findings regarding their relationship: Firstly, familiarity and AoA are associated with each other, as evidenced by a negative correlation between familiarity and AoA rating scores (Bonin, Laroche, & Méot, 2021; Juhasz, Lai, & Woodcock, 2015; Su, Li, & Li, 2023). Familiar compound words are more likely to be rated as earlier-acquired compounds and vice versa. Secondly, both of them are reliable predictors of lexical decision times (Bonin, Laroche, & Méot,

2021; Juhasz, Lai, & Woodcock, 2015). Thirdly, adding familiarity and AoA into the statistical model produced a significant increase in the amount of variance explained compared to the baseline model which only includes corpus-based frequency estimates (Juhasz, Lai, & Woodcock, 2015). This means that rating-based subjective frequency measures were able to capture some important and unique features of compound words, which were not reflected by corpus-based objective frequency measures.

However, despite the increasingly widespread use of these two rating-based frequency measures in psycholinguistic studies, there are two major mysteries still remain unsolved. Firstly, whether familiarity and AoA reflect the same underlying constructs? With respect to this question, inspection of the literature reveals that, compared to AoA, familiarity is more strongly related to orthographic properties of compound words, such as number of strokes (Su, Li, & Li, 2023). On the other hand, AoA is more strongly associated with semantic properties of compound words, such as imageability and sensory experience (Bonin, Laroche, & Méot, 2021; Juhasz, Lai, & Woodcock, 2015). Secondly, how they are related to corpus-based measures? Although scarce, previous studies showed that familiarity were strongly associated with whole-word frequency rather than constituent frequencies, whereas AoA were weakly associated with both whole-word and constituent frequencies (Bonin, Laroche, & Méot, 2021; Juhasz, Lai, & Woodcock, 2015; Su, Li, & Li, 2023). These results indicate that although both familiarity and AoA have proven to be effective measures of personal linguistic experience, they may tap in some different aspects of word features.

2.3 Why Japanese verb-verb compounds?

Japanese is rich in verb-verb compounds. According to Compound Verb Lexicon (National Institute for Japanese Language and Linguistics, 2015), which is an online database for Japanese verb-verb compounds, at least 2700 verb-verb compounds are commonly used in contemporary Japanese. Besides this quantitative feature, Japanese verb-verb compounds also possess some peculiar structural features which might play special roles in processing.

Firstly, Japanese verb-verb compounds are all concatenated words. Specifically, no orthographical segmentation cues, such as spaces or hyphens, can be inserted between the two constituents of Japanese verb-verb compounds. This is distinctly different from English compound words, which can be presented either with or without a space or hyphen. In addition, unlike Greek and Polish, in which compound words are generally formed by inserting a linking vowel between their corresponding constituents, Japanese verb-verb compounds are usually formed by melting two monomorphemic words together (i.e., by eliminating the verbal ending of their corresponding first constituents). For example, to form a Japanese verb-verb compound 食べ始める "start eating", we need to transform its first constituent 食べる "eat" into the continuous form 食べ by eliminating the verbal ending る. In other words, compound verbs are orthographically far more like monomorphemic words in Japanese than in other languages.

Secondly, although Japanese verb-verb compounds are similar in orthographic structure, there are two distinct association patterns between the first and second constituents. One pattern, the so-called syntactic verb-verb compound, is highly productive because principally it allows flexible combination of all monomorphemic verbs, the other pattern, the so-called lexical verb-verb compound, is highly restricted because it is based on the transitivity of the constituents (影山 1993). In case of lexical verb-verb compounds, for example, a transitive verb 拭く "wipe" can be combined with another transitive verb 消す "to make something disappear", but can never be combined with an intransitive verb, such as 消える "disappear". In contrast, with respect to syntactic verb-verb compounds, a transitive verb 拭く "wipe" can be combined with either a transitive verb 始める "start" or an intransitive verb 始まる "started". In other words, regular word-formation rules and exceptions co-exist in Japanese verb-verb compounds.

