Beyond linear order: The role of argument structure in speaking

Shota Momma

University of California, San Diego

Victor S. Ferreira

University of California, San Diego

Author Note

Shota Momma, Department of Psychology, University of California, San Diego
This research was supported in part by the National Institutes of Health grant R01-HD051030 to Victor Ferreira.

Correspondence concerning this article should be addressed to Shota Momma, Department of Psychology, University of California, San Diego, 9500 Gilman Drive. La Jolla, CA, 92093. Contact: smomma@ucsd.edu

Abstract

In speaking, the time-course of sentence planning may reflect both the linear and the hierarchical structures of sentences. The current study examines how speakers plan sentences in which two words that form hierarchical dependency relationships - arguments and verbs - appear far apart in linear distance, to investigate how linear and hierarchical aspects of sentences simultaneously shape sentence planning processes. The results of five extended picture-word interference experiments, complemented by the temporal analysis of codability effects, suggest that speakers retrieve sentence-final verbs before the articulation of their sentence-initial patient or theme arguments, but not agent arguments, and before retrieving sentence-medial nouns inside modifiers. These results suggest that the time-course of sentence planning reflects hierarchically-defined dependency relationships over and above linear structure.

Keywords: Sentence production, Advance planning, Grammatical encoding, Argument structure, extended Picture-word interference

Beyond linear order: The role of argument structure in speaking

Introduction

Sentences are both linearly and hierarchically organized. In sentence production, the time-course of sentence planning may reflect both these linear and hierarchical structures. On the one hand, speakers may formulate sentences following surface word-order, to synchronize the retrieval and articulation of words as much as possible, thereby minimizing potential working memory cost (Christiansen & Chater, 2016; De Smedt, 1996; V. S. Ferreira & Dell, 2000; Iwasaki, 2010; Levelt, 1989; Slevc, 2011; Van Nice & Dietrich, 2003). On the other hand, speakers may 'look-ahead' to words that are arbitrarily distant, to encode hierarchically-defined dependency relationships (e.g., Momma, Slevc, & Phillips, 2016, 2018). The current study examines how speakers plan sentences in which two word classes that form hierarchical dependency relationships - arguments and verbs - appear far apart in linear distance, to investigate how linear and hierarchical aspects of sentences simultaneously shape sentence planning processes.

The relationship between word-order and sentence planning

Sentences are specifically ordered sequences of words. Because, unlike hierarchical relationships, word-order is an observable property of linguistic input, it is reasonable to assume that when developing their language skills speakers learn to plan words according to the surface word-order of their language. In accordance with this learning consideration, many models of sentence production assume that the time-course of lexical retrieval mirrors the surface word-order of sentences (Christiansen & Chater, 2016; Dell, Oppenheim, & Kittredge, 2008; De Smedt, 1996; V. S. Ferreira & Dell, 2000; Iwasaki, 2010; Kempen & Hoenkamp, 1987; Levelt, 1989; Slevc, 2011; Van Nice & Dietrich, 2003). That is, speakers are often assumed to retrieve words sequentially, from 'beginning to end.'

Another motivation for sequential retrieval is based on well-known limitations of human working memory (Christiansen & Chater, 2016; De Smedt, 1996; Iwasaki, 2010;

Kempen & Hoenkamp, 1987). Human working or short-term memory is considered to be capacity-limited (Miller, 1956) or susceptible to similarity-based interference at some stages of processing (Baddeley & Dale, 1966; Gillund & Shiffrin, 1984). Holding retrieved words in memory while preparing other words for retrieval and production can be cognitively costly because the retrieved words interfere with processing of other words until the retreived words can be spoken. To the extent that this is so, it would be more cognitively economical for speakers to retrieve and produce sentences one word at a time, on a *just-in-time* basis.

In accordance with these learning and memory considerations, the sequential retrieval assumption is widely held, implicitly or explicitly, by many models of sentence production. For example, Levelt (1989), based on the model proposed by Kempen and Hoenkamp (1987), describes a model of sentence production in which the order of lemma retrieval roughly follows the linear order of words as spoken. De Smedt (1996) formalized this property using the notion of information cascading across levels of processing - the idea that information at a higher level processing is sent to the next level of processing without delay. When applied across all processing levels - from message generation to articulation - the consequence of information cascading is the tight synchronization between planning and articulation, which, in ideal situations, results in sequential lexical retrieval. Iwasaki (2010) adopted this idea and argued that even the lemmas of sentences in head-final languages could be planned sequentially. These models are generally consistent with the approach proposed by V. S. Ferreira and Dell (2000). They described a model of sentence production in which speakers produce sentences following what they called *the principle of immediate mention*. According to this view, speakers aim to say available words immediately. All these models are motivated by the same memory consideration discussed above: They seek to minimize potential memory cost by maximizing the isomorphism between the retrieval and articulation of words. Dell et al. (2008), based on Gordon and Dell (2003), adopted a model of sentence production in which words are "activated incrementally, in sequence: if word *i* precedes word *j*, then word *i* is activated first (e.g., Dell, 1986; MacKay, 1982)". Note that these sequential models are not radically incremental models (see Christianson & Ferreira, 2005). Radically incremental models do not use overall message- or syntactic-level representations of a sentence to control lexical retrieval processes. In contrast, sequential models can and often do use higher-level representations to control lexical retrieval processes (see Bock & Ferreira, 2013; Chang, Dell, & Bock, 2006; Dell, 1986; Dell et al., 2008; V. S. Ferreira & Dell, 2000; Levelt, 1989).

The relationship between grammatical dependencies and sentence planning

One of the defining characteristics of human language is that sentences are hierarchically organized (Chomsky, 1986b; Goldberg, 1995; Kaplan, Bresnan, et al., 1982; Sag & Pollard, 1987; Steedman, 2000). Hierarchical structures are essential in defining semantic and (morpho-) syntactic relationships between words in sentences. A prominent example involves semantic and syntactic dependencies between arguments and verbs. Arguments and verbs dependency relationships, in the sense that the syntactic, semantic and morphological status of arguments often depends on verbs. For example, the relational meaning, grammatical function, morphological and syntactic case, syntactic and semantic categories, and even the presence or absence of arguments depends on the lexical properties of verbs. Consistent with these linguistic considerations, some prominent production models (e.g., Bock & Levelt, 1994; F. Ferreira, 2000) assume that verb lemmas must be selected to encode arguments grammatically. This claim is subsumed under the *verb-guidance hypothesis*, which claims that verbs are critically involved in determining the syntactic status of arguments. Critically for current purposes, the verb guidance hypothesis predicts that verbs should be selected before the encoding (and therefore the articulation) of their arguments, even when they appear last in a sentence. That is, under the verb guidance hypothesis, the time-course of sentence planning reflects argument-verb dependencies over and above surface word-order.

