Relative frequency
and the holistic processing of morphology
Evidence from a corpus of vernacular Japanese

Kevin Heffernan and Yo Sato
Kwansei Gakuin University / Satoama Language Services

This study presents apparent-time changes in the morphology of the expression mitai-na ‘similar to’. Based on apparent-time data, we argue that the morphological boundary between mitai and the attributive morpheme -na in the phrase mitai-na has disappeared, and that this complex phrase is now processed as a monomorphemic form. We suggest that relative frequency is the key to understanding the results.

We further supplement our argument with data on the standardization of the adverbial adjective form in the Kansai dialect. Young speakers overwhelmingly use the standard form of adverbials for all adjectives except two: yō ‘a lot, well’ and hayō ‘quickly, early’ (instead of Standard Japanese yoku and hayaku). The three linguistic forms that display unusual behavior (mitai-na and the adverbial forms of yō and hayō) all have a high relative frequency. We conclude that when a complex form occurs more frequently than its components (high relative frequency), then it behaves as a monomorphemic unit. The irregular adverbial forms are leftover from an obsolete system, in the same way that many English irregular past forms are leftover from the Germanic strong verb system. In contrast, the irregular form mitai-na emerged from and competes with the regular inflection paradigm for mitai, illustrating a previously undocumented path for the diachronic emergence of irregular morphology.

Keywords: Japanese, collocation, lexical bundle, holistic processing, entrenchment, apparent time

1. Introduction

Our capacity to formulate thought and our capacity to speak are in a constant race to keep up with each other. If our rate of speech outruns our rate of thinking, then
Kevin Heffernan and Yo Sato

We must pause our speech. These pauses are realized as filler words, such as “umm” (Biber, Johansson, Leech, Conrad, & Finegan, 1999). This is a well-understood phenomenon. But what happens when our rate of thinking outruns our rate of speaking? In such situations, speakers resort to an increased use of frequently-occurring lexical bundles (Biber et al., 1999; Kuiper, 1996; Wray, 2002).

A lexical bundle is a combination of two or more linguistic units (words, morphemes, etc.) that occur together frequently, such as the English expressions beautiful day and supporting evidence and the phrases I don’t think and a kind of. \(^1\) The concept of the lexical bundle has been examined from a range of theoretical perspectives, such as register (Biber et al., 1999), first- (Bannard & Lieven, 2012) and second-language acquisition (Chen & Baker, 2010; Juknevičienė, 2009), and language processing (Tremblay, Derwing, Libben, & Westbury, 2011).

But how exactly does a lexical bundle facilitate our ability to speak faster? Currently, there are two competing answers to this question: the holistic processing theory (Biber et al., 1999; Conklin & Schmitt, 2012; Kuiper, 1996; Miller, 1956; Pawley & Syder, 1983; Tremblay & Baayan, 2010; Wray, 2002) and the lexical priming theory (Durrant & Doherty, 2010; Hoey, 2005; Meyer & Schvaneveldt, 1971). Advocates of the first theory argue that highly frequently occurring combinations meld into a single, holistic form which can then be processed very quickly. Advocates of the second theory argue that the different parts of highly frequently occurring combinations form special cognitive bonds that mutually facilitate the processing of each part. Both of these theories describe the way we produce and comprehend morphologically complex words and phrases, and both of these theories have a large body of supporting evidence. We explain in detail in Section 2.

Our objective is to contribute to our understanding of both theories by focusing on the boundary between the two. Our argument begins with the claim that complex forms are initially processed (comprehended and produced) holistically by very young children (Bannard & Lieven, 2012; Tomasello, 2003, 2009; Wray, 2002). The young learner eventually reanalyzes and reorganizes most of the complex forms into separable units, which prime each other (Hoey, 2005). However, as we argue in this paper, there is one scenario in which this shift does not occur. Specifically, if a complex phrase occurs more frequently than its individual mor-

---

1. Researchers refer to frequently occurring sequences of lexical units by a variety of names such as ‘lexical bundle’, ‘collocation’, and ‘idiomatic expression’. Bider et al. (1999) differentiate ‘collocation’ from ‘lexical bundle’ based on the number of words in the sequence. However, such a distinction is problematic for research on languages such as Japanese, for which the concept of a word is itself problematic. We instead use the morpheme as our most basic linguistic unit, and we collectively refer to any sequence of frequently occurring morphemes, both lexical and grammatical, as a lexical bundle, regardless of the size of the bundle.
phemes do elsewhere, then reanalysis fails to occur. We repeat and elaborate this point about the relative frequency of morphemes throughout the paper.

Our data comes from the *Corpus of Kansai Vernacular Japanese* (Heffernan, 2012). The Kansai dialect of Japanese is ideal for studying stylistic variation because it has a rich collection of lexical and morphological variants that range along a non-regional polite/non-regional non-polite/regional continuum. For example, speakers of the Kansai dialect of Japanese use not only the Standard Japanese sentence-final particles such as *yo* and *ne*, but also use sentence-final particles such as *nen* and *de*, which have a strong regional flavor. Furthermore, unlike some of the other dialects in Japanese, the Kansai dialect is healthy. Although many traditional Kansai dialect forms such as *sakai* 'because' are obsolescing, young speakers are innovating new linguistic forms, such as the verbal intensifier *meccha* 'very'. In this paper, we look at the obsolescence of morphological pattern called *onbin* 'convenient pronunciation', which we explain with examples below.

We present two studies in this paper. The first is an apparent-time study of ongoing changes in the Japanese lexical bundle *mitai-na* 'similar to', an example of non-regional variation. We demonstrate that the morphological patterning of *mitai-na* in vernacular Japanese has shifted from regular to irregular, and the result is that *mitai-na* is now a single unit. Finally, we argue that the key to understanding the shift is the relative frequency of the whole expression to its parts.

We support our claim for the role of relative frequency in the holistic processing of morphology with data from a phonological process that occurs in the Kansai dialect, called *onbin*. In general, the *onbin* forms are obsolete: the younger generation almost entirely uses the Standard Japanese forms instead – except for two words. These two words are the adverbial forms for the adjectives *yoi* 'good; wel; a lot' and *hayai* 'early; quick'. We argue that these two words alone have failed to standardize because they are processed holistically. We again base our argument on the relative frequency of the forms.

To date, almost all work on collocation and lexical bundles has been done on European languages. To our knowledge, this is the first look at collocations in vernacular Japanese. The Japanese language is particularly well-suited to the study of irregular morphology because it is a prototypical agglutinative language. There is almost always a clear one-to-one mapping between morpheme and (grammatical) meaning. Because of this one-to-one mapping, any irregular morphology stands out.