Thirdly, in Japanese verb-verb compounds, the orthographic-semantic associations are not always

systematic. Most of the Japanese verb-verb compounds are polysemous words. For example, a verb-verb compound 仰ぎ見る may refer either to "look up" or "look up to someone as". In addition, many Japanese verb-verb compounds are composed of homonymic constituent verbs. For example, a monomorphemic verb つく may refers either to "attached" (i.e., in 追いつく "catch up") or "to imagine something" (i.e., in 思いつく "come up with"). In other words, the transparency of the semantic relationship between the Japanese verb-verb compound and its corresponding constituents is relatively low.

To sum up, the aforementioned quantitative and structural features make Japanese verb-verb compounds to be an ideal testing ground for exploring the relationship between different frequency measures, since the morphological productivity of the compound words (Frost, Kugler, Deutsch, & Forster, 2005; Velan & Frost, 2011), the orthographic segmentation cues (Rastle, Davis, & New, 2004), and the homonymy of the constituents (Bertram, Laine, Baayen, Schreuder, & Hyönä, 2000; Bertram, Schreuder & Baayen, 2000) are assumed to be influential factors of compound processing. However, to the best of our knowledge, this relationship has not yet been investigated in Japanese verb-verb compounds and no normative data are available for Japanese verb-verb compounds. Thus, the current study aims to fill these gaps by providing normative data for 100 Japanese verb-verb compounds and investigates how different frequency measures are related to each other.

3. Method

3.1 Participants

A total of forty-five native speakers of Japanese participated in our study (14 males and 31 females; mean age, 24.27 years; range, 18-54 years). All participants had normal or corrected-to-normal vision. They received ¥4,500 as compensation for their participation. Participants performed familiarity and AoA ratings individually. This study involving human participants was reviewed and approved by the Sophia University Ethics Committee for Research on Human Subjects.

3.2 Materials

A total of 100 Japanese verb-verb compounds were selected for familiarity and AoA rating tasks from the database BCCWJ (Balanced Corpus of Contemporary Written Japanese; National Institute for Japanese Language and Linguistics, 2012). The selection was performed by the author.

To be included in the stimulus list, a word had to have a "monomorphemic verb + monomorphemic verb" structure. The verb-verb compounds ranged in word length from four to six characters (mean word length = 4.46, *SD* = 0.54) and syllable length from four to seven (mean syllable length = 5.26, *SD* = 0.8). The word lengths of first constituents ranged from two to four characters (mean word length = 2.34, *SD* = 0.5), and the word lengths of second constituents ranged from two to five characters (mean word length = 2.78, *SD* = 0.79). The syllable lengths of first constituents ranged from two to five (mean syllable length = 2.78, *SD* = 0.79), and the syllable lengths of second constituents ranged from two to four (mean syllable length = 2.8, *SD* = 0.64).

Two types of corpus-based objective frequency estimates were derived from the BCCWJ corpus by using *Chunagon* interface (about 104 million words). Surface frequency was operationalized as the token frequency per million with which the particular word-form appears in the corpus (spelling variants were not included). Surface frequency ranged from 2 to 1 per million for verb-verb compounds (mean surface frequency = 0.37, *SD* = 0.43), ranged from 9 to 1 per million for first constituent (mean surface frequency = 1.89, *SD* = 1.71), and ranged from 501 to 1 per million for second constituents (mean surface frequency = 37.49, *SD* = 62.93).

Lemma frequency was operationalized as the summed frequency of all inflectional and spelling variants of a particular word. For instance, lemma frequency of a verb-verb compound 食べ始める "start eating" was the summed frequency of its inflectional (e.g., 食べ始める "start eating", 食べ

始めた “started eating”, 食べ始めよう “let’s start eating”, 食べ始めたら “once (I) started eating”) and spelling variants (e.g., 食べはじめる, たべはじめる, たべ始める, 食べはじめた). Lemma frequency ranged from 17 to 1 per million for verb-verb compounds (mean lemma frequency = 2.07, *SD* = 2.58), ranged from 98 to 1 per million for first constituents (mean lemma frequency = 10.93, *SD* = 13.48), and ranged from 820 to 1 per million for second constituents (mean lemma frequency = 166, *SD* = 229.28).