Strong versions of the verb-guidance hypothesis are challenged in the literature

(Allum & Wheeldon, 2007; Iwasaki, 2010). Notably, Schriefers, Teruel, and Meinshausen (1998) conducted a series of experiments showing that verbs in German verb-final clauses do not need to be retrieved before speakers start to produce the subject of a sentence. Based on such evidence, it has been argued that verb lemmas are not necessary to encode the grammatical structures of sentences (Allum & Wheeldon, 2007; Iwasaki, 2010; Schriefers et al., 1998). This verb-independent encoding of arguments, sometimes referred to as the *conceptual guidance hypothesis*, serves to reduce speakers' memory cost. If verbs are not needed to encode their arguments, speakers need not hold the yet-to-be-spoken verbs in memory while retrieving intervening words until those verbs can be spoken.

However, there are theoretical and empirical reasons to think that verbs may still be critical in grammatical encoding, but selectively when encoding object arguments. In some linguistic theories, object arguments are considered to be the only true argument of verb roots (Kratzer, 1996, 2003). More generally, most linguistic theories assume contrasts between object and subject arguments with respect to their relationships with verbs. For example, object arguments and verbs in exclusion of subject arguments constitute verb phrases, but subject arguments and verbs in exclusion of object arguments do not (except in some grammatical theories like categorial grammar, Steedman, 2000). Object arguments, but not subject arguments, are selected (i.e., subcategorized) by verbs (Chomsky, 1965; Haegeman, 1991). Selectional restrictions of verbs that are irreducible to conceptual structures determine properties of object arguments (Grimshaw, 1990). It is common that object arguments and verbs, in exclusion of subject arguments, form idioms, but idioms that involve subject arguments and verbs in exclusion of object arguments are rare (Marantz, 1981). Subject arguments exist independently of verbs in English and many other languages (there is even a formal principle encoding this contrast *Extended Projection Principle*, Chomsky & Keyser, 1982). Object arguments are easier to extract from embedded clauses in long-distance dependencies due to licensing from verbs (namely, *theta marking*, Chomsky, 1986a). Generally speaking, it is generally assumed in linguistic theory that object arguments hold some special relationship to verbs that subject arguments do not. Consistent with this theoretical consideration, Momma et al. (2016) argued that Japanese speakers retrieve verbs before starting to speak object-initial sentences, but not before subject-initial sentences. Thus, both theoretical linguistic considerations and psycholinguistic evidence suggest that verbs may be critical for the grammatical encoding of their arguments, but only their object arguments.

More recent investigations have found that speakers retrieve verbs not just before the surface objects of sentences but also before underlying objects (semantic objects) of sentences. Momma, Slevc, and Phillips (2015) showed that speakers plan verbs before speaking the subjects of passive sentences (which are semantically objects and hence are internal arguments) but not before the subject of an active sentence. Building on this finding, Momma et al. (2018) contrasted the timing of verb retrieval in two types of intransitive sentences: sentences headed by *unaccusative* verbs and sentences headed by *unergative* verbs. Unaccusative verbs are a type of intransitive verb whose sole argument is a theme or patient argument (e.g., *boil*, in that in *The octopus boils*, the octopus is not doing the boiling, but being boiled). Unergative verbs are intransitive verbs whose sole argument is an agent argument (e.g., *swim*, as in *The octopus swims*, the octopus is doing the swimming). Importantly, the subject argument of an unaccusative sentence, though appearing as the surface subject, exhibits semantic and syntactic properties typical of object arguments in many languages, including in English (see Alexiadou, Anagnostopoulou, & Everaert, 2004; Levin & Hovav, 1995; Perlmutter, 1968 among many others) and therefore is considered an internal argument in some linguistic theories (e.g., see Perlmutter, 1978). Momma and colleagues found that speakers retrieve verbs before speaking the subject of unaccusative verbs, but not before the subject of unergative verbs. Thus, the generalization about the timing of verb retrieval processes seems to be that speakers retrieve verbs before starting to speaking their *internal* arguments or their semantic objects - including stereotypical objects, passive subjects, and unaccusative subjects. Henceforth, we use the term *internal argument* for expository purpose, without theoretical commitments about whether the semantic object should also be treated as syntactic complement of the verb in the underlying representation of a sentence.

Looking ahead to downstream verbs

The time-course of lexical retrieval processes can be guided by the linear structure of sentences, or by verb-argument dependencies. However, these two hypothesized factors can conflict with each other. Namely, arguments and verbs can appear (in principle) arbitrarily far apart in linear distance. For example, an indefinite number of words can intervene between the subject noun head and its unaccusative verb: *The octopus that John found in the grocery store next to the gas station... is boiling*. If speakers retrieve words sequentially, and if they need to retrieve verbs before speaking their internal arguments, it is predicted that speakers start speaking unaccusative sentences (and also passive sentences) only after they retrieved all the words in the sentence up to the verb (see the *head principle* by Martin & Freedman, 2001 for a related view). This predicts that speakers need to allocate processing time proportional to the number of words that intervene between an unaccusative verb's subject and that verb before they can start speaking an unaccusative subject. This prediction is intuitively implausible, and as far as we know has no empirical support.

There are at least two possible ways to avoid this potential need for extensive buffering of words. The first is to weaken the role of verbs during grammatical encoding. Advance verb retrieval may only occur when a verb appears close in linear distance to its internal arguments. For example, speakers may have some rough estimate of the complexity of an utterance (cf. Griffin 2001; Yamashita and Chang 2001), and retrieve verbs in advance only when the estimated complexity of the utterance is simple. When the utterance is estimated to be complex, they may start speaking without retrieving the verbs in advance, assuming that the sentence can be continued coherently and grammatically. Under this view, speakers should not need to retrieve verbs to encode their internal arguments. This view requires sentence production models to have two different ways to encode internal arguments: verb-dependently (when the sentence is simple) and verb-independently (when the sentence is complex).