As stated, our conclusion is that complex morphological forms are processed holistically if the complex phrase occurs more frequently than the individual morphemes do elsewhere. Many researchers have previously concluded that specific morphologically complex forms are processed holistically, but their claim has primarily been based on the processing (either listening comprehension or
production) speeds as reflected by reaction times in psycholinguistics experiments (Arnon & Snider, 2010; Conklin & Schmitt, 2012; Pawley & Syder, 1983; Tremblay & Baayan, 2010). The gist of their argument is as follows. When an experiment participant processes a multi-part form, he or she must first retrieve all of the parts from lexical memory, and then combine those parts. This two-part process is longer than the retrieval of a single, whole form. Therefore, forms that are stored holistically are processed faster than forms that are stored in parts.

Experimental studies on the processing of idiomatic expressions such as ‘kick the bucket’ support this claim. Idiomatic expressions can only be understood as a whole unit. Furthermore, they are processed faster than non-idiomatic expressions (see Tabossi, Fanari, & Wolf, 2009, for an overview). Similarly, highly frequent non-idiomatic expressions such as ‘bride and groom’ are also processed faster than less frequently occurring phrases (Siyanova-Chanturia, Conklin, & van Heuven, 2011). These observations lead us to conclude that since idiomatic expressions – which are stored holistically in lexical memory – are processed quickly, and highly frequent non-idiomatic expressions are processed quickly, then the latter must also be stored holistically.

Others, such as Siyanova-Chanturia (2015), disagree with this claim. They argue that a faster processing time alone does not necessarily reflect holistic storage. Schmid (2017, p. 18) summarizes the current state of the debate as follows:

The evidence collected so far seems to be quite conclusive as regards the holistic storage and processing of prototypical, that is, noncompositional idioms. In contrast, the extent to which other, less fixed and more transparent combinations are indeed processed as chunks has turned out to be much less easy to determine.

This paper contributes to this ongoing investigation of holistic processing by presenting a case in which a “transparent combination” seems to be processed as a chunk – that is, in a holistic manner.

The remainder of this paper is laid out as follows. The next section introduces holistic processing and lexical priming in more detail. We then present an apparent-time study of ongoing changes in the morphology of the phrase mitai+na. We follow with a study of the standardization of onbin in the Kansai region. Finally, we conclude with the implications of our results for our understanding of morphological processing.
2. Conversing under pressure: two competing theories

2.1 The holistic processing of complex forms

We begin by introducing the claim that certain types of language are processed in a holistic manner. According to this argument (Biber et al., 1999; Conklin & Schmitt, 2012; Kuiper, 1996; Miller, 1956; Pawley & Syder, 1983; Tremblay & Baayan, 2010; Wray, 2002), holistic processing of expressions stems from the limitations of our working memory. Our working memory processes speech as we engage in conversation. However, working memory has limited capacity: The greatest number of units that working memory can reliably process is approximately seven, give or take a few (Conklin & Schmitt, 2012, p. 47; Cowan, 2010; Miller, 1956; Tremblay et al., 2011, pp. 571–572). Our brain overcomes this limitation by using a much more abundant resource – long term memory – to store holistic representations of frequently occurring sequences. This “eases the burden on working memory during initial processing and subsequent recall” (Conklin & Schmitt, 2012, p. 52), as these whole units are easily retrieved and used without the need to compose them online through grammatical sequencing.

But why is there a need to ease the processing burden? Biber et al. (1999) claim that this need is a consequence of our relatively slow processing of linguistic information: “Time pressure [during conversation] makes it more difficult for speakers to exploit the full innovative power of grammar and lexicon” (Biber et al., 1999, p. 1049). It takes time to innovate, time that we do not have as we rapidly rattle off strings of words during conversation. They argue that instead of innovating, speakers rely on prefabricated word sequences stored in memory as a whole unit, ready to go.

Kuiper’s (1996) study of the speech of sportscasters during horse races and the speech of auctioneers during auctions presents clear evidence on the impact of time pressure. Kuiper found that the faster the speech, the greater the repeated use of a limited set of formulaic phrases such as ‘Oh what a beauty’ and ‘X dollars for that?’ Kuiper (1996, p. 72) concluded that “the occurrence of formulaic features … varies according to the pressure on the speaker’s working memory and processing capacity”: As the pressure increases, the reliance on a small set of frequently used phrases also increases.

In general, psycholinguists equate cognitive burden with processing time. Two types of phrase that have consistently been shown to be processed faster are idioms and frequently occurring phrases (e.g., Tabossi et al., 2009; Tremblay et al., 2011). But why are idioms and frequently occurring phrases so special? Several researchers have posited holistic processing as the link between idioms and frequently occurring phrases. If a morphologically complex form is processed in a
holistic manner, then speakers comprehend and produce the phrase as if it consisted of a single, monomorphemic form (Conklin & Schmitt, 2012; Tremblay & Baayen, 2010).

Besides the processing time behavioral evidence, there is also neurocognitive evidence supporting holistic processing (but see Siyanova-Chanturia, 2015, for a counter argument). Tremblay and Baayen (2010) used ERP (event-related potentials) imaging to investigate the processing of four-word lexical bundles such as ‘in the middle of’. ERP studies record brain waves that are generated in response to particular events, such as seeing a word or picture on a computer screen. ERP is an ideal tool for examining brain activity as it is capable of recording events with a very fine temporal resolution (down to the level of the millisecond). Tremblay and Baayen focused on the 200–250 millisecond range with the objective of distinguishing between holistic processing and online computation. They argue that if a measure of collocation strength shows a correlation with brain activity in this range, then it must be holistic processing, as it is improbable that a speaker is able to retrieve four words and then perform the necessary grammatical computation to integrate them within 200 to 250 milliseconds. They found such an effect and, based on this, concluded that their results provide “evidence to the effect that four-word sequences are retrieved in a holistic manner” (Tremblay & Baayen, 2010, p. 170).

Yet attempts to clearly establish the link between holistic processing and frequency have not been completely convincing (Siyanova-Chanturia, 2015). For example, Tabossi, Fanari, and Wolf (2009) attempted to test directly the link between idioms and holistic processing, and concluded that familiarity, not idiomaticity, leads to the faster processing of idioms.

