

RESEARCH

Showcasing the interaction of generative and emergent linguistic knowledge with case marker omission in spoken Japanese

Kevin Heffernan, Yusuke Imanishi and Masaru Honda

Kwansei Gakuin University, 2-1 Gakuen, Sanda, 669-1337, JP

Corresponding author: Kevin Heffernan (kevin@kwansei.ac.jp)

Linguists debate the nature of grammatical knowledge. Many argue it is innate knowledge of syntactic structure that we use when generating utterances; others argue it emerges from linguistic experience, and forms exemplars for modeling novel utterances. Yet, still others argue that grammatical forms are processed in parallel by both types of knowledge (innate or otherwise), and crucially, that these two processing routes compete with each other.

Our objective is to support the dual route argument with a corpus study illustrating the interaction of these two types of knowledge. We interpret two proxies as indicators of these two types of knowledge: syntactic complexity for generative knowledge and dispersion for emergent knowledge. Previous psycholinguistic work has shown that increased syntactic complexity correlates with increased judgment reaction times. Conversely, increased dispersion correlates with decreased judgment reaction times. If these two processing mechanisms compete, then we predict an interaction: specifically, we hypothesize that the more dispersed a linguistic form is, the less influence syntactic complexity has.

We conducted a mixed-effects logistic regression analysis on case marker omissions in Japanese. Our results show that a casual speech style, a dispersed object–verb pair, and a syntactically simple noun phrase for the object correlate with increased case marker omission. More importantly, syntactic complexity and dispersion interact: as dispersion decreases, the estimated coefficient for syntactic complexity increases. These results support the claim that generative knowledge and emergent knowledge compete during language processing.

Keywords: grammatical case; Japanese; emergent grammar; dispersion; frequency

1 The emergent–generative debate

Linguists are currently debating the nature of our grammatical knowledge. Many have argued that such knowledge is innate and that we use this knowledge to generate utterances; others have argued that grammatical knowledge emerges from linguistic experience, and that such knowledge forms exemplars for modeling novel utterances. Generative grammar advocates emphasize the way that utterances are generated by combining individual words as needed (Hauser et al. 2002; Newmeyer 2010, and many others). In contrast, emergent grammar advocates emphasize that prefabricated sequences of abstract and concrete linguistic forms are recycled (Thompson 2002; Ellis 2008; Beckner et al. 2009; Bybee 2010; Dąbrowska 2014, and many others; see MacWhinney & O’Grady 2015 and references therein).

Consider Dąbrowska’s (2010) example (9a), repeated here as (1a). This sentence illustrates the long-distance dependency between the *wh* word in the main clause and the trace in the object position for the verb *do* in the subordinate clause. Dąbrowska points

out that word choice in sentences with long-distance dependencies in natural conversation is constrained; that is, the main clause auxiliary in such sentences is *do* over 90% of the time, the subject is *you* about 90% of the time, and the verb is *think* or *say* about 90% of the time. Dąbrowska gave several more examples that illustrated these tendencies, and we have repeated two of them as (1b) and (1c). She concludes that utterances with long-distance dependencies followed specific templates most of the time, and emerged frequently because “they are psychologically more basic than non-prototypical ones... prototypical questions are produced more fluently, judged to be more acceptable, remembered better and acquired earlier by children” (Dąbrowska 2010: 63).

- (1) Dąbrowska (2010)
- a. What do you think you are doing?
 - b. Who do you think you are?
 - c. What do you think that means?

Such skewed distributions occur in other patterns. For example, Hopper (2015: 320) reported that in pseudo-cleft constructions such as *what I have to do is...*, the verb was either *do* or *happen* a majority of the time. Based on such observations, a new model of grammar has emerged that emphasizes the reuse of prefabricated speech fragments (Thompson 2002; Ellis 2008; Bybee 2010; Dąbrowska 2014; 2015; Ellis et al. 2015). The main idea of this approach is summarized concisely as “grammar may best be understood as combinations of reusable fragments” (Thompson 2002: 141).

But others disagree. Newmeyer (2010: 15, emphasis in original) stated, “I have no problem with the idea that many of the more commonly-used phrases are stored in memory, but the idea that *a grammar* might be a stock of fragments strikes me as utterly implausible.”

Yet, the evidence for emergent grammar continues to accumulate. After reviewing universal grammar examples in language acquisition, Ibbotson & Tomasello (2016: 74) concluded that “the notion of universal grammar is plain wrong”; rather, it was felt that “[i]n its place, research on usage-based linguistics can provide a path forward for empirical studies of learning, use, and historical development” (Ibbotson & Tomasello 2016: 75).

At the same time, the evidence for generative grammar continues to accumulate. A recent neurophysiological brain activity study (Nelson et al. 2017) claims to have demonstrated the neurophysiological reality of phrase structure. In the authors’ opinion, their results “strongly motivate the view of language as being hierarchically structured and make it increasingly implausible to maintain the opposite nonhierarchical view” (Nelson et al. 2017: E3676).

And the debate continues.

Thompson, Newmeyer, and many others, agree that commonly-used fragments of speech are stored in memory, and that those fragments are reused during language production. What they do not agree on is the importance of such phrases. The proponents of the emergent grammar approach argue that such phrases, along with more abstract templates derived from those phrases—which are collectively called exemplars—are the end result of the acquisition of grammar knowledge. Through cumulative language experiences, we track and store the relative importance of each exemplar, perhaps as knowledge reflected by the exemplar’s distinctiveness (Adelman et al. 2006) or relative weight (McClelland 2015). Exemplar knowledge and the respective knowledge of distributions form our grammar knowledge.

In contrast, proponents of the generative grammar approach argue that grammatical knowledge is derived from Universal Grammar and an associated set of parameters. They

emphasize the role of syntactic structure—particularly its hierarchical nature—during language processing.

Exemplar grammar models explain many linguistic usage patterns; for example, more predictable words, phrases, and patterns are shorter (Bell et al. 2009; Seyfarth 2014), articulated less clearly (Bell et al. 2003), and show increased omission of optional functional words such as the complementizer *that* (Jaeger 2010). On the other hand, exemplar models struggle to explain grammatical knowledge that is clearly not based on previous language experience, such as grammatical innovations by children during language acquisition which do not have target language exemplars (Singleton & Newport 2004; Culbertson & Newport 2015).

Still others take a third approach that attempts to strike a balance between a completely generative approach and a completely usage-based approach. A characteristic example is the processing determinism approach by O’Grady (2008; 2015). This approach emphasizes the role of processing pressures during language formation, and “assigns a significantly smaller role to the input than is common in usage-based theories and no role at all to Universal Grammar” (O’Grady 2015: 8). O’Grady argues that there are two sources that shape language acquisition. The first source is external factors that arise from the language environment, such as frequency of occurrence. The second source is “internal pressures that stem from the burden that particular computational operations place on working memory” (O’Grady 2015: 9). Note that the computational operations are not necessarily innate by design, and O’Grady argues against the need for faculty-specific principles that cannot be learned from experience. What differentiates this approach from an emergent grammar approach is the role of experience-based knowledge: in the words of O’Grady (2008: 623), such knowledge “does in fact have a very important role to play in understanding how language works—but only if we acknowledge that its effects are modulated by the efficiency-driven processor that is at the heart of the language faculty.” It is this style of approach that we support with this work.

In summary, evidence supports both sides of the emergent–generative grammar debate, with many researchers arguing in favor of one approach. Still others argue for an approach that combines elements of the emergent approach with elements of the generative approach. Our goal is to contribute to this debate by supporting the third approach. We demonstrate that both emergent knowledge and generative knowledge play a role in speech processing, and more importantly, that these two types of knowledge compete. For that purpose, we present a corpus study of the variable omission of object case markers in spoken Japanese.

Standard Japanese obligatorily marks the relationship between the noun and the verb with a case marker immediately following the noun phrase, as in examples (2a) and (3a). However, in casual spoken Japanese, the case marker is variably omitted without any change in meaning, as in examples (2b) and (3b).

- (2) a. Gohan-**o** tabe-ta.
meal.ACC eat.PAST
‘(I) ate lunch.’
b. Gohan-**∅** tabe-ta.
- (3) a. Gakkō-**ni** it-ta.
school.DAT go.PAST
‘(I) went to school.’
b. Gakkō-**∅** it-ta.

Previous work on variable case marker omission in Japanese has focused on either the accusative case or the nominative case (Fujii & Ono 2000; Matsuda 2000; Fry 2003; Yoshizumi 2016, *inter alia*). In this study, however, as the focus is on object–verb pairs, we do not examine the nominative case. However, we expand previous work by extending the scope of our study to the dative case, as seen in the examples in (3).