The alternative possibility is for sentence formulation mechanisms to retrieve verbs before speaking internal arguments, without retrieving words that occur in between, in order to robustly encode the dependency relationship between internal arguments and verbs. Under this view, preserving the correspondence between the order of mention and the order of lexical retrieval processes is not the prioritized goal of speakers. Speakers instead prioritize establishing the linguistic dependency between arguments and verbs, and they do so by systematically violating the correspondence between order of mention and order of planning. The advantage of this approach is the simplification of the models of grammatical encoding, in that internal arguments can always be encoded in a verb-dependent fashion, so there is no need to assume two different ways to encode internal arguments.

Figures 1 and Figure 2 illustrate the two mechanistic strategies for producing sentences with unaccusative verbs, assuming phrase-by-phrase planning and articulation of sentences (note that this assumption is adopted simply for illustrative purpose). Figure 1 illustrates a strategy of the production mechanism that constructs the sentence linearly, without relying on verb guidance to encode internal arguments. Figure 2, in contrast, illustrates the non-linear strategy of a production mechanism that prioritizes encoding the dependency between internal arguments and verbs. The current study examines which mechanistic strategy is adopted by speakers. To do so, we build on the previous finding that speakers retrieve unaccusative verbs before starting to speak their internal arguments (Momma et al., 2018). We investigate the production of sentences in which internal arguments and unaccusative verbs are separated by another phrase that is not directly relevant to verb-argument dependency (e.g., *the octopus below the lemon is boiling*, see the description of the study design below). If lexical retrieval is indeed obligatorily sequential, speakers should not retrieve sentence-final unaccusative verbs without first retrieving all of the words in any intervening phrase. If, on the other hand, lexical retrieval is not obligatorily sequential, speakers may retrieve sentence-final unaccusative verbs in advance, without first retrieving all of the words in the intervening phrase.

Probing the time-course of lexical retrievals

Before we describe the details of the experimental design, we discuss how the time-course of lexical retrieval in sentence production can be studied experimentally. The timing of retrieving a particular lemma in a sentence can be studied by characterizing the temporal profiles of semantic interference effects in what has been called the extended picture-word interference (ePWI) paradigm. ePWI is an extension of the widely used picture-word interference paradigm (Lupker, 1979). In a standard PWI paradigm, speakers produce a single word as a response to a picture stimulus, as they either hear or see a distractor word. Distractor presentation timing varies, but distractors are usually presented temporally close to the picture presentation onset. The classic finding using this paradigm is the *semantic interference effect*: speakers are slower to name pictures given distractors that are conceptually similar to the pictured referent (Lupker, 1979; Roelofs, 1992; Schriefers, Meyer, & Levelt, 1990; Vigliocco, Vinson, Lewis, & Garrett, 2004). This effect is considered to reflect the selection of lemmas, abstract linguistic representations that carry semantic and syntactic (but not phonological) information about the word (Kempen & Huijbers, 1983; Levelt, Roelofs, & Meyer, 1999). Interference does not seem to reflect conceptual preparation processes because an analogous interference effect is absent in picture categorization tasks (Schriefers et al., 1990). ePWI is an extended version of this paradigm, in which speakers produce not just a single word but a phrase or a sentence (Hwang & Kaiser, 2014; Meyer, 1996; Momma et al., 2016, 2018; Schnur, 2011; Schriefers et al., 1998). Traditionally, ePWI has been used to test whether a particular word in a sentence is retrieved before the onset of the utterance that contains that word. However, this method can be further extended to measure the relative timing at which speakers retrieve particular words in sentences. Specifically, it is possible to measure the production duration of each word of a speakers' utterance and to measure the effect of distractor words on production duration. We use this word-by-word production

