Journal of Memory and Language 83 (2015) 152–178

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Communicative efficiency in language production: Optional case-marking in Japanese

Chigusa Kurumada [a,*], T. Florian Jaeger [b]

[a] University of Rochester, Department of Brain and Cognitive Sciences, United States
[b] University of Rochester, Departments of Brain and Cognitive Sciences, Linguistics, and Computer Science, United States

ARTICLE INFO

Article history:
Received 19 November 2013
revision received 8 March 2015

Keywords:
Language production
Communication
Robust information transfer
Production ease
Optional case-marking
Japanese

ABSTRACT

Grammatical encoding is one of the earliest stages in linguistic encoding. One broadly accepted view holds that grammatical encoding is primarily or exclusively affected by production ease, rather than communicative considerations. This contrasts with proposals that speakers' preferences during grammatical encoding reflect a trade-off between production ease and communicative goals. In three recall sentence production experiments, we investigate Japanese speakers' production of optional object case-marking. Case-marking conveys information about the intended sentence interpretation, facilitating comprehension, but it also increases production effort. We find that Japanese speakers are more likely to produce case-marking when the properties of the sentence would otherwise bias comprehenders against the intended interpretation. Experiment 1 observes this effect based on the animacy of the object. Experiments 2 and 3 find the same effect based on the plausibility of the intended grammatical function assignment, even when animacy is held constant. We discuss how speakers might achieve this type of trade-off. In addition to evidencing the role of communicative pressures during even the earliest stages of language production, the results inform linguistic typology, where similar patterns have been observed in obligatory (differential) case-marking.

Introduction

One of the central questions in research on language production is the extent to which language production is affected by our communicative goals. While there is little doubt that those goals affect *what* message we wish to convey, it is less clear to what extent communicative goals affect *how* we convey that message. Specifically, the question still under debate is to what extent the linguistic encoding processes underlying language production are affected by speakers' goal to successfully convey their message.

On the one hand, it is certainly true that some aspects of the planning and decision processes involved in linguistic encoding are affected by the goal to be understood. The fact that we tend to write and speak in a language intelligible to our intended audience illustrates this quite clearly. Similarly, when conversing on a windy mountain peak, we tend to speak louder than when conversing in a quiet room. On the other hand, it is less clear to what extent communicative goals affect linguistic encoding beyond broad adjustments to language choice and speech styles. These encoding processes are generally assumed to involve several largely automatic stages (Bock & Levelt, 1994; Levelt, 1989).

Here, we focus on one of the earliest stages in linguistic encoding, *grammatical encoding*. In particular, we focus on the assembly of a sentence's morpho-syntactic structure. Grammatical encoding is of particular interest because, according to the predominant view in our field, it is primarily or exclusively affected by *production ease*

* Corresponding author.
E-mail address: ckurumada@bcs.rochester.edu (C. Kurumada).

http://dx.doi.org/10.1016/j.jml.2015.03.003
0749-596X/

(e.g., Arnold, 2008; Arnold, Wasow, Asudeh, & Alrenga, 2004; Ferreira, 2008; Ferreira & Dell, 2000; Lam & Watson, 2010; MacDonald, 2013). Indeed, there is now broad agreement that pressures inherent to linguistic encoding affect speakers' preferences during grammatical encoding. This includes, for example, pressures stemming from the problem of retrieving lexical information in time for its use and the linear assembly of these pieces of information (e.g., Bock & Levelt, 1994; MacDonald, 2013).

The perhaps best documented consequence of these pressures is a preference for grammatical structures that "permit quickly selected lemmas to be mentioned as soon as possible" (Ferreira & Dell, 2000, 299). Such availability-based production has received broad cross-linguistic support (reviewed in Jaeger & Norcliffe, 2009). For example, speakers prefer to order constituents referring to easily retrievable referents earlier in the sentence (e.g., Bock & Irwin, 1980; Bock & Warren, 1985; Branigan, Pickering, & Tanaka, 2008; Bresnan, Cueni, Nikitina, & Baayen, 2007; Ferreira & Yoshita, 2003; Prat-Sala, 2000; Tanaka, Branigan, McLean, & Pickering, 2011). Similarly, speakers are more likely to produce optional elements, such as disfluencies (Clark & Fox Tree, 2002; Shriberg & Stolcke, 1996) and optional function words (e.g., English complementizer *that*, Ferreira & Dell, 2000; Jaeger, 2010b; Jaeger & Wasow, 2006; Roland, Elman, & Ferreira, 2006), and to lengthen words (Fox Tree & Clark, 1997) when upcoming material is not yet available to continue production.

The goal of the present paper is to investigate whether communicative goals can affect grammatical encoding. In particular, we ask whether a preference for robust information transmission affects grammatical encoding, beyond effects that can be attributed to production ease. A number of mutually related accounts share the idea that language or language production are affected by the goal to convey information robustly or even efficiently (e.g., Aylett & Turk, 2004; Genzel & Charniak, 2002; Gibson et al., 2013; Jaeger, 2006, 2010b; Levy & Jaeger, 2007; Lindblom, 1990a; Piantadosi, Tily, & Gibson, 2011, 2012; Zipf, 1949). One aspect that has so far been lacking is a clearer link between these approaches and more traditional psycholinguistic accounts. In an attempt to reduce this gap, we pursue our question within a framework outlined by the second author and collaborators (e.g., Buz & Jaeger, 2012; Jaeger, 2010a, 2013). The central prediction investigated below is, however, shared by most of the accounts just cited. We refer to this perspective as the *ideal speaker* framework, to highlight its relation to ideal observers (Geisler, 2003; Jacobs, 2002), which have proven insightful in understanding language comprehension (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015; Levy, 2011; Levy, Bicknell, Slattery, & Rayner, 2009; Norris & McQueen, 2008). Since the perspective provided by the ideal speaker and similar frameworks is crucial for the experiments we present below, we briefly summarize the core assumptions of these approaches. We focus on the conceptual components and leave the formalization to another place.

The first assumption we are making is that the linguistic signal the speaker intends to produce will be at least partially degraded by noise. This noise originates from multiple sources, including noise during the planning and execution of linguistic encoding and articulation, noise in the environment, and noise in the perceptual system of the listener. This makes comprehension a problem of inference under uncertainty or inference over noisy input. Optimal solutions to this problem take advantage of predictions based on the statistics of the input. Indeed, the computational properties of actual language comprehension closely resemble those expected under such a model. This includes evidence from brain potentials (DeLong, Urbach, & Kutas, 2005; Dikker & Pylkkänen, 2013; Kutas & Hillyard, 1984; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; for recent reviews, see Kuperberg, 2013; Van Petten & Luka, 2012), eye-movements during reading (Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008; Staub & Clifton, 2006), spoken sentence comprehension (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and self-paced reading time data (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994; Smith & Levy, 2013; Trueswell, Tanenhaus, & Kello, 1993). All these works point to a language comprehension system that heavily relies on prediction of the signal (see also Farmer, Brown, & Tanenhaus, 2013; Kuperberg, 2013; MacDonald, 2013; Pickering & Garrod, 2013).

Most relevant to the current purpose, these studies provide evidence that language comprehension becomes more difficult (e.g., less accurate and slower) when the observed signal is unexpected given prior expectations. Comprehenders are also more likely to misunderstand or misremember sentences when they have unexpected meanings (Ferreira, 2003). These costs of unexpected form and meaning are the price to pay for a comprehension system that, on average, infers intended messages robustly despite noisy input.

It is important to understand the consequences of noisy input. In the presence of noise, a rational comprehender should maintain uncertainty over the input. For example, comprehenders should not base their expectations about upcoming structure with absolute certainty on what they *believe* to have comprehended so far. This seems to be indeed observed, both during spoken and written word recognition (e.g., "right-context" effects, Dahan, 2010; also Levy et al., 2009) as well as syntactic processing (cf., "local coherence", Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Kukona, Cho, Magnuson, & Tabor, 2014; Tabor, Galantucci, & Richardson, 2004). This view also correctly predicts that sentences can be misunderstood if their properties bias comprehenders towards an unintended parse, even if the grammatical properties of the sentence rule out the unintended interpretation (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003; Ferreira & Patson, 2007).

The second assumption of the ideal speaker framework is that linguistic signals differ in the degree to which they support an intended inference. In other words, some linguistic signals will make it more likely that a comprehender will infer the intended message, compared to other linguistic signals. More specifically, though, we assume that there are typically different ways of realizing the same or a near meaning-equivalent message – an assumption shared with most psycholinguistic work – and that these options differ in the extent to which they support the intended inference (in the current context). For example, a word can be articulated with more or less acoustic detail and this will affect its recognition (Smiljanic & Bradlow, 2011; Connine, 2004; Pitt, 2009).[1] Similarly, some syntactic forms will increase the probability of successful communication, compared to other structural choices. As a convenient shorthand, we will talk about more or less *robust* linguistic forms or signals, when referring to this idea.

The third assumption of the ideal speaker framework is that the systems underlying language production are organized so as to achieve robust or even efficient information transmission by aiming to produce signals that increase the probability of successfully conveying the intended message. This assumption, or some form of it, is shared with many similar accounts (Jaeger, 2006, 2010b; Levy & Jaeger, 2007; Lindblom, 1990a; Piantadosi et al., 2011), including stronger claims about audience design (Clark, 1996; Temperley, 2003). This prediction has received support from studies finding that contextually predictable linguistic units (i.e., those that are more easily inferable) tend to be more reduced (e.g., Aylett & Turk, 2004; Gahl & Garnsey, 2004; Jaeger, 2010b; Resnik, 1996; Wasow, Jaeger, & Orr, 2011; for a review, see Jaeger & Buz, submitted for publication), although several of these findings have been claimed to be compatible with production ease accounts (e.g., Arnold, 2008; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). One of the goals of the current work is to test this prediction.

Fourth, this bias for robust information transfer is assumed to compete with the pressures and cognitive demands inherent to linguistic encoding (i.e., production ease). There is ample evidence that linguistic encoding such as retrieving and linearizing lexical information is resource demanding (for overviews, see Ferreira, 2008; MacDonald, 2013).[2] The ideal speaker framework assumes that the pressures for alleviating such encoding difficulties, and perhaps production *effort* in general, are weighted against a bias for robust information transfer (for related ideas, see Jaeger, 2006; Piantadosi et al., 2011; Zipf, 1949).

Fifth and finally, the ideal speaker framework proposes that speakers continuously and implicitly *learn* to balance these competing pressures (see Jaeger, 2013; Jaeger & Ferreira, 2013). A detailed discussion of these learning processes is beyond the scope of this paper. For the current purpose, it is sufficient to assume that speakers are sensitive to feedback about the likely communicative success of their past productions and that this feedback is integrated into subsequent production (Jaeger, 2013; Buz, Tanenhaus, & Jaeger, submitted for publication). In contexts in which the intended message is easily inferable, reduction (i.e., the use of less robust signals) will have no negative consequences and will be less likely to lead to negative feedback. In other contexts, however, speakers will receive at least implicit negative feedback (e.g., confusion on a listener's face or a failure to act according to the intended message). This feedback is hypothesized to affect subsequent productions. There are related alternative accounts that share with the ideal speaker framework that communication is assumed to shape speakers' production preferences, but deviate from the ideal speaker in that they do not assume speakers to learn from the perceived communicative success of their previous productions (e.g., Guy, 1996; Pierrehumbert, 2001, 2006). These accounts and the ideal speaker framework are mutually compatible. We return to these accounts and their relation to the ideal speaker framework in the discussion.

With these preliminaries in mind, we return to the goals of the present work. We primarily focus on the third and fourth assumptions of the ideal speaker framework. The ideal speaker framework predicts that speakers should provide more robust signals – usually requiring more effort and more time – only in contexts where communication difficulty is otherwise likely (and has therefore previously been experienced). We investigate whether grammatical encoding reflects such a bias for robust information transmission in addition to well-documented effects of production ease. To this end we present three recall experiments on Japanese speakers' production of optional case-marking in transitive sentences. We begin by introducing the phenomenon of optional case-marking and motivate its particular relevance for our goal. We then spell out the predictions of the ideal speaker framework and contrast them with the predictions of availability-based accounts of grammatical encoding. There are other production ease accounts that we will consider in the discussion (for a recent review, see MacDonald, 2013). For now, we focus on availability accounts as those have received most attention in previous work and are therefore best understood.

Optional case-marking in Japanese

Japanese allows both subject-object-verb (SOV) and object-subject-verb (OSV) orders. While constituent order carries a considerable amount of information (e.g., SOV is more frequent than OSV), constituent order alone leaves comprehenders with a comparatively high degree of uncertainty about the intended message. Japanese further lacks subject-verb agreement. This distinguishes Japanese from English, where morphological markers on the verb (e.g., *-s* for third person singular in the present tense) carry

[1] To be precise, more acoustic detail and exaggerated forms are not always expected to lead to better recognition. For example, there is evidence that highly frequent word forms are *expected* to be phonetically reduced (Sumner, Kurumada, Gafter, & Casillas, 2013). This does not change, however, that for each intended meaning, there are forms that are more or less well suited in supporting their recognition in the current context. Reducing the variability around these targets therefore will, on average, support robust information transfer.

[2] It should be noted, however, that experiments that assess production difficulty in isolated word and sentence production might over-estimate the production difficulty experienced during typical conversational speech (Pickering & Garrod, 2013; Tanenhaus, 2013). Indeed, while studies based on conversational speech have generally confirmed, for example, availability-based production (cf. Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Jaeger & Wasow, 2006; Roland et al., 2006; Wasow, 2002), some of the effects of production difficulty clearly observed in isolated production can be swamped by other factors in conversational speech (cf. Carbary, Frohning, & Tanenhaus, 2010; Healey, Purver, & Howes, 2014).

information about the grammatical function assignment intended by the speaker. As in many other languages with flexible constituent order, Japanese encodes additional information about grammatical function assignment via case-marking (de Hoop & Malchukov, 2008). This case-marking is realized morphologically with post-positional particles, as in (1): *-ga* is the nominative subject marker and *-o* is the accusative object marker.[3]

(1)	Taro-(ga)	Hanako-(o)	mi-ta.
	Taro-(SBJ)	Hanako-(OBJ)	see-PAST.
	Taro saw Hanako.		

What makes Japanese of particular interest for the current purpose is that case markers are *optional* in many environments. That is, Japanese speaker can often omit the case markers. In casual conversation the rate of omission is high (e.g., 32% for the nominative marker *-ga* and 51% for the accusative marker *-o*, Fry, 2003). Similarly high rates of case marker omission have been observed in casual speech in Korean (Lee, 2006).

Previous work on case-marking in Japanese has focused on semantic, pragmatic, and information structural constraints (e.g., Hinds, 1982; Kuno, 1973; Matsuda, 1995; Minashima, 2001; Onoe, 1987; Shibatani, 1990; Shimojo, 2006; Yatabe, 1999). For example, in certain linguistic environments, case-marking seems to be categorically omitted or retained (e.g., Kuno, 1973; Onoe, 1987; Shimojo, 2006). Crucially though – as also confirmed by the experiments we present below, there are environments in which case-marking seems to be optional. In this sense, Japanese case-marking resembles *that*-mentioning in English complement clauses (Ferreira & Dell, 2000), which is subject to obligatory constraints in some environments.

Here we consider optional case-marking as a window into grammatical encoding. We restrict our investigation to optional *object* case-marking, although the arguments presented here also apply to the optional subject-marking (for preliminary support for this prediction, see, e.g., Fry, 2003; Lee, 2006, described below).