We proceed as follows. This study was motivated by Ullman’s (2004; 2008; 2016) declarative/procedural model of language. Therefore, we begin with a review of it. We then look more closely at Japanese case marker omissions, and present specific hypotheses. We subsequently present our corpus study on Japanese case marker omission, after which we discuss the implications of our results on the emergent–generative grammar debate.

2 The dual route model of language processing

Michael Ullman, a neurobiologist who researches language and memory, argues that language processing depends primarily on the two memory systems in the brain: declarative memory and procedural memory (2004; 2008; 2016). Briefly, declarative memory underlies the rapid learning and linking of different bits of knowledge across multiple contexts and modalities, such as the flavor of strawberries with the word strawberry. This knowledge is time- and context-specific. Declarative knowledge is learned rapidly, after as little as a single exposure, but strengthens with additional exposures. Declarative memory plays a crucial role in the acquisition of words, including their phonological forms, meanings, and subcategorization patterns, as well as relationships between words. Declarative memory also stores our knowledge of multi-word phrases, idioms, collocations, and irregular inflectional patterns, such as their frequency and patterns of use. Furthermore, declarative memory is capable of acquiring and abstracting such patterns into schemas such as those posited by construction grammar (Hoffman & Trousdale 2013).

On the other hand, procedural memory underlies the learning and processing of implicit sequences and rules. These rules are most likely probabilistic in nature, rather than deterministic. Procedural memory plays a key role in hierarchical and probabilistic sequencing across all linguistic subdomains, from syntax to phonology. Procedural memory also facilitates structural processing with a mechanism for predicting.

These two systems both overlap and compete. To some extent, both memory systems acquire the same or analogous knowledge. Thus, declarative memory may contain the past tense form *walked* while procedural memory contains the rule-based computation to derive the corresponding form *walk* + *ed*. During processing, both forms are accessed in parallel, with the faster form winning out: “Access to a stored representation which has similar mappings to one which could be composed compositionally... would block completion of the latter computation” (Ullman 2004: 247).

The concept of blocking may turn out to be exaggerated. Nevertheless, the concept of parallel processing routes working at different speeds has been supported by recent work in morphology (Schmidtke et al. 2017; Lõo et al. 2018), and it is this aspect of Ullman’s model that we test in this paper.

Schmidtke et al. (2017) conducted a series of lexical decision and eye movement experiments on English and Dutch morphologically derived (*teach* + *er*), pseudo derived (*corn* + *er*), and simple words. They measured a variety of frequency, morphological, orthographic, and semantic variables, and then used a statistical technique called survival analysis to determine the earliest point in time at which a given variable exerts an influence over their response variables (judgment reaction times and eye fixation durations). They found that in general the first variables to begin to show effects were frequency, followed by semantic, then finally the morphological and orthographic. Their results show that speakers make use of frequency information long before they start making use of morphological

information. They point out that their results support “dual- or multi-route theories of morphological processing, which argue for a parallel and interactive use of properties associated with whole words and their morphemes” (Schmidtke et al. 2017: 18).

Lõo et al. (2018) conducted a large-scaled word naming study of Estonian nouns. They measured the delay between stimulus presentation and beginning of articulation (production latency). They placed particular emphasis on three factors: whole-word frequency, inflectional paradigm size, and morphology family size. They broke up the production latency data into quantiles, so that the fastest 10% of the responses formed the first quantile, the next fastest 10% the second quantile, etc. A quantile regression analysis then showed at what point each of these three factors had the strongest effect. They found that the effect of whole-word frequency peaked in strength at the fourth quantile, followed by morphological family size in the sixth quantile, and then finally inflectional paradigm size in the eighth quantile. Their results show that whole-word frequency has its strongest effect when words are processed relatively rapidly, whereas inflectional paradigm size has its strongest effect on words that are processed relatively slowly.

The results from both of these studies mesh neatly with Ullman’s claim for parallel processing routes. Furthermore, it seems that declarative knowledge (frequency and semantic information) is fast whereas procedural knowledge (inflectional paradigm patterns) is slow. If declarative and procedural knowledge compete, then we expect to see an interaction during language production. Specifically, we expect that the more a linguistic form relies on declarative knowledge, the less of a role procedural knowledge plays. Since declarative knowledge is fast, if it suffices then it will do so quickly. If it does not prove to be enough, then and only then will the slower procedural knowledge fully engage. In the rest of this article, we test this expectation by looking for such an interaction in spoken Japanese.

Many others similarly argue for the engagement of both types of knowledge, from a range of research areas such as phonetics (Pierrehumbert 2016), phonology (Guy 2014), derivational and inflectional morphology (Baayen et al. 1997; Vannest et al. 2005; Kuperman et al. 2009; Bakker et al. 2013), theoretical syntax (Reuland 2010), first language acquisition (O’Grady 2015), language evolution (Reuland 2010; Miyagawa 2017), and formulaic language (van Lancker Sidtis 2012). Some of these researchers propose models that a quite similar to the declarative/procedural model. For example, in spite of the notably different terminology, O’Grady’s external factors and internal factors resemble declarative knowledge and procedural knowledge.

2.1 Dispersion as a proxy of declarative knowledge

One of the most consistent and robust effects found in psycholinguistics studies has been the frequency effect: words and phrases that occur more frequently are processed faster, as shown by word judgment reaction times (Forster & Chambers 1973; Ellis 2002; Brysbaert et al. 2018). Each time a linguistic form is heard, it is imprinted in declarative memory, and the more often the imprinting occurs, the easier it is to recall that word or phrase.

However, several researchers have argued that the correlation between lexical frequency and the processing speed of a word is confounded by dispersion (McDonald & Shillcock 2001; Adelman et al. 2006; Perea et al. 2013). Dispersion refers to the number of different contexts that word is seen in. In practice, context tends to be defined as a single block of text or speech, such as a novel, textbook, or conference presentation, in which the target word appears one or more times. Research on memory has shown that the memory benefit from repeated exposure to a stimulus diminishes unless the context of the stimulus is changed (e.g., Verkoeijen et al. 2004). Work on dispersion has repeatedly

shown that it is more predictive of word judgment reaction times than word frequency. This effect remained even after considering the effect of the covariates such as ambiguity, word length, and orthographic neighborhood size.

Since more dispersedly-used words are processed faster, we take dispersion as a proxy measure of the cost to process in declarative memory. Declarative memory has the capacity to store and process multiword phrases (Bannard & Matthews 2008; Tremblay et al. 2011). With each such processing, the cost of doing so by declarative memory reduces. If such a phrase is processed frequently, then the cost of processing in declarative memory becomes lower than the cost of processing in procedural memory. In extreme cases, multipart expressions phonetically and cognitively merge together to form single units (Bybee & Scheibman 1999; see Heffernan & Sato 2017 for Japanese examples). We expect the most dispersed object–verb pairs to show the least evidence of processing in procedural memory.

2.2 Grammatical complexity as a proxy of procedural knowledge

We indirectly observe processing by procedural memory by building on the observation that “sentences that have more complex syntactic structures are more difficult and time consuming to understand” (Caplan & Waters 1999: 79; see also Scontras et al. 2015). Several theories have been posited to account for the increased processing costs of complex sentences. We espouse the Dependency Locality Theory (Gibson 2000). This theory equates differences in sentence comprehension performances with differences in the cost associated with integrating the syntactic structure of an incoming word into the current structure held in memory. For our purposes, it is sufficient to note that the cost is related to the size of the structure held in memory. Gibson measures integration costs in arbitrary energy units, with one unit equal to one discourse referent. As he was primarily concerned with the different costs incurred when comprehending subject-gap relative clauses and object-gap relative clauses, his theory does not assign costs to non-discourse referents such as determiners and adjectives. Yet determiners and adjectives must also be held in memory before integration. Our preliminary analysis of our data (see Table 1) by modifying constituents showed a clear difference between bare nouns, nouns modified by determiners, and nouns modified by more complex structures such as relative clauses. For the sake of simplicity, in this paper we equate the complexity of a noun phrase with the number of constituents modifying the projection of the head. Thus, in the phrase *the flower*, one constituent (the DP) modifies the head, giving the phrase a complexity value of one. In comparison, the bare noun *flower* has a complexity value of zero, *this flower* has a complexity value of one, and *freshly picked flowers* has a complexity of two. We give specific Japanese examples in Section 4.

Table 1: Token counts and omission rates by modifier type.

Modifier type	Token counts	Omission rate	Reclassification scheme
Bare	10,370	73.2%	0
Determiner	757	65.7%	1
I-adjective	239	62.3%	1
Genitive NP	949	55.8%	2
Na-adjective	101	54.5%	2
Relative clause	258	51.9%	2
Na-adjectival phrase	118	43.2%	2

Our primary hypothesis is that declarative knowledge and procedural knowledge compete during language processing. We are taking dispersion as a proxy of declarative knowledge. Similarly, we take the number of modifying constituents as a proxy of procedural memory. In other words, we hypothesize that dispersion and number of modifying constituents interact. We elaborate how once we have introduced previous work on case particle omission in Japanese.