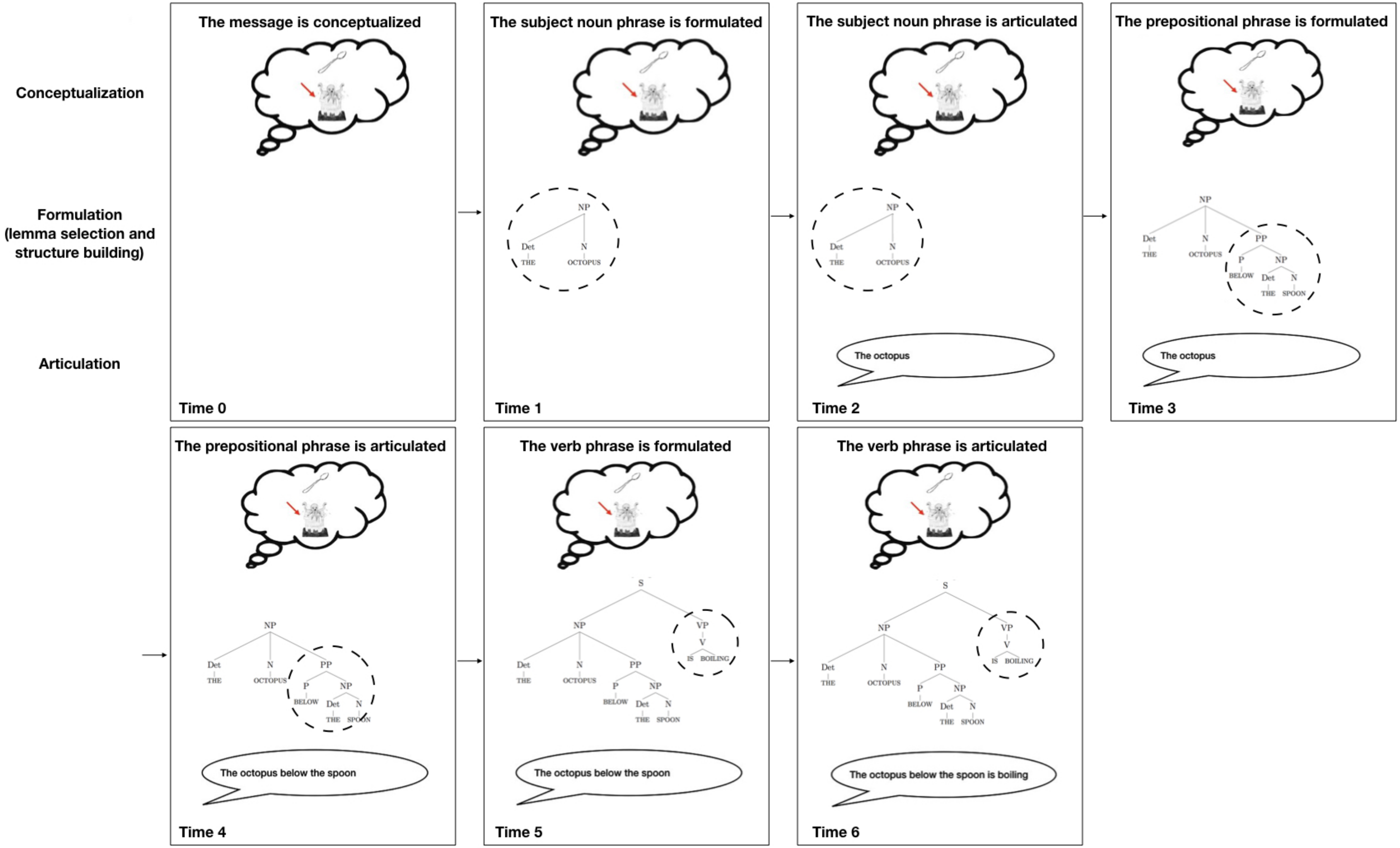

Figure 1. Illustration of how a linear sentence planning mechanism generates the unaccusative sentence, *The octopus below the spoon is boiling*. Each panel represents a time step of the sentence planning process. The top row in each panel represents conceptual representations, which are assumed to be represented as a whole before sentence formulation begins. The middle row represents the syntactic representations. The bottom row represents the articulated part of the target sentence. The dotted circle represents the locus of speakers' attention. Note, in Time 2, formulation process commit to and articulate a subject NP without any certainty that the verb it depends on represents all NP properties (e.g., case) that have presumably been encoded into the NP.

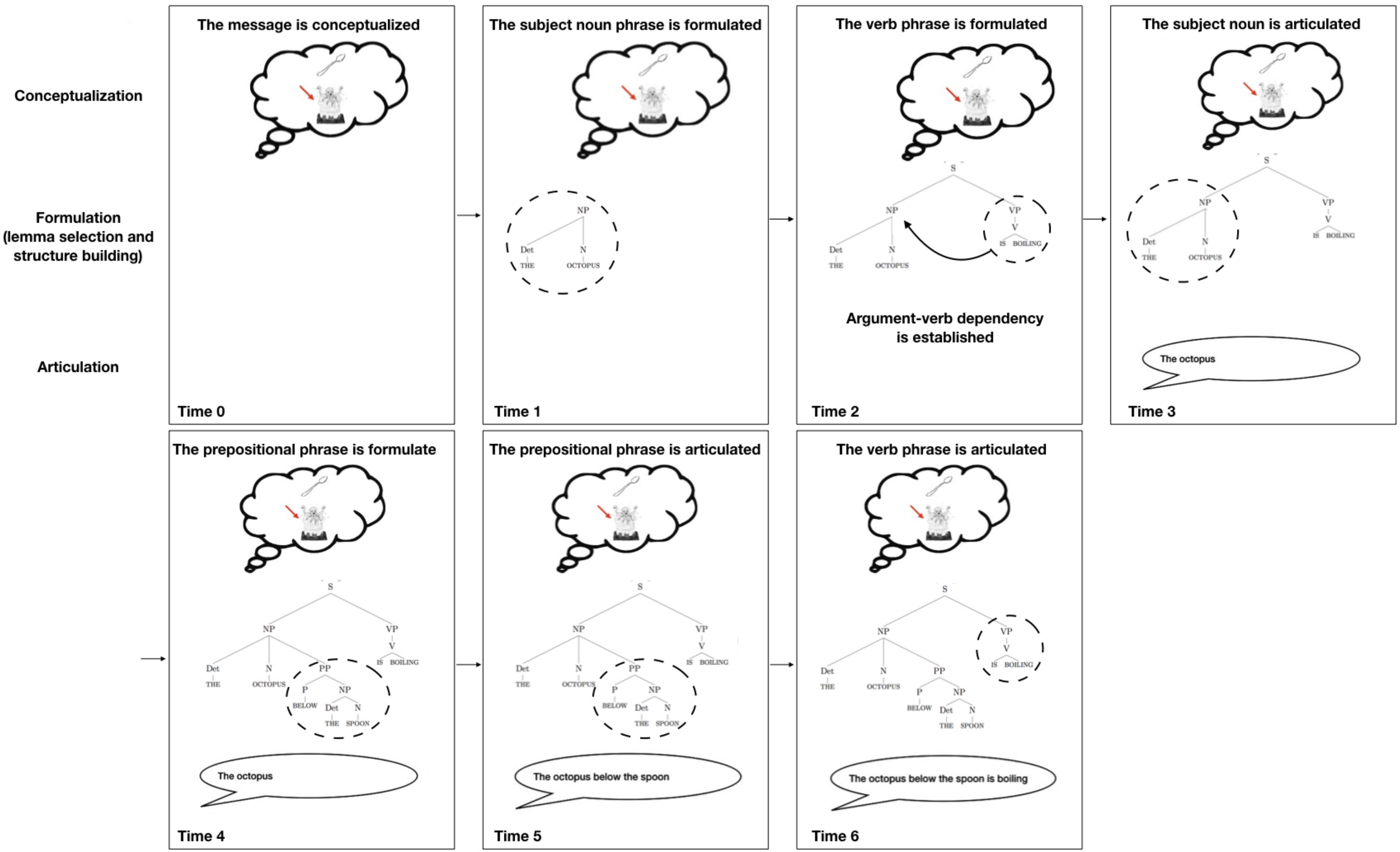

Figure 2. Illustration of how a dependency-based sentence planning mechanism works for the unaccusative sentence, *The octopus below the spoon is boiling.* Each panel represents a time step of the sentence planning process. The top row in each panel represents conceptual representations, which are assumed to be represented as a whole before sentence formulation begins. The middle row represents the syntactic representations. The bottom row represents the articulated part of the target sentence. The dotted circle represents the locus of speakers' attention.

duration measurement in the current experiments to characterize the temporal profile of semantic interference effect to assess the relative timing of sentence-medial noun planning versus sentence-final verb planning.