Predictions

Under the reasonable assumption that the object case-marker *-o* is more likely to be successfully perceived as *-o* than to be misperceived as, for example, the subject case-marker *-ga*, producing the object case-marker will tend to be informative about the intended grammatical function assignment (henceforth GF-assignment). That is, producing *-o* will on average increase the probability that listeners will infer the GF-assignment intended by the speaker, thereby facilitating robust information transfer. Indeed, the presence of case-marking has been found to facilitate comprehension (Yamashita, 1997; Miyamoto, 2002).

[3] Here and below, we are using the following abbreviations in the aligned translations: SBJ – subject, OBJ – object, LOC – locative, GEN – genetive, PAST – past tense morpheme, SFP – sentence-final particle.

At the same time, producing case-marking adds to speakers' production effort as it increases the amount of signal the speaker needs to encode and articulate. For the case of optional *-o* marking, changes in both effort and the likelihood of communicative success are likely small, leading to relatively subtle trade-offs. However, even subtle trade-offs might be learnable over time. We thus hypothesize that speakers can learn trade-offs between (a) the production of *-o* and (b) the additional information that *-o* provides about GF-assignment beyond other cues. Hence, an ideal speaker account of grammatical encoding predicts that speakers should be more likely to produce the object case-marker when the intended GF-assignment is otherwise less inferable. This is the prediction we test below (see also Hinds (1982) and Kuno (1973), who propose that omission is *possible* when the grammatical function assignment is recoverable).

Consider the example in (2a), which is intended to have an SOV interpretation. SOV is the preferred constituent order, so that listeners will a priori be biased towards the intended GF-assignment. After processing the first noun *sensei* (teacher), the subject case-marker *-ga*, if correctly perceived, biases listeners further towards the intended GF-assignment. Next, the listener processes the second noun (*shobosha* (fire-engine)). In this example, the noun is inanimate, a property more common for referents of grammatical objects (cf., Silverstein, 1976; Comrie, 1981). In short, all properties of the sentence in (2a) bias comprehenders towards the intended interpretation. While object case-marking would further increase the probability of successful information transfer (this follows from the assumption of the noisy channel), it would add *little*.

(2)	a.	Sensei-ga	shobosha(-o)	mi-ta.
		teacher-SBJ	fire-engine(-OBJ)	see-PAST.
		The teacher saw a fire-engine.		
	b.	Sensei-ga	seito(-o)	mi-ta.
		teacher-SBJ	student-(OBJ)	see-PAST.
		The teacher saw a student.		

Example (2a) contrasts with (2b), in which the second noun refers to a human referent (a student). Incremental sentence comprehension is known to be sensitive to such animacy cues (Gennari & Macdonald, 2008; Mak, 2002; Traxler, 2002). Comprehenders will thus experience more difficulty while processing (2b), compared to (2a). Object case-marking, however, would ameliorate this problem, providing more additional information in (2b), compared to (2a). Given that the production of the case-marker is associated with the same effort in both examples, the framework pursued here predicts that the utility of case-marking is higher in (2b), compared to (2a). Speakers should therefore be more likely to produce object case-marking in (2b), compared to (2a). Availability accounts make no such prediction. In both (2a) and (2b) the material following the object and its case-marker is identical (the verb, *mi-ta* (saw)).

Experiment 1

Experiment 1 investigates whether Japanese speakers' preference to produce object case-marking is sensitive to the animacy of the direct object. We used a spoken-recall paradigm (Ferreira & Dell, 2000), and manipulated the animacy of the direct object (human vs. inanimate) and the presence of a direct object case-marker (*-o* present vs. absent) in a 2-by-2 Latin Square design. If speakers are more likely to mark the unexpected, they should be more likely to produce object case-markers when the object is human (atypical) compared to when it is inanimate (typical).

Previous evidence that speaks to this question comes from corpus-based studies of conversational Japanese and Korean (Fry, 2003; Lee, 2006). Lee (2006) finds that Korean speakers' preference in optional subject and object case-marking follows the predictions of the ideal speaker account: Korean speakers were more likely to produce case-marking on objects with atypical properties (e.g., human rather than inanimate, definite rather than indefinite) as well as subjects with atypical properties (e.g., inanimate rather than human, indefinite rather than definite).

In a corpus study of conversational Japanese, Fry (2003) replicates the results from Korean only partially. Fry finds that atypical subjects (inanimate and indefinite) are significantly more likely to be case-marked than typical subjects (animate and definite subjects). He does not, however, find any effects of animacy on object case-marking. Both studies (Fry, 2003; Lee, 2006) were conducted on relatively small data sets, so that it is possible that the failure to find an animacy effect for Japanese object case-marking is due to problems with statistical power. Another possibility is that the null effect for Japanese is real and that the *significant* effects of animacy in Korean were due to uncontrolled confounds. Both studies were conducted over heterogenous samples of conversational speech, in which animacy in all likelihood correlates with a variety of other referential properties (e.g., definiteness, givenness, topicality, etc.) as well as word order (Branigan et al., 2008). Neither of the studies employed statistical methods that control such potential confounds (unlike more recent corpus-based research, Bresnan et al., 2007; Jaeger, 2010b; Roland et al., 2006). Finally, it is possible that subject case-marking, but not object case-marking, is affected by animacy or that Japanese differs from Korean with regard to object case-marking. All but the first of these possibilities would be in conflict with the hypothesis that speakers are more likely to mark the unexpected. Experiment 1 addresses these possibilities.

Methods

Participants

23 native speakers of Japanese in the Stanford and the Rochester areas participated in this study. Participants were all born in Japan and had spent no more than five years in the United States. They received $7 for their participation.

Materials

We constructed 48 fillers and 24 items in each of their four conditions (i.e., 96 item stimuli). An example item in its four conditions is given in (3). Items always were transitive sentences presented in the dominant Subject-Object-Verb order. Subject referents were always human and object referents were also human when they were animate. The subject noun was always case-marked, thereby avoiding ambiguity about the intended GF-assignment. The object was always followed by an adverbial phrase (e.g., *near the train station*), followed by a verb (e.g., *saw*). The elements following the object NP were held constant. This was done in order to avoid potential effects on the use of case-markers due to the incremental planning of upcoming material, thereby ruling out availability-based production as an explanation for potential effects of animacy (Bock & Warren, 1985; Ferreira & Dell, 2000). The structure of items in Experiment 1 is illustrated in Fig. 1. The complete list of items is given in Appendix A.

(3) a. *Animate, -o present:*

Sensei-ga	**seito-o**	ekimae-de	mi-ta-yo.
teacher-SBJ	student-OBJ	station-LOC	see-PAST-SFP

b. *Animate, -o absent:*

Sensei-ga	**seito**	ekimae-de	mi-ta-yo.
teacher-SBJ	student	station-LOC	see-PAST-SFP

The teacher saw a student near the train station.

c. *Inanimate, -o present:*

Sensei-ga	**shobosha-o**	ekimae-de	mi-ta-yo.
teacher-SBJ	fire-engine-OBJ	station-LOC	see-PAST-SFP

d. *Inanimate, -o absent:*

Sensei-ga	**shobosha**	ekimae-de	mi-ta-yo.
teacher-SBJ	fire-engine	station-LOC	see-PAST-SFP.

The teacher saw a fire-engine near the train station.

Fillers were length-matched sentences with a subject, an intransitive verb, and a longer adverbial phrase. An example is given in (4). To mask the experimental manipulation of the object case-marker, half (24) of the filler sentences included omission of either the nominative marker *-ga* or the locative-marker *-ni*.

(4)

kameraman-ga	Hawai-no	umi-ni	mogut-ta-yo.
cameraman-SBJ	Hawaii-GEN	ocean-LOC	dive-PAST-SFP

The cameraman went diving in the Hawaiian ocean.

Both fillers and items were always in the past tense. There was no lexical overlap between any of the stimuli seen by a participant. Stimuli were grouped into pairs so that there were 12 filler-filler pairs and 24 item-filler pairs

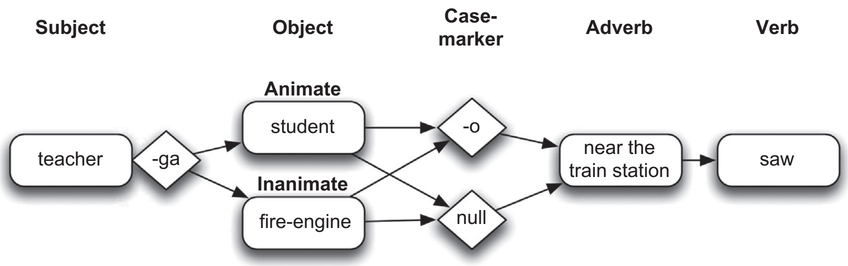

Fig. 1. Schematic illustration of the item design in Experiment 1.

(i.e., items were always paired with a filler), totaling 36 pairs. All stimuli were recordings of the same female native speaker of Japanese. In addition, the same speaker recorded a recall prompt for each sentence. For both items and fillers, this recall prompt was always the verb in the same tense as in the recall stimulus. To minimize effects of prosodic focus marking, we avoided placing phrasal-level accent on any of the nouns or case-markers (cf. Yatabe, 1999, who finds these accents to affect optional subject case-marking). [4]

Each participant was assigned to one of four lists, which together formed a Latin Square design (i.e., each list contained each item in exactly one of its conditions and all conditions equally often). A pseudo-random order of stimuli pairs was created, so that items of the same condition were maximally far apart from each other without making the order of conditions predictable and while interleaving filler-filler and filler-item pairs. The order of items and fillers within a pair and the order of pairs were held constant across lists.

Procedure

Each trial consisted of an encoding phase and a recall phase. During encoding, participants listened to a pair of sentences with a 3-second interval between them, and were instructed to remember them. Each sentence was heard only once. During recall, they heard the verb of one of the two sentences (the recall cue) and spoke the full sentence corresponding to that verb. Subsequently, the second recall cue was played prompting production of the other sentence. Across filler and item trials, the sentence encoded first was also recalled first for half of the trials. In the other half of the trials, the order was reversed. To reduce literal recall, target items were never recalled directly after encoding (Ferreira & Dell, 2000). That is, if a target item was encoded first, it was recalled either first or second. If it was encoded second, it was always recalled second. Across participants the recall order for each pair of stimuli was held constant.

The experiment started with four filler-filler practice trials, followed by the 36 trial pairs described above. Participants were tested individually in a sound-proof booth. All the instructions were given in written and spoken Japanese. The experiment took approximately 40 min to complete.

[4] Case-marking can also shift the locus of a lexical pitch accent (McCawley, 1968). Among the 24 target nouns we used in our stimuli, only one ("musume" *daughter*) was subject to such accent shift. Exclusion of this stimulus does not change any of the effects reported below.

Scoring

All 552 recorded item responses were transcribed and coded by a research assistant who is a native speaker of Japanese. When the coder was unsure if the case-marker was present or absent, the item was sent to another native speaker. When the second coder was not able to determine the presence, the item was excluded from the analysis (2% of the data).

Of the 23 participants, 3 always produced case-marking, 0 never produced case-marking, and 2 perfectly recalled case-marking in the input sentences. Since there is no evidence that these participants treated case-marking as an alternation, they were excluded from analysis (following Jaeger, Furth, & Hilliard, 2012a, 2012b). Exclusion did not affect the results reported below.

Of the remaining 432 responses, incomplete responses and responses with wrongly recalled subjects, objects, or verbs were considered recall errors and excluded (10.6%). This rate is comparable to, or lower than in, previous sentence recall experiments (e.g., Ferreira & Dell, 2000; Norcliffe & Jaeger, 2015). Responses including omission of, or close synonyms for, the adverbial phrases were not excluded. Participants always produced the subject in initial and the verb in the final position. Changes in the order of the object and adverbial phrase were not regarded as an error. This left 386 responses for analysis.

Although the rate of recall errors was numerically somewhat higher for animate objects without case-marker in the recall stimulus, tests revealed no statistically significant differences in recall rates by condition (see Table 1, $\chi^2(3) = 1.6, p > .6$).

Analysis

A mixed logit regression (Jaeger, 2008) with the full 2 (animacy of object) $\times$ 2 (presence of object case-marker in the stimulus) factorial design was employed to analyze the presence over absence of a case-marker in participants' responses.

All experiments were analyzed using the maximum random effect structure justified by the data based on

Table 1
Percentage of recall errors by condition in Experiment 1.

	Object animacy	
	Animate (%)	Inanimate (%)
Recall stimulus		
-*o* present	9.3	8.3
-*o* absent	13.0	12.0

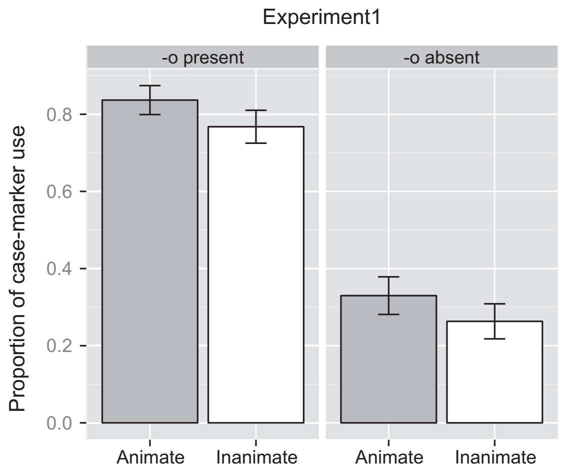

Fig. 2. Proportion of object case-marking in Experiment 1 by object animacy (animate vs. inanimate) and presence of case-marking in the recall stimulus (*-o* present vs. absent). The error bars show 95% confidence interval.

model comparison. For Experiment 1, this analysis contained random by-subject and by-item intercepts as well as by-item slopes for the presence of the case-marker in the stimulus. All predictors were ANOVA (i.e., sum) coded and there were no signs of collinearity in any of our analyses (fixed effect correlations *r*s < .3).

Results

Participants were more likely to produce object case-marking if the original stimulus contained object case-marking ($\hat{\beta} = 2.2, z = 6.4, p < .0001$). This recall effect was very strong: Speakers produced the case-marker approximately 80% of the time when the stimulus contained one, compared to about 30% of the time when the stimulus did not contain case-marking.

The overall high rate of *-o* is likely due to at least two reasons. First, case-markers are considered obligatory in written and more formal registers and some participants might have considered the context of the experiment as formal. In spontaneous conversation, case omission is high (as much as 51% for object case-markers, cf. Fry (2003, p. 8)). Second, critical stimuli always contained a subject case-marker. More generally, the clear majority of NPs in the experiment contained case-markers. This might have directly or indirectly (e.g., by evoking a more formal register) primed object case-marking.

Crucially, participants were more likely to produce case-marking if the object referent was animate ($\hat{\beta} = .40, z = 2.3$, $p < .05$). The two effects did not interact ($p > .7$). Fig. 2 summarizes the two main effects.