Researchers now use a cautionary tone when talking about holistic processing. Tremblay et al. (2011, p. 595) caution that “it is possible that discriminating between holistic retrieval and speeded online computation cannot be achieved by considering behavioral data alone”. Similarly, Arnon and Snider (2010, p. 78) point out that the “distinction between ‘stored’ and ‘computed’ material is further blurred by recent findings on the processing of idioms”. Wray (2012, p. 233) points out that “the question of whether a processing advantage in terms of speed indicates holistic storage or simply the faster mapping of components” is still an important issue that needs to be addressed. Siyanova-Chanturia (2015, p. 13) concludes that “more empirical evidence is needed if one wants to address the question of holistic storage and processing”.

On the other hand, if we consider the etymological origins of words, then clearly holistic processing has occurred in some extreme cases. Such loss of internal structure is seen in diachronic changes in English that have led to the expressions ‘goodbye’ from ‘God be with ye’, and ‘because’ from ‘by cause’, to name
but two examples (Bybee, 2010; Schmid, 2017). The modern day phrases result from the extreme loss of phonemic content and morphological constituency, and a complete shift in semantic meaning. They are now new words that are, just like any other monomorphemic word, indisputably processed holistically.

Bybee argues (2010, Chapter 3) that the degree to which a complex linguistic form is treated as a single unit lies along a “chunking” continuum. At one end of the continuum lie combinations of words that are seldom used together, and therefore do not form a chunk. In the middle of the continuum lie idiomatic expressions. Bybee (2010, p. 36) gives the examples of ‘lend a hand’ and ‘pick and choose’. Such expressions are “easily accessible as wholes” while still maintaining their internal morphology (ibid.). At the other end of the continuum are grammaticalized phrases and words such as ‘because’ that “lose their internal structure and identifiability of their constituent parts” (ibid.). Following Bybee (2010), we interpret holistic processing as the lack of internal morphological structure.

2.2 Lexical priming

An alternative explanation for the faster processing of highly frequent lexical bundles is lexical priming. The concept of lexical priming was first documented in a psycholinguistics experiment in which participants judged whether or not pairs of words presented on a screen were actual words (Meyer & Schvaneveldt, 1971). Words that were semantically related to each other (e.g., ‘bread’ and ‘butter’) were judged significantly faster than pairs of words that were not related to each other (e.g., ‘bread’ and ‘nurse’). Meyer and Schvaneveldt (1971, p. 232) propose that the “passive ‘spread of excitation’ to other nearby locations facilitate[s] later retrieval from them”. This phenomenon has been studied extensively and extended to multi-word phrases and even to other areas of linguistic processing, such as the priming of grammatical structures (Hoey, 2005).

Linguists interpret the priming effect as the neurological activation of a linguistic form (word, phrase, grammatical structure, etc.) simultaneously activating a cloud of related linguistic forms (Bybee, 2010; Hoey, 2005). When one of these forms that has been activated by association is later processed, the processing is facilitated by the pre-activation. This is seen, for example, in reduced reaction times during psycholinguistic lexical decision tasks.

The concept of lexical priming has been extended to lexical bundles (Durrant & Doherty, 2010). Durrant and Doherty demonstrated that experiment participants respond faster to the lexical decision task of a two-word collocation when primed by the first word even when the two words are not conceptually associated with one another.
We are thus left with two models for the processing of frequently occurring lexical bundles. On the one hand, many researchers have claimed that they are processed in a holistic manner, similar to monomorphemic words. On the other hand, others argue that the processing effects seen in the behavioral experiments can be accounted for by a model of lexical priming. In this case, the initial part of the lexical bundle primes the latter part of the lexical bundle, leading to faster processing.

The objective of this paper is to reconcile these two approaches by positing a boundary between them. We argue that lexical bundles are for the most part processed through priming. However, they are processed in a holistic manner under certain specific conditions.

3. An apparent-time study of changes in the morphology of the phrase mitai-na

3.1 The phrase mitai

The expression mitai is a na-adjective. The Japanese term for na-adjective is keiyō-dōshi, which literally means ‘adjectival verb’. The Japanese name reflects that a na-adjective is followed by a copula (i.e., a verb) when it forms the predicate of a sentence, as in the case of kirei ‘pretty’ in example (1). In many ways, they behave as nouns. However, one major difference is that nouns take the genitive case marker -no when modifying a following noun, whereas na-adjectives use the attributive -na when modifying a following noun, as in the case of kirei-na in example (2). In English literature, they are variously referred to as adjectival verbs, adjectival nouns, na-adjectives, and na-nominals.

(1) Tanaka-san-wa kirei desu.
Tanaka-hon-top pretty cop
‘Ms. Tanaka is pretty.’

2. The following glosses are used:

<table>
<thead>
<tr>
<th>ADV</th>
<th>adverbal</th>
<th>PAST</th>
<th>past tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATR</td>
<td>attributive</td>
<td>POST</td>
<td>postposition</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
<td>POT</td>
<td>potential</td>
</tr>
<tr>
<td>COND</td>
<td>conditional</td>
<td>REPR</td>
<td>representative example</td>
</tr>
<tr>
<td>CONJ</td>
<td>conjunction</td>
<td>SBJV</td>
<td>subjunctive mood</td>
</tr>
<tr>
<td>EXIST</td>
<td>verb indicating existence</td>
<td>SFP</td>
<td>sentence-final particle</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
<td>STAT</td>
<td>stative aspect</td>
</tr>
<tr>
<td>HON</td>
<td>honorific</td>
<td>SUB</td>
<td>subject</td>
</tr>
<tr>
<td>NEG</td>
<td>negative</td>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>NLZ</td>
<td>nominalizer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The expression *mitai-na* consists of the word *mitai* and the attributive morpheme *-na*. The word *mitai* is used to show approximate equivalence (Matsutomo, Miyamoto, & Wakuri, 2007), similar to the English expression ‘seems like’, as illustrated by example (3).

(3) *Kanojo-no hanashi-wa kodomo mitai ne.*
she-GEN talking- TOP child like SFP
‘Her speech is like that of a child.’

(Matsutomo et al., 2007; Example 1 from the *mitai-da* entry)

Prescriptive grammars such as Matsutomo et al. (2007) indicate that the attributive morpheme *-na* must follow *mitai* when modifying a noun, as in the examples given in (4).

(4) a. *oshiro mitai-na ie*
castle like-ATR house
‘a house like a castle’

(Matsutomo et al., 2007; Example 2 from the *mitai-da* entry)

b. *haru-ga ki-ta mitai-na atataka-sa*
spring-SUB come-PAST like-ATR warm-NLZ
‘a warmth like spring has arrived’

(Matsutomo et al., 2007; Example 4 from the *mitai-da* entry)

We focus mainly on the expression *mitai-na*. In this regard, we are not the first to do so. The phrase *mitai-na* is undergoing grammaticalization, and as part of this process, the semantics of this specific combination have expanded. Several researchers have explored in detail the discourse function of *mitai-na*. Such work focuses on the degree to which *mitai-na* softens the tone of an utterance, adds vagueness, or is used as a quotative marker. Recent examples of such research include Horasawa (2011), Horio (2014), Hoshino (2009), Iwasaki (2012), Maeda (2004) and Ohba (2009). Contrary to the prescriptive grammars, all of these researchers give examples of *mitai-na* used at the end of a phrase, and not followed by a noun. Example (5), adapted from example (86) in Horio (2014), illustrates this usage.