Two types of rating-based subjective frequency estimates of verb-verb compounds were collected by using questionnaires. Familiarity was operationalized in terms of ratings of how often a word is used or heard in daily communication. Familiarity of verb-verb compounds was collected by using a 7-point Likert scale. Response scales ranged from *very familiar with the verb-verb compound* (1) to *very unfamiliar with the verb-verb compound* (7). AoA was operationalized as the age at which the particular compound verbs were acquired. AoA of verb-verb compounds was collected by using a 12-point scale with 0.5-year age bands. Response scales ranged from *word acquired at 1-year-old* (1) to *word acquired after 7-year-old* (7). For each rating task, a form was created in which the order of word presentation was randomized across participants.

3.3 Procedure

Participants first gave their consent. They were instructed to rate their level of familiarity or age of acquisition to each verb-verb compound intuitively. The familiarity rating task was provided first for each item. At the top of the form, participants were provided with an instruction consisting of the main question and a description of the Likert scale categories. In the header of each page, a brief statement reiterating the scale was also provided. Target words were presented on the left-hand side of the page and participants made their ratings by marking the point with a circle. Secondly, the AoA rating task was provided and the procedure was identical to the familiarity rating task.

3.4 Statistical analysis

All statistical analyses were conducted using SPSS version 24.0 (SPSS Inc., Chiago, IL, USA). The Pearson correlation analysis was used for assessing the correlations between different frequency estimates. The item-level hierarchical multiple and multiple regression analyses were used for evaluating the predictive power of objective frequency estimates on subjective rating scores. For all analyses, a value of $p < .05$ was considered significant. For each corpus-based objective frequency estimate, we added 1 to the raw frequency value, then calculated the frequency per million, and finally calculated the $\log_{10}$ of the frequency of the per million value.

4. Results

The mean ratings and standard derivations, and ranges for each of the rating-based subjective frequency estimates are presented in Table 1.

Table 1 Descriptive statistics for the two rating-based subjective frequency estimates

Estimate	Minimum	Maximum	Mean	SD
Familiarity	6.87	2.22	5.58	1.16
AoA	6.96	4.48	6.24	0.62

SD = standard deviation. AoA = age of acquisition. All ratings are on a 1 to 7 scale.

4.1 Correlation analyses

Table 2 presents the correlations between the two rating-based subjective and the six corpus-based objective frequency estimates. Familiarity was positively correlated with whole-word surface

frequency [r (100) = .491, p = .000] and whole-word lemma frequency [r (100) = .664, p = .000], and weakly correlated with first-constituent surface frequency [r (100) = .252, p = .011] as well as first-constituent lemma frequency [r (100) = .267, p = .007]. However, it was not found to be significantly correlated either with second-constituent surface frequency [r (100) = .057, p = .576] or with second constituent-lemma frequency [r (100) = .043, p = .669]. These results suggest that verb-verb compounds that with higher whole-word frequency or with more frequent first-constituents tended to be rated as being more familiar by native speakers of Japanese.

On the other hand, AoA was marginally negatively correlated with whole-word surface frequency [r (100) = -.258, p = .01] and whole-word lemma frequency [r (100) = -.412, p = .000]. A significant but weak negative correlation was observed between AoA and first-constituent surface frequency [r (100) = -.212, p = .034], as well as between AoA and first-constituent lemma frequency [r (100) = -.269, p = .007]. However, no significant correlation was found either between AoA with second-constituent surface frequency [r (100) = -.043, p = .674], or between AoA and second-constituent lemma frequency [r (100) = -.043, p = .673]. These results suggest that verb-verb compounds that with higher whole-word frequency or with more frequent first-constituents tended to be rated as being of earlier-acquired words by native speakers of Japanese.

Finally, AoA were strongly negatively correlated with familiarity [r (100) = -.599, p = .000], indicating that verb-verb compounds learned early in life tended to be rated as more familiar by native speakers of Japanese.