Discussion

These results indicate that animacy of the direct object is a predictor of Japanese speakers' production of case-marking. This replicates for Japanese what has been found in corpus-based studies on Korean speech (Lee, 2006). It also suggests that the null effect found by Fry (2003) was likely caused by data sparsity or confounding due to other factors influencing object case-marking. One reason for this might be that the effects of animacy on case-marking are relatively small. In Experiment 1, animate objects were about 8.4% more likely to be case-marked than inanimate objects (odds of 1.7 compared to 1.2).[5]

The availability of upcoming material (e.g., Bock & Warren, 1985; Branigan et al., 2008; Ferreira & Dell, 2000) is an unlikely explanation of the observed effect of animacy on object case-marking: The design of Experiment 1 held the material following the object (i.e., the adverb and the verb) constant. The ideal speaker framework, on the other hand, correctly predicts the animacy-based effects: *animate* referents are less expected to be the grammatical object of a sentence. Case is a cue to grammatical function assignment, making signals with case-marking more robust. While we are not aware of previous work that has investigated effects of animacy on *comprehension* in exactly the type of structure investigated here, previous work has provided evidence that animacy serves as a cue to sentence structure (Gennari & Macdonald, 2008; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Mak, 2002; Traxler, 2002; Traxler, Williams, Blozis, & Morris, 2005). These works have found that comprehension slows down when the animacy of referents biases comprehenders against the intended interpretation. Unexpected alignment of animacy and thematic roles, as in sentences in which the undergoer is higher on the animacy scale than the agent, can also lead comprehenders to misunderstand or misremember sentences, even when their grammatical structure rules out such interpretations. For example, Ferreira (2003) had comprehenders listen to sentences like "The mouse ate the cheese" or "The cheese ate the mouse". After each sentence, participants were prompted to identify the agent, patient or other thematic roles in the sentence. Ferreira found that sentences with implausible animacy assignments were considerably more likely to be misinterpreted. These findings strongly suggest that sentences with atypical (and thus less expected) alignment between grammatical function and animacy are harder to process.

[5] We note that a few properties of Experiment 1 biased *against* our hypothesis, which might have reduced the effect of animacy. For example, one anonymous reviewer proposed that the overall frequency of nouns, f(NOUN), or their co-occurrence frequency with the object case-marker, f(NOUN-O), in participants' previous experience might transfer into our experiment. Google counts (listed in Appendix B) suggest that this would bias against our hypothesis: web-based estimates of f(NOUN), f(NOUN-O), and f(NOUN-O $\mid$ NOUN) were significantly higher for inanimates than for animate nouns. We note that these counts also confirm our assumption that the inanimate nouns used in Experiment 1 are more likely to be objects, compared to animate nouns (as Google counts are primarily based of written registers for which case-marking is more common than in conversational speech, p(NOUN is object $\mid$ NOUN) $\approx f$(NOUN-O $\mid$ NOUN)). Finally, for clarity's sake, we note that none of these counts directly assesses the critical correlation tested by our experiments, namely whether p(NOUN-O $\mid$ NOUN is an object) $\propto p$(NOUN is an object $\mid$ NOUN).

For Experiment 1 this implies that the stimuli with human objects are likely subtly harder to process than those with inanimate objects. At the same time, it has been found that case-marking can guide comprehenders' expectations during sentence processing (Kamide et al., 2003; Kim, 1999; Yamashita, 1997). For example, Kamide et al. (2003, Experiment 3) find that Japanese listeners incrementally integrate the information provided by case-marking to predict upcoming material. Hence, producing case when the grammatical function assignment is less expected given the properties of the message – in this case, the animacy of the object – strikes an efficient balance between production ease and robust information transfer.

There are alternative (though related) interpretations of our findings that deserve discussion (e.g., exemplar-based accounts, Pierrehumbert, 2001, and similarity-based interference accounts of grammatical encoding, MacDonald, 2013). We return to these accounts in the general discussion. For now, we note that the ideal speaker framework predicts the animacy-based effect on optional case-marking observed here for Japanese and in previous work for Korean (Lee, 2006). Like previous research on optional case-marking (e.g., Fry, 2003; Lee, 2006), Experiment 1 leaves open whether only object properties (such as the animacy of the object referent) can affect object case-marking. If the perspective provided by ideal speaker framework is on the right track, we might expect other variables that affect the inferability of the intended GF-assignment to also affect case-marking. Experiment 2 tests this hypothesis.

Experiment 2

In Experiment 2, we used sentences in which both the subject and object were animate, so that both arguments made plausible subjects. These pairs of arguments were grouped with two verbs. The two arguments and verbs were chosen so as to manipulate the plausibility (and thus inferability) of the intended GF-assignment. For example, in (5), the verb *arrest* and the two arguments *police officer* and *criminal* allow both the plausible sentence in (5a) and the implausible sentence in (5b).

(5) a. *High plausibility:*

Keisatsu-ga	hannin(-o)	yonaka-ni	taihoshi-ta-yo.
policeofficer-SBJ	criminal-OBJ	night-at	arrest-PAST-SFP.

The police officer arrested the criminal in the middle of the night.

b. *Low plausibility:*

Hannin-ga	keisatsu(-o)	yonaka-ni	taihoshi-ta-yo.
criminal-SBJ	policeofficer-OBJ	night-at	arrest-PAST-SFP.

The criminal arrested the police officer in the middle of the night.

The same two arguments combined with the verb *attacked*, as in (6), make the opposite GF-assignment comparatively more plausible (this was confirmed by a norming study described below).

(6) a. *High plausibility:*

Hannin-ga	keisatsu(-o)	yonaka-ni	osot-ta-yo.
criminal-SBJ	policeofficer-OBJ	night-at	attack-PAST-SFP.

The criminal attacked the police officer in the middle of the night.

b. *Low plausibility:*

Keisatsu-ga	hannin(-o)	yonaka-ni	osot-ta-yo.
policeofficer-SBJ	criminal-OBJ	night-at	attack-PAST-SFP.

The police officer attacked the criminal in the middle of the night.

As in Experiment 1, the first argument was always the subject indicated by *-ga* and all material following the object was held constant within items. The design of Experiment 2 is illustrated in Fig. 3 (Experiment 3, described below, employed the same design).

Thus, the design of Experiment 2 formed a 2-by-2-by-2 design, crossing the presence of case-marking in the recall stimulus (*-o* present or absent), the identity of the verb, and the plausibility of the intended GF-assignment given the arguments and the verb (high vs. low).

Although verb identity was part of the design, it was a nuisance factor, for which no predictions were made. It merely served to facilitate the plausibility manipulation while holding other syntactic, semantic, and referential properties of the sentence constant. The ideal speaker framework predicts that the same noun is more likely to be case-marked when its GF is less inferable given the combination of verb and arguments (i.e., when the intended GF-assignment is less plausible).

Methods

Participants

32 native speakers of Japanese were recruited from the same populations as in Experiment 1 and received the same payment. The larger number of participants was motivated by the small effect sizes in Experiment 1. There was no overlap in participants between the experiments.

Stimuli

As in Experiment 1, the 24 items were combined with 48 length-matched fillers in a Latin-square design. There was no lexical overlap between any of the items and fillers. The order of stimuli was held constant across lists. The grouping of stimuli into pairs was the same as in Experiment 1. All stimuli were recorded by a female native speaker of Japanese. As in Experiment 1, the recall cue was the verb. The 24 items consisted of subject, object, adverb, and verb (in that order). Each item consisted of eight conditions, resulting from the $2 \times 2 \times 2$ design described above. The complete list of items is available in Appendix C. Items were created based on two norming studies, which we describe next.

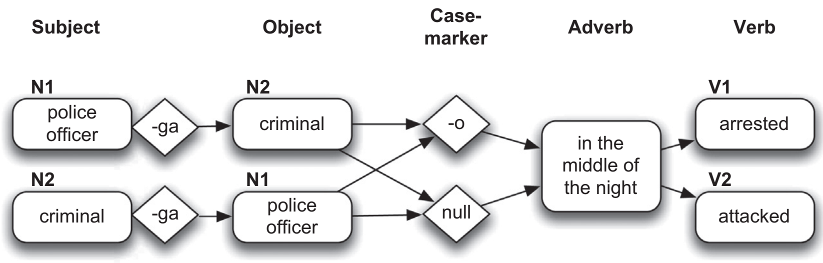

Fig. 3. Schematic illustration of the item design in Experiments 2 and 3.

Norming studies

We conducted two norming studies, both of which assess the plausibility of different GF-assignments. A total of 80 native speakers of Japanese were recruited over the web and completed the norming surveys online (40 in each study). Participants all had grown up in a Japanese monolingual environment until at least the age of 18. Both studies consisted of two blocks and took about 20 min to complete. All instructions and stimuli were given in written Japanese. The resulting norms are summarized in Appendix D.

Norming Study 1

Block 1 served to elicit estimates of the inferability of each of the two possible GF-assignments *given only the two arguments*. These estimates provided us with a manipulation check early in the process of stimuli creation. Participants saw the 24 animate–animate noun pairs used in our items (e.g., police officer-criminal). Participants rated the relative frequency of the two possible grammatical function assignments by distributing a total of 10 points across the variants (e.g., "The police officer (did something to) the criminal" vs. "The criminal (did something to) the police officer"). The results are included in Table 4.

Block 2 served to provide us with different verbs that made either of the two GF-assignments possible for each pair of nouns particularly plausible. Participants saw the same noun pairs as in Block 1 under both possible case-assignments (e.g., "police-officer-SBJ criminal-OBJ" and "criminal-SBJ police-officer-OBJ"). For both of these variants, they were asked to provide the most natural verb to complete the sentence.

For each pair of nouns we chose two of these verbs, so that one of the verbs was frequently mentioned under one GF-assignment and the other verb was frequently mentioned under the other GF-assignment. For example, for the *police officer-criminal* pair, we chose *(to) arrest*, which was frequently mentioned as a natural continuation for, "police-officer-SBJ criminal-OBJ" and we chose *(to) attack*, which was frequently mentioned as a natural continuation for "criminal-SBJ police-officer-OBJ". We then added an adverbial phrase to create stimuli with the "Noun-SBJ Noun-(OBJ) Adverb Verb" structure.

Norming Study 2

Participants were asked to rate the plausibility of the stimuli created based on Norming Study 1. Block 1 served to elicit estimates of the inferability of each of the two possible GF-assignments *given the two arguments, the adverb, and the verb*. This is the measure most crucial for our design. For example, participants would see the two sentence "police-officer-SBJ criminal-OBJ in-the-middle-of-the-night-LOC arrested" vs. "criminal-SBJ police-officer-OBJ in-the-middle-of-the-night-LOC arrested" and, as in Norming Study 1, distributed a total of 10 points across the two variants. Two lists were created using Latin-square design, so that each participant saw all 24 items with only one of the two verbs. All stimuli contained both subject and object case-marking (see Table 2).

As intended by our design, stimuli in the high plausibility condition were rated as more plausible (mean = .76, *SD* = .12) than stimuli in the low plausibility conditions (mean = .24, *SD* = .12). This difference was significant ($t(46) = 15.24, p < .0001$). (The resulting norms are listed in Table 4.)

Block 2 served to elicit estimates of the predictability of the two different verbs in each item, given the two arguments, the adverb, and the GF-assignment. Participants again saw 24 pairs of sentences with the same two arguments. This time, the GF-assignment was identical across the two sentences, but the sentences differed in their verb (e.g., "police-officer-SBJ criminal-OBJ in-the-middle-of-the-night-LOC *arrested*" vs. "police-officer-SBJ criminal-OBJ in-the-middle-of-the-night-LOC *attacked*'). Two lists were created using Latin-square design, so that each participant saw a pair of arguments under one GF-assignment. (The resulting norms are listed in Table 5.)

Procedure

The recall procedure was identical to that of Experiment 1. Since the meaning of some of the stimuli (e.g., *The criminal arrested the police officer*) were by design less plausible, participants were instructed to listen to sentences carefully and recall them faithfully even when they were "somewhat surprising".

Scoring

All 768 responses were transcribed and coded by two native speakers of Japanese. Of the 32 participants, 1 always produced case-marking and was excluded from analysis. This did not affect the results reported below.

Table 2
Percentage of recall errors by condition in Experiment 2.

	Plausibility	
	Low (%)	High (%)
Recall stimulus		
-*o* present	15.1	9.1
-*o* absent	11.3	8.1

Of the remaining 744 responses, 81 (10.9%) contained recall errors. This left 663 responses for our analysis. Numerically, the error rate was higher for sentences in the low plausibility condition (14.5% vs. 8.6%) but the difference was not statistically significant ($\chi^2(3) = 5.4, p > .14$).

Analysis

The same statistical procedure as in Experiment 1 was employed to assess the effect of plausibility on case-marking against the full $2 \times 2 \times 2$ factorial design of case-marking in the stimulus (present vs. absent), plausibility of GF-assignment (high vs. low) and verb identity.

Results

We found the expected recall effect: when the original stimulus contained case-marking, speakers were significantly more likely to produce case-marking in their response ($\hat{\beta} = 1.5, z = 9.7, p < .0001$). Crucially, there also was a significant main effect of plausibility: participants were more likely to produce case-marking in low plausibility condition, compared to the high plausibility condition ($\hat{\beta} = 0.4, z = 2.9, p < .02$). The two main effects did not interact ($p > .4$). They are illustrated in Fig. 4. Neither verb identity nor any of its interactions affected case-marking (all $ps > .17$).

Discussion

Experiment 2 conceptually replicates and extends the results of Experiment 1. Speakers were more likely to produce case-marking when the intended GF-assignment was otherwise less inferable. Unlike in Experiment 1, the two arguments of stimuli in Experiment 2 were always matched in terms of their animacy. Instead, we manipulated the inferability of the GF-assignment based on the entire sentence (i.e., the two nouns, the adverb, and the verb). Plausibility of intended interpretations has previously been found to affect comprehension speed (Garnsey et al., 1997) and accuracy (Ferreira, 2003). For example, Ferreira (2003) found that listeners were less accurate in answering questions about implausible, compared to plausible, sentences even when both arguments were of the same or similar animacy. The current findings are thus expected under the ideal speaker framework. By producing case-marking on the object speakers provide more cues to the intended sentence meaning, yielding a more robust signal. By doing so more often when the properties of the sentence would otherwise bias listeners away from the intended interpretation, speakers strike a balance between production effort and robust information transfer.

Fig. 5 further illustrates the effect of plausibility on case-marking. We used the mean plausibility rating score for each sentence from Norming Study 2 (Block 1) as an estimate of the sentence's GF-assignment inferability (henceforth, GF-inferability). Plausibility ratings were normalized (so that they ranged from 0 to 1) and log-transformed (following other work that has shown that log-transformed, rather than raw, predictability affects language processing, Smith & Levy, 2013, and production, Jaeger, 2006). As GF-inferability based on the full sentence increases, case-marking on the object becomes less common. When this continuous estimate of GF-inferability is substituted for the binary plausibility predictor in the analysis reported above, it is a significant predictor of object case-marking ($\hat{\beta} = -.43, z = -3.3, p < .001$). This effect did not interact with the recall effect ($p > .19$).

Interestingly, the continuous GF-inferability measure derived from the plausibility norms explained variance in case-marking over and above the binning of stimuli into low vs. high plausibility: adding the continuous predictor to the main analysis significantly improved the model (model comparison over difference in deviance, $\chi^2_{\Delta_{dev}}(1) = 11.0, p < .0001$). This further confirms that plausibility seems to be driving Japanese speakers' preference in optional object case-marking.

In sum, both the factorial analysis and the analysis of the continuous GF-inferability measure suggest that properties of the sentence beyond just the object referent can affect optional object case-marking. Is it possible that this result is, after all, due to object typicality (i.e., properties of the object referent) rather than the plausibility of the intended GF-assignment? Recall that Experiment 2 was designed to rule out explanations of the observed effects in terms of the typicality of the object referent (by having each referent occur equally often as the subject and object of the sentence). It is, however, conceivable that asymmetric data loss would have introduced a potential confound. To assess this possibility, we derived a second set of GF-inferability estimates based on only the two arguments (Norming Study 1, described above). Although these norms, too, were correlated with case-marking in the predicted direction, this correlation was weak (as intended by our design) and not a significant predictor of case-marking ($\hat{\beta} = -0.15, z = -.3, p > .7$).