Current experiments

Hypotheses and predictions. The critical hypotheses are threefold. The first hypothesis is that speakers selectively retrieve sentence-final unaccusative verbs, but not sentence-final unergative verbs, before the onset of articulation of the subject nouns of those unaccusative verbs. This hypothesis is derived from the previous study by Momma et al. (2018), and also from linguistic considerations regarding the verb-argument relation discussed above.

The second and third hypotheses are that speakers retrieve (a) sentence-final unergative (but not unaccusative) verbs and (b) sentence-medial nouns inside subject-modifying adjunct phrases (henceforth adjunct nouns) on a just-in-time basis. These hypotheses are compatible with the widely accepted idea that planning and articulation can interleave within a single utterance.

These hypotheses generate three specific predictions in these ePWI experiments. The prediction based on the first hypothesis is that speakers should be slower to start speaking unaccusative sentences given a distractor word that is semantically related to unaccusative verbs. A distractor that is related to an unaccusative verb should slow the retrieval of that verb; if the subject of the unaccusative verb cannot be articulated until the unaccusative verb is retrieved, then subject retrieval should also be slowed, delaying the onset of sentence production. The prediction based on the second hypothesis is that speakers should not be slower to begin speaking unergative sentences given a distractor word that is semantically related to the unergative verb. Instead, they should elongate the production time of the words just preceding the unergative verbs. The prediction based on the third hypothesis is that speakers should not be slower to start speaking unaccusative or unergative sentences given a distractor word that is semantically related to an intervening adjunct noun. Instead, they should elongate the production time of the words immediately preceding the adjunct nouns.

These three hypotheses together form a broader hypothesis that the time-course of sentence planning reflects verb-argument dependencies over and above the surface word-order of sentences.

Overview of the experiments. We report five experiments that involved different variants of ePWI tasks. In the first four experiments (Experiment 1, 2a, 2b, and 3), we used the ePWI paradigm to investigate when speakers retrieve sentence-final verbs and sentence-medial nouns (see the method sections below for more details), building on the previous study by Momma et al. (2018). Given pictures as in Figure 3 and the instruction to describe the entity indicated by the red arrow, speakers consistently produced sentences like:

(1) The octopus below the spoon is boiling [unaccusative]

(2) The octopus below the spoon is swimming [unergative]

To measure when speakers retrieve the relevant lemmas, we presented visual distractor words that were conceptually related to either the noun head inside a subject modifying adjunct (e.g., *knife*) or to the verb (e.g., *melt*). The timing of the interference effect due to the related distractors (as compared to unrelated distractors) should correspond to when speakers retrieve the relevant lemmas. Experiments 1, 2a and 2b used essentially the same task structure, except that the timing of distractor presentation was different (150 ms before the picture onset in Experiment 1, at the same time as picture onset in Experiment 2a, 300 ms after the picture onset in Experiment 2b). Experiment 3 is a replication of the temporal pattern of verb interference effects in Experiment 1, 2a, and 2b, with two different timings of distractor presentation.

Finally, in the last experiment, participants performed a standard ePWI task using the same set of event pictures (without object pictures) and distractor words as in Experiment 1-3. But in the last experiment, speakers produced simple sentences, where the adjunct modifying the subject noun phrase was removed, like *The octopus is boiling/swimming*. This allowed Experiment 4 to serve as a test for the effectiveness of the distractor verbs used in Experiment 1-3.

Experiment 1

The goal of Experiment 1 was to test whether speakers selectively retrieve unaccusative verbs (e.g., *boil*) before the onset of sentence-initial subject nouns and whether speakers retrieve adjunct nouns (e.g., *spoon*) and unergative verbs (e.g., *swim*) later in their utterances, on a *just-in-time* basis. Based on the hypothesis that speakers retrieve unaccusative verbs before the production of their subject arguments, but retrieve unergative verbs and adjunct nouns on a just-in-time basis, three predictions were generated. First, an unaccusative verb interference effect, but not an unergative verb interference effect, should be observed before the onset of the subject noun. Second, an unergative verb interference effect should be observed in the regions just preceding the unergative verbs (Momma et al., 2018). Third, an adjunct noun interference effect should be observed in the regions neighboring the noun but not in the sentence onset.

Method

Participants. Sixty undergraduate students at the University of California, San Diego participated in the experiment for course credit. Two participants were replaced because their first language was not English, according to self-report. Another participant was also replaced because of low accuracy (< 50 % in at least one of the conditions). All remaining participants reported that they learned English as their first language. Informed consent was obtained from each participant before the experimental session.

Materials. Twenty-four event pictures with a person or animal and twenty-four object pictures were combined to yield forty-eight pictures like Figure 3. Twelve of the event pictures corresponded to unaccusative sentences (e.g., *the octopus below the spoon is boiling*). The remaining twelve event pictures corresponded to unergative sentences (e.g., *the octopus below the lemon is swimming*). The unaccusative and unergative verbs were mostly chosen based on transitivity alternation tests and were confirmed using additional tests that are known to correlate with unaccusativity (see https://osf.io/a7vxw/ for the list of verbs with the tests used to classify them). The average log frequency (per million) was well matched between unaccusative and unergative verbs (unaccusative: M = 8.90, SD = 1.03; unergative: M = 8.61, SD = 1.68; $t(22) = 0.27$, $p = .61$). Each of the twenty-four complex pictures contained two events (e.g., *boiling* and *swimming* events) sharing the same event participant (e.g., *octopus*), and two object pictures (e.g., *lemon* and *spoon*) all taken from the UCSD International Picture Naming (IPNP: Szekely et al., 2004) database. In half of the pictures, event pictures were placed on the top half of the display. In another half of the pictures, event pictures were placed on the bottom half of the display.

For each picture, the positioning of the two object entities (e.g, *spoon* and *lemon*) was switched to yield another twenty-four complex pictures. These two versions of the complex pictures were distributed across two different experimental lists, so the sets of words preceding the critical verbs (e.g., *the octopus below the spoon/lemon*) were identical (across subjects) between unaccusative and unergative conditions. For each version of the pictures, a red arrow pointed to one of the action pictures, so participants say either unergative or unaccusative sentences depending on the action that the red arrow pointed to. Furthermore, based on these two experimental picture sets, we created two versions of the lists with different random orders of trials. This yielded four different stimulus lists. Finally, based on these four different stimulus lists, we created an additional four stimulus lists by reversing the order of trials of each list. Thus, we used a total of eight different stimulus lists, and participants were distributed roughly evenly across these lists (the list assignment is not entirely even because some participants were excluded from the analysis due to high error rates). The entire set of target sentences is available online (https://osf.io/a7vxw/).