Table 2 Correlation between rating-based and corpus-based ($\log_{10}$) frequency estimates

	Fam	AoA	W-surface	W-lemma	1st surface	1st lemma	2nd surface	2nd lemma
Fam	1	-.559***	.491***	.664***	.252*	.267**	.057	.043
AoA		1	-.258**	-.412***	-.212*	-.269**	-.043	-.043
W-surface			1	.795***	.280**	.178	.390***	.357***
W-lemma				1	.352***	.325***	.172	.155
1st-surface					1	.794***	.134	.159
1st-lemma						1	.070	.094
2nd-surface							1	.975***
2nd-lemma								1

Fam = whole-word familiarity. AoA = whole-word age of acquisition. W-surface = whole-word surface frequency. W-lemma = whole-word lemma frequency. 1st-surface = first-constituent surface frequency. 1st-lemma = first-constituent lemma frequency. 2nd-surface = second-constituent surface frequency. 2nd-lemma = second-constituent lemma frequency. Frequency is the log-transformed per million frequency from the BCCWJ corpus (National Institute for Japanese Language and Linguistics, 2012). * $p < .05$. ** $p < .01$. *** $p < .001$.

4.2 Regression analyses

To further examine whether the characteristics of the verb-verb compounds and their corresponding constituents influence the familiarity and AoA rating scores, firstly we conduct three-step hierarchical item-level multiple regression analyses for both familiarity and AoA ratings, with rating scores as dependent measures. Length variables (Whole-word length, First-constituent word length, Second-constituent word length, Whole-word syllable length, First-constituent syllable length, Second-constituent syllable length) were entered in Step 1, Constituent-based frequency variables (First-constituent surface frequency, First-constituent lemma frequency, Second-constituent surface frequency, Second-constituent lemma frequency) were entered in Step 2, and Compound-based

frequency variables (Whole-word surface frequency, Whole-word lemma frequency) were entered in Step 3. The results of these hierarchical regression analyses are shown in Table 3.

Table 3 Standardized regression coefficients (β) and significance from Step 1 to 3 of the regression analyses for familiarity and AoA ratings (Hierarchical regression)

Predictors	Familiarity ratings	AoA ratings
Step 1: Length variables		
R-square	.123	-.009
Step 2: Constituent-based frequency variables		
R-square	.203	.045
	$\Delta R^2 = .081$	$\Delta R^2 = .089$
Step 3: Compound-based frequency variables		
R-square	$.427^{***}$	$.156^{**}$
	$\Delta R^2 = .293$	$\Delta R^2 = .117$

Length variables = Whole-word length, First-constituent word length, Second-constituent word length, Whole-word syllable length, First-constituent syllable length, Second-constituent syllable length. Constituent-based frequency variables = First-constituent surface frequency, First-constituent lemma frequency, Second-constituent surface frequency, Second-constituent lemma frequency. Compound-based frequency variables = Whole-word surface frequency, Whole-word lemma frequency. Frequency is the log-transformed per million frequency from the BCCWJ corpus (National Institute for Japanese Language and Linguistics, 2012). * $p < .05$. ** $p < .01$. *** $p < .001$.

As shown in Table 3, Length variables did not account for significant variance across the rating scores in both familiarity and AoA ratings. Likewise, Frequency variables at the constituent level also did not account for significant variance in both familiarity and AoA rating scores. However, Frequency variables at the whole-word level accounted for significant additional unique variance in the two rating-based estimates (29.3% and 11.7% for rating scores in familiarity and AoA rating respectively). These results indicate that compound-based frequency variables play determinant and exclusive role in subjective ratings.

For further investigate the influence of corpus-based objective frequencies on subjective ratings, we conduct multiple regression analyses for both familiarity ratings and AoA ratings, with rating scores as dependent measures, and including the following predictors: Whole-word surface frequency, Whole-word lemma frequency, First-constituent surface frequency, First-constituent lemma frequency, Second-constituent surface frequency, Second-constituent lemma frequency. The results of these multiple regression analyses are shown in Table 4.