This suggests that argument properties are insufficient to explain the results of Experiment 2. Instead, case-marking seems to be affected by the plausibility of the GF-assignment given both the arguments and the verb. This would mean that speakers have access to lexical (or at least event structure) information of the verb before they produce or omit the case-marker. In a language like English, where the verb typically precedes the object, this is hardly surprising (cf., Lindsley, 1975). On the other hand, for verb-final languages like Japanese, it is less clear how early

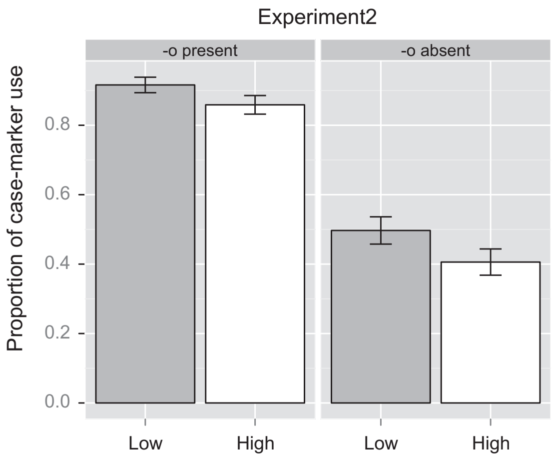

Fig. 4. Proportion of object case-marking in Experiment 2 by plausibility of GF-assignment (high vs. low) and presence of case-marking in the recall stimulus (-*o* present vs. absent). The error bars show 95% confidence interval.

different aspects of the verb meaning are planned (Iwasaki, 2011). This raises the question as to what extent the results of Experiment 2 depend on the type of recall prompt presented. Both Experiments 1 and 2 employed the verb as prompt, which arguably made the verb highly available during sentence recall. If the current results were aided by the use of verb prompt, they might not generalize to more incremental language production in everyday language use. To address this issue, Experiment 3 attempts to replicate Experiment 2 while employing a different recall prompt.

Experiment 3

The design and procedure of Experiment 3 were identical to Experiment 2. However, instead of the verb, Experiment 3 uses the subject noun of all stimuli as the recall prompt. This allows us to test if the plausibility effect we observed in Experiment 2 was due to the high availability of the verb.

Methods

Participants

26 native speakers of Japanese were recruited from the same populations as in Experiments 1 and 2 and received the same payment. There was no overlap in participants between any of the experiments.

Stimuli

The stimuli were identical to those of Experiment 2, with one exception. We replaced one item from Experiment 2 because, unlike all other items, it contained two proper names. Separate norms were obtained for this one item following the same procedure as in Experiment 2. All results reported below hold with and without this item included in the analysis. The new recall cues (the subject nouns) were recorded by the same speaker who recorded the stimuli for Experiment 2.

Procedure

The experimental procedure was identical to those of Experiment 2.

Scoring

Scoring followed the same procedure and was conducted by the same annotators as in Experiment 2. The recordings of one participant were lost due to a technical problem. Of the 25 remaining participants, 1 always produced case-marking and was excluded from analysis. Exclusion did not affect the results reported below. Of the remaining 576 responses, 109 (18.9%) contained recall errors and were excluded. This left 467 responses for analysis.

As shown in Table 3, the rate of recall errors was somewhat higher than in Experiment 2. As in Experiment 1 and 2, the rate of recall errors did not significantly differ across conditions ($\chi^2(3) = 5.6, p > .14$).

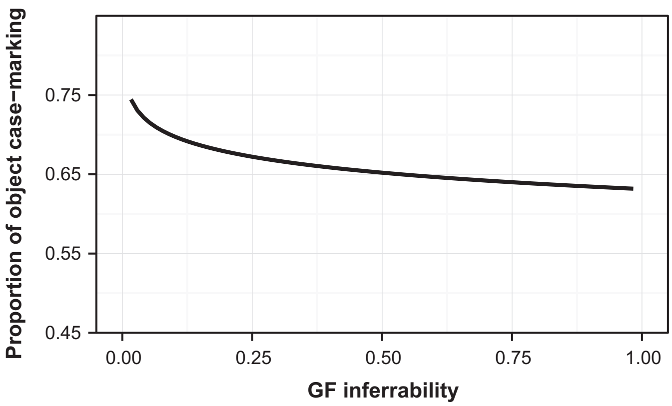

Fig. 5. Visualization of the effect of GF-inferability on speakers' case-marking preferences in Experiment 2, based on the mixed logit analysis. The proportion of responses with case-marking is plotted against the inferability of the intended GF-assignment. GF-inferability was derived by normalizing and log-trasforming the plausibility norms obtained in Norming Study 2.

Table 3
Percentage of recall errors by condition in Experiment 3.

	Plausibility	
	Low (%)	High (%)
Recall stimulus		
-*o* present	23.6	18.1
-*o* absent	13.2	20.8

Analysis

We conducted the same analysis as in Experiment 2.

Results

Unsurprisingly, we again found a significant main effect of recall: when the original stimulus contained object case-marking, speakers were more likely to produce it ($\hat{\beta} = 1.4, z = 5.7, p < .0001$). Crucially, we also replicated the main effect of plausibility: in the low plausibility condition speakers were more likely to produce object case-marking, compared to the high plausibility condition ($\hat{\beta} = -.5, z = 3.5, p < .001$). As in Experiments 1 and 2, the two effects did not interact ($p > .4$). Neither verb identity nor any of its interactions significantly affected case-marking (all *ps* > .2).

Discussion

The results of Experiment 3 replicate those of Experiment 2: We again find that speakers prefer to produce case-marking when the intended meaning of the sentence is otherwise less inferable (see Fig. 6).

As in Experiment 2, further analyses investigated the effect of (normalized and log-transformed) plausibility ratings from Norming Study 2, when substituted for the design factor of low vs. high plausibility. We again found that this continuous measure of the inferability of the GF-assignment based on the sentence's properties significantly affected case-marking: speakers were less likely to produce case-marking, the more inferable the GF-assignment was ($\hat{\beta} = -0.36, z = -2.7, p < .01$). Like in Experiment 2, this effect did not interact with the recall effect ($p > .9$). Unlike in Experiment 2, the improvement in predicting case-marking associated with the addition of the continuous measure to the model employed in the main analysis did not reach significance ($\chi^2_{\Delta_{dev}}(1) = .1, p > .7$). As in Experiment 2, plausibility ratings based on only the two arguments did not reach significance, as intended by our design ($p > .7$). This supports our interpretation that Experiments 2 and 3 exhibit effects on case-marking beyond argument properties.

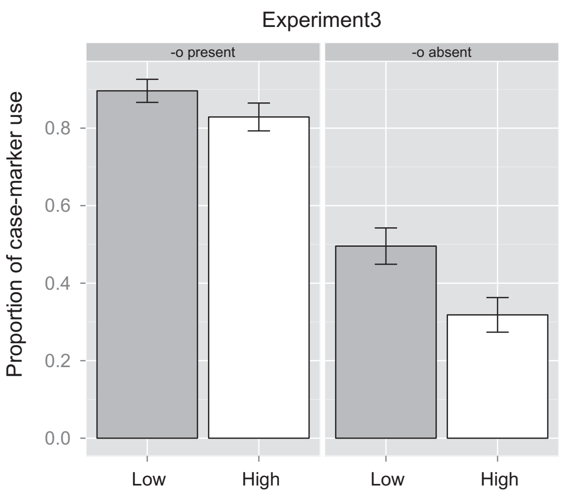

Fig. 6. Proportion of object case-marking in Experiment 3 by plausibility of GF-assignment (high vs. low) and presence of case-marking in the recall stimulus (-*o* present vs. absent). The error bars show 95% confidence interval.

General discussion

In three spoken recall experiments, we found that Japanese speakers are more likely to produce the optional object case-marker when the intended GF-assignment is less inferable from the other properties of the sentence. Experiment 1 found this effect based on the animacy of the object. This replicates previous findings from conversational Korean (Lee, 2006) and suggests that previous failures to find animacy effects on case-marking in Japanese were due to lack of power (Fry, 2003). Together with these studies, Experiment 1 suggests that the semantic properties of arguments – including, but not limited to, animacy – affect the rate of case-marking in languages in which case-marking is optional. For example, Fry (2003) and Lee (2006) found effects of definiteness, person, and previous mention on the rate of case-marking.

Experiment 2 and 3 found the same tendency to produce case-marking for unexpected GF-assignments even when the animacy of both arguments was held constant. These experiments provide evidence that the overall plausibility, or inferability, of the intended GF-assignment can affect speakers' decision to produce case-marking. Two findings in recent research are of particular interest in this context. First, Ahn and Cho (2007) compared case-marking rates in Korean, depending on the constituent order of the sentence. They found that both subject and object case-marking were optional in the canonical SOV order, but close to obligatory in the infrequent OSV order (which accounts for about 2% of sentences in Korean, Kim (2008)). The same preference to produce more case-marking in the OSV order was also found in a recent series of artificial language learning experiments. Fedzechkina, Jaeger, and Newport (2013) had monolingual English speakers learn a verb-final language with optional case-marking. They found that learners produced more case-marking in the OSV order, compared to the SOV order (Fedzechkina et al., 2013). In light of our findings, one possible interpretation of these findings is that case-marking becomes highly preferred when word order strongly biases against the intended interpretation.

Second, Lee and Kim (2012) report that omission of subject case-marking is acceptable even in Korean OSV sentences, provided the subject referent has properties that bias listeners towards the intended interpretation. Lee and Kim manipulate whether the subject referent is mentioned in the previous discourse or not. In two acceptability judgment experiments, Lee and Kim find that the omission of subject case-marking is *preferred*, compared to its retention, even in OSV sentence, when the subject has previously been mentioned.

These and our results are predicted by the hypothesis that language production reflects a trade-off between production ease and the goal to be understood. More specifically, these findings also suggest that even grammatical encoding reflects this trade-off (contrary to, e.g., Arnold, 2008; Ferreira, 2008; Ferreira & Dell, 2000; Lam & Watson, 2010; MacDonald, 2013). This prediction was derived in the introduction to this paper from general considerations about communication and language production. For brevity's sake we will continue to refer to this set of assumptions and the perspective on language production as the ideal speaker framework.

While the effects of GF-inferability observed in the current experiments were robust and in predicted directions, the estimated effect on the log-odds of case-marking is relatively small (though quite comparable to those observed by previous work on optional case-marking, Fry, 2003; Lee, 2006). We believe that several factor contribute to this. First, the recall effect (i.e., an overall tendency for participants to produce what they heard) was quite large, leaving relatively little room for variation in participants' production. Strong recall effects are common among sentence recall tasks (e.g., Ferreira & Dell, 2000; Norcliffe & Jaeger, 2015). Additionally, optional case-marking in Japanese is strongly affected by register, further limiting the variability that can be accounted for by GF-inferability. Indeed, some participants never omitted case-markers (see exclusions for Experiments 1–3). There was similarly strong item-specific variation. While the effects observed here hold across participants and items, strong participant- and item-specific effects further limit the influence of other effects. Finally, it is worth noting that case-marking is one of several means by which Japanese speakers encode GF-assignments. In fact, compared to other means such as word order alternation, the absence or presence of case-marking affects both the effort and the likelihood of robust communication only subtly. It might be useful to put our effect sizes in the context of other reduction effects. For example, instances of the same function words are produced with about 20 ms shorter duration in context that make them most predictable, compared to contexts that make them least unpredictable (5th vs. 95th percentile of predictability, cf. Bell et al., 2003). Like the effects observed here and in previous work on optional case-marking, this effect is relatively small in comparison to, for example, fluctuations in speech rate, which can easily double a word's duration. One reason for this might be that the pressures affecting word pronunciation, optional case-marking and other alternations are to a large extent conventionalized (for related ideas and discussion, see, e.g., Bresnan & Hay, 2006; Bybee, 2002; Fox & Thompson, 2007; Jaeger, 2006; Johnson, 1997; Pierrehumbert, 2001) – an idea that is quite compatible with the ideal speaker framework, as we discuss in more detail below.

Before we discuss the consequences of these results for psycholinguistic and linguistic theory, we address potential alternative explanations of our results in terms of production ease. To anticipate the outcome of this discussion, the results of our experiments are unlikely to be reducible to production ease, although we find evidence that there are *independent* effects of production ease on optional case-marking that are (mostly) orthogonal to the effects of GF-inferability observed in the main analyses reported above.

Can our findings be accounted for by production ease alone?

The experiments reported above were designed to test the ideal speaker framework while ruling out explanations in terms of availability-based production. To that end, the materials following the object (i.e., adverb and the sentence-final verb) were held constant in all three experiments. It is, however, possible that the materials preceding the verb, which were manipulated in our experiments, made the sentence-final verb more or less predictable. Availability has mostly been investigated in terms of *conceptual* accessibility, focusing on ease of retrieval of referential, rather than form information. As a consequence, the contribution of predictability to the availability of upcoming material is relatively underexplored (but see Frank & Jaeger, 2008; Gahl, Jurafsky, & Roland, 2004; Jaeger, 2006, 2010b; Kuperman & Bresnan, 2012). Existing findings draw a somewhat mixed picture. Some studies have found that the predictability of upcoming syntactic structures can affect the phonetic duration of preceding words (Gahl et al., 2004; Kuperman & Bresnan, 2012). Evidence that the predictability of upcoming words affects the omission of optional function words is also mixed (Levy & Jaeger, 2007; Jaeger, 2010b). For morphological processes (such as Japanese case-marking), too, the evidence is mixed. For morpho-syntactic reduction, previous work on English found no effects of predictability of upcoming material (Frank & Jaeger, 2008). For the omission of optional bound head-marking morphology, however, there is some evidence from Yucatec Maya that the availability of upcoming material can affect speakers' decision to produce or omit the morpheme (Norcliffe & Jaeger, 2015).

Experiments 2 and 3 allow us to address this question. As described above, the norming studies we conducted also elicited estimates of the expectedness of the verb, following the subject, object, and adverb in the sentence. We normalized those ratings and log-transformed them, creating an admittedly coarse-grained estimate of the verb's predictability given the preceding material. Unsurprisingly, this measure was relatively highly correlated with the GF-inferability measure we derived from the plausibility ratings (Experiments 2 and 3: both *r*s = .80). We thus asked which of these two measures explained Japanese speakers' preference in case-marking better. Following the same mixed logit regression approach as in the analyses reported above for GF-inferability, we first fit a comparable model for the verb predictability measure. Then we compared the GF-inferability model and the verb predictability model against a model that contained both effects (and their respective interactions with the recall effect as well as all random by-subject and -item slopes for the recall effect; larger models would not converge). This analysis revealed that GF-inferability marginally significantly added to a model with verb predictability (Experiment 2: $\chi^2(2) = 5.6$, $p < .08$; Experiment 3: $\chi^2(2) = 4.3, p < .11$) but not vice versa (Experiment 2: $p > .6$; Experiment 3: $p > .9$). In short, while it is possible that the availability of upcoming material (e.g., Bock & Warren, 1985; Ferreira, 1996; Ferreira & Dell, 2000) *also* contributes to the observed case-marking patterns, GF-inferability seems to affect case-marking at least as strongly.

In a recent review of the literature, MacDonald (2013) identifies two additional production preferences, she attributes to production ease: *plan reuse* (e.g., syntactic priming, Bock, 1986; Pickering & Branigan, 1998) and *reduce interference*. Syntactic priming is an implausible explanation for our experiments: experimental lists held constant what fillers preceded the critical trials (all lists employed the same order of stimuli). Additionally, lexical material was not repeated across stimuli, ruling out any form of lexical priming.