For each picture, the related distractor words were first chosen based on intuition, and then relatedness was verified using a cosine similarity measure from Latent Semantic Analysis (Landauer & Dumais, 1997). For verbs, we specifically chose distractors in such a way that (a) the average relatedness between the subject nouns and the verb distractors did not differ between unaccusative and unergative conditions,

(b) the average relatedness between the distractors and the target verbs in unaccusative and unergative conditions are approximately equal. For related verb distractors, the mean cosine similarity is .28 ($SE = 0.04$) in unaccusative conditions, and .32 ($SE = 0.05$) in unergative conditions. Thus, the cosine similarity measures were well matched between unaccusative and unergative conditions ($t(22) = 0.14$, $p > .5$). The related verb distractors were re-paired with other pictures within the same verb type to yield unrelated distractors. This procedure ensures that the set of related and unrelated distractor words are identical, so no first-order properties of distractor words (e.g., frequency, word length, imaginability, etc.) can explain any potential differences between the related and unrelated conditions. The mean cosine similarity between the target and the unrelated distractors was comparable between unergative verbs ($M = .10$, $SE = 0.01$) and unaccusative verbs ($M = .09$, $SE = 0.01$, $t(22) = 0.23$, $p > .5$). The cosine similarity difference between related and unrelated verb distractors was statistically reliable in the unaccusative conditions ($t(11) = 4.23$, $p < .001$) and in the unergative conditions ($t(11) = 4.21$, $p < .001$). For noun distractors, we also chose related distractors based on intuition and then verified the relatedness judgment using LSA. The target-distractor noun pairs were identical between unaccusative and unergative conditions. For related noun distractors, the mean cosine distance between the target noun and the distractor nouns was .37 ($SE = 0.04$). Again, these related distractors were re-paired to create unrelated target-distractor pairs. For unrelated noun distractors, the mean cosine distance between the target and distractor nouns was .08 ($SE = 0.02$). The cosine similarity difference between related and unrelated noun distractors was statistically reliable ($t(23) = 6.65$, $p < .0001$).

Design. In Experiment 1, we manipulated three independent variables. We manipulated the type of verbs used in target utterances (VerbType: unaccusative vs. unergative), the type of the distractor words (DistractorType: Noun vs. Verb), and the relatedness of distractor words to the target (Relatedness: Related vs. Unrelated). Thus, the experiment adopted a 2 (VerbType) x 2 (DistractorType) x 2 (Relatedness) within-subject design. As discussed above, speakers uttered sentences that contained

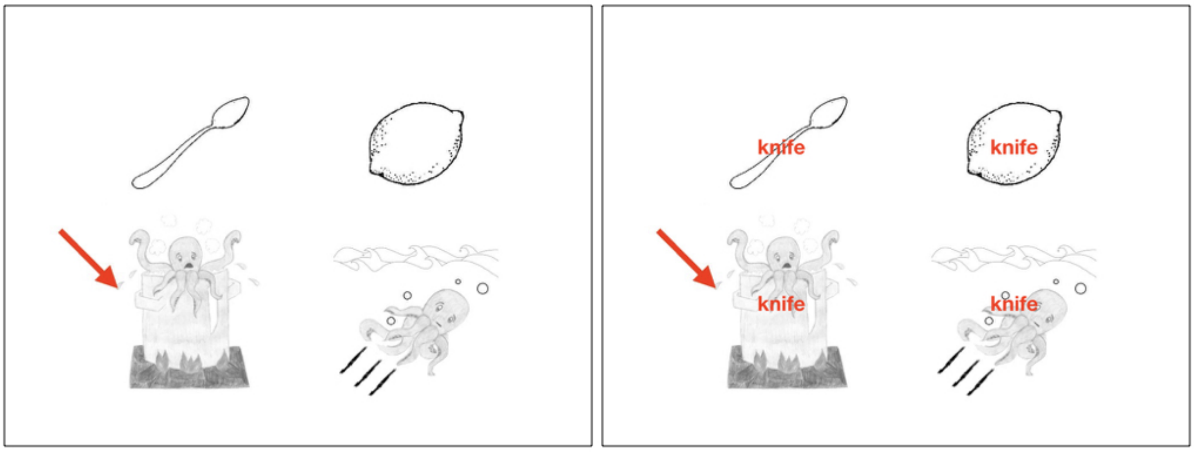

Figure 3. Two example picture stimuli used in Experiment 1-4 with (right) and without (left) distractor words superimposed. Note that the redundant presentation of distractors on each component of the picture was to prevent speakers from visually ignoring the distractors. These pictures both elicited *The octopus below the spoon is boiling*. When the red arrow points to the octopus in the bottom-right corner, it elicited *The octopus below the lemon is swimming*. In half the stimulus lists, the position of two actions is switched. Across different picture stimuli, the action pictures corresponding to unaccusative verbs appeared both in the right and left the side of the screen roughly equally often.

either an unergative verb (e.g., *the octopus below the spoon is swimming*) or an unaccusative verb (e.g., *the octopus below the spoon is boiling*) while seeing a noun distractor that was related (e.g., *knife*) or unrelated (e.g., *apple*), or verb distractor that was related (e.g., *melt/run*) or unrelated (e.g., *fall/smile*). No distractor-target pair started with the same syllable or rhymed. Following Momma et al. (2018), we used a stimulus onset asynchrony (SOA) of -150 ms. That is, the distractor words appeared 150 ms before the onset of the picture presentation.

Procedure. First, participants studied event pictures with the corresponding event and object descriptions using a picture booklet containing all experimental picture components (e.g., just one of quadrant of the full scene shown in Figure 3). Each picture was presented with the written target sentence, and participants were instructed to study the picture until they felt comfortable describing each picture using the target sentence. This familiarization process was to ensure that speakers say the sentences needed in the experimental session and is similar to a common practice in single-word picture-word interference studies (e.g., Schriefers et al., 1990), and it may even be a necessary procedure for obtaining reliable semantic interference effects (Collina, Tabossi, & De Simone, 2013). Following this familiarization session, participants practiced describing the complex pictures that they saw in the following experimental session. In this practice session, participants were instructed to first find a red arrow in the picture, and describe which of the two participants (indicated by the red arrow) is doing what action. After one practice for each picture, participants proceeded to the experimental session.