3 Japanese case marker omission

In this section, we first review four recent studies of variable Japanese case marker omission. We then present our specific case marker omission hypotheses.

3.1 Previous work on case marker omissions in Japanese

Fujii & Ono (2000) examined accusative case marker omissions in 40 minutes of speech made up of seven short conversations. They found that speakers tended to omit accusative case markers in the following situations:

- In idiomatic expressions, such as *me no iro kaeru* ‘get serious; lit., change eye color’
- With direct objects that were demonstratives (*kore* ‘this’), indefinite pronouns (*nanka* ‘something’), and interrogative pronouns (*dare* ‘who’)
- With direct objects that occurred adjacent to the verb
- With direct objects with specific referents in the conversation.

From these observations, they concluded that first, speakers used the accusative case marker as a rhetorical device to draw attention to important information, such as the discourse topic, something newsworthy, or part of a contrast, and second, speakers used the accusative case marker to “facilitate the processing of information that may require some cognitive effort on the part of the listener” (Fujii & Ono 2000: 28). We build on this second conclusion in this study.

Matsuda (2000) investigated variable accusative case marking in 37 conversations between Tokyo Japanese speakers using variable rule analysis. Similar to Fujii & Ono, Matsuda found the strongest effect for verb adjacency. He also found a significant effect for object type, with objects showing greater case marker omission in the following order: nouns modified by relative clauses < non-interrogative pronouns < other nouns < interrogative pronouns. Matsuda also investigated the effect of an object having a previous referent in the conversation, but did not find a significant effect.

Fry (2003) investigated ellipses and case marker omissions in short excerpts from 120 telephone conversations between Japanese residing in the United States and close acquaintances residing in Japan, and found that accusative case marker omissions significantly increased with interrogative pronouns, shorter utterance lengths, and objects adjacent to the verb; however, speaker gender, dialect, word length, and object animacy were not found to be significant.

Yoshizumi (2016) examined case marker omissions for subjects and direct objects in conversational interviews with 16 heritage Japanese speakers and 16 native Kansai Japanese speakers. Using variable rule analyses, she found that a focus particle and a direct object that occurred adjacent to the verb significantly correlated with omissions; however, gender, age, the presence of sentence-final particles, and whether the clause contained an object–verb pair in a subordinate or a main clause were not found to be significant.

These four studies all focused on accusative case markers. We build on these studies by expanding the scope of our investigation to also include dative case markers. As

introduced in the example sentences in (3), speakers also variably omit the dative case marker in conversation.

3.2 Our case marker omission hypotheses

Our hypotheses are grounded on the claim that case marker omission in spoken Japanese correlates with processing effort. Previous work on case particle omissions in Japanese supports this assumption: case markers tend to be omitted when the object is closer to the verb, when the object is not modified by a relative clause, and in shorter utterances. This assumption is also supported by Fedzechkina et al. (2017), a miniature artificial language study on case marker omission. They contrasted the learners acquiring an artificial language with fixed order against one with variable order, from which it was found the first group omitted case significantly more frequently. Fedzechkina et al. argue that their results reflect the learners' attempts to balance effort with information transmission. Interestingly, the learners consistently produced case markers more often in OSV sentences than in SOV sentences. The authors posited several possible explanations for this imbalance, but needless to say all their explanations involved the role of case markers in reducing cognitive processing effort.

Our primary hypotheses are as follows. Our first hypothesis builds on the observation that phrases that occur more frequently (that is, more dispersed) are processed faster, which we take as an indication of less cognitive effort. Similarly, case particles tend to be omitted from utterances that require less cognitive effort. Therefore, we hypothesize that increased case marker omission correlates with increased dispersion.

Second, given that more syntactically complex constituents require more processing time, we hypothesize that increased case marker omission correlates with an increase in the number of constituents modifying the object of the object–verb pair. Matsuda's (2000) results support this claim.

Third, we hypothesize that the above two correlations interact. If, as argued by Ullman, rapid processing by declarative memory preempts processing by procedural memory, then we should find a stronger correlation with number of modifier constituents for object–verb pairs that have low dispersion values. First consider object–verb pairs with high dispersion values. A high dispersion value indicates that an expression is used widely in a variety of contexts. Such phrases are presumably processed rapidly by declarative memory. During processing, the faster declarative memory processing wins out, and the slower procedural processing is aborted. Since number of modifying constituents is our proxy measure of processing in procedural memory, and procedural processing does not complete in these cases, then the correlation between case marker omission and number of modifying constituents should be weak.

Now consider object–verb pairs with low dispersion values. In these cases, declarative memory processing is relatively slow, and therefore the faster procedural memory wins out. If procedural memory is sensitive to syntactic complexity, then we should see a strong correlation between the number of modifying constituents and case marker omission for these cases.

Besides dispersion and number of modifying constituents, we also consider the following factors: speaker gender, speaker age, and speech style. Previous work on accusative case marker omissions did not find gender to be a significant factor, and we also predict that gender does not correlate with case marker omissions. However, we still include this factor for thoroughness. Similarly, we predict that age does not correlate with case marker omissions.

Given that case marker omissions occur in casual speech (Tsutsui 1984), we predict that case marker omissions correlate with speech style; specifically, speakers who use a more vernacular style omit case more often.

As mentioned, our study includes not only transitive verbs, but also unaccusative and unergative verbs. Due to the lack of studies on variable case marker omission for verbs other than transitive verbs, this aspect of our study is exploratory in nature, and it is difficult to a priori hypothesize about case marker omission patterns in other verb types.

4 Methodology

For our data, we used the *Corpus of Kansai Vernacular Japanese*, which consists of 150 sociolinguistic interviews conducted by university students attending a private university in the Kansai region of Japan. The interviewees were either family members or close acquaintances of the interviewers. Both the interviewers and the interviewees were Kansai Japanese native speakers. Interviewers were instructed to speak in a casual manner using the local Japanese vernacular. They talked about a wide range of freely chosen topics such as school life, dating, work, family, history, and tragic events, for approximately one hour. Each interview was transcribed, checked for accuracy, parsed at the morpheme level, and tagged with part of speech information using MeCab (Kyoto University Graduate School of Informatics & Nihon Telegraph and Telephone Corporation 2013). Collectively, the interviews comprise 1.71 million lines of data (morphemes plus punctuation), of which 1.06 million lines were produced by the interviewees. We only examine the interviewees' speech (hereafter, speakers). Table 2 shows the age and gender distribution of the speakers.

Our initial investigation revealed a highly skewed distribution of the linguistic contexts. For example, bare nouns occurred over one hundred times more frequently than nouns modified by na-adjectives (see below for examples of these categories). To maximize the volume of rare linguistic categories such as nouns modified by na-adjectives, we extracted all applicable data using a semi-automated process. Every token that met the criteria described below was first extracted and classified with a Python programming language script. We then hand checked three hundred randomly selected tokens, and modified the script to account for consistent errors. We repeated this process until the data was as error free as possible.

4.1 Extraction criteria

For practical reasons, we limited the study to the most frequent verbs only. We classified the 500 most frequently used verbs by verb type (transitive, unaccusative, or unergative), excluding any verbs not of these types, such as verbs of existence. We limited the list to transitive verbs that naturally occur with direct objects marked with the accusative case marker *o*, and unaccusative verbs and unergative verbs that naturally occur with indirect objects marked with the dative case marker *ni*. For example, since the stative transitive verb *wakaru* 'understand' does not occur naturally with the accusative case marker *o*, we omitted it.

In line with previous research on case marker omission (Matsuda 2000; Yoshizumi 2016), we omitted Sino-Japanese verbal nouns such as *benkyō* 'study' when they occurred with the verb *suru* 'do, play.' We included other nouns with *suru*, such as *tenisu* 'tennis.' We controlled for the distance between the object and the verb by only extracting the

Table 2: Speaker distribution by gender and age.

	High school student	University student	Young adult (25 ~ 39 yrs)	Middle-aged adult (40 ~ 59 yrs)	Elderly adult (60 + years)
Women	14	13	14	18	18
Men	11	14	14	16	18

objects immediately adjacent to the verb. Finally, we did not include objects modified by focus particles.

We extracted nouns that immediately preceded a verb on our list of frequent verbs. We excluded numerals and adverbial nouns such as *toki* ‘time, when’ and *asa* ‘morning.’ We also excluded pronouns and proper nouns, as these words tend to not be modified by relative clauses. We treated compound nouns such as *borantia-kei* ‘activities such as volunteering’ as single nouns.