Reduce interference (MacDonald, 2013) refers to a hypothesized strategy of speakers to structure their utterances so as to avoid similarity-based interference. Such similarity-based interference has been found to affect *comprehension* (Gordon, Hendrick, Johnson, & Lee, 2006; Lewis & Nakayama, 2001; Van Dyke & McElree, 2011). Similarly, there is some evidence that similarity between elements in a sentence affects sentence planning (Bock, 1987; Ferreira & Firato, 2002; Gennari, Mirković, & Macdonald, 2012; Jaeger et al., 2012a, Jaeger, Furth, & Hilliard, 2012b; Smith & Wheeldon, 2004). For example, Bock (1987) found that speakers had a tendency to delay the production of words that shared phonological segments with a preceding prime word until later in the sentence (see also Jaeger et al., 2012a). Ferreira and Firato (2002) found that speakers were more likely to produce the optional complementizer *that* in English complement clauses if the complement clause subject shared semantic properties with referents mentioned earlier in the sentence (see also Gennari et al., 2012; but see Jaeger & Wasow, 2006, 176). Similarly, several studies have found that speakers prefer to avoid double *that* sequences in complement or relative clauses (e.g., *I believe (that) that is true*, Jaeger, 2012; Lee & Gibbons, 2007; Walter & Jaeger, 2005). These findings suggest that similarity between elements in a sentence can cause production difficulty and that this can affect speakers' preferences between several near meaning-equivalent realizations of their message – such as producing or not producing the optional *that*. It is an open question whether speakers actively seek to reduce such interferences (as proposed by Dell, Oppenheim, & Kittredge, 2008; MacDonald, 2013) or whether similarity-based effects on, for example, constituent ordering are a consequence of delayed lexical and phonological planning (Bock, 1987; for discussion, see Jaeger et al., 2012b). If similarity-based effects on sentence production are indeed due to attempt to reduce interference, the next question is whether it is only interference during production or also anticipated interference in comprehension that drives this strategy. These questions deserve further attention in future work.

Regardless of the answer though, it is of interest to understand whether the results of our experiments could be due to similarity-based interference. To address this question, we conducted two follow-up analyses. First, we investigated the distribution of recall errors across plausibility conditions. Recall errors provide a measure of potential similarity-based interference during the comprehension of the stimulus as well as its storage, retrieval, and production. For Experiment 1, similarity-based interference in memory should lead to increased recall errors for animate objects (recall that subject referents were always animate in Experiment 1). This prediction was not supported ($\chi^2(1) = 0.02, p > .8$). For Experiments 2 and 3 the prediction of a similarity-based interference account is less clear. In those experiments both the subject and the object always were animate referents. Without further assumptions, the reduce interference hypothesis therefore neither predicts any effects on recall errors nor any effects on case-marking – contrary to what we observed. There was a marginal effect of plausibility in Experiment 2 ($\chi^2(1) = 3.5, p < .06$), such that there were marginally fewer recall errors in the plausible condition. There was no effect of plausibility on recall errors in Experiment 3 ($p > .8$). In short, recall errors revealed at best weak evidence for similarity-based interference in our experiments.

Second, we tested whether there were any signs of similarity-based interference in the production data. To this end, responses that included filled pauses (e.g., *ah*, *eeto*), false starts, repetitions, elongated pronunciation of a word followed by a pause, or self-directed speech (e.g., *Nanda-kke* "Cannot remember") were annotated as disfluent. Disfluencies are well-known to indicate production difficulty (e.g., Fox Tree & Clark, 1997; Clark & Fox Tree, 2002; Shriberg & Stolcke, 1996; for Japanese, see Watanabe, Den, Hirose, Miwa, & Minematsu, 2006; Yamashita & Kondo, 2009). On average about 14–17% of the response contained a disfluency (Experiment 1: 21%; Experiment 2: 14.3%; Experiment 3: 15.2%). Crucially, disfluency rates did not differ by animacy or plausibility condition (Experiment 1: $p > .5$; Experiment 2: $p > .9$; Experiment 3: $p > .08$). In short, we find little evidence that object animacy or plausibility caused interference in memory or affected production difficulty.

We thus tentatively conclude that none of the three most commonly discussed contributors to production ease are sufficient to explain the effect of our GF-inferability manipulations in Experiments 1–3. This leaves open whether there are *independent* effects of production difficulty on optional case-marking. This would mirror the well-documented effect of production ease on grammatical encoding. It would also be compatible with the ideal speaker framework, which explains speakers' preferences as a trade-off between production pressures and the goal to be understood.

Are there any effects of production ease on case-marker omission?

To investigate whether optional bound morphology, such as the optional case-marking in Japanese, can be affected by production difficulty, in addition to the effects of our GF-inferability manipulations, we conducted additional follow-up analyses. We calculated the average rate of recall mistakes across participants for each unique stimulus (the unique combination of item and design conditions). This provides a measure of the difficulty of successful recall. We included this measure and the presence of a disfluency in the response as an additional control predictor in the same models we employed in our main analyses (no random slopes were included for the control predictors; for Experiments 2 and 3 we removed verb identity, which had no effect on case-marking). In all analyses disfluent responses were more likely to contain object case-marking, although the effect was significant only in Experiments 1 and 3 (Experiment 1: $\hat{\beta} = .4, z = 2.2$, $p < .05$; Experiment 2: $p > .7$; Experiment 3: $\hat{\beta} = .9$, $z = 2.4, p < .02$). The average rate of recall errors also emerged as a predictor of case-marking in Experiments 1 and 3 (Experiment 1: $\hat{\beta} = 2.8, z = 1.94, p < .06$, Experiment 3: $\hat{\beta} = 2.3, z = 2.7, p < .01$; Experiment 2: *ps* > .17).

This suggests that production difficulty *does* affect optional case-marking, consistent with a large body of previous work on grammatical encoding (for recent reviews, see Ferreira, 2008; MacDonald, 2013), and, specifically the mention of optional elements (e.g. Ferreira & Dell, 2000; Ferreira & Firato, 2002; Jaeger, 2005, 2010b; Race & MacDonald, 2003; Roland et al., 2006).[6] The effect of average recall error rates on the mention of optional case-marking further supports recent findings of similarity-based interference effects on production preferences (Gennari et al., 2012).

Importantly, the main effects of GF-inferability (Object animacy in Experiment 1 and plausibility in Experiments 2 and 3) remained significant after controlling the recall errors and disfluencies (Experiment 1: $p < .05$, Experiment 2: $p < .02$; Experiment 3: $p < .01$). Additionally, fixed effect correlations between either of the two measures of production ease and GF-inferability were low (*r*s < .3). Taken together this shows that – at least for our experiments – production ease and communicative goals both have independent effects on optional case-marking. This echoes previous work on the reduction or omission of optional bound morphology, which has found effects of both production ease and robust information transmission (Frank & Jaeger, 2008; Norcliffe & Jaeger, 2015). For example, Norcliffe and Jaeger (2015) investigate Yucatec Maya – a head-marking, rather than dependent-marking, language spoken in parts of Mexico and Guatemala – and find that speakers' preference to realize optional morphology is affected by both the redundancy of the morphology and by the availability of upcoming material.

All of this raises the questions as to how the trade-off between production ease and communicative goals comes to affect language production. That is the question we turn to next.

How does the trade-off between production effort and robust information transfer enter language production?

As we outlined in the introduction, a number of previous studies have attempted to provide accounts for phonetic, phonological and morphological reduction with an emphasis on the roles of repeated experiences. One class of accounts focuses on the role of 'chunking' (Boyland, 1998; Haiman, 1994) or 'automatization' (Bybee, 2002, 2006; Bybee & Scheibman, 1999): "a frequently repeated stretch of speech becomes automated as a processing unit" (Bybee & Scheibman, 1999, 577). This automatization is assumed to result in joint storage of frequently co-occurring units in memory. Hence, automatization accounts attribute reduction to *repetition* or *frequency of use* (e.g., Bybee & Scheibman, 1999, 576).

The original purpose of these proposals was to account for effects of language use on diachronic changes to word's phonetic, phonological, and morphological representations, including changes within the life-span of a speaker. Without further assumptions, these accounts predict that increasing use of a word will lead to uniform reduction of that word. This is, however, not what is observed: reduction is not random or uniform across a word, but rather seems to be quite targeted (e.g., Buz, Jaeger, & Tanenhaus, 2014; Johnson, 1998; Kirov & Wilson, 2012). For example, recent studies have found hyper-articulation of specifically those segments of words that made the pronunciation contrastive with contextually salient alternatives, without hyper-articulation of other segments (Buz et al., 2014; Kirov & Wilson, 2012).

Another potential problem in extending automatization accounts to optional case-marking is that these accounts tend to focus on the frequency of occurrence of surface forms (Bybee & Scheibman, 1999; rather than predictability or inferability of forms with the meaning encoding by them). This is problematic when trying to account for results such as ours: we varied the inferability of GF-assignment while holding constant the frequency of the case-marker (which always was the object case-marker). Similarly, as discussed above (Footnote 5), an account that attributes speakers' preference to the co-occurrence frequency of nouns with the object case-marker will make the wrong predictions for our experiments.

In the remainder of this section, we focus on two broad classes of accounts, that can potentially capture the type of pattern observed in Experiments 1–3. In both accounts, the bias towards the conservation of production effort is counteracted by the consequences that reduction or omission

[6] We note that the *cause* for the observed effects of fluency and average recall error rates will require further studies. For example, while it is broadly accepted that overt disfluencies are a consequence of production difficulty, it is still a matter of debate whether producing a disfluency – rather than silence – also serve to facilitate information transfer (cf. Arnold, Hudson-Kam, & Tanenhaus, 2007; Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Clark & Fox Tree, 2002; Jaeger, 2013).

Table 4
Items used in Experiments 2 and 3 along with the normalized G(rammatical) F(unction)-assignment plausibility ratings. Columns 1–4 describe the stmuli. Column 5 provides a ratio of norming judgments elicited in Norming Study 1 Block 1. For example, for the item in the first row it shows the ratio of points given for "*elephant-ga* (SBJ) *ant-o* (OBJ)" and points for "*ant-ga* (SBJ) *elephant-o* (OBJ)". Columns 6–9 provide results of Norming Study 2 Block 1, in which GF-assignment plausibility was rated based on the two arguments (Columns 1–2), adverbial (not shown, but see Appendix C) and a verb (Columns 3–4). For the item in the first row, Columns 6 and 7 show the relative plausibility of the N1-*ga* N2-*o* and N2-*ga* N1-*o* when the verb is V1 (i.e., *The elephant stamped on the ant with full force* vs. *The ant stamped on the elephant with full force*). Columns 7 and 8 show the relative predictability of the same GF-assignment for V2 (i.e., *The elephant bit the ant with full force* vs. *The ant bit the elephant with full force*).

N(oun)1	N(oun)2	V(erb)1	V(erb)2	GF-assignment plausibility ratings with two nouns	GF-assignment plausibility ratings with verbs			
				N1-*ga* N2-*o*/ N2-*ga* N1-*o*	N1-*ga* N2-*o* given V1	N2-*ga* N1-*o* given V1	N1-*ga* N2-*o* given V2	N2-*ga* N1-*o* given V2
elephant	ant	stamped on	bit	0.66	0.9	0.1	0.22	0.78
police officer	criminal	arrested	attacked	0.59	0.97	0.03	0.2	0.8
reader	author	searched for	criticized	0.48	0.81	0.19	0.37	0.63
king	soldier	sent	assaulted	0.6	0.98	0.02	0.3	0.7
teacher	student	scolded	spoke to	0.56	0.69	0.31	0.47	0.53
life guard	child	rescued	teased	0.70	0.84	0.16	0.31	0.69
CEO	secretary	discharged	punched	0.57	0.9	0.1	0.44	0.56
painter	model	drew	seduced	0.66	0.81	0.19	0.44	0.56
clerk	customer	have X wait	yelled at	0.52	0.67	0.33	0.26	0.74
mother	baby	tended	called	0.6	0.75	0.25	0.48	0.52
monkey	dog	scratched	chased	0.5	0.66	0.34	0.47	0.53
citizen	politician	chose	deceived	0.54	0.97	0.03	0.19	0.81
protagonist	antagonist	beat	waited for	0.54	0.86	0.14	0.14	0.86
shop owner	part-time worker	give advice to	calm down	0.63	0.91	0.09	0.25	0.75
grandma	grandchild	held	troubled	0.66	0.84	0.16	0.23	0.77
actress	housekeeper	boss around	helped	0.62	0.83	0.17	0.47	0.53
coach	player	trained	throw X into the air	0.58	0.96	0.04	0.12	0.87
chef	apprentice	fired	infuriate	0.62	0.98	0.02	0.25	0.75
helper	elderly	visited	called out	0.55	0.9	0.1	0.19	0.81
Jaian	Nobita	bullied	waylaid	0.44	0.85	0.15	0.39	0.61
security guard	shoplifter	caught	fled from	0.62	0.91	0.09	0.14	0.86
comedian	audience	entertained	booed at	0.62	0.9	0.1	0.3	0.7
manager	new worker	cheered up	feared	0.63	0.91	0.09	0.32	0.68
station officer	drunkard	woke up	hit	0.53	0.94	0.06	0.25	0.75
doctor	patient	treated	awaited	0.64	0.81	0.19	0.38	0.62

has on the likelihood of successful communication. The two accounts differ, however, in how the latter bias comes into effect.

According to the first class of accounts, communicative effects on production preferences – such as the preference to produce or omit a case-marker – occur across repeated exchanges *between* interlocutors. An example of this type of account is the exemplar-based account of phonetic reduction proposed by Pierrehumbert (2001). Comprehenders are hypothesized to store more or less pre-processed percepts along with their linguistic labels in memory (e.g., the acoustic realization of a word along with its lexical identity). Each time a speaker produces an instance of a word, they sample a pronunciation based on the cloud of exemplars previously experienced or produced. The sampled pronunciation is assumed to be perturbed by random noise and a preference for conservation of effort, the "leniting bias". That is, pronunciations tend to be more reduced than the mean of previously experienced exemplars. As a consequence, the more often a word is produced, the more the cloud of exemplars will shift towards increasingly reduced forms, creating a positive correlation between frequency of use and degree of reduction. Counteracting this bias for reduction is the production-perception loop *between interlocutors*: if a word is reduced in a way that makes it too confusable, listeners will not be able to correctly understand it, so that the exemplar will not be stored (Pierrehumbert, 2002, 110). As a consequence, this particular exemplar will not affect the listener's sampling process in subsequent productions of that target word. This principle – applied across a population of speakers – prevents exemplars that are too confusable (e.g., being reduced in a wrong context or being reduced in a wrong way) from being stored and hence from spreading through the language. That is, communicatively bad productions are effectively 'weeded out' through processes operating during comprehension (see also Guy, 1996).

This in principle allows exemplar-based models to account for the fact that reduction does not apply

Table 5

Items used in Experiments 2 and 3 along with the normalized verb predictability ratings. Columns 1–4 describe the stimuli. Columns 5–8 provide verb predictability ratings elicited in Norming Study 2 Block 2. Verb predictability was rated conditional on the two arguments (Columns 1–2), adverbial (not shown, but see Appendix C) and a G(rammatical) F(unction) assignment. For example, for the item in the first row, Columns 5 and 6 show the relative predictability of the verbs *stamped on* and *bit* after *elephant-ga ant-o with full force*. Columns 7 and 8 show the relative predictability of the same two verbs for the reversed GF-assignment (i.e., *ant-ga elephant-o with full force*).