Table 4 Standardized regression coefficients (β) and significance from models predicting familiarity and AoA ratings (Multiple regression)

Predictors	Familiarity ratings	AoA ratings
Adjusted R^2	.414	.150
Whole-word surface frequency	-.057	.125
Whole-word lemma frequency	$.701^{***}$	$-.498^{**}$
First-constituent surface frequency	-.037	.076
First-constituent lemma frequency	.085	-.194
Second-constituent surface frequency	.073	-.003
Second-constituent lemma frequency	-.118	-.012

AoA = age of acquisition. Frequency is the per million frequency from the BCCWJ corpus (National

Institute for Japanese Language and Linguistics, 2012). * $p < .05$. ** $p < .01$. *** $p < .001$.

As shown in Table 4, with approximately 70% of explained variance, familiarity rating scores were predicted to a large extent by the whole-word lemma frequency, with verb-verb compounds rated as more familiar as the whole-word lemma frequency becomes higher. However, familiarity rating scores were not significantly predicted by either whole-word surface frequency or constituent-based surface/lemma frequencies. Likewise, whole-word lemma frequency was also a significant predictor of AoA rating scores. The higher whole-word lemma frequency the verb-verb compounds with, the more they were rated as early acquired. Similar to what occurred for familiarity, AoA rating scores were not significantly predicted by either whole-word surface frequency or constituent-based surface/lemma frequencies. In other words, the whole-word lemma frequency is the single best determinant for both familiarity and AoA rating scores.

5. Discussion

As familiarity and AoA become widely used factors in psycholinguistic studies, there has been an intense surge of interest in the investigation of their construct validity (Bonin, Laroche, & Méot, 2022; Juhasz, Lai, & Woodcock, 2015; Song & Li, 2021). However, due to data limitations, the relationship between familiarity and AoA was not yet validated in Japanese verb-verb compounds. In this article, we provided the first normative dataset of Japanese verb-verb compounds. This dataset contains two types of subjective frequency measures (familiarity, AoA) for 100 Japanese verb-verb compounds as well as two types of objective frequency measures (surface frequency, lemma frequency) for 100 Japanese verb-verb compounds and their corresponding constituents. Familiarity ratings were made on a scale ranging from 1 = *very familiar* to 7 = *very unfamiliar* (7-point Likert scale) and AoA ratings were made on a scale ranging from 1 = *1-year-old* to 7 = *after 6-year-old* (12-point scale with 0.5-year age bands) by 45 native speakers of Japanese. Surface and lemma frequencies were calculated on the basis of a corpus of approximately 104 million words gathered from modern Japanese texts. To find out whether these two frequency measures related to one another we further performed correlation and regression analyses on the aforementioned dataset. The major findings were as follows.

With respect to correlation analyses, first of all, we found that the two rating-based subjective frequency measures were strongly correlated with each other. Compound words that were judged to be familiar in daily communication were estimated to be acquired earlier in life. Secondly, we further found that the two rating-based subjective frequency measures have similar but not identical patterns of intercorrelations with corpus-based objective frequency measures. For familiarity, rating scores were strongly associated with both whole-word surface and whole-word lemma frequencies, and weakly associated with first-constituent surface and first-constituent lemma frequencies. Likewise, for AoA, rating scores were weakly related to whole-word surface frequency and first-constituent frequencies, and more closely associated with whole-word lemma frequency. However, neither of them was associated with second-constituent frequencies. On the other hand, the major findings of regression analyses shed light on whether and to what extent objective frequency variables influence subjective frequency ratings: We found that whole-word lemma frequency was the best single predictor of familiarity and AoA rating scores. This result indicates that the total probability of encountering a compound word rather than the probability of encountering a particular word-form determines native speakers' impression of the compound word.