We coded each token by verb category: transitive, unergative, or unaccusative. Three highly frequent verbs accounted for almost 35% of the tokens: *suru* ‘do,’ *iku* ‘go,’ and *naru* ‘become.’ We coded each of these verbs separately. Altogether, we found 12,792 object–verb pairs formed from 4,799 nouns and 500 verb lemmas, or 7,721 unique object–verb pairs. The average case particle omission rate is 70.5%. Table 3 summarizes our data by verb category.

4.2 Token classification

4.2.1 Dispersion

For our measure of dispersion, we used the DP measure derived by Stefan Th. Gries (2008). See Gries (2008) for advantages over other alternatives, and details of the calculation. We briefly summarize the calculation procedure here. We expect each speaker to produce an equal proportion of the number of occurrences of a specific expression. Thus, if we have data from ten speakers, then we expect each speaker to produce one tenth of the tokens for any given expression. We then calculate the observed proportion of tokens produced by each speaker. If two out of ten speakers produced half of the tokens each, then the observed proportion for each of those two speakers is 0.5, and 0 for the rest of the speakers. We next sum the absolute values of the difference between the expected proportion and the observed proportions. Finally, we divide the sum by 2. This procedure produces a value that theoretically ranges from zero to one, with zero representing perfect dispersion (i.e., every speaker produced exactly the expected proportion of tokens), to approximately one.

We calculated DP values for each of the 7,721 object–verb pairs based on the number of speakers that used that object–verb pair during their interview. Our highest value is 0.993 for an object–verb pair used by only one speaker in our corpus. Our values ranged from 0.560 to 0.913. Table 4 lists the fifteen most dispersed object–pairs in our data. The DP scores are extremely skewed: over half of our tokens ($N = 7,179$) have the highest value (i.e., they were only used by one speaker). In order to compensate for the extreme sparsity of data in the lower range, we grouped the data into four levels of dispersion, with one representing low dispersion and four representing high dispersion. We give the

Table 3: Token counts and case marker omission rates by verb category.

Verb category	Token counts	Omission rates
Transitive	5,588	78.9%
Unergative	1,493	62.8%
Unaccusative	1,396	65.8%
<i>suru</i> ‘do’	1,134	89.0%
<i>iku</i> ‘go’	1,910	73.2%
<i>naru</i> ‘become’	1,271	26.5%
Total	12,792	70.5%

Table 4: Fifteen most dispersed object–verb pairs, their DP values, and case particle omission rates.

Expression	Number of speakers	DP value
<i>toko-ni iku</i> 'go to a place (that is...)'	66	0.560
<i>gakkō-ni iku</i> 'go to school'	42	0.720
<i>gohan-o taberu</i> 'eat a meal'	37	0.753
<i>ie-ni kaeru</i> 'return home'	36	0.760
<i>daigaku-ni iku</i> 'go to university'	34	0.773
<i>kanji-ni naru</i> 'have a feeling'	33	0.780
<i>ryokō-ni iku</i> 'go on a trip'	33	0.780
<i>hanashi-o kiku</i> 'listen to a story'	31	0.793
<i>ki-ni naru</i> 'be concerned'	25	0.833
<i>daigaku-ni hairu</i> 'enter university'	23	0.847
<i>hanashi-ni naru</i> 'become that way'	23	0.847
<i>baito-o suru</i> 'do part-time work'	22	0.853
<i>ki-o tsukeru</i> 'take care'	21	0.860
<i>terebi-o miru</i> 'watch TV'	20	0.867
<i>Tōkyō-ni iku</i> 'go to Tokyo'	20	0.867

Table 5: Token and type counts, omission rates, and DP value ranges for our four levels of dispersion.

Dispersion level	Token counts	Type counts	Omission rate	DP range
1	7,179	6,551	68.9%	0.993
2	2,289	856	70.9%	0.980 ~ 0.987
3	1,943	268	73.2%	0.920 ~ 0.973
4	1,381	46	74.1%	0.560 ~ 0.913

specific DP value ranges for each dispersion level along with token and type counts in Table 5. Table 5 also lists the average case marker omission rate by dispersion level. The case marker omission rate increases by roughly 1% with each increase in dispersion level. This result supports our first hypothesis, which is increased dispersion correlates with increased case marker omission.

4.2.2 Syntactic complexity

We initially coded the object noun by the constituent type it was modified by, which was either a determiner (*kono* 'this'), an i-adjective¹ (*takai* 'expensive'), a na-adjective (*kirei-na* 'pretty'), a noun phrase with a genitive case marker (*sensei-no* 'teacher's'), a relative

¹ Japanese has two types of adjectives: i-adjectives and na-adjectives. The i-adjective behaves as a prototypical adjective. It gets its name from the dictionary form, which ends in the vowel /i/. Similarly, the na-adjective gets its name from the attributive form, which is followed by the *-na* suffix. The Japanese name for na-adjectives is *keiyō-dōshi*, which means 'adjectival verb.' The Japanese name reflects the fact that a na-adjective is followed by a copula (i.e., a verb) when it forms the predicate.

clause, or a na-adjectival phrase (*sensei-no-yō-na* ‘like the teacher’).² When the preceding constituent did not directly modify the object, then we coded the object as bare. The examples given in (4), taken from the corpus, illustrate each of these categories. The relevant noun is highlighted.

- (4) **bare**
 Mazu **mizo-o** hora-na’akan. (KSJ/078/m/ma)³
 first ditch.ACC dig.DEO
 ‘First, (you) have to dig a ditch.’
- determiner**
 Ano-**hito-o** midashi-ta hito-ga sugoi yo ne. (KSJ/001/f/uni)
 that person.ACC spot.PAST person.NOM amazing SFP SFP
 ‘The person who scouted her is amazing, don’t you think?’
- i-adjective**
 Nanka attsui **kōhī** nomi-tai na. (KSJ/081/f/hs)
 like hot coffee drink.DES SFP
 ‘(I) want to drink, like, hot coffee.’
- na-adjective**
 Hen-na **nihongo-ni** nat-ta. (TKC/004/m/uni)
 strange Japanese.DAT become.PAST
 ‘My Japanese has become strange.’
- genitive noun phrase**
 sempai-no **shashin** ∅ atsume-totte (KSJ/118/f/hs)
 senior students.GEN picture.ACC collect.NF
 ‘collecting pictures of the senior students’
- relative clause**
 Yakyū mitai-ni ugoi-teru **bōru** ∅ utsu wake-nai. (KSJ/063/m/ma)
 baseball like.ADV move.PROG ball.ACC hit the case that.NEG
 ‘It is not like baseball, where you hit a moving ball.’
- na-adjectival phrase**
 Hito ∅ ira-n yō-na **jidai-ni** nat-ton nen. (KSJ/022/m/el)
 person.NOM need.NEG like.ATR era.DAT become.past SFP
 ‘It has become so that (the company) no longer needs employees.’

² One of the reviewers pointed out that we have not taken into consideration the possibility that what has been omitted is the topic marker *wa*, and not a case marker. The reviewer noted that topic marker omission tends to happen more frequently in certain contexts, such as nouns modified by demonstratives (Miyamoto et al. 1999). Miyamoto et al. removed from their analysis nouns all proper nouns, demonstratives, and nouns that have been previously mentioned in the discourse.

In order to assess the extent of this problem, we extracted all cases of noun + *wa* + verb, and classified the noun according to our schema in Table 1. We found 363 tokens, with a distribution comparable to that seen in Table 1. Thus, all the cases of case marker omission are prone to this problem. Without examining each token one by one in context, we have no way of assessing what was omitted. We acknowledge this shortcoming as a drawback of the corpus methodology that we have opted to use for this study. Unlike Miyamoto et al., we have opted to not exclude any data for this reason. First, we have already taken into consideration whether or not an object was previously mentioned in the discourse. Second, we are primarily concerned with cognitive processing, and less so with the theoretical concept of case. Third, the relatively infrequent rate of occurrence of *wa* compared to an overt case marker in the specific context that we examined (363 versus 3,774) suggests that the scale of this problem is small.

³ Examples taken from the corpus are followed by brackets containing the following speaker characteristics: corpus code, either KSJ, KYT, or TKC; identifier number; gender, either female (f) or male (m); and age cohort, either high school student (hs), university student (uni), young adult (ya), middle-aged adult (ma), or elderly adult (el).

Initial analysis of the case marker omission rates (Table 1) shows that we can simplify this classification scheme. If we consider the rates of case marker omission (the second column of Table 1), then we can reclassify the data into three groups corresponding to zero, one, or more than one modifying constituent. We therefore reclassified the tokens by the number of modifying constituents as either ZERO, ONE, or TWO or more (the third column of Table 1). From Table 1 we see that case particle omission rate decreases as the number of modifying constituents increases. This result supports our second hypothesis.