(5) *Oya-ni-shite mi-reba nattoku ik-anai mitai-na.*
parent-POST see-COND convince go-NEG like-ATR
‘But isn’t it like, if you consider it from the perspective of a parent, then it is not convincing?’
We are not concerned with the discourse function of mitai-na, only its form. Maeda (2004, p. 57) briefly addresses the question of why the discourse marker is mitai-na and not mitai-da, or just mitai: “後ろに来る名詞と述語部分を省略することから、文末「みたいな。」が生じたものであろう” [The expression mitai-na comes at the end of an utterance because the following noun and the predicate of the sentence are most likely omitted]. Maeda is suggesting that when the phrase mitai-na appears not to be followed by a noun, then there is an omitted noun. An anonymous reviewer further clarified that the omitted noun is a null noun, and it is the silent counterpart to the overt noun kanji ‘feeling’ in the phrase mitai-na kanji.

This explanation accounts for utterances such as example (5). The concept of a null noun has been posited by Kayne (2005, Chapter 10) to explain the unusual syntactic patterning of number, color and time expressions in English. In these cases, the null noun forms the head of the constituent, turning the phrase into a noun phrase, as demonstrated by constituency tests. But when we examine the syntactic patterning of mitai-na, we see a different picture. Positing a null noun after mitai-na predicts that both of the following examples should be grammatical, and not just the example with the overt noun kanji:

(6)  Hatsukoi mitai-na kanji-ga suru.
     first love like-atr feeling-sub do
     ‘It feels like falling in love for the first time.’

(7)  *Hatsukoi mitai-na ∅-ga suru.
     first love like-atr NULL-sub do

The ungrammaticality of example (7) seems to refute the null noun explanation. We instead claim that mitai-na is treated as a single unit by speakers, and that the -na morpheme no longer acts as an attributive morpheme.

3.2 Methodology

The data were extracted from the Corpus of Kansai Vernacular Japanese (see Heffernan, 2012, for a brief introduction). This corpus currently consists of 124 conversational interviews. The interviews were conducted by university students native to the Kansai region of Japan. All interviewees grew up in and are still currently living in the Kansai region. All interviewers and interviewees self-reported that they are native speakers of Kansai Japanese. All interviewees were either relatives or close acquaintances of the interviewers. The interviewers were instructed to conduct the interview for approximately 50 minutes in vernacular Japanese. The interviews took place in a casual setting chosen by the interviewee.
Collectively, the 124 interviews consist of 6,778 minutes of speech, or 2.3 million transcribed characters (Japanese moji). All of the transcribed data have been parsed into morphemes and tagged with part of speech information, yielding approximately 1.4 million lines of tagged data. Each morpheme is also tagged with speaker information. Only the data of the interviewees (hereafter referred to as speakers) are analyzed. Table 1 shows the age and gender distribution of the speakers.\(^3\)

Two types of expression were extracted from the interview data. Firstly, a total of 3,564 tokens of the na-adjective mitai were extracted, along with the morpheme that followed mitai. These tokens were coded based on the part of speech of the following morpheme. If the following morpheme was the attributive -na, then one more following morpheme was extracted. In these cases, we also coded the tokens according to the part of speech of the third element. The example sentences given in (8) illustrate each of these cases, with example (8b) containing the attributive -na. The extracted part of each sentence is underlined. Token (8a) was coded as “copula”. Token (8b) was coded as “sentence-final particle”. In the case where the extracted phrase was followed by punctuation, as in example (8c), then the token was coded as “bare”. See Table 2 for the list of grammatical contexts for which we coded.

Table 1. Age and gender distribution of the speakers

<table>
<thead>
<tr>
<th></th>
<th>High school students (15–18 yrs)</th>
<th>University students (19–23 yrs)</th>
<th>Young adults (25–39 yrs)</th>
<th>Middle-aged adults (40–59 yrs)</th>
<th>Elderly adults (60+ yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Women</td>
<td>11</td>
<td>10</td>
<td>12</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

(8) a. Kore gaikokugo mitai yan ke.
this foreign language like COP-SBJV SFP
‘This is like a foreign language, isn’t it?’ (KSJ/043/m/us)\(^4\)

b. Kore hoshii na, demo kaw-are-hen wa, mitai-na sa.
this want SFP CONJ buy-POT-NEG SFP like-ATTR SFP
‘(I was) like I want this, but I cannot afford it.’ (KSJ/058/m/hs)

3. The age boundaries of the age cohorts were chosen to reflect three major life transitions, namely, becoming a university student, graduating from university and getting a job, and retiring.

4. Examples taken from the corpus are followed by brackets containing the following speaker characteristics: corpus code, either KSJ or TKC; identifier number; gender, either female (f) or male (m); and age group, either high school student (hs), university student (us), young adult (ya), middle-aged adult (ma), or elderly adult (el).
   Ahh, been-while like-ATR healthy like-ATR
   ‘(He was) like ahh, long time no see. How are you? Like that.’
   (KSJ/079/f/ya)

Secondly, a total of 1,507 occurrences of na-adjectives other than *mitai* followed by the attributive *-na* were extracted. The following morpheme was also extracted. Examples are given in (9), with the extracted part underlined. Again, we coded each token based on the part of speech of the third element. As we mentioned above, prescriptive grammar requires this element to be a noun. Both of the examples in (9) were coded as “noun”.

(9) a. *Iro-ga na, kirei-na iro ya-ro.*
   color-SUB SFP pretty-ATR color COP-SBJV
   ‘The color is, well, a pretty color, isn’t it?’  (KSJ/098/m/el)

b. *Demo, shūshoku-wa taihen-na jiki yat-ta.*
   CONJ job hunting-TOP difficult-RLZ period COP-PAST
   ‘Job hunting sure was a tough time.’  (KSJ/047/f/ma)

Table 2 lists the distribution of the *mitai* tokens by the following morpheme. For the sake of brevity, we also list the collocation strength for each context. Collocation strength is measure of the strength of association between a linguistic form and a context. We discuss the derivation of the strength values below. Looking at the token counts, we see that the following morpheme in the majority of cases was the attributive *-na* (*N* = 3066). The second most commonly occurring morpheme in this position was a copula, as shown in Table 2.