The predominant effect of whole-word lemma frequency on subjective frequency ratings in the present study challenges the adequacy of existing Full-parsing models (e.g., Libben, Derwing, & de Almeida, 1999; Taft & Forster, 1976) of compound processing, which assume obligatory morphological decomposition during compound processing. On the basis of the coefficient patterns induced in familiarity and AoA ratings, we would claim that native speakers of Japanese tended to

process verb-verb compounds as a whole unit when they were asked to rate the lexical properties of those words explicitly in un-speeded task settings. This is consistent with the predictions of Full-listing accounts which hypothesize that compound words are activated directly via their whole-word representations during processing (e.g., Butterworth, 1983; Bybee, 1995).

The lack of effects of constituent frequencies on lexical ratings could be attributed to the orthographic and semantic features of Japanese verb-verb compounds. As mentioned in the review section, Japanese verb-verb compounds are words with low orthographic decomposability and low transparency of orthographic-semantic association. Both of these two features were assumed to influence readers' perception of the constituents of the compound words (Gagné, Spalding, Spicer, Wong, Rubio, & Cruz, 2020). For example, spaced compound words are more likely to be processed as separate units than closed compound words (Ji, Gagné & Spalding, 2011), and semantically opaque compounds were more likely to be processed as a whole unit compared to semantically transparent compounds (Marelli & Luzzatti, 2012). In other words, it might be the degree of salience of the constituents affects whether the compound word will be processed as a whole unit or as separate units, and for compound words with less salient constituents, such as Japanese verb-verb compounds, holistic processing route rather than analytic processing route became the preferable route to processing.

Another interesting result of the present study is the absence of effect of whole-word surface frequency on the subjective ratings. This result leads to the conclusion that Japanese verb-verb compounds possess their own lemma in the mental lexicon, and inconsistent with the previous results reported in English and French compound nouns, in which the subjective rating scores of compound words were co-determined by whole-word and constituent surface frequencies (Bonin, Laroche, & Méot, 2022; Juhasz, Lai, & Woodcock, 2015). This discrepancy may be due to the fact that unlike noun-noun compounds, verb-verb compounds are rarely used in their original form, and instead are commonly used in inflected forms, such as past-tense or passive forms. Thus, for compound verbs, whole-word surface frequency might not be powerful enough to predict or explain lexical behaviors, and instead, whole-word lemma frequency might be a better indicator of lexical processing.

6. Conclusion

Taken as a whole, over the past two decades, there has been an explosion of interest in using databases of behavioral performance measures to investigate questions concerning mechanisms of compound processing (for English, see Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007; for French, see Ferrand, New, Brysbaert, Keuleers, Bonin, Méot, Augustinova & Pallier, 2010; for Chinese, see Sun, Hendrix, Ma, & Baayen, 2018; for Dutch, see Keuleers, Diependaele, & Brysbaert, 2010). Those databases allow a more complete picture of how compound words are processed in the mental lexicon and serve as a useful tool for cross-language comparisons.

However, the majority of available behavioral data are established for Indo-European languages, and there is a lack of normative and behavioral data for Japanese. In this study, we established two psycholinguistic norms for 100 Japanese verb-verb compounds derived from native speakers of Japanese. By examining whether characteristics of the verb-verb compounds and their corresponding constituents influence ratings of familiarity and AoA of verb-verb compounds, we found that whole-word lemma frequency is the best single predictor of the rating scores, and there was only a weak association between first-constituent frequencies and subjective rating scores. These findings support the holistic accounts of compound processing (e.g., Butterworth, 1983; Bybee, 1995). Our interpretation of these findings is that the low degree of salience of the constituents inhibits the morphological decomposition and facilitates the lexical storage.

As a first step toward constructing a new database of behavioral performance measures for Japanese

compound words, this study develops our understanding of compound processing and the underlying organization of mental lexicon. However, further work is required to evaluate whether familiarity and AoA are powerful predictors of behavioral performance data (i.e., lexical decision times, word naming times) of Japanese verb-verb compounds, and to improve and expand the current dataset to incorporate other lexical (i.e., morphological family size, morphological family frequency) and psycholinguistic variables (i.e., semantic transparency, concreteness, imageability).

Acknowledgements

This research was supported by JSPS KAKENHI Grant number 22K20032.