4.2.3 Other factors

We classified each token for the following social characteristics: age group (five groups, see Table 2), gender (male or female), and speech style index (as described in Heffernan & Hiratuka 2017). Briefly, the speech style index is a value that theoretically ranges from zero to one and which indicates the degree to which the speaker used Standard Japanese during the interview (with a higher value indicating a greater use of Standard Japanese). The value was derived by combining the following seven measures:

- Proportion of standard versus regional copula variants (for example, *da* vs. *ya* ‘be’);
- Proportion of standard versus regional verbal negative suffixes (for example, *tabe-nai* vs. *tabe-hen* ‘not eat’);
- Proportion of standard versus regional verbs of existence (*iru* vs. *oru* ‘be’);
- Proportion of nasalization of a verb final *-ru* (for example, *taberu* vs. *taben* ‘eat’);
- Proportion of non-regional versus regional sentence-final particles (for example, *yo* vs. *de*);
- Proportion of non-regional versus regional adverbial intensifiers (for example, *erai* vs. *kekko* ‘very’);
- Proportion of standard *ii* ‘good’ versus regional variant *ee* ‘good.’

Each of these seven variables was scaled between zero and one, with each speaker’s average indicating the speech style index for that speaker.

We coded whether or not the object had been previously mentioned, as Fujii & Ono (2000) found that case marker omissions tended to occur with objects that had direct referents in the conversation.

Finally, if we observed a case marker directly following the noun then we coded the case marker as INCLUDED, otherwise as OMITTED.

5 Mixed-effects analysis

Our objective of this section is twofold. First, we want to confirm the correlations between case marker omission rate, dispersion, and number of modifying constituents, while taking into consideration other covariates such as speech style. Second, we want to test our third hypothesis, which was that declarative knowledge and procedural knowledge interact. For this purpose, we conducted a mixed-effects analysis with a logit link using the lme4 package in the R environment (R Core Team 2015). Mixed-effects modeling is ideal for linguistic data because such models consider interspeaker variability when determining the factor weights (Johnson 2009; Gries 2015).

We conducted our analysis following the step-up strategy outlined in West et al. (2014: section 2.7.2). This approach begins with a model that has a fixed intercept as the only fixed parameter, with the intercepts for SPEAKER and WORD (object-verb pair) added as random effects. From there we added each of the other fixed effect covariates, beginning

with those factors that had shown an obvious effect in our preliminary data analysis. After each addition beyond the initial model with just the random effects, we conducted a likelihood ratio test comparing the new model with the old model. If the two models were found to be significantly different, then we retained the added factor; otherwise, we removed it. We repeated this process until we had either added or rejected each of the fixed effect covariates.

Once we had our initial model with random and fixed effects, we then tested for interactions between the fixed effects. Again, we added each interaction term one at a time, and compared the revised model with the new model. We present the final model as Table 6. Note that GENDER, AGE, and all the interactions except one were not chosen as significant.

The tables have two important information sources that allow for easy comparison between covariates. The first is the relative magnitude and sign of the parameter estimates, with a positive sign indicating a positive correlation. The second information source is the magnitude of the p value, with a smaller p value indicating a stronger correlation between case marker omissions and the factor. Comparing the covariates in this way confirms the correlations seen in Tables 1 and 5: increased case particle omission correlates with fewer modifying constituents and increased dispersion. As expected, SPEECH STYLE shows a strong correlation with case particle omission, as well as whether or not the object was previously mentioned. These results agree with the previous research on case particle omission in Japanese.

Crucially, the interaction between dispersion and the number of modifying constituents reached statistical significance, supporting our third hypothesis, that declarative and procedural memory compete with each other during language processing. In order to better understand this interaction, we plotted the interaction using the *interplot* package in R (Solt & Hu 2015) as Figure 1. The *interplot* package plots the changes in the coefficient of one variable in an interaction conditional on the value of the other variable. The plot also includes simulated 95% confidence intervals of these coefficients. Figure 1 shows the

Table 6: Linear mixed-effects model for case marker omission.

	Estimate	Standard error	Z	p
Intercept	-2.282	0.234	-9.756	<2.00e-16
Dispersion	-0.1953	0.044	-4.493	7.03e-06
Num. modifying constituents	0.777	0.083	9.365	<2.00e-16
Speech style	3.797	0.477	7.953	1.82e-15
Verb = unergative	0.418	0.118	3.542	0.0004
Verb = transitive	-0.966	0.095	-10.141	<2.00e-16
Verb = <i>iku</i> 'go'	-0.073	0.118	-0.625	0.5323
Verb = <i>naru</i> 'become'	2.282	0.135	16.852	<2.00e-16
Verb = <i>suru</i> 'do'	-1.935	0.156	-12.392	<2.00e-16
Previously mentioned = yes	-0.147	0.056	-2.646	0.0081
Dispersion × Num. modifying constituents	-0.119	0.041	-2.915	0.0036

Notes: The dependent variable (case particle omission) was coded as 1 = inclusion, 0 = omission. The random intercepts of the individual speakers (s.d. = 0.682) and the object-verb pairs (s.d. = 1.337) were added to the model. Not chosen as significant: age, gender, and all other interaction terms other than the last entry of the table.

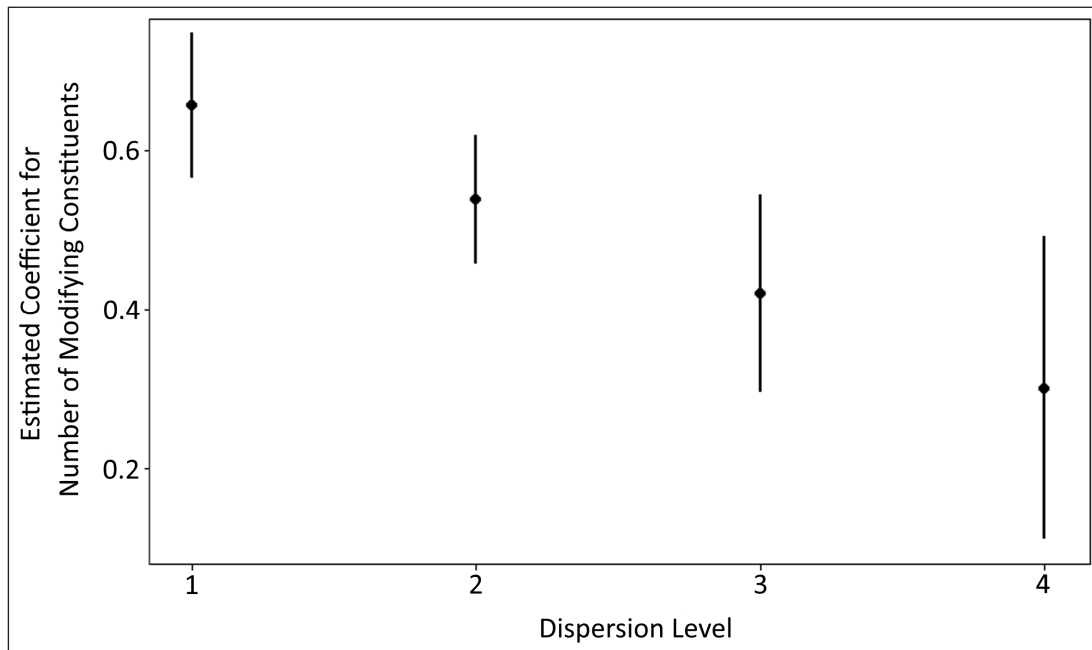


Figure 1: Estimated coefficients and 95% confidence intervals for number of modifying constituents by dispersion level.

changes in the estimated coefficient for the number of modifying constituents for each level of dispersion. As the dispersion of the object–verb pair increases, the estimated coefficient decreases, indicating a gradual weakening of the impact of the number of constituents on case particle omission.

The 95% confidence interval also appears to decrease in size as dispersion decreases. However, this change simply reflects the lower number of tokens in the more dispersed subsets (confidence interval size is proportional to the inverse of the square root of the number of tokens). This result is consistent with Ullman’s claim that declarative memory and procedural memory compete, with declarative memory preempting procedural memory for the most dispersed expressions. This result is also consistent with recent studies in morphology (Schmidtke et al. 2017; Lõo et al. 2018) showing that frequency (i.e., declarative knowledge) impacts morphological processing sooner than morphological factors (i.e., procedural knowledge).

Before concluding the paper, we first discuss the other covariates. The correlation between speech style and case marker omission is perhaps not surprising. Half a century of work on stylistic variation, beginning in the 1960s with work by Labov (1966/2006), has shown that speakers subtly control socially meaningful linguistic variation in order to dynamically portray specific social identities (Schilling-Estes 2002).