Each experimental trial started with a fixation cross for 500 ms. After the fixation cross, a full scene as in as in Figure 3 was presented with a short click sound that was later used to identify the onset of the picture in the audio recording. In this experiment, distractor appeared on the screen 150 milliseconds before the picture is presented. The entire experiment session was audio recorded, and research assistants then transcribed the audio recording.

Analysis. The transcriptions and audio files corresponding to each individual trial were aligned using an automatic text-to-speech forced alignment algorithm (P2FA, Yuan & Liberman, 2008). From the output of the text-to-speech alignment algorithm, the speech onset latency and the production time of each word (duration and any potential pause before the onset of the next word) were extracted.

Based on these data, we computed two measurements that reflect the production time of two regions of interest: the onset latency of subject noun head (i.e., the onset latency of the sentence plus the duration of the sentence-initial determiner), and the total production time of the two words that precede the critical word (i.e., the total production time of the adjunct noun head and auxiliary regions in the verb distractor conditions, and the total production time of the preposition and the second determiner regions in the noun distractor conditions). We focused on these two regions of interest because speakers use different forms of *the* (e.g., *the* vs. *thee*) in order to signal a suspension of speech (Fox Tree & Clark, 1997), so that the simple speech onset latency may not be a suitable measure of the processing cost associated with advance sentence planning. Therefore, the measurement that is likely to be more suitable for estimating the processing cost associated with advance sentence planning is the combined measure of speech onset latency and the production time of the initial determiner.

Second, we defined the second region of interest to be the total production time of the two words prior to the critical word. This decision was based on the hypothesis that speakers retrieve unergative verbs and adjunct nouns on a just-in-time basis. Previous studies suggest that it takes roughly 250-450 milliseconds from the initiation of the lemma selection process (the process of interest) to the initiation of the articulation in picture naming studies investigating single-word production processes (Indefrey & Levelt, 2004). Because it takes roughly 100-250 milliseconds on average to produce a single syllable, the average temporal locus of the semantic interference effect can be estimated to be roughly 2-3 syllables before the production onset of the word in question, if speakers retrieve words on a just-in-time basis. In the current experiments, 2-3 syllables before the critical words correspond to the adjunct noun head for verbs and preposition for nouns. However, this estimate assumes fluent speech with a uniform speech rate. In order to accommodate minor disfluencies, changes in speech rate, and some random variability in when words are retrieved, the total production time of the two words preceding the critical word were used to assess whether speakers retrieve unergative verbs and adjunct nouns on just-in-time basis.

For each region of interest, any trials containing overt hesitations (e.g., *um*) and errors (i.e., utterances deviating from the target sentences) were not included in the analysis. Additionally, production onsets of more than 5000 ms and production times of more than 1500 ms of any word (which indicates a long pause) were not included in the analysis. Outliers, defined as data points greater than three standard deviations away from each participant's mean on log-transformed data, were removed (but for the Inverse Gaussian analyses we report below, we did not apply this procedure because the Inverse Gaussian distribution accommodates slow trials).

For each region of interest, separately for each distractor type (noun vs. verb), a mixed effects model was constructed with Relatedness, VerbType, and the Relatedness x VerbType interaction as fixed effects. All categorical variables (Relatedness, VerbType, and Region) were centered (i.e., -0.5 vs. 0.5). The random effect structures in all models were maximal (Barr, Levy, Scheepers, & Tily, 2013), unless otherwise noted. In case of convergence failure, the random effects structures were simplified by successively removing the by-participant or by-item slopes that accounted for the least amount of variance, until model convergence.

In case a semantic interference effect or interaction between Relatedness and VerbType was found in either region of interest, we conducted a subsequent analysis that compares the size of the observed semantic interference effect with the magnitude of the semantic interference effect in the average of all the preceding or following regions, depending on the region of interest. This analysis was conducted to assess the time-course of semantic interference effects.

In all analyses, p-values were based on normal approximations, but the significance pattern did not change when p-values were based on maximum likelihood ratio tests.

Also, for each region of interest, we conducted a generalized mixed effects model to fit the production time using the Inverse-Gaussian distribution (also known as Wald distribution), using the identity link function (Lo & Andrews, 2015). The rationale for conducting this analysis in addition to the stereotypical Gaussian analysis is that production time (both onset latency and production duration) are highly skewed, and there is an *a priori* reason to believe that verb semantic interference effects in sentence production resides in slow trials (Momma et al., 2016). Transforming production times and fitting them using a Gaussian distribution may obscure the true underlying pattern, because log-transformation of production time is likely to disproportionately 'shrink' effects that reside in slow trials. Thus, Gaussian mixed effects analysis based on the transformed production time is likely to be inadequate for reliably detecting potential semantic interference effects. Therefore, we report the results of three types of analyses: (a) Gaussian mixed effects models on log-transformed production time, (b) Gaussian mixed effects models on (trimmed) raw production times and finally (c) generalized mixed models with the Inverse Gaussian distribution using the identity function on untrimmed raw production times.

Results

11.96% of all data points (689 of 5760 trials) came from erroneous trials (i.e., deviated from target sentences), and so were excluded from the subsequent analyses. The word-by-word mean production time and accuracy rate by each condition is summarized in Table 1. Figure 4 is a difference plot showing the semantic interference effect for each word by distractor type.