References

Barbara J. Juhasz (2018). "Experience with compound words influences their processing: An eye movement investigation with English compound words", *Quarterly Journal of Experimental Psychology*, 71:1, pp.103-112.

Barbara J. Juhasz, Yun-Hsuan Lai, and Michelle L. Woodcock (2011). "A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience", *Behavior Research Methods*, 47, pp.1004-1019.

Brain Butterworth (1983). "Lexical representation", In Brain Butterworth (Ed.), *Language Production vol.2: Development, writing and other language processes* (pp.257-294), London: Academic Press.

Ching Chu Sun, Peter Hendrix, Jianqiang Ma, and Rolf H. Baayen (2018). "Chinese lexicon database (CLD): A large-scale lexical database for simplified Mandarin Chinese", *Behavior Research Methods*, 50, pp.2606-2629.

Chi-Shing Tse, Melvin J. Yap, Yuen-Lai Chan, Wei Ping Sze, Cyrus Shaoul, and Dan Lin (2017). "The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words", *Behavior Research Methods*, 49, pp.1503-1519.

Christina L. Gagné, Thomas L. Spalding, Patricia Spicer, Dixie Wong, Beatriz Rubio, and Karen Perez Cruz (2020). "Is buttercup a kind of cup? Hyponymy and semantic transparency in compound words", *Journal of Memory and Language*, 113, 104110.

Dangui Song, and Degao Li (2021). "Psycholinguistic norms for 3,783 two-character words in simplified Chinese", *SAGE Open*, 11, 215824402110544. https://doi.org/10.1177/21582440211054495.

Daniel Schmidtke, Christina L. Gagné, Victor Kuperman, Thomas L. Spalding, and Benjamin V. Tucker (2018). "Conceptual relations compete during auditory and visual compound word recognition", *Language, Cognition and Neuroscience*, 33:7, pp.923-942.

David A. Balota, Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman (2007). "The English Lexicon Project", *Behavior Research Methods*, 39:3, pp.445-459.

David A. Balota, Michael J. Cortese, Susan D. Sergent-Marshall, Daniel H.Spieler, and Melvin J. Yap (2004). Visual word recognition of single-syllable words", *Journal of Experimental Psychology: General*, 133:2, pp.283-316.

Emmanuel Keuleers, Kevin Diependaele, and Marc Brysbaert (2010). “Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords”, *Frontiers in Psychology*, 1, 174. doi:10.3389/fpsyg.2010.00174.

Emmanuel Keuleers, Marc Brysbaert, and Boris New (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles”, *Behavior Research Methods*, 42, pp.643-650.

Fritz Günther, Marco Marelli, and Jens Bölte (2020). “Semantic transparency effects in German compounds: A large dataset and multiple-task investigation”, *Behavior Research Methods*, 52, pp.1208-1224.

Gary Libben, Bruce L. Derwing, and Roberto G. de Almeida (1999). “Ambiguous novel compounds and models of morphological parsing”, *Brain and Language*, 68:1-2, pp.378-386.

Gary Libben, Christina L. Gagné, and Wolfgang U. Dressler (2020). “The representation and processing of compounds words”, In Vito Pirrelli, Ingo Plag, and Wolfgang U. Dressler (Eds.), *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon* (pp.336-352), Berlin, Boston: De Gruyter Mouton.

Georgios Tagalakis, and Mark T. Keane (2006). “Familiarity and relational preference in the understanding of noun-noun compound”, *Memory & Cognition*, 34:6, pp.1285-1297.

Hélène Giraudo, and Jonathan Grainger (2000). “Priming complex words: Evidence for supralexical representation of morphology”, *Psychonomic Bulletin & Review*, 8:1, pp.127-131.

Joan Bybee (1995). “Regular morphology and the lexicon”, *Language and Cognitive Processes*, 10:5, pp.425-455.

Hadas Velan, and Ram Frost (2011). “Words with and without internal structure: What determines the nature of orthographic and morphological processing?”, *Cognition*, 118:2, pp.141-156.