How exactly the impetus to perform socially meaningful work interacts with cognitive load is a question that still remains largely unexplored. However, we are aware of at least one study that addresses this question. Heffernan & Hiratuka 2017 examined stylistic variation in spoken Japanese negative verbal suffix choice. Japanese has both standard and vernacular variants for the verbal negative suffixes. Heffernan & Hiratuka observed a strong correlation between negative verbal suffix choice and other markers of vernacular speech—except for the verbs *shiru* ‘know,’ *wakaru* ‘understand,’ and *iru* ‘need.’ They argue that since the negated forms of these verbs (i.e., *don’t know*, *don’t understand*, and *don’t need*) occur relatively more frequently in spoken Japanese than their uninflected forms, they are processed in declarative memory as a single cognitive unit. Generally speaking,

verbs tend to occur most frequently in their non-past, non-negative form. This form presumably forms the base of an inflectional paradigm, and in most cases, the negative form of the verb is derived in procedural memory. However, in the cases of verbs *shiru* ‘know,’ *wakaru* ‘understand,’ and *iru* ‘need,’ the faster processing in declarative memory blocks the slower processing in procedural memory—more evidence that declarative knowledge and procedural knowledge compete during processing.

Finally, consider the results of the VERB CATEGORY factor group. The ranking of the specific verb categories by their estimated coefficients in the order in which the verb type favors case marker omission yields (5). As there has been little previous research on case marker omissions in the dative case, this work is groundbreaking. Our results show that the accusative case marker favors omissions more often than the dative case marker.

(5) suru ‘do’ > transitive > iku ‘go,’ unaccusative > unergative > naru ‘become’

We tentatively suggest that this difference could be captured by generative grammar theory. Imanishi (2017) proposes that the gradient accusative case marker omission pattern seen in spoken Japanese parallels the categorical pattern of pseudo noun incorporation in Niuean (Massam 2001). Massam argues that pseudo noun incorporation only applies to a constituent in the complement position of a VP head. Crucially, dative-marked elements in unaccusative and unergative sentences appear in an adjunct position and not in a complement position. Therefore, there is a noticeable parallel between the gradient case marker omission patterns in Japanese and pseudo noun incorporation in Niuean, further supporting Imanishi’s claim that the two phenomena are theoretically similar.

As pointed out by one of the anonymous reviewers, this conclusion bears some resemblance to the behavior of accusative case marking in Old Japanese, as discussed in Miyagawa (1989; 2012) and Yanagida (2007). Miyagawa (1989) provides an in-depth analysis of the accusative case marking for direct objects of eighth-century Old Japanese, which can be summarized as follows. The object of the attributive form of the verb must bear the accusative case marker *-o*, whereas the object of the conclusive form need not (= zero marking) so long as it is adjacent to the verb (Miyagawa 1989: 212). However, as Yanagida (2007) has discovered, there are exceptions to Miyagawa’s generalization that the object of an attributive form must bear morphological case. Yanagida found 55 examples of exceptions in the *Man’yōshū*, an anthology of Japanese poems composed during the seventh and eighth centuries. As she points out, 54 out of these counterexamples contain objects that can be analyzed as bare nouns. Miyagawa (2012) suggests that these zero-marked objects are licensed by head incorporation (Baker 1988; but see Yanagida 2007 for a different analysis). In this analysis, we see a categorical contrast between bare nouns and modified nouns, with only bare nouns undergoing noun incorporation (and hence case marker omission). Such an analysis closely resembles Imanishi’s treatment of spoken Japanese, only really differing by a matter of degree (categorical versus gradient). But that is often the case when comparing records of written language such as the *Man’yōshū* with records of spoken language such as the *Corpus of Kansai Vernacular Japanese*.

6 Conclusions

Our study examined case marker omissions in object–verb pairs, with the objective of contributing to the emergent–generative grammar debate. We specifically looked for evidence of an interaction between emergent knowledge and generative knowledge, as defined in Ullman’s declarative/procedural language model (2004; 2008; 2016). We took

the degree of dispersion of the object–verb pair within the corpus as a proxy of declarative knowledge, and the syntactic complexity of the object (as measured by the number of modifying constituents) as a proxy of procedural knowledge.

Our study is built on the assumption that increased case marker omission reflects reduced cognitive processing effort (Fedzechkina et al. 2017). Since more dispersed words require less cognitive effort (McDonald & Shillcock 2001), we hypothesized that increased dispersion correlates with increased case marker omission. Similarly, since more syntactically complex expressions require more cognitive effort (Caplan & Waters 1999), we hypothesized that increased syntactic complexity of the object correlates with reduced case particle omission. Furthermore, motivated by Ullman’s (2016) assertion that declarative knowledge and procedural knowledge compete during processing, we hypothesized that dispersion level and number of modifying constituents interact. Specifically, we expected the impact of the number of modifying constituents to decrease as dispersion level increased.

In order to test these hypotheses, we conducted a mixed-effects analysis of case marker omissions. After taking into consideration the effects of the speakers, the object–verb pairs (i.e., word effects), speech style, verb type, and whether or not the object had been previously mentioned, we found support for all three of our hypotheses.

Some recent work in variationist linguistics continues to debate the role of lexical frequency. For example, Walker (2012) examined t/d-deletion in conversational data from 47 speakers and concluded (Walker 2012: 410) that “once we take into account the contribution of a small set of highly frequent lexical items,” then frequency no longer correlates significantly with variation. Similarly, Bayley et al. (2013) examined a variety of factors such as gender, verbal tense and subject person/number, and the frequency of the verb form in subject pronoun omissions in conversational data from 29 Spanish speakers, and found that verb form frequency had a relatively small effect. They did find, however, that correlations with the factors are generally stronger for non-frequent verb forms. In a way, we are building on these results by examining similar questions, but replacing lexical frequency with dispersion. Furthermore, we also found a similar, albeit more constrained, result to that of Bayley et al.: the correlation with one of our factors was stronger for the less dispersed object–verb forms.

Our results do not support the usage-based theory of grammar. In this approach, the cognitive organization of language is based directly on experience with language, the basic units of syntactic structure replaced by constructions (Beckner et al. 2009). Such an approach could account for the correlations between our measure of syntactic complexity—the number of modifying constituents—by appealing to the different rates of exposures a speaker has to modifying constituents; a determiner modifying a noun occurs more frequently than a relative clause modifying a noun. However, such an approach does not explain the interaction between dispersion and modifying constituents since in a usage-based grammar approach one is not dependent on the other.

Our results also do not support a model of grammar that ignores user experience, since we found a strong correlation between dispersion and case marker omissions. Contrary to Walker (2012) and Bayley et al. (2013), we conclude that morphosyntactic variationist studies need to continue to take into account some measure of community-level usage rates of specific expressions, such as dispersion.

Our study has limitations. We say nothing about the innateness of Universal Grammar. Ullman (2004; 2016) argues that procedural memory, which underlies the computational system which gives structure to language, also subserves other cognitive functions. O’Grady (2008) similarly argues for general cognitive functions underlying the language faculty. Our results are compatible with such an approach, or with an innateness approach.

We also have not considered the role of prediction in our study (Gahl & Garnsey 2004; Levy 2008; Kuperberg & Jaeger 2016). These researchers account for differences in cognitive processing with a model of prediction based on language experience. A listener dynamically processes linguistic input as each new word is heard by simultaneously predicting multiple candidate structural representations of the partial input. She then ranks each representation based on its probability of occurrence. The difficulty of cognitively processing a new word corresponds to how closely that new word matches the predicted representations, and the cost to rerank the candidate representations when predictions are not met.

Previous research has demonstrated that measures of prediction correlate with the acoustic duration of words (Bell et al. 2009; Seyfarth 2014), fixation times in eye tracking studies (Levy 2008), and reaction times in lexical decision studies (Balling & Baayen 2012) and event-related potentials (Van Petten & Luka 2012). However, we have avoided a measure of expectation in this study. As O'Grady points out (2015: 9), any measure of rate of occurrence such as dispersion will confound any measure of parsimonious processing such as predictability. Thus, adding both dispersion and predictability risks multicollinearity problems; we opted for dispersion. But what about using a measure of prediction as a proxy of procedural knowledge? Ullman allocates predictive linguistic processing to procedural knowledge, suggesting that a measure of predictability might make a better proxy for procedural knowledge than our rather simple measure of syntactic complexity. We leave this development for future research.

What is the next step? We have empirically supported Ullman's claim that declarative and procedural knowledge interact, but we still know very little about how. Reuland (2010) speculates that the declarative/procedural interface along with the limited capacity of working memory define key characteristics of syntax, such as the locality restrictions on *wh* movement in English known as island constraints. For example, English *wh* words cannot move out of an adverbial clause (= adjunct island). Consider the relevant example discussed by Reuland, reproduced here as (6a). Reuland points out that, in contrast, pronominal binding does not have such a restriction, as shown by (6b). We have no difficulty interpreting "him" in terms of "every boy" in the matrix clause. Reuland explains the asymmetry as follows. Incomplete linguistic input is temporarily stored in working memory while the declarative/procedural interface parses it into a larger chunk. That chunk is then shifted to a longer term memory, freeing up working memory for the next chunk. Reuland speculates that in the case of (6a), the *wh* word must stay in working memory until its dependency is resolved. The limited capacity of working memory eventually overloads, leading to the ungrammaticality of the sentence. The expression "every boy" does not jam up working memory because it does not necessarily have a dependency; although "every boy" can bind the pronominal "him," it may also refer to some other entity in the larger discourse. Such dependencies are presumably resolved in declarative memory.