N(oun)1	N(oun)2	V(erb)1	V(erb)2	Verb predictability scores			
				V1 given N1-*ga* N2-*o*	V2 given N1-*ga* N2-*o*	V1 given N2-*ga* N1-*o*	V2 given N2-*ga* N1-*o*
elephant	ant	stamped on	bit	0.9	0.1	0.26	0.74
police officer	criminal	arrested	attacked	0.85	0.15	0.14	0.86
reader	author	searched for	criticized	0.5	0.5	0.36	0.64
king	soldier	sent	assaulted	0.83	0.17	0.13	0.86
teacher	student	scolded	spoke to	0.51	0.49	0.23	0.77
life guard	child	rescued	teased	0.74	0.26	0.25	0.75
CEO	secretary	discharged	punched	0.54	0.46	0.11	0.89
painter	model	drew	seduced	0.59	0.41	0.2	0.8
clerk	customer	have X wait	yelled at	0.77	0.23	0.25	0.74
mother	baby	tended	called	0.63	0.37	0.07	0.93
monkey	dog	scratched	chased	0.5	0.5	0.23	0.77
citizen	politician	chose	deceived	0.83	0.17	0.1	0.9
protagonist	antagonist	beat	waited for	0.82	0.18	0.33	0.67
shop owner	part-time worker	give advice to	calm down	0.44	0.56	0.24	0.76
grandma	grandchild	held	troubled	0.9	0.1	0.3	0.7
actress	housekeeper	boss around	helped	0.66	0.34	0.3	0.7
coach	player	trained	throw X into the air	0.75	0.25	0.17	0.83
chef	apprentice	fired	infuriate	0.71	0.29	0.29	0.71
helper	elderly	visited	called out	0.88	0.12	0.2	0.8
Jaian	Nobita	bullied	waylaid	0.72	0.28	0.46	0.54
security guard	shoplifter	caught	fled from	0.89	0.11	0.25	0.75
comedian	audience	entertained	booed at	0.79	0.21	0.25	0.75
manager	new worker	cheered up	feared	0.69	0.31	0.37	0.63
station officer	drunkard	woke up	hit	0.75	0.25	0.25	0.75
doctor	patient	treated	awaited	0.69	0.31	0.2	0.8

uniformly, but rather in targeted ways (Johnson, 1998; Kirov & Wilson, 2012). For example, Kirov and Wilson (2012) have speakers produce target words that form minimal pairs with another word (e.g., *pin* vs. *bin*). The minimal pair competitor (e.g., *bin*) either was or was not co-present on the screen. Kirov and Wilson found that speakers were more likely to hyper-articulate the acoustic cue to the target word – in this case, voice onset time – when the competitor was present (see also Baese-Berk & Goldrick, 2009). Crucially, the hyper-articulation seems to be focused on the segment that distinguishes the target and the competitor (e.g., the/p/ in *pin*). With some assumptions necessary to extend the model of phonetic production to one of morphological production, a qualitatively similar approach should in principle be able to account for context-specific preferences to omit or produce optional case-markers, such as the plausibility effects observed in Experiments 2 and 3.

The second class of accounts does not limit communicative effects to repeated exchanges between interlocutors. The ideal speaker framework that we outlined in the introduction is an example of such accounts. It is in spirit closely related to the chunking, automatization, and exemplar-based proposals discussed so far. Like these accounts, the ideal speaker can change as a function of previous experience. Unlike these previous accounts, however, we propose that speakers' preferences during linguistic encoding can also be affected by previous experience with their own productions, rather than only through the production-perception loop between interlocutors. Specifically, the ideal speaker framework holds that speakers adapt their productions based on the perceived communicative success of previous productions, thereby *learning* to decide whether to produce an optional element or not in a given context (Jaeger, 2013; Jaeger & Ferreira, 2013).

This learning is assumed to be mostly implicit and at least to some degree context-specific (e.g., including adaptations specific to certain types of audiences, such as clear speech in the presence of noise, Smiljanic & Bradlow, 2005, and foreign directed speech, Scarborough, Brenier, Zhao, Hall-lew, & Dmitrieva, 2007; Smiljanic & Bradlow, 2011). In this view, language users can use perception of their own utterances and feedback from their interlocutors. Interlocutor feedback can be implicit (e.g., facial expressions signaling confusion, failure to react in an expected way) or explicit (requests for clarification) and it can be immediate or delayed (e.g., when a misunderstanding becomes apparent later). If speakers indeed adjust their productions based on the perceived communicative success on previous production, this learning mechanism would explain how speakers learn when to reduce and when to produce more signal.

To determine which account can more accurately predict language users' behavior, it is important for future work to investigate whether and how effectively speakers can learn from the perceived communicative success of their utterances. We briefly discuss existing evidence that speaks to this question. Evidence for speakers' ability to adapt their productions based on recent experience with their own articulations has come from perturbation studies (e.g., Houde, 1998; Tourville, Reilly, & Guenther, 2008; Villacorta, Perkell, & Guenther, 2007). These studies have shown that speakers can adjust their articulation if they perceive their own productions to deviate from the intended target. Crucially, these adaptations take place in the absence of an interlocutor that reproduces the words back to the speaker.

While broadly compatible with exemplar-based accounts (Johnson, 1997; Pierrehumbert, 2001), findings like these show that, at least for articulation, an account that solely relies on processes taking place across interlocutors is not *sufficient* to capture speakers' ability to learn from their own productions (for further evidence that speakers can learn from their own productions, see Dell, Reed, Adams, & Meyer, 2000; Goldrick, 2004; Warker & Dell, 2006).

There is also preliminary evidence that even non-verbal feedback from interlocutors can affect subsequent productions (Buz et al., submitted for publication; Roche, Dale, & Kreuz, 2010; Schertz, 2013; Stent, Huffman, & Brennan, 2008). For example, Stent et al. (2008) had participants produce utterances that were analyzed by an automatic speech recognizer (or so participants were told). On some trials, the speech recognizer failed to recognize participants' speech. Stent and colleagues observed hyper-articulation following those trials. Similarly, Buz et al. (submitted for publication) observed hyper-articulation following communicative failure. In their experiment, participants were instructed to pronounce one of the three words displayed on a computer screen, instructing an interlocutor to click on that word. Replicating Baese-Berk and Goldrick (2009), they found that participants hyper-articulated in the presence of a minimal pair neighbor (e.g., *pin* and *bin*). Crucially, this effect was magnified when the (confederate) interlocutor had selected the wrong word on previous trials (see also Schertz, 2013). This suggests that speakers can adapt their productions based on non-verbal feedback even in the absence of the production-perception loop between interlocutors (of the type assumed in Guy, 1996; Pierrehumbert, 2002).

In summary, previous studies provide preliminary evidence that speakers can in principle adjust their productions if they receive evidence that such adjustment is necessary to increase the chance of communicative success. It is thus possible that speakers' preference for morphological or syntactic omission, including optional case-marking in Japanese, are affected through a similar learning process. We consider further research on this question a particularly promising venue for future work.

The ideal speaker trade-off and linguistic typology

The type of trade-off between the conservation of effort (sometimes referred to as 'economy') and robust information transfer that we have investigated here for language production has also received considerable attention in the linguistic literature. Many languages seem to have grammaticalized categorical constraints that reflect this trade-off. One example that is particularly relevant for the current discussion are *differential object marking* (DOM) systems. Unlike the type of optional case-marking system studied here, DOM systems categorically mark some types of objects and categorically do not mark other types (Aissen, 2003).

Strikingly, the types of objects for which case-marking is obligatory in DOM systems tend to be the same as those that are more likely to be case-marked in optional case-marking system (Aissen, 2003; Bossong, 1991; Comrie, 1981; de Hoop & Malchukov, 2008; de Swart, 2007, 2011; Malchukov, 2008; Silverstein, 1976). That is, objects with referential properties that are typical for grammatical objects (e.g., inanimate and indefinite objects) can be categorically unmarked while atypical types of objects (e.g., animate and definite objects) are obligatorily case-marked. Though cross-linguistically more variable, the same tendency of marking the atypical is observed for differential subject marking (cf., Aissen, 2003). In the linguistic literature, this generalization has been discussed in terms similar to the framework pursued here. In the words of Aissen (2003, 438), "the overt marking of atypical objects facilitates comprehension where it is most needed, but not elsewhere. DOM systems are thus relatively economical". Others state that DOM systems seem to use the case-marker "primarily to avoid ambiguity" (Comrie, 2011, chap. 98, referring to Bromley, 1981).

The two dominant accounts attribute these types of grammatical patterns to either *frequency of use* (e.g., Bybee, 2006; Bybee & Hopper, 2001; Haspelmath, 2008; for an overview, see also Diessel, 2007) or the trade-off between *economy* and *iconicity* (e.g., Aissen, 2003). The ideal speaker framework can be seen as incorporating aspects of both frequency-based approaches and the economy vs. iconicity tradition. Like the latter (see, in particular, Zipf, 1949), the ideal speaker attributes reduction to a trade-off between a force for conservation of effort and a need for a sufficiently distinguishable linguistic signal. Like the former, reduction is expected to be affected by linguistic distributions–just that is not frequency of use per se, but rather its effect on comprehenders' expectations that drive reduction (an assumptions shared with, and supported by, e.g., Gibson et al., 2013; Jaeger, 2006; Levy & Jaeger, 2007; Piantadosi et al., 2011).

Regardless of which of these accounts ultimately explains differential case-marking and related grammatical phenomena, the current results together with those of previous work (Fry, 2003; Lee, 2006; Lee & Kim, 2012) suggest that optional case-marking system are sensitive to the same type of factors that affect categorical case systems. This provides support for functionally motivated accounts of differential case-marking (Aissen, 2003). This link between gradient preference patterns in production and categorical patterns in grammar (see also Bresnan, Dingare, & Manning, 2001) further grounds the hypothesis that at least some of cross-linguistical generalizations have functional motivations and stem from language use (e.g., Bates & MacWhinney, 1982; Bornkessel-Schlesewsky & Schlesewsky, 2009; Givón, 1998; Hawkins, 1994, 2004).

Further support for this conclusion comes from a study directly related to the current experiments. Fedzechkina, Jaeger, and Newport (2012) had monolingual speakers of English learn a miniature artificial language resembling Japanese in several aspects: the miniature language was verb-final with SOV and OSV constituent order and optional case-marking. Unlike Japanese, however, the input language that participants were exposed to did *not* condition case-marking on the animacy of the arguments. Intriguingly, the monolingual English participants – speakers of a language with no productive case-marking system – induced the same pattern into their productions that we have seen here. After acquisition of the language was completed, participants preferably marked animate objects and inanimate subjects with case, compared to inanimate objects and animate subjects. That is, the learners in the study by Fedzechkina and colleagues seem to induce the same type of preferential case-marking pattern we observed in our experiments.

Conclusion

We set out to test the hypothesis that communicative goals – specifically, the goal to be understood – can affect the processes underlying grammatical encoding. Grammatical encoding is one of the earliest processes in the translation of a pre-linguistic message into a linguistic signal. Work over recent years has proposed that all of language production, including grammatical encoding, is organized so as to facilitate robust information transfer, trading off with the pressures inherent to the linguistic encoding processes (Jaeger, 2013).

The results of our recall experiments on Japanese optional case-marking showed that Japanese speakers were more likely to produce case-marking when the grammatical function assignment indicated by the case-marker would otherwise be hard to infer. This preference contributes to robust information transfer by providing additional cues to the intended sentence interpretation when they are most needed. In addition, we observed independent effects on case-marking that are compatible with production ease accounts of grammatical encoding (e.g., Ferreira & Dell, 2000; MacDonald, 2013). Optionality and variation in linguistic forms thus seem to be exploited to satisfy both the demands inherent to the grammatical planning of utterances and communicative goals.

Our findings thus add to a growing body of evidence (e.g., Jaeger, 2010b; Kravtchenko, 2014; Resnik, 1996; Wasow et al., 2011) that not only late processes in production, such as phonetic and phonological reduction (Aylett & Turk, 2004; Bell et al., 2009; Lindblom, 1990b; Pellegrino, Coupe, & Marsico, 2011; Scarborough, 2012) can be affected by communicative goals but also early processes, such as grammatical encoding, contrary to recent proposals (Ferreira, 2008; MacDonald, 2013).

Acknowledgments

The authors are grateful to Matthew Goldrick, Maryellen C. MacDonald, Janet Pierrehumbert, Peter de Swart, Scott Seyfarth and anonymous reviewers for their comments on earlier versions of this paper. The authors would also like to thank the following people: Andrew Watts for scripting the experiment; Xin Shan for testing participants; Makiko Fujimoto and Mai Kato for stimuli recording and data transcription; audiences at the Center of Language and Cognition at the University of Groningen, the Beckman Institute at the University of Illinois, the 2012 DGfS Workshop on *Information Density and Linguistic Variation*, the 2013 *CUNY Sentence Processing Conference* and the 2013 *Annual Meeting of the Cognitive Science Society* for feedback on presentations on this work. This work was funded by the Japan Society for the Promotion of Sciences postdoctoral research fellowship to CK and NSF CAREER IIS-1150028 and a 2012 Provost Award to TFJ. The views expressed here are those of the authors and not necessarily those of the funding agencies.

Appendix A

Stimuli sentences for Experiment 1.

1. Ekiin-ga [maigo/boshi](-o) homu-de mitsuketa-yo.
“A station stuff found a [stray child/hat] on a platform”

2. Arubaito-ga [tencho/pen](-o) jimusho-de sagashita-yo.
“A part-time worker searched for a [supervisor/pen] in an office”

3. Henshucho-ga [kisha/kiji](-o) chorei-de hometa-yo.
“A chief editor praised a [writer/article] in a morning meeting”

4. Gadoman-ga [chikan/manekin](-o) depato-de taoshita-yo.
“A security guard knocked down a [groper/mannequin] in a department store”

5. Shicho-ga [chiji/sesaku](-o) intabyu-de hihanshita-yo.
“A mayor criticized a [governor/policy] in an interview.”

6. Goto-ga [rojin/konbini](-o) kokudo-de osotta-yo.
"A burglar attacked a(n) [elderly/convenience store] on a state highway."

7. Sakka-ga [musume/jitaku](-o) rikonchotei-de tebanashita-yo.
"A writer lost a [daughter/house] in a lawsuit.

8. Onnanoko-ga [isha/chusha](-o) byoin-de kowagatta-yo.
"A girl feared a [doctor/needle] in a hospital room"

9. Bucho-ga [gakusei/kikaku](-o) kaigi-de saiyoshita-yo.
"A manager accepted a(n) [applicant/proposal] in a meeting."

10. Shashinka-ga [moderu/yuyake](-o) kaigan-de satsueishita-yo.
"A photographer shot a [model/sunset] on a beach."

11. Sensei-ga [haiyu/shobosha](-o) shotengai-de mikaketa-yo.
"A school teacher saw a(n) [actor/fire-engine] in a mall."

12. Daiku-ga [deshi/shinsha](-o) nomikai-de jimanshita-yo.
"A carpenter bragged about a(n) [apprentice/car] at a party."

13. Terorisuto-ga [mihari/bakudan](-o) kukou-de haichishita-yo.
"A terrorist placed a(n) [watch/bomb] in an airport."

14. Kodomo-ga [ojisan/hikoki](-o) koen-de yubisashita-yo.
"A child pointed to a(n) [airplane/elderly] in a park."

15. Maneja-ga [senshu/boru](-o) benchi-de kazoeta-yo.
"An assistant counted [players/balls] in a dugout."

16. Otokonoko-ga [tomodachi/kokuban](-o) kyoushitsu-de tataita-yo.
"A boy hit a [friend/blackboard] in a classroom."

17. Kyouju-ga [joshu/ronbun](-o) gakkai-de shokaishita-yo.
"A professor introduced a(n) [assistant/paper] at a conference."