<table>
<thead>
<tr>
<th>Grammatical construction</th>
<th>Frequency</th>
<th>Collocational strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the attributive <em>-na</em></td>
<td>3,066</td>
<td>infinity*</td>
</tr>
<tr>
<td>Before a copula (<em>ya, yan, desu, da</em>, etc.)</td>
<td>176</td>
<td>0</td>
</tr>
<tr>
<td>Before the postposition <em>-ni</em></td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>Bare</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>Before a sentence-final particle</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>All others</td>
<td>39</td>
<td>n/a</td>
</tr>
</tbody>
</table>

* See footnote 5
3.3 Apparent-time results

Rate of usage
We begin investigating the ongoing changes in mitai-na by first looking at the rate of usage. The rate of usage of mitai-na was calculated for each speaker as the number of occurrences per 1,000 morphemes. Two values were calculated: the rate of mitai-na and the rate of all other tokens of mitai, other than mitai-na. The speaker rates were then averaged for each age group.

Figure 1 presents the results. As can be seen in Figure 1, younger speakers use mitai-na notably more frequently than older speakers. However, this increase only occurs with the expression mitai-na, and the increase seems to level off among those speakers who are between 25 and 40 years old. In contrast, the rate of other usages of mitai shows very little difference between age groups.

We next look at what follows the expression mitai-na. As described in Section 3, the only purpose of the relativizer morpheme -na is to indicate that mitai modifies a following noun. Thus, we expect the vast majority of the tokens of mitai-na to be followed by a noun phrase. The proportion of mitai-na tokens followed by a noun phrase by age group is presented in Figure 2. Indeed, for the eldest group of speakers, over 70 percent of the mitai-na tokens precede a noun phrase. However, this ratio decreases as the speaker’s age become younger, and less than half of the tokens produced by speakers who are young adults precede a noun phrase. The decrease in the number of tokens preceding a noun phrase

![Graph showing the rate of usage of mitai by following morpheme and age group.](image-url)

Figure 1. Rate of usage of mitai by following morpheme and age group. A Chi-squared analysis comparing the token counts against the counts of the high school students shows a significant difference for the elderly group ($\chi^2(2) = 303.7, p < .001$), and the middle-aged group ($\chi^2(2) = 174.8, p < .001$), but not for the young adults ($\chi^2(2) = 0.70, p > .10$).
Kevin Heffernan and Yo Sato

is mirrored by an increase in the number of tokens that immediately precede a phrase boundary (Examples 5 and 8c).

The combined results of Figures 1 and 2 show that the phrase mitai-na has undergone linguistic change between the elderly speakers and the young adult speakers. The change seen in Figure 2 is particularly meaningful, as it implies that the morpheme -na, when following mitai, no longer functions as an attributive modifying a following noun.

It is possible that the results seen in Figure 2 reflect a more general change in the na-adjectives. In order to investigate this possibility, we next examined na-adjectives other than mitai. We calculated the proportion of na-adjective + attributive -na tokens (Example 9) by following morpheme and age group. This results are presented in Figure 3. This figure shows that usage rates do not change from one age group to the next, ruling out the possibility of a more general change going on with na-adjectives in general.

Comparing Figures 2 and 3 for just the elderly speakers reveals that the results are nearly identical. In other words, the usage of mitai-na among the elderly speakers mirrors the usage of other na-adjectives. To the elderly speakers, mitai-na is just another na-adjective, as reported in the descriptive grammars.
Figure 3. Rate of usage of na-adjectives, other than mitai, with attributive -na by following morpheme and age group. A Chi-squared analysis comparing the token counts of the elderly group against the high school students shows that they are not significantly different ($\chi^2(2) = 0.28, p > 0.10$).

The ratio of mitai-na versus other na-adjectives plus -na
We compared the number of occurrences of mitai-na with the number of occurrences of other na-adjectives plus the -na morpheme for each age group. Figure 4 shows the value of the first number divided by the second number. This ratio is the relative frequency of the attributive -na morpheme in two environments: following mitai and following a na-adjective other than mitai. Note that if the ratio is greater than one, then the -na suffix occurs more frequently after mitai than after the combined total of all other na-adjectives. That is indeed the case for the speakers under 40 years of age. A one-way between-subjects ANOVA was conducted on the ratios for the high school students ($M = 1.30, SD = 0.82$), university students ($M = 1.13, SD = 1.23$), young adults ($M = 1.26, SD = 1.24$), middle-aged adults ($M = 0.34, SD = 0.27$), and elderly adults ($M = 0.19, SD = 0.20$). Overall, the groups show significant differences, $F(4, 119) = 11.52, p < .001$. Tukey’s post hoc tests showed that the two oldest groups significantly differ from the three youngest groups ($p < .01$), but the two oldest groups and the three youngest groups do significantly differ amongst themselves ($p > .1$). We argue below that the relative frequency of -na is key to understanding its morphological nature in the speech of the younger speakers.
Kevin Heffernan and Yo Sato

Figure 4. The ratio of the token frequency of the na attributive suffix (mitai versus all other na-adjectives), by age group. The ratio was calculated by dividing the token count for mitai-na by the token count for all of na-adjectives, for that age group.

3.4 What is -na?

So what is the -na morpheme following mitai in the speech of the younger speakers? We will not actually answer that question until Section 5. Instead, let us first consider what -na is not.

Based on the results seen in Figure 2, we have concluded that -na is no longer an attributive marking the relationship between a preceding na-adjective and a following noun phrase.

It is possible that -na is the sentence-final particle na. This particle is the most frequently occurring particle in the corpus data, and it occurs not only at the end of utterances but also inside utterances at constituent boundaries. Thus, it seems to be a likely candidate. However, there are two pieces of evidence that show that -na is not the sentence-final particle na.

Speakers often end utterances with multiple sentence-final particles, such as the combinations ka na and kedo ne. However, there is a strict order hierarchy, and na occurs last. This is borne out by the corpus data. The speakers produced over 8,000 tokens of the sentence-final particle na, and not one of them is followed by another sentence-final particle. In contrast, of the 3,066 tokens of mitai-na produced by the speakers, 56 are followed by a sentence-final particle, such as example (8b). Furthermore, fifteen of those 56 tokens are the sentence-final particle na, resulting in na occurring twice in a row, as in example (10). No Japanese sentence-final particles occur twice in a row, leading us to the conclusion that the
-na following mitai in example (10) is not the sentence-final particle na. Before we address the nature of the na suffix, we first introduce the onbin study.