- (6) Reuland (2010)
- a. *What did John think Mary would be unhappy about after she ruined (what)?
 - b. Every boy thought that Mary would be unhappy after she ruined his apartment.

Reuland's ideas about the declarative/procedural interface lead to more predictions. Again we expect to see gradient interactions. In a similar discussion about *wh* movement in Japanese, Saito & Fukui (1998) give examples that are ungrammatical as well as examples that are only slightly so. We make two predictions. First, we should be able

to come up with a full range of grammatical violations, from mild to severe. Second, the severity of the violation should interact with some indicator of declarative knowledge, such as dispersion. That is, the more an expression is dispersed, the more speakers should show tolerance for grammatical violations due to wh movement. There should also be specific violation types that are clearly unacceptable regardless of dispersion level. Further exploring the interaction between declarative knowledge and procedural knowledge in this way should lead to a better understanding of intuitions about grammaticality. The methodology used by Lõo et al. (2018) seems ideally suited to such a task.

Gries (2012: 477) laments, “Linguistics is fundamentally a divided discipline, as far as theoretical foundations and empirical methodologies are concerned.” Certainly there will always be researchers narrowly focused on one or another aspect of language. But we hope that with this study, along with other advocates of dual route models, we are building a bridge over this theoretical gap.

Abbreviations

We use the following glosses: ACC = accusative case, ADV = adverbial, ATR = attributive, DAT = dative case, DEO = deontic obligation, DES = desiderative mood, GEN = genitive case, NF = nonfinite, NOM = nominative case, NONP = non-past, NEG = negative, PAST = past tense, PER = perfective aspect, PROG = progressive aspect, SFP = sentence-final particle, SPEC = speculative mood.

Acknowledgements

We are very grateful to the two anonymous reviewers and Herald Baayen for their valuable comments and suggestions. We also thank the audiences at NWAV AP 4 (National Chung Cheng University), and Exploiting Parsed Corpora: Applications in Research, Pedagogy, and Processing (National Institute for Japanese Language and Linguistics) for their insightful feedback. Finally, we thank the many research assistants who transcribed, checked, and rechecked the corpus data. Any shortcomings or errors in the data or analysis are our own.

Competing Interests

The authors have no competing interests to declare.

References

- Adelman, James S., Gordon D. A. Brown & José F. Quesada. 2006. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17(9). 814–823. DOI: <https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Baayen, R. Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal of Memory and Language* 37(1). 94–117. DOI: <https://doi.org/10.1006/jmla.1997.2509>
- Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago, IL: University of Chicago Press.
- Bakker, Iske, Lucy J. MacGregor, Friedemann Pulvermüller & Yury Shtyrov. 2013. Past tense in the brain's time: Neurophysiological evidence for dual-route processing of past-tense verbs. *NeuroImage* 71. 187–195. DOI: <https://doi.org/10.1016/j.neuroimage.2012.12.065>
- Balling, Laura Winther & R. Harald Baayen. 2012. Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition: International Journal of Cognitive Science* 125(1). 80–106. DOI: <https://doi.org/10.1016/j.cognition.2012.06.003>

- Bayley, Robert, Kristen Greer & Cory Holland. 2013. Lexical frequency and syntactic variation: A test of a linguistic hypothesis. *University of Pennsylvania Working Papers in Linguistics* 19(2). Article 4.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59(Suppl. 1). 1–26. DOI: <https://doi.org/10.1111/j.1467-9922.2009.00533.x>
- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113(2). 1001–1024. DOI: <https://doi.org/10.1121/1.1534836>
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. DOI: <https://doi.org/10.1016/j.jml.2008.06.003>
- Brysbaert, Marc, Paweł Manderla & Emmanuel Keuleers. 2018. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1). 45–50. DOI: <https://doi.org/10.1177/0963721417727521>
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge, England: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511750526>
- Bybee, Joan & Joanne Scheibman. 1999. The Effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 37(4). 575–596. DOI: <https://doi.org/10.1515/ling.37.4.575>
- Caplan, David & Gloria S. Waters. 1999. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22(1). 77–94. DOI: <https://doi.org/10.1017/S0140525X99001788>
- Culbertson, Jennifer & Elissa L. Newport. 2015. Harmonic biases in child learners: In support of language universals. *Cognition: International Journal of Cognitive Science* 139. 71–82. DOI: <https://doi.org/10.1016/j.cognition.2015.02.007>
- Dąbrowska, Ewa. 2010. The mean lean grammar machine meets the human mind: Empirical investigations of the mental status of linguistic rules. In Hans-Jörg Schmid, Sandra Handl & Susanne Handl (eds.), *Cognitive foundations of linguistic usage patterns: Empirical approaches*, 151–170. Berlin, Germany: Mouton de Gruyter.
- Dąbrowska, Ewa. 2014. Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics* 25(4). 617–653. DOI: <https://doi.org/10.1515/cog-2014-0057>
- Dąbrowska, Ewa. 2015. What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology* 6(852). 1–17. DOI: <https://doi.org/10.3389/fpsyg.2015.00852>
- Ellis, Nick C. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24(2). 143–188. DOI: <https://doi.org/10.1017/S0272263102002024>
- Ellis, Nick C. 2008. Phraseology: The periphery and the heart of language. In Fanny Meunier & Sylviane Grainger (eds.), *Phraseology in foreign language learning and teaching*, 1–13. Amsterdam, Netherlands: Benjamins.
- Fedzechkina, Maryia, Elissa L. Newport & T. Florian Jaeger. 2017. Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science* 41(2). 416–446. DOI: <https://doi.org/10.1111/cogs.12346>
- Forster, Kenneth I. & Susan M. Chambers. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12(6). 627–635. DOI: [https://doi.org/10.1016/S0022-5371\(73\)80042-8](https://doi.org/10.1016/S0022-5371(73)80042-8)

- Fry, John. 2003. *Ellipsis and wa-marking in Japanese conversation*. New York, NY and London, England: Routledge.
- Fujii, Noriko & Tsuyoshi Ono. 2000. The occurrence and non-occurrence of the Japanese direct object marker *o* in conversation. *Studies in Language* 24(1). 1–39. DOI: <https://doi.org/10.1075/sl.24.1.02fuj>
- Gahl, Susanne & Susan Marie Garnsey. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language: Journal of the Linguistic Society of America* 80(4). 748–775. DOI: <https://doi.org/10.1353/lan.2004.0185>
- Gibson, Edward. 2000. Dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O’Neil (eds.), *Image, language, brain: Papers from the First Mind Articulation Project Symposium*, 95–126. Cambridge, MA: MIT Press.
- Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4). 403–437. DOI: <https://doi.org/10.1075/ijcl.13.4.02gri>
- Gries, Stefan Th. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36(3). 477–510. DOI: <https://doi.org/10.1075/sl.36.3.02gri>
- Gries, Stefan Th. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125. DOI: <https://doi.org/10.3366/cor.2015.0068>
- Guy, Gregory R. 2014. Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. *Lingua: International Review of General Linguistics* 142. 57–65. DOI: <https://doi.org/10.1016/j.lingua.2012.07.007>
- Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598). 1569–1579. DOI: <https://doi.org/10.1126/science.298.5598.1569>
- Heffernan, Kevin & Yo Sato. 2017. Relative frequency and the holistic processing of morphology: Evidence from a corpus of vernacular Japanese. *Asia-Pacific Language Variation* 3(1). 67–94. DOI: <https://doi.org/10.1075/aplv.3.1.04hef>
- Heffernan, Kevin & Yusuke Hiratuka. 2017. Morphological relative frequency impedes the use of stylistic variants: Evidence from a corpus of vernacular Japanese. *Asia Pacific Language Variation* 3(2). 200–231. DOI: <https://doi.org/10.1075/aplv.16009.hef>
- Hoffman, Thomas & Graeme Trousdale. (eds.) 2013. *The Oxford handbook of construction grammar*. Oxford, England: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780195396683.001.0001>
- Hopper, Paul J. 2015. An emergentist approach to grammar. In Brian MacWhinney & William O’Grady (eds.), *The handbook of language emergence*, 314–327. Hoboken, NJ: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781118346136.ch14>
- Ibbotson, Paul & Michael Tomasello. 2016. Language in a new key. *Scientific American* 315(5). 71–75. DOI: <https://doi.org/10.1038/scientificamerican1116-70>
- Imanishi, Yusuke. 2017. Pseudo noun incorporation and its kin. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *M.I.T. Working Papers in Linguistics*, Vol. 80, *A pesky set: Papers for David Pesetsky*, 427–436. Cambridge, MA: MITWPL.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1). 359–383. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00108.x>
- Kuperberg, Gina R. & T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1). 32–59. DOI: <https://doi.org/10.1080/23273798.2015.1102299>