18. Eigyoman-ga [aidoru/pasokon](-o) purezen-de urikonda-yo.
"A salesperson pitched a [pop idol/computer] in a presentation."

19. Tarento-ga [repota/shuppansha](-o) saiban-de uttaeta-yo.
"A celebrity accused a [reporter/publisher] in a lawsuit."

20. Kangoshi-ga [akachan/karute](-o) byoshitsu-de tewatashita-yo.
"A nurse handed a [baby/card] in a hospital room."

21. Kankokyaku-ga [gaido/basu](-o) hoteru-de matta-yo.
"A tourist awaited a [tour guide/bus] in a hotel."

22. Toushu-ga [shikaku/daruma](-o) senkyosen-de okutta-yo.
"A party leader sent a(n) [assassin candidate/daruma doll] during an election."

23. Shacho-ga [shokuin/shataku](-o) shisha-de shobunshita-yo.
"A CEO disposed of a(n) [employee/company housing] at a branch office."

24. Keikan-ga [hannin/takushi](-o) kosaten-de tsukamaeta-yo.
"A police officer caught a [criminal/cab] at an intersection."

Appendix B

Google counts of nouns used in Experiment 1. Search queries were entered in the Japanese orthography with Chinese characters whenever possible.

Google counts of nouns used in Experiment 1

Animate	Counts with any or no case-marker	Counts with -o	Noun-o ratio	Inanimate	Counts with any or no case-marker	Counts with -o	Noun-o ratio
1 maigo (stray child)	7590000	107000	1.41%	1 booshi (hat)	22,800,000	5,360,000	23.51%
2 tencho (supervisor)	74800000	2560000	3.42%	2 pen (pen)	12,100,000	2,970,000	24.55%
3 kisha (writer)	34900000	1440000	4.13%	3 kiji (article)	1,080,000,000	185,000,000	17.13%
4 chikan (groper)	29600000	1810000	6.11%	4 manekin(mannequin)	4,470,000	189,000	4.23%
5 chiji (governor)	46,000,000	1,470,000	3.20%	5 seisaku (policy)	186,000,000	7,100,000	3.82%
6 goutou (burglar)	8,430,000	708,000	8.40%	6 konbini (conveniece store)	165,000,000	1,720,000	1.04%
7 musume (daughter)	113,000,000	21,100,000	18.67%	7 jitaku (house)	128,000,000	5,320,000	4.16%
8 isha (doctor)	28,600,000	1,610,000	5.63%	8 chusha (needle)	9,330,000	2,780,000	29.80%
9 gakusei (student)	331,000,000	4,150,000	1.25%	9 kikaku (proposal)	200,000,000	9,940,000	4.97%
10 moderu (model)	228,000,000	15,200,000	6.67%	10 yuyake (sunset)	6,010,000	506,000	8.42%
11 haiyu (actor)	30,000,000	1,530,000	5.10%	11 shobosha (fire engine)	2,630,000	369,000	14.03%
12 deshi (apprentice)	17,000,000	675,000	3.97%	12 shinsha (new car)	40,000,000	2,890,000	7.23%
13 mihari (watch)	2,500,000	317,000	12.68%	13 bakudan (bomb)	16,200,000	3,120,000	19.26%
14 ojiisan (grandpa)	3,950,000	136,000	3.44%	14 hikooki (airplane)	39,600,000	3,000,000	7.58%
15 senshu (player)	91,400,000	11,100,000	12.14%	15 boru (ball)	87,700,000	11,200,000	12.77%
16 tomodachi (friend)	211,000,000	9,180,000	4.35%	16 kokuban (blackboard)	6,660,000	502,000	7.54%
17 joshu (assistant)	22,400,000	794,000	3.54%	17 ronbun (research paper)	31,800,000	13,100,000	41.19%
18 aidoru (idol)	82,200,000	3,920,000	4.77%	18 pasokon (PC)	205,000,000	18,400,000	8.98%
19 repota (reporter)	7,080,000	289,000	4.08%	19 shuppansha (publisher)	19,700,000	7,580,000	38.48%
20 akachan (baby)	75,600,000	5,490,000	7.26%	20 karute (card)	7,220,000	828,000	11.47%
21 gaido (tour guide)	284,000,000	8,020,000	2.82%	21 basu (bus)	256,000,000	8,680,000	3.39%
22 shikaku (assassin candidate)	3,150,000	241,000	7.65%	22 daruma (daruma doll)	7,520,000	139,000	1.85%
23 shokuin (employee)	54,800,000	4,750,000	8.67%	23 shataku (company housing)	12,000,000	179,000	1.49%
24 hanin (criminal)	9,180,000	3,310,000	36.06%	24 takushi (taxi)	47,700,000	3,310,000	6.94%
mean	74,840,833	4,162,792	7.31%		108,060,000	12,257,583	12.66%

Appendix C

Stimuli sentences for Experiment 2 and 3. Sentence 20 was replaced with 25 in Experiment 3.

1. [zo/ ari] chikaraippai [hunduketa/kanda]-yo.
"[elephant/ant] [stamped on/bit] with full force"

2. [keisatsu/ hannin] yonaka-ni [taihoshita/osotta]-yo.
"[police officer/criminal] [arrested/ attacked] at night"

3. [dokusha/ sakka] netto-de [kensakushita/ hihanshita]-yo.
"[reader/author] [searched for/ criticized] on the internet"

4. [osama/ heitai] totsuzen [hakenshita/ shugekishita]-yo.
"[king/soldier] [sent/ assaulted] out of the blue"

5. [sensei/ seito] rouka-de [shikatta/ yobitometa]-yo.
""[teacher/ student] [scolded/ spoke to] in the hallway"

6. [kanshiin/ kodomo] puru-de [tasuketa/ karakatta]-yo.
"[life saver/ child] [rescued/ teased] in the swimming pool"

7. [shacho/ hisho] ikinari [kaikoshita/ nagutta]-yo.
"[CEO/ secretary] suddenly [discharged/ punched]"

8. [gaka/ moderu] atorie-de [egaita/ yuwakushita]-yo.
"[painter/ model] [drew/ seduced] in the studio"

9. [tenin/ kyaku] reji-de [mataseta/ donaritsuketa]-yo.
"[clerk/ customer] [have X wait/ yelled at] at the cashier"

10. [okasan/ akachan] nando-mo [ayashita/ yonda]-yo.
"[mother/ baby] [tended/ called] repeatedly"

11. [saru/inu] ushiro-kara [hikkaita/ oikaketa]-yo.
"[monkey/ dog] [scratched/ chased] from behind"

12. [kokumin/ seijika] senkyo-de [eranda/ damashita]-yo.
"[citizen/ politician] [elected/ deceived] in the election"

13. [shujinkoo/warumono] hune-de [taijishita/machiuketa]-yo.
"[protagonist/ antagonist] [beat/ waited for] on a boat"

14. [tencho/arubaito] shizukani [chuishita/nadameta]-yo.
"[shop owner/part-time worker] quietly [gave advice to/ calmed down]"

15. [obaasan/mago] koen-de [dakkoshita/ komaraseta]-yo.
"[grandma/ grand child] [held/ troubled] in the park"

16. [joyu/otetsudaisan] daidokoro-de [kokitsukatta/ tetsudatta]-yo.
"[actress/ housekeeper] [bossed around/ helped] in the kitchen"

17. [kantoku/senshu] gurando-de [shigoita/doageshita]-yo.
"[coach/ player] [trained/ tossed into the air] in the field"

18. [itamae/minarai] toto [kubinishita/okoraseta]-yo.
"[chef/ apprentice] finally [fired/ infuriated]"

19. [herupa/otoshiyori] mainichi [homonshita/ yobidashita]-yo.
"[helper/ elderly] [visited/ called out] everyday"

20. [Jaian/Nobita] akiti-de [ijimeta/machibuseta]-yo.
"[Jaian/ Nobita] [bullied/ waylaid] in a playground"

21. [gadoman/manbikihan] chushajo de [tsukamaeta/ hurikitta]-yo.
"[security guard/ shoplifter] [caught/ fled from] in a parking garage"

22. [rakugo-ka/kankyaku] gekijo-de [warawaseta/ yajitta]-yo.
"[comedian/ audience] [entertained/ booed at] in a theater"

23. [kacho/shinjin] kaigi-de [hagemashita/ kowagatta]-yo.
"[manager/ new worker] [cheered up/ feared] in a meeting"

24. [ekiin/yopparai] homu-de [okoshita/ tataita]-yo.
"[station officer/ drunkard] [woke up/ hit] at a platform"

25. [isha/kanja] byoshitsu-de [chiryoshita/ matta]-yo.
"[doctor/ patient] [treated/ awaited] in a hospital room"

Appendix D

The norming results obtained for Experiment 2 and 3 are displayed in Tables 4 and 5.

References

Ahn, H.-D., & Cho, S. (2007). Subject-Object asymmetries of morphological case realization. *Language and Information, 11*, 53–76.

Aissen, J. (2003). Differential object marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory, 21*, 435–483.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10585516>.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*, 495–527. http://dx.doi.org/10.1080/01690960801920099. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/01690960801920099>.

Arnold, J. E., Hudson-Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of*

Memory and Language, 51, 55–70. http://dx.doi.org/10.1016/j.jml.2004.03.006.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*, 527–554. http://dx.doi.org/10.1080/01690960802299378.Mechanisms.

Bates, E. & MacWhinney, B. (1982). Functionalist approaches to grammar. In *Language acquisition: The state of the art* (pp. 173–218).

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92–111. http://dx.doi.org/10.1016/j.jml.2008.06.003.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*, 1001–1024. http://dx.doi.org/10.1121/1.1534836.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575–586.

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence. *Journal of Memory and Language, 26*, 119–137.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior, 19*, 467–484.

Bock, J. K. & Levelt, W. J. M. (1994). Language production, grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984).

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass, 3*, 19–58. http://dx.doi.org/10.1111/j.1749-818X.2008.00099.x.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123–147.

Bossong, G. (1991). Differential object marking in Romance and beyond. In D. Wanner & D. A. Kibbee (Eds.), *New analyses in Romance linguistics: Selected papers from the XVIII linguistic symposium on Romance languages* (pp. 143–170). Amsterdam, Philadelphia: Jorn Benjamins.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2*, 1–12.

Boyland, J. T. (1998). A corpus study of would + have + past-participle. In *Selected papers from the 12th international conference on historical linguistics, Manchester, August 1995* (pp. 1–17).

Branigan, H., Pickering, M., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua, 118*, 172–189. http://dx.doi.org/10.1016/j.lingua.2007.02.003. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0024384107000241>.

Bresnan, J., Dingare, S. & Manning, C. D. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In M. Butt & T. H. King (Eds.), *LFG 01 conference* (pp. 1–20).

Bresnan, J., Cueni, A., Nikitina, T. & Baayen, R. H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94).

Bresnan, J., & Hay, J. (2006). Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua, 118*, 245–259. http://dx.doi.org/10.1016/j.lingua.2007.02.007. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S002438410700 0289>.

Bromley, H. M. (1981). *A grammar of lower Grand Valley Dani*. Unpublished doctoral dissertation, Australian National University.

Buz, E., Tanenhaus, M. K. & Jaeger, T. F. (submitted for publication). Miscommunication leads to context specific hyperarticulation.

Buz, E., & Jaeger (2012). Effects of phonological confusability on speech duration. In *Poster presented at the 25th CUNY sentence processing conference*. NY: New York University.

Buz, E., Jaeger, T. F., & Tanenhaus, M. (2014). Contextual confusability leads to targeted hyperarticulation. In *Proceedings of the 34th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*, 215–221.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711–733.

Bybee, J., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure*. John Benjamins Publishing Company.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics, 37*, 575–596.

Carbary, K. M., Frohning, E. E., & Tanenhaus, M. K. (2010). Context, syntactic priming, and referential form in an interactive dialogue task: Implications for models of alignment. In *Proceedings of the 32nd annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407. http://dx.doi.org/10.1006/cogp.2001.0752. <http://www.ncbi.nlm.nih.gov/pubmed/11368528>.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*, 804–809. http://dx.doi.org/10.1016/j.cognition.2008.04.004. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2582186&tool=pmcentrez&rendertype=abstract>.

Comrie, B. (1981). *Language universals and linguistic typology: Syntax and morphology* (Vol. 59). Chicago: University of Chicago Press. http://dx.doi.org/10.2307/413584. Retrieved from <http://www.jstor.org/stable/413584?origin=crossref>.

Comrie, B. (2011). Alignment of case marking of full noun phrases. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. Retrieved from <http://wals.info/chapter/98>. Accessed 29.10.13.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review, 11*, 1084–1089. http://dx.doi.org/10.3758/BF03196741.

Dahan, D. (2010). The time course of interpretation in speech comprehension. *Current Directions in Psychological Science, 19*, 121–126. http://dx.doi.org/10.1177/0963721410364726. Retrieved from <http://cdp.sagepub.com/lookup/doi/10.1177/0963721410364726>.

de Hoop, H. & Malchukov, A. L. (2008). *Case-marking strategies* (Vol. 39). http://dx.doi.org/10.1162/ling.2008.39.4.565.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes, 23*, 583–608. http://dx.doi.org/10.1080/01690960801920735.Saying.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1355–1367. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11185769>.

DeLong, K. a., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121. http://dx.doi.org/10.1038/nn1504. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16007080>.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*, 193–210. http://dx.doi.org/10.1016/j.cognition.2008.07.008. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18930455>.

de Swart, P. (2007). *Cross-linguistic variation in object marking*. Ph.D. thesis, Radboud University Nijmegen, LOT Publications.

de Swart, P. (2011). Sense and simplicity: Bidirectionality in differential case marking. In A. Benz & J. Mattausch (Eds.), *Bidirectional optimality theory* (pp. 125–150). Amsterdam: John Benjamins.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology, 25*, 108–127. http://dx.doi.org/10.1016/j.newideapsych.2007.02.002. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0732118X07000207>.

Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language, 127*, 55–64. http://dx.doi.org/10.1016/j.bandl.2012.08.004. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23040469>.

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences, 36*, 31–32.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication.

Proceedings of the National Academy of Sciences of the United States of America, 1–6. http://dx.doi.org/10.1073/pnas.1215776109. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23071337>. <http://www.pnas.org/cgi/doi/10.1073/pnas.1215776109>.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2013). Communicative biases shape structures of newly acquired languages. In M. Knauff, N. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society (CogSci13)* (pp. 430–435). Austin, TX: Cognitive Science Society.

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language, 35*, 724–755. http://dx.doi.org/10.1006/jmla.1996.0038. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X96900388>.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164–203. http://dx.doi.org/10.1016/S0010-0285(03)00005-7. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0010028503000057>.

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation, 49*, 209–246. http://dx.doi.org/10.1016/S0079-7421(08)00006-6. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2731310&tool=pmcentrez&rendertype=abstract>.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340. http://dx.doi.org/10.1006/cogp.1999.0730. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10888342>.

Ferreira, V. S., & Firato, C. E. (2002). Proactive interference effects on sentence production. *Psychonomic Bulletin & Review, 9*, 795–800. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1950737&tool=pmcentrez&rendertype=abstract>.

Ferreira, F., & Patson, N. D. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass, 1*, 71–83.

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research, 32*, 669–692. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14653013>.

Fox, B. A., & Thompson, S. A. (2007). Relative clauses in English conversation relativizers, frequency, and the notion of construction. *Studies in Language, 31*, 293–326.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing the as thee to signal problems in speaking. *Cognition, 62*, 151–167. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9141905>.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 939–944). Austin, TX: Cognitive Science Society.