Hien Pham, and Harald Baayen (2015). “Vietnamese compounds show an anti-frequency effect in visual lexical decision”, *Language, Cognition and Neuroscience*, 30:9, pp.1077-1095.

Hongbo Ji, Christina L. Gagné, and Thomas L. Spalding (2011). “Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds”, *Journal of Memory and Language*, 65, pp.406-430.

James Myers, Yu-chi Huang, and Wenling Wang (2006). “Frequency effects in the processing of Chinese inflection”, *Journal of Memory and Language*, 54:3, pp.300-323.

Kaitlin Falkauskas, and Victor, Kuperman (2015). “When experience meets language statistics: Individual variability in processing English compound words”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41:6, pp.1607-1627.

Kathleen Rastle, Matthew H. Davis, and Boris New (2004). “The broth in my brother’s brothel: Morpho-orthographic segmentation in visual word recognition”, *Psychonomic Bulletin & Review*, 11:6, pp.1090-1098.

Ludovic Ferrand, Boris New, Marc Brysbaert, Emmanuel Keuleers, Patrick Bonin, Alain Méot, Maria Augustinova, and Christophe Pallier (2010). “The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords”, *Behavior Research Methods*, 42:2, pp.488-496.

Marc Brysbaert (2017). “Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings ‘corrected’ for frequency”, *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 70:7, pp.1129-1139.

Marco Marelli, and Claudio Luzzatti (2012). “Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency”, *Journal of Memory and Language*, 66:4, pp.644-664.

Marcus Taft, and Kenneth I. Forster (1976). “Lexical storage and retrieval of polymorphemic and polysyllabic words”, *Journal of Verbal Learning and Verbal Behavior*, 15, pp.607-620.

Mercedes Güemes, Carolina Gattei, and Alejandro Wainselboim (2019). “Processing verb-noun compound words in Spanish: Evidence from event-related potentials”, *Cognitive Neuropsychology*, 36:5−6, pp.265-281.

Mohammad Momenian, Shuk K. Cham, Jafar Mohammad Amini, Narges Radman, and Brendan Weekes (2021). “Capturing the effects of semantic transparency in word recognition: A cross-linguistic study on Cantonese and Persian”, *Language, Cognition and Neuroscience*, 36:5, pp.612-624.

Patrick Bonin, Betty Laroche, and Alain Méot (2021). “Psycholinguistic norms for a set of 506 French compound words”, *Behavior Research Methods*, 54, pp.393-413.

Ram Frost, Tamar Kugler, Avital Deutsch, and Kenneth I. Forster (2005). “Orthographic structure versus morphological structure: Principles of lexical organization in a given language”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31:6, pp.1293-1326.

Raymond Bertram, Matti Laine, Harald R. Baayen, Robert Schreuder, and Jukka Hyönä (2000). “Affixal homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language”, *Cognition*, 74, B13-B25.

Raymond Bertram, Robert Schreuder, and Harald R. Baayen (2000). “The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26:2, pp.489-511.

Robert Schreuder, and Harald R. Baayen (1995). “Modeling morphological processing”, In Laurie B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp.131-154). Hillsdale, New Jersey: Lawrence Erlbaum.

Victor Kuperman, and Julie A. Van Dyke (2013). “Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers”, *Journal of Experimental Psychology: Human Perception and Performance*, 39:3, pp.802-823.

Yongqiang Su, Yixuan Li, and Hong Li (2023). “Familiarity ratings for 24,325 simplified Chinese words”, *Behavior Research Methods*, 55, pp.1496-1509.

影山太郎 (1993).『文法と語形成』ひつじ書房

姚一佳 (2020).「語彙判断課題における日本語母語話者と学習者の複合動詞アクセス」『ことばの科学研究』21, pp.81-100.

関連 URL

コーパス検索アプリケーション『中納言』　https://chunagon.ninjal.ac.jp/

『複合動詞レキシコン』　https://vvlexicon.ninjal.ac.jp

『現代日本語書き言葉均衡コーパス』　https://clrd.ninjal.ac.jp/bccwj/