(10) Tomodachi oru kara it-te-ta mitai-na na.
friend EXIST because go-STAT-PAST like-ATR SFP ‘Like, (I) went because my friend was there.’ (KSJ/062/f/hs)

4. An apparent-time study of changes in the use of onbin

This section presents an apparent-time study of changes in the usage rates of the onbin form of adverbial adjectives. We begin with a brief introduction of the onbin form. We then present the methodology used and the results of the study. We conclude by pointing out similarities between the results presented in this section and the results presented in the previous section.

4.1 The onbin form

The adverbial of an i-adjective is formed by replacing the word’s final /i/ sound with /ku/. Thus the adverbial form of haya-i ‘quick’ is haya-ku ‘quickly’. In the regional varieties of Japanese spoken in western Japan, /u/ is added in place of the Standard Japanese /ku/ morpheme. Furthermore, if the vowel immediately preceding the /u/ is /i/, then the two vowels coalesce to form a long /yu/. Similarly, if the preceding vowel is /a/, then the two vowels coalesce to form a long /o/. Thus, the onbin variant of haya-ku is hayo-o.

The examples given in (11) illustrate the Standard Japanese adverbial forms. The example sentences given in (12) illustrate onbin forms.

(11) a. Ōki-ku kawara-hen kedo sa.
large-ADV change-NEG SFP SFP ‘(They) are not really all that different.’ (TKC/023/m/us)
b. Sore yo-ku aru.
THAT often-ADV EXIST ‘That is often the case.’ (KSJ/015/f/ma)

(12) a. Hayo-o ki-tari oso-o ki-tari suru hito toka.
early-ADV come-REPR late-ADV come-REPR do person such as ‘(There were) people who did things such as show up early or show up late.’ (TKC/028/f/el)
b. Demo sukuno-o nat-ta yo.
CONJ few-ADV become-PAST SFP ‘However, (it) has become fewer.’ (TKC/031/m/ma)
4.2 Methodology

This study uses the same corpus as described in Section 3.2. Two types of data were extracted. Firstly, a total of 9,065 tokens spread across 241 lexical types of uninflected (i.e., dictionary form) i-adjectives were extracted. Secondly, 2,533 tokens spread across 125 lexical types of the adverbial form of i-adjectives were extracted. Each adverbial form was classified as either Standard Japanese or onbin. Altogether, the onbin form occurred 879 times.

4.3 Results

Table 3 lists the most frequently occurring onbin forms. For the sake of brevity, we have also included the collocation strengths in this table. The onbin forms of three adjectives stand out as occurring frequently relative to their corresponding Standard Japanese counterparts: yoi, hayai, and hidoi. However, seven out of ten of the hidoi tokens were not only produced by the same speaker, but were also produced in the same expression: hidoo natta 'has become terrible'. These tokens were excluded from further consideration. We separate out the other two words, yoi and hayai, in our presentation of the apparent-time results.

Table 3. Frequencies of ten adjectives by linguistic form, the ratio of the adverbial form to the uninflected form, and collocation strength between the adjective and the adverbial onbin construction

<table>
<thead>
<tr>
<th>i-adjective</th>
<th>Adverbial form</th>
<th>Uninflected form</th>
<th>Ratio of adv. form to uninfl. form</th>
<th>Collocational strength</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>onbin</td>
<td>SJ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yoi</td>
<td>659</td>
<td>310</td>
<td>24</td>
<td>40.38:1</td>
</tr>
<tr>
<td>hayai</td>
<td>81</td>
<td>74</td>
<td>98</td>
<td>1.58:1</td>
</tr>
<tr>
<td>warui</td>
<td>26</td>
<td>60</td>
<td>331</td>
<td>0.26:1</td>
</tr>
<tr>
<td>shindo</td>
<td>12</td>
<td>25</td>
<td>160</td>
<td>0.23:1</td>
</tr>
<tr>
<td>osoi</td>
<td>11</td>
<td>39</td>
<td>56</td>
<td>0.28:1</td>
</tr>
<tr>
<td>hidoi</td>
<td>10</td>
<td>4</td>
<td>50</td>
<td>0.20:1</td>
</tr>
<tr>
<td>takai</td>
<td>9</td>
<td>33</td>
<td>210</td>
<td>0.89:1</td>
</tr>
<tr>
<td>kitsui</td>
<td>7</td>
<td>14</td>
<td>67</td>
<td>0.31:1</td>
</tr>
<tr>
<td>oï</td>
<td>5</td>
<td>39</td>
<td>574</td>
<td>0.01:1</td>
</tr>
<tr>
<td>kowai</td>
<td>4</td>
<td>7</td>
<td>352</td>
<td>0.01:1</td>
</tr>
</tbody>
</table>

Note. SJ = Standard Japanese; adv. = adverbial; uninfl. = uninflected. The ratio is calculated based on the sum of the two adverbial forms.
Figure 5 shows the usage rate of the *onbin* forms of adverbial adjectives by age group. In general, the youngest speakers show lower rates of *onbin* usage than the oldest speakers, suggesting that the *onbin* form is gradually falling out of use. The Standard Japanese form is now being used instead. That is, when speakers produce adverbial forms, they inflect the base form – except in the case of *yoi* and *hayai*. These two words seem to be resisting standardization.

What is so special about these two words? Anticipating the discussion in the next section, we suggest that it is the relative frequency of the adverbial forms. Table 3 lists the ratio of the sum of the two adverbial forms to the uninflected form for each of the i-adjectives in the table. Two i-adjectives show a relative frequency greater the one – and these are also the same two forms that show a high rate of *onbin* usage. Again, we see that the forms that are displaying unusual behavior have a high relative frequency.

![Figure 5](image)

**Figure 5.** The proportion of tokens that are *onbin* by age group. A Chi-squared analysis comparing the token counts against the counts of the high school students shows a significant difference for the elderly group ($\chi^2(2) = 6.01, p = .049$), and the university students ($\chi^2(2) = 9.11, p < .010$), but not for the middle-aged group ($\chi^2(2) = 2.31, p > .10$).

5. **Measuring strength of association**

In the previous sections, we have argued that speakers treat the form *mitai-na* in a manner that is different to other na-adjectives (contrary to prescriptive grammar). We have also pointed out the unusual behavior of the adverbial forms of the adjectives *yoi* and *hayai*. In order to highlight further the unusual nature of these three forms, we now turn to collostructional analysis (specifically, collexeme analysis,
Gries, Hampe, & Schönefeld, 2005; Stefanowitsch & Gries, 2003). The collexeme metric measures the strength of association between specific lexical items and specific grammatical constructions. We use this tool to assess the strength of association between the word mitai inflected with the attributive morpheme, and between the words yoi and hayai and the adverbial onbin construction. We calculate the strength with the Fisher exact test on a two-by-two table representing the single and joint frequencies of the lexical item and the grammatical construction under investigation (see Stefanowitsch & Gries, 2003, for details).