- Kuperman, Victor, Robert Schreuder, Raymond Bertram & R. Harald Baayen. 2009. Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance* 35(3). 876–895. DOI: <https://doi.org/10.1037/a0013484>
- Kyoto University Graduate School of Informatics & Nihon Telegraph and Telephone Corporation. 2013. MeCab. Computer Software. 0.996. (<http://taku910.github.io/mecab/>) (Accessed July 12, 2013).
- Labov, William. 1966/2006. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition: International Journal of Cognitive Science* 106(3). 1126–1177. DOI: <https://doi.org/10.1016/j.cognition.2007.05.006>
- Lõo, Kaidi, Juhani Järviö, Fabian Tomaschek, Benjamin V. Tucker & R. Harald Baayen. 2018. Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 28(1). 71–97. DOI: <https://doi.org/10.1007/s11525-017-9318-7>
- MacWhinney, Brian & William O’Grady. (eds.) 2015. *The handbook of language emergence*. Hoboken, NJ: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781118346136>
- Matsuda, Kenjirō. 2000. Tōkyō hōgen joshi o no shiyō ni kakawaru gengoteki shoyōin no sūryōteki kenshō [A Quantitative Verification of Linguistic Factors Affecting the Use of the Case Particle o in Tokyo Japanese]. *Kokugogaku* 51(1). 61–76.
- McClelland, James L. 2015. Capturing gradience, continuous change, and quasi-regularity in sound, word, phrase, and meaning. In Brian MacWhinney & William O’Grady (eds.), *The handbook of language emergence*, 53–80. Hoboken, NJ: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781118346136.ch2>
- McDonald, Scott A. & Richard C. Shillcock. 2001. Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech* 44(3). 295–322. DOI: <https://doi.org/10.1177/00238309010440030101>
- Miyagawa, Shigeru. 1989. *Structure and case marking in Japanese*. New York, NY: Academic Press.
- Miyagawa, Shigeru. 2012. The Old Japanese accusative revisited: Realizing all the universal options. In Shigeru Miyagawa (ed.) *Case, argument structure, and word order*, 259–273. New York, NY and London, England: Routledge.
- Miyagawa, Shigeru. 2017. Integration Hypothesis: A parallel model of language development in evolution. In Shigeru Watanabe, Michel A. Hofman & Toru Shimizu (eds.), *Evolution of the brain, cognition, and emotion in vertebrates*, 225–247. Tokyo, Japan: Springer Japan.
- Miyamoto, Edson T., Kenneth Wexler, Takako Aikawa & Shigeru Miyagawa. 1999. Case-dropping and unaccusatives in Japanese acquisition. In Annabel Greenhill, Heather Littlefield & Cheryl Tano (eds.), *BUCLD 23: Proceedings of the 23rd annual Boston University conference on language development*, 443–445. Somerville, MA: Cascadilla Press. DOI: https://doi.org/10.1007/978-4-431-56559-8_11
- Nelson, Matthew J., Imen El Karoui, Kristof Giber, Xiaofang Yang, Laurent Cohen, Hilda Koopman, Sydney S. Cash, Lionel Naccache, John T. Hale, Christophe Pallier & Stanislas Dehaene. 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences* 114(18). E3669–E3678. DOI: <https://doi.org/10.1073/pnas.1701590114>
- Newmeyer, Frederick J. 2010. What conversational English tells us about the nature of grammar: A critique of Thompson’s analysis of object complements. In Kasper Boye &

- Elisabeth Engberg-Pedersen (eds.), *Language usage and language structure*, 3–44. Berlin, Germany and New York, NY: De Gruyter Mouton.
- O’Grady, William. 2008. Innateness, universal grammar, and emergentism. *Lingua: International Review of General Linguistics* 118(4). 620–631. DOI: <https://doi.org/10.1016/j.lingua.2007.03.005>
- O’Grady, William. 2015. Processing determinism. *Language Learning* 65(1). 6–32. DOI: <https://doi.org/10.1111/lang.12091>
- Perea, Manuel, Ana Paula Soares & Montserrat Comesaña. 2013. Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology* 116(1). 37–44. DOI: <https://doi.org/10.1016/j.jecp.2012.10.014>
- Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2(1). 33–52. DOI: <https://doi.org/10.1146/annurev-linguistics-030514-125050>
- R Core Team. 2015. R: A language and environment for statistical computing. computer program. (<http://www.R-project.org/>) (Accessed Feb. 22, 2016).
- Reuland, Eric. 2010. Imagination, planning, and working memory: The emergence of language. *Current Anthropology* 51(S1). S99–S110. DOI: <https://doi.org/10.1086/651260>
- Schilling-Estes, Natalie. 2002. Investigating stylistic variation. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 375–401. Malden, MA and Oxford, England: Blackwell.
- Schmidtke, Daniel, Kazunaga Matsuki & Victor Kuperman. 2017. Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(11). 1793–1820. DOI: <https://doi.org/10.1037/xlm0000411>
- Scontras, Gregory, William Badecker, Lisa Shank, Eunice Lim & Evelina Fedorenko. 2015. Syntactic complexity effects in sentence production. *Cognitive Science* 39(3). 559–583. DOI: <https://doi.org/10.1111/cogs.12168>
- Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition: International Journal of Cognitive Science* 133(1). 140–155. DOI: <https://doi.org/10.1016/j.cognition.2014.06.013>
- Singleton, Jenny L. & Elissa L. Newport. 2004. When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology* 49(4). 370–407. DOI: <https://doi.org/10.1016/j.cogpsych.2004.05.001>
- Solt, Frederick & Yue Hu. 2015. Interplot: Plot the effects of variables in interaction terms. (<https://CRAN.R-project.org/package=interplot>) (Accessed June 12, 2017).
- Thompson, Sandra A. 2002. “Object complements” and conversation: Towards a realistic account. *Studies in Language* 26(1). 125–163. DOI: <https://doi.org/10.1075/sl.26.1.05tho>
- Tsutsui, Michio. 1984. *Particle ellipses in Japanese*. Urbana, IL: University of Illinois.
- Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition: International Journal of Cognitive Science* 92(1–2). 231–270. DOI: <https://doi.org/10.1016/j.cognition.2003.10.008>
- van Lancker Sidsis, Diana. 2012. Two-track mind: Formulaic and novel language support a dual-process model. In Miriam Faust (ed.), *The handbook of the neuropsychology of language*, 342–367. Hoboken, NJ: Wiley-Blackwell.
- Van Petten, Cyma & Barbara J. Luka. 2012. Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology* 83(2). 176–190. DOI: <https://doi.org/10.1016/j.ijpsycho.2011.09.015>

- Vannest, Jennifer, Thad A. Polk & Richard L. Lewis. 2005. Dual-route processing of complex words: New fMRI evidence from derivational suffixation. *Cognitive, Affective, & Behavioral Neuroscience* 5(1). 67–76. DOI: <https://doi.org/10.3758/CABN.5.1.67>
- Verkoeijen, Peter P., Remy M. Rikers & Henk G. Schmidt. 2004. Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(4). 796–800. DOI: <https://doi.org/10.1037/0278-7393.30.4.796>
- Walker, James A. 2012. Form, function, and frequency in phonological variation. *Language Variation and Change* 24(3). 397–415. DOI: <https://doi.org/10.1017/S0954394512000142>
- West, Brady T., Kathleen B. Welch & Andrzej T. Galecki. 2014. *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: CRC Press. DOI: <https://doi.org/10.1201/b17198>
- Yanagida, Yuko. 2007. Miyagawa's (1989) exceptions: An ergative analysis. In Yoichi Miyamoto (ed.), *Proceedings of FAJL 4: Formal Approaches to Japanese Linguistics* (MIT Working Papers in Linguistics 55), 265–276. Cambridge, MA: MIT Press.
- Yoshizumi, Yukiko. 2016. *A Canadian perspective on Japanese-English language contact*. Ottawa, Canada: University of Ottawa dissertation.

How to cite this article: Heffernan, Kevin, Yusuke Imanishi and Masaru Honda. 2018. Showcasing the interaction of generative and emergent linguistic knowledge with case marker omission in spoken Japanese. *Glossa: a journal of general linguistics* 3(1): 72.1–24, DOI: <https://doi.org/10.5334/gjgl.500>

Submitted: 04 August 2017 **Accepted:** 14 April 2018 **Published:** 19 June 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS The Open Access logo, which is a stylized 'O' with a person icon inside.