Fry, J. (2003). *Ellipsis and wa-marking in Japanese conversation*. Routledge. Retrieved from <http://books.google.com/books?id=fDuAngEACAAJ>.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language, 80*, 748–775.

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers, 36*, 432–443. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15641433>.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58–93. http://dx.doi.org/10.1006/jmla.1997.2512. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X97925122>.

Geisler, W. S. (2003). Ideal observer analysis. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 825–837). MIT Press.

Gennari, S. P., & Macdonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161–187. http://dx.doi.org/10.1016/j.jml.2007.07.004. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2735264&tool=pmcentrez&rendertype=abstract>.

Gennari, S. P., Mirković, J., & Macdonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology, 65*, 141–176. http://dx.doi.org/10.1016/j.cogpsych.2012.03.002. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22537914>.

Genzel, D. & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the association for computational linguistics (ACL)* (pp. 199–206).

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science, 24*, 1079–1088. http://dx.doi.org/10.1177/0956797612463705. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23649563>.

Givón, T. (1998). The functional approach to grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 41–66). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language, 51*, 586–603. http://dx.doi.org/10.1016/j.jml.2004.07.004. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X04000841>.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1304–1321.

Guy, G. R. (1996). Form and function in linguistic variation. In G. R. Guy, C. Feagin, D. Schiffrin, & J. Baugh (Eds.), *Towards a social science of language: Papers in honor of William Labov* (pp. 221–252). Amsterdam: John Benjamins.

Haiman, J. (1994). Ritualization and the development of language. In W. Pagliuca (Ed.), *Perspectives on grammaticalization* (pp. 1–28). Amsterdam: John Benjamins.

Haspelmath, M. (2008). Creating economical morphosyntactic patterns in language change. In J. Good (Ed.), *Language universals and language change* (pp. 185–214). Oxford: Oxford University Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford Univ Press.

Healey, P. G., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS One, 9*. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0098598>.

Hinds, J. (1982). *Ellipsis in Japanese*. Carbondale.

Houde, J. F. (1998). Sensorimotor adaptation in speech production. *Science, 279*, 1213–1216. http://dx.doi.org/10.1126/science.279.5354.1213. Retrieved from <http://www.sciencemag.org/cgi/doi/10.1126/science.279.5354.1213>.

Iwasaki, N. (2011). Incremental sentence production: Observations from elicited speech errors in japanese. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.). *Processing and producing head-final structures* (Vol. 38, pp. 131–151). Netherlands: Springer. http://dx.doi.org/10.1007/978-90-481-9213-7_7.

Jacobs, R. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6*, 345–350. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12140085>.

Jaeger, T. F. (2005). Optional that indicates production difficulty: Evidence from disfluencies. In *Proceedings of DiSS05, disfluency in spontaneous speech workshop* (pp. 1–6). Aix-en-Provence, France.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Unpublished doctoral dissertation, Stanford University.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2613284&tool=pmcentrez&rendertype=abstract>.

Jaeger, T. F. (2010a). The probabilistic base of the language production system. In *Invited lecture at the 6th international workshop on language production*. Edinburgh, UK.

Jaeger, T. F. (2010b). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*, 23–62. http://dx.doi.org/10.1016/j.cogpsych.2010.02.002. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2896231&tool=pmcentrez&rendertype=abstract>.

Jaeger, T. F. (2012). Phonological optimization and syntactic variation: The case of optional that. In *Proceedings of the 32nd annual meeting of the Berkeley Linguistic Society* (pp. 175–187). Ann Arbor, MI: Sheridan Books.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology, 4*, 1–4. http://dx.doi.org/10.3389/fpsyg.2013.00230. Retrieved from <http://www.frontiersin.org/Language_Sciences/10.3389/fpsyg.2013.00230/full>.

Jaeger, T. F. & Buz, E. (submitted for publication). Signal reduction and linguistic encoding. In E. M. Fernandez & H. S. Cairns (Eds.), *Handbook of Psycholinguistics*. Wiley-Blackwell.

Jaeger, T. F., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences, 36*, 31–32. http://dx.doi.org/10.1017/S0140525X12002592. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23790009>.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012a). Incremental phonological encoding during unscripted sentence production. *Frontiers in Psychology, 3*, 1–22. http://dx.doi.org/10.3389/fpsyg.2012.00481. Retrieved from <http://www.frontiersin.org/Language_Sciences/10.3389/fpsyg.2012.00481/abstract>.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012b). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1439–1449. http://dx.doi.org/10.1037/a0027862.

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass, 3*, 866–887. http://dx.doi.org/10.1111/j.1749-818X.2009.00147.x. Retrieved from <http://doi.wiley.com/10.1111/j.1749-818X.2009.00147.x>.

Jaeger, T. F., & Wasow, T. (2006). Processing as a source of accessibility effects on variation. In R. Cover & Y. Kim (Eds.), *Proceedings of the 31st annual meeting of the Berkeley Linguistic Society* (pp. 169–180). Ann Arbor, MI: Sheridan Books.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics, 50*, 101–113.

Johnson, K. (1998). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29–54).

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Memory and language the time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156. http://dx.doi.org/10.1016/S0749-596X(03)00023-8.

Kim, Y. (1999). The effects of case marking information on Korean sentence processing. *Language and Cognitive Processes, 14*, 687–714. http://dx.doi.org/10.1080/016909699386239. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/016909699386239>.

Kim, T. (2008). *Subject and object markings in conversational Korean*. Unpublished doctoral dissertation, State University of New York at Buffalo.

Kirov, C., & Wilson, C. (2012). The specificity of online variation in speech production. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *The 34th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*, 148–203. http://dx.doi.org/10.1037/a003869.

Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in russian. In *36th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 40*, 326–347.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119*, 23–42. http://dx.doi.org/10.1016/j.cognition.2010.12.002. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3047461&tool=pmcentrez&rendertype=abstract>.

Kuno, S. (1973). *The structure of the Japanese language*. MIT Press.

Kuperberg, G. R. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension*. Baltimore: Paul Brookes Publishing.

Kuperberg, G. R., Kreher, D., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language, 100*, 223–237. http://dx.doi.org/10.1016/j.bandl.2005.12.006. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16546247>.

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory & Language, 66*, 588–611.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38*, 1137–1146. http://dx.doi.org/10.3758/MC.38.8.1137. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3057424&tool=pmcentrez&rendertype=abstract>.

Lee, H. (2006). Parallel optimization in case systems: Evidence from case ellipsis in Korean. *Journal of East Asian Linguistics, 15*, 69–96. http://dx.doi.org/10.1007/s10831-005-3004-1. Retrieved from <http://www.springerlink.com/index/10.1007/s10831-005-3004-1>.

Lee, M.-W., & Gibbons, J. (2007). Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition, 105*, 446–456. http://dx.doi.org/10.1016/j.cognition.2006.09.013. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17097626>.

Lee, H., & Kim, N. (2012). Non-canonical word order and subject-object asymmetry in korean case ellipsis. In S. Müller (Ed.), *Proceedings of the 19th international conference on head-driven phrase structure grammar* (pp. 427–442). Stanford, CA: CSLI Publications. Retrieved from <http://cslipublications.stanford.edu/HPSG/2012/lee-kim.pdf>.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

Levy, R. (2011). Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. In *Proceedings of the 49th annual meeting of the association for computational linguistics* (pp. 1–11).

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 21086–21090. http://dx.doi.org/10.1073/pnas.0907664106. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2795489&tool=pmcentrez&rendertype=abstract>.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS) 19* (pp. 849–856). Cambridge, MA: MIT Press.

Lewis, R. & Nakayama, M. (2001). Syntactic and positional similarity effects in the processing of Japanese embeddings. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 85–111).

Lindblom, B. (1990a). Explaining phonetic variation: A sketch of the H& H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439).

Lindblom, B. (1990b). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication, 6*, 220–230.

Lindsley, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology, 7*, 1–19.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 1–16. http://dx.doi.org/10.3389/fpsyg.2013.00226. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3636467&tool=pmcentrez&rendertype=abstract>.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703. http://dx.doi.org/10.1037//0033-295X.101.4.676. Retrieved from <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.101.4.676>.

Mak, W. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*, 50–68. http://dx.doi.org/10.1006/jmla.2001.2837. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X01928372>.

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua, 118*, 203–221. http://dx.doi.org/10.1016/j.lingua.2007.02.005. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0024384107000265>.

Matsuda, K. (1995). *Variable zero-marking of (o) in Tokyo Japanese*. (Doctoral dissertation, University of Pennsylvania). Retrieved from <http://repository.upenn.edu/dissertations/AAI9615088>.

McCawley, J. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Minashima, H. (2001). On the deletion of accusative case markers in Japanese. *Studia Linguistica, 55*, 176–191. http://dx.doi.org/10.1111/1467-9582.00078.

Miyamoto, E. (2002). Case markers as clause boundary inducers in Japanese. *Journal of Psycholinguistic Research, 31*, 307–347. http://dx.doi.org/10.1023/A:1019540324040.

Norcliffe, E. J., & Jaeger, T. F. (2015). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition*. http://dx.doi.org/10.1017/langcog.2014.39.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395. http://dx.doi.org/10.1037/0033-295X.115.2.357.

Onoe, K. (1987). Shugo ni wa mo ga mo tsukaenai bun ni tsuite (On sentences which reject either wa or ga for the grammatical subject). *Kokugogaku, 150*, 48.

Pellegrino, F., Coupe, C., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language, 87*, 539–558.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 3526–3529. http://dx.doi.org/10.1073/pnas.1012551108. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3048148&tool=pmcentrez&rendertype=abstract>.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280–291. http://dx.doi.org/10.1016/j.cognition.2011.10.004. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22192697>.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633–651. http://dx.doi.org/10.1006/jmla.1998.2592. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X9892592X>.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*, 1–19. http://dx.doi.org/10.1017/S0140525X12001495. Retrieved from <http://www.journals.cambridge.org/abstract_S0140525X12001495>.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–158). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. *Laboratory Phonology, 7*, 101–139.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics, 34*, 516–530. http://dx.doi.org/10.1016/j.wocn.2006.06.003. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S009544700600043X>.

Pitt, M. A. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology. Human Perception and Performance, 35*, 896–910. http://dx.doi.org/10.1037/a0013160.

Prat-Sala, M. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language, 42*, 168–182. http://dx.doi.org/10.1006/jmla.1999.2668. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X99926682>.

Race, D. S., & MacDonald, M. C. (2003). The use of that in the production and comprehension of object relative clauses. In *Proceedings of the 25th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition, 61*, 127–159. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8990970>.

Roche, J., Dale, R., & Kreuz, R. J. (2010). The resolution of ambiguity during conversation: More than mere mimicry? In *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 206–211). Austin, TX: Cognitive Science Society.

Roland, D., Elman, J. L., & Ferreira, V. S. (2006). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition, 98*, 245–272. http://dx.doi.org/10.1016/j.cognition.2004.11.008. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16399264>.

Scarborough, R. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua, 122*, 164–176. http://dx.doi.org/10.1016/j.lingua.2011.06.006. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0024384111001033>.

Scarborough, R., Brenier, J., Zhao, Y., Hall-lew, L. & Dmitrieva, O. (2007). An acoustic study of real and imagined foreigner-directed speech. In *ICPhS XVI* (pp. 2165–2168).

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics, 41*, 249–263.

Shibatani, M. (1990). *The languages of Japan (Cambridge Language Surveys)*. Cambridge University Press.

Shimojo, M. (2006). Properties of particle omission revisited. *Toronto Working Papers in Linguistics, 26*, 123–140.

Shriberg, E. & Stolcke, A. (1996). Word predictability after hesitations: A corpus-based study. In *The 4th international conference on spoken language processing*. Philadelphia, PA, USA. Retrieved from <http://www.isca-speech.org/archive/icslp_1996/i96_1868.html>.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian National University.

Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America, 118*, 1677. http://dx.doi.org/10.1121/1.2000788. Retrieved from <http://link.aip.org/link/JASMAN/v118/i3/p1677/s1&Agg=doi>.

Smiljanic, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentednessa. *The Journal of the Acoustical Society of America, 130*, 4020. http://dx.doi.org/10.1121/1.3652882.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319. http://dx.doi.org/10.1016/j.cognition.2013.02.013. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23747651>.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 675–686. http://dx.doi.org/10.1037/0278-7393.30.3.675. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15099135>.

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either…or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 425–436. http://dx.doi.org/10.1037/0278-7393.32.2.425. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1479855&tool=pmcentrez&rendertype=abstract>.

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50*, 163–178.

Sumner, M., Kurumada, C., Gafter, R. J., & Casillas, M. (2013). Phonetic variation and the recognition of words with pronunciation variants categorization and pronunciation variants. In *Proceedings of the annual meeting of the 35th annual conference of the Cognitive Science Society* (pp. 3486–3491). Austin, TX: Cognitive Science Society.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50*, 355–370. http://dx.doi.org/10.1016/j.jml.2004.01.001. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X04000087>.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language, 65*, 318–330. http://dx.doi.org/10.1016/j.jml.2011.04.009. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X11000489>.

Tanenhaus, M. K. (2013). All P's or mixed vegetables? *Frontiers in Psychology, 4*, 234. http://dx.doi.org/10.3389/fpsyg.2013.00234. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3647112&tool=pmcentrez&rendertype=abstract>.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language, 79*, 464–484.

Tourville, J., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*, 1429–1443. http://dx.doi.org/10.1016/j.neuroimage.2007.09.054. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3658624&tool=pmcentrez&rendertype=abstract>.

Traxler, M. J. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*(1), 69–90. http://dx.doi.org/10.1006/jmla.2001.2836. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X01928360>.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204–224. http://dx.doi.org/10.1016/j.jml.2005.02.010. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0749596X05000306>.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528–553. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8501429>.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467. http://dx.doi.org/10.1037/0278-7393.31.3.443. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15910130>.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*, 247–263. http://dx.doi.org/10.1016/j.jml.2011.05.002. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3171743&tool=pmcentrez&rendertype=abstract>.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*, 2176–2190. http://dx.doi.org/10.1016/j.ijpsycho.2011.09.015. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22019481>.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America, 122*, 42306–42319. http://dx.doi.org/10.1121/1.2773966. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17902866>.

Walter, M. A. & Jaeger, T. F. (2005). Constraints on optional that: A strong word form OCP effect. In *Proceedings from the annual meeting of the Chicago Linguistic Society* (pp. 505–519).

Warker, J., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 387–398. http://dx.doi.org/10.1037/0278-7393.32.2.387. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16569154>.

Wasow, T. (2002). *Postverbal bahavior*. Stanford: CSLI Publications.

Wasow, T., Jaeger, T. F., & Orr, D. M. (2011). Lexical variation in relativizer frequency. In H. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in grammar* (pp. 175–195). De Gruyter.

Watanabe, M., Den, Y., Hirose, K., Miwa, S. & Minematsu, N. (2006). Factors affecting speakers' choice of fillers in Japanese presentations. In *INTERSPEECH 2006 – ICSLP, ninth international conference on spoken language processing, Pittsburgh* (pp. 1256–1259).

Yamashita, H. (1997). The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research, 26*, 163–188.

Yamashita, H. & Kondo, T. (2009). Disfluency in the production of long-before-short order: Analysis of corpus of spontaneous Japanese. In *Proceedings of the MAPLL2009* & *thought and language (IEICE)*.

Yatabe, S. (1999). Particle ellipsis and focus projection in Japanese. *Language, Information, Text, 6*, 79–104. Retrieved from <http://phiz.c.u-tokyo.ac.jp/yatabe/particle.pdf>.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.