We require four frequencies for our calculations.

(13) i. The frequency of lexeme L in the specific grammatical construction G.
ii. The frequency of L in grammatical constructions other than G.
iii. The number of occurrences of G other than with L.
iv. The count of all other grammatical contexts other than G, and other than those with L.

The first three items are straightforward; the fourth is not. Consider the strength of association between hayai and the adverbial onbin construction. In this case, the first item in (13) is simply the number of occurrences of the adverbial onbin form of hayai. The second item is all other occurrences of hayai, such as the Standard Japanese adverbial form hayaku, the base form hayai, the negative form hayakunai, etc. The third item is all occurrences of the onbin adverbial form of all adjectives, excluding the adjective hayai.

That brings us to the fourth item. An adverbial form modifies a verb. Therefore, we use the number of main verbs produced by the speakers as a close approximation of the number of relevant grammatical contexts. Once we have these counts, we next create a two-by-two frequency table, such as Table 4.

Table 4. The single and joint frequencies of hayai and the adverbial onbin construction

<table>
<thead>
<tr>
<th>Adjective</th>
<th>Adverbial onbin</th>
<th>All others</th>
</tr>
</thead>
<tbody>
<tr>
<td>hayai</td>
<td>81</td>
<td>156</td>
</tr>
<tr>
<td>all others</td>
<td>798</td>
<td>115,385</td>
</tr>
</tbody>
</table>

Finally, we calculate the strength metric as the negative logarithm to the base ten of the p-value of the one-sided Fisher-Yates Exact test for this table (Gries et al.,
2005), using the statistical software package R.\(^5\) This yields the result of 99.8. A higher value indicates a stronger association between the specific lexeme and the construction.

We considered the strength of association from two perspectives. First we calculated the strength of association between a single lexeme – in this case, *mitai* – and various grammatical constructions. Our results are listed in Table 2.

We also calculated the strength of association between a single grammatical construction – in this case, the adverbial *onbin* form – and different adjectives. Our results are listed in Table 3.

In both cases, we see a notable jump in the strength values between the three forms under investigation and the other forms; in the case of *mitai*, that jump is extreme. The remarkable difference between the strength scores for the three holistic forms and the other forms further supports our conclusions. Clearly, speakers are treating these three forms in a manner that differs from how they treat the other forms.

6. The role of relative frequency in the formation of morphological paradigms

The objective of the paper is to investigate the holistic processing of complex linguistic forms. We defined holistic processing as the lack of internal morphological structure. Based on apparent-time data, we argued that the morphological boundary between *mitai* and the attributive *-na* in the phrase *mitai-na* has disappeared, and therefore the phrase is now processed in a holistic manner.

We further supplemented the argument with data on the standardization of the adverbial adjective form. We showed that the adverbial forms for two words are resisting standardization. We have suggested that relative frequency is the key
to understanding the results, and pointed out that the linguistic forms that display unusual behavior have a high relative frequency.

As discussed above, researchers have previously argued that holistic processing occurs because the highly frequent repeated use of the same phrase leads to the different parts fusing together. Those researchers presented faster processing times as evidence. Others argue that the faster processing time results can be accounted for by a model of lexical priming.

This paper attempts to reconcile this rift by bringing to the table new evidence. Specifically, we consider the inflectional paradigm of the na-adjective mitai. Prescriptive grammars dictate that, as a na-adjective, mitai takes the attributive suffix -na only when modifying a following noun. As shown in Figure 2, clearly this is no longer the case in the vernacular Japanese of speakers under the age of 40 years. Fewer than half of the mitai-na tokens that they produced are actually followed by a noun. The na-adjective inflectional paradigm no longer applies to the word mitai.

Now consider the expression mitai-na. This expression occurs at the end of utterances, and may be followed by a sentence-final particle in that position. Thus, the expression mitai-na behaves as a single unit that is processed holistically.

In order to understand why mitai-na is processed as a single unit, we now turn to recent work on language acquisition. The first steps of language acquisition occur at the phrase level, not the word/morpheme level (Bannard & Lieven, 2012; Tomasello, 2003, 2009; Wray, 2002, Chapter 6). Children learn to say ‘There ya go’ without understanding the individual components of the phrase (Tomasello, 2009, p. 73). It is only after a long process of abstraction involving first entrenchment through repetition, and then reanalysis through analogy, that the child finally acquires the individual parts of complex phrases.

Returning back to the example of mitai-na, according to the above claim that the phrase is acquired first, we can infer that Japanese children acquire the attributive -na by first acquiring a collection of na-adjectives plus -na plus noun phrase. It is only after children have reanalyzed the complex phrases into parts that they acquire the more abstract schema consisting of na-adjective plus -na plus noun phrase.

The reanalysis of a whole into its parts seems not to occur with the phrase mitai-na. But what would block the reanalysis? We posit that it is relative frequency. Specifically, when a complex form occurs more frequently than its components, then the reanalysis into parts is blocked, or at least impeded. Thus, the key is relative frequency.

The importance of relative frequency has been pointed out repeatedly. Bybee (1985) discusses the relationship between irregular forms and high frequency. Irregular forms, such as the past tense form ‘went’ for the verb ‘go’, and the plural form ‘men’ for the word ‘man’, tend to be frequently occurring words. Bybee suggests that the frequency of the irregular form must be considered in two
dimensions. The first dimension is absolute frequency of the form. The second dimension is the relative frequency “within the paradigm” (Bybee, 1985, p. 120).

Corbett, Hippisley, Brown, and Marriott (2001) extended Bybee’s claim to the plural paradigm of nouns in Russian. They conducted a frequency-based corpus study on over 5,000 different noun lexemes, and examined each lexeme’s plural paradigm for inflectional irregularity and prosodic irregularity. Their results show that “relative frequency of occurrence in the plural appears to be important where non-prosodic irregularity is concerned” (Corbett et al., 2001, p. 218).

Hay (2001, pp. 1059–1060), based on her results from a simple experiment that elicited impressionistic judgments of morphologically complex words, concluded:

The literature widely assumes a direct link between high lexical frequency, non-decomposability [i.e., holistic processing], and nontransparency. However, such a direct link is in fact not predicted by models of morphological processing. Rather, processing models appear to predict a link between nondecomposition and relative frequency – the frequency of the derived form relative to its base.

Haspelmath (2008) argues that some of the motivating factors behind the claim for the role of iconicity in language are in fact better explained in terms of relative frequency. Among these is the principle of economy: “The more predictable a sign is, the shorter it is” (Haspelmath, 2008, p. 2). Haspelmath points out that frequency implies predictability and, based on this, argues that the principle of iconicity is better expressed as the more frequent a word is, the shorter it is. He then extends this principle to morphologically complex forms: more complex forms are less frequent than less complex forms. Crucially to our discussion, Haspelmath clarifies: “It is important to recognize that the relevant type of frequency … is relative frequency, not absolute frequency” (Haspelmath, 2008, p. 9).

Heffernan and Hiratsuka (in press) investigated the relationship between stylistic variation and verbal negative suffix choice in the Corpus of Kansai Vernacular Japanese. In vernacular Japanese, the verbal negative suffix varies by region, and differs from standard, informal Japanese. Heffernan and Hiratsuka found strong correlations between six different stylistic variables and verbal negative suffix choice – except for a few specific contexts, such as following the verb stem shir-‘know’. Speakers overwhelmingly use a single variant, in this case, shira-n ‘not know’. They point out that the common characteristic of the unusual forms such as shira-n is their high relative frequency compared to the other forms in their paradigms. They argue that the high relative frequency of these forms has led to them being processed in a holistic manner.

How then does a high relative frequency of a derived form lead to holistic processing? Bybee (2008, p. 225), based on her work on the diachronic emergence of new morphemes in languages around the world, explains:
Token frequency plays a role in morphology by making the high frequency forms of a paradigm the anchoring points for other forms. Lower frequency forms can be analyzed and learned in terms of these robust forms, creating a relationship of dependency.

Let us now consider the processing of the three forms (mitai-na, and the adverbiaal forms of adjectives yoi and hayai) that we focused on in this paper. All three of these forms have a high relative frequency. The phrase mitai-na occurs more frequently than its components do separately. The adverbial forms for yoi (that is, yoku and yō) occur more frequently than the uninflected, dictionary form yoi. Similarly, the adverbial forms for hayai (that is, hayaku and hayō) occur more frequently than the uninflected, dictionary form hayai. Thus, young Japanese children do not develop a robust relationship of morphological dependency for these forms, and they continue to be processed in a holistic manner.

In the case of mitai-na, holistic processing suppresses the morphological characteristics of the -na morpheme. In the case of the standardization of the adverbial forms, it is the morphological rule governing the addition of the -ku suffix that is being standardized. But this rule is not applied to the adverbial forms of yoi and hayai at the time of processing because they are processed in a holistic manner, and not as a morphologically complex expression consisting of adjective plus suffix.

7. Conclusion

We have highlighted three forms that show morphological irregularity in vernacular Japanese: mitai-na, and the adverbiaal forms of the adjectives yoi and hayai. As Japanese is prototypically agglutinative, the morphological behavior of these forms is conspicuous. We pointed out that these forms display a high relative frequency compared to their constituents/uninflected forms. Based on this observation, we argued that their high relative frequencies impeded the formation of a morphological relationship and the inclusion of these forms in morphological paradigms. Thus, we concluded that these forms are processed as if they were monomorphemic forms.

The irregular adverbial forms are leftover from an obsolete system, in the same way that many English irregular past forms are leftover from the Germanic strong verb system. In contrast, the irregular form mitai-na emerged from and competes with the regular inflection paradigm for mitai, illustrating a previously undocumented path for the diachronic emergence of irregular morphology. However, this is only one path, and there are most likely other factors. One such factor that has been investigated repeatedly is absolute frequency, but at this point
the relationship between absolute frequency and irregular morphology remains unclear. For example, Michel et al. (2011) examined the regularization of the past tense forms in written English between 1800 and 2000. They found that lower frequency past tense forms tend to regularize but higher frequency ones do not. In contrast, Fratini, Acha, and Laka (2014) did not find a clear relationship between absolute frequency and irregularity of verb forms in a corpus study of contemporary Spanish. Clearly, more research needs to be done if we wish to sort out these seemingly conflicting results. One possible avenue of such research is to reexamine the relationship between irregular verb forms and frequency by expanding the scope of the studies to include relative frequency.

One more avenue of research that needs to be pursued is research on non-European languages. The vast majority of frequency-related research has been done on European languages, in particular English. Studies such as this one that look at the same linguistic phenomenon through a different lens are able to corroborate and refine results obtained by the research done on European languages, and we hope to see more of such studies in the future.

Acknowledgements

We would like to thank the audiences at NWAV-AP2 (Chiayi, Taiwan) and JLVC2016 (Tokyo, Japan), and the two anonymous reviewers for helpful comments and suggestions; any remaining shortcomings are our own. We are also indebted to the many students who have spent countless hours gathering and editing the corpus data that now forms the Corpus of Kansai Vernacular Japanese. Anyone interested in using the corpus for their own research should contact the first author.

References


Relative frequency and the holistic processing of morphology


Tremblay, Antoine, & Baayen, Harald (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In David Wood (Ed.), *Perspectives on formulaic language*: 

拙論では表現「みたい」の形態における通時的変化について論じる。経時データに基づき、当該表現においては、語幹「みたい」と体言接続語尾「な」の境界が消失し、単一の複合語が形成されつつあることを示す。データ分析にあたっては相対頻度がその論拠を構成する。

さらに、相対頻度の重要性を裏付けるため、関西弁における形容詞語尾の標準語化を取り上げる。同方言では、若年層において連用語尾の標準語化が進み、方言語尾「う」よりも「く」が多用されるが、「はやい」と「よい」に限っては「う」が残り、標準語化がさほど見られない。これは「みたい」同様、「はよう」と「よう」が残り、標準語化がさほど見られない。これは「みたい」同様、「はよう」と「よう」の相対頻度が高いことと関連していると考えられる。これらの現象から、形態素の連鎖の相対頻度が各構成要素と比較して高くなると、形態的融合に繋がると論じる。つまり、「はよう」と「よう」における古い形態の残存は、英語動詞の不規則過去形がゲルマン語強変化の名残であるのと同様の現象であると考えられる。一方「みたい」については、本来の活用パラダイムが崩れ、逆に一定の活用形が頻度の高まりによって単一語化する過程と見られ、この点で不規則形態が生じる過程について新鮮な観点を提供するものといえる。

Address for correspondence

Kevin Heffernan
Kwansei Gakuin University
Sanda City, Hyogo
Japan
kevin@kwansei.ac.jp