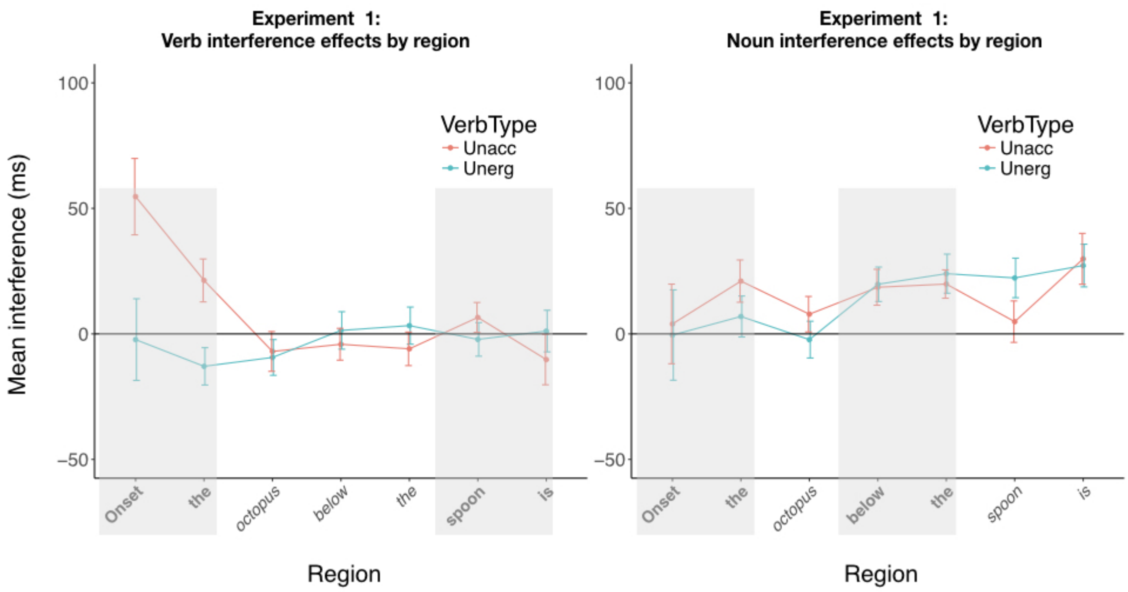

Figure 4. Interference effects (in raw production time) on region-by-region production time by VerbType in the verb distractor condition (left) and the noun distractor condition (right) in Experiment 1. Error bars represent standard error of the mean. Gray areas represent the regions of interest for each type of distractor types.

The effect of verb distractors on production time. The production times in the two regions of interest by each condition are shown in Figure 5. When speaking unaccusative sentences with verb distractors, speakers were about 75 ms slower to start speaking the subject head noun in the related distractor condition than in the unrelated distractor condition. In comparison, when speaking unergative sentences with verb distractors, speakers were 15 ms *faster* to start speaking the subject noun head in the related distractor condition than in the unrelated distractor condition. As can be seen

Table 1
The by-subject mean raw production time for each word in Experiment 1, along with the error rate by each condition.

DistractorType	VerbType	Relatedness	Onset	The	octopus	below	the	spoon	is	accuracy %
Verb	Unacc	Related	1254	179	507	402	140	546	259	88
Verb	Unacc	Unrelated	1200	158	514	406	146	539	269	88
Verb	Unerg	Related	1217	161	495	395	150	530	243	87
Verb	Unerg	Unrelated	1219	174	504	394	146	532	241	90
Noun	Unacc	Related	1244	191	514	424	162	558	287	85
Noun	Unacc	Unrelated	1240	170	506	405	142	553	257	88
Noun	Unerg	Related	1210	172	494	418	171	555	256	88
Noun	Unerg	Unrelated	1211	165	496	399	147	533	229	90

in Table 2, the analysis of the first region of interest revealed a two-way interaction between Relatedness and VerbType in all analyses. There was no main effect of Relatedness or VerbType (*ps* $> .5$).

Furthermore, this semantic interference effect on the unaccusative verbs weakened or disappeared later in utterances. Based on an analysis that compared the effect of Relatedness in the unaccusative conditions between the first region of interest and the average production time of all the subsequent regions, there was a significant interaction between Relatedness and Region in all analyses (Gaussian on log-scale: $\hat{\beta} = 0.06$, $SE = 0.02$, $t = 2.73$, $p = .01$; Gaussian: $\hat{\beta} = 82.17$, $SE = 37.15$, $t = 2.21$, $p = .004$; Inverse Gaussian: $= 49.49$, $SE = 22.38$, $z = 2.21$, $p = .027$). The main effect of Region was highly significant in all analysis (all *ps* $< .001$), showing only that some regions took longer to articulate than others. The main effect of Relatedness was not significant (*ps* $> .1$).

However, Experiment 1 provides no evidence as to when speakers plan unergative verbs. As can be seen in Table 2 and Figure 5, the main effect of Relatedness, VerbType, and the interaction between Relatedness and VerbType were not significant in the second region of interest (all *ps* $> .33$). Thus, we found no evidence that speakers were affected by related verb distractors at any point in the unergative conditions.

The effect of noun distractors on production time. The analysis of the noun distractor conditions in the first region of interest (i.e., subject noun onset latency) yield no evidence that speakers retrieved the adjunct noun head before starting to speak the subject head noun. The production times in the two regions of interest in each condition are shown in Figure 6. Speakers were not reliably faster or slower to start speaking the subject head noun with the related distractor nouns. As can be seen in Table 3, the main effect of Relatedness, VerbType and the interaction between Relatedness and VerbType were not significant (all *ps* $> .37$). In contrast, speakers were about 41 ms slower to articulate the second region of interest with related distractor nouns. As can be seen in Table 3, the main effect of Relatedness in the second region of interest was highly significant in all analysis (Gaussian on log-scale: $\hat{\beta}$

Table 2
Results of mixed effects model analyses on the two regions of interest in the verb distractor conditions in Experiment 1.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.19	0.03	250.94	$<$ .001***
Relatedness	-0.01	0.01	-0.36	.395
VerbType	-0.01	0.03	-0.27	.792
Relatedness x VerbType	0.06	0.03	1.98	.048*
Gaussian				
Intercept	1399	42	32.96	$<$ .001***
Relatedness	-24	22	-1.06	0.280
VerbType	-7	37	-0.17	0.865
Relatedness x VerbType	103	46	2.24	.025*
Inverse Gaussian				
Intercept	1607	49	32.68	$<$ .001***
Relatedness	-10	27	-0.39	.695
VerbType	-22	40	-0.55	.579
Relatedness x VerbType	125	39	3.16	.002**
Second region of interest				
Gaussian (log scale)				
Intercept	6.63	0.03	252.50	$<$ .001***
Relatedness	0.01	0.01	0.73	.464
VerbType	-0.04	0.05	-0.89	.371
Relatedness x VerbType	-0.01	0.03	-0.39	.696
Gaussian				
Intercept	795	21	38.33	$<$ .001***
Relatedness	6	13	0.50	.615
VerbType	-34	37	-0.91	.360
Relatedness x VerbType	-8	25	-0.31	.759
Inverse Gaussian				
Intercept	882	30	20.25	$<$ .001***
Relatedness	-3	15	-0.17	.864
VerbType	-40	42	-0.97	.334
Relatedness x VerbType	-5	22	-0.21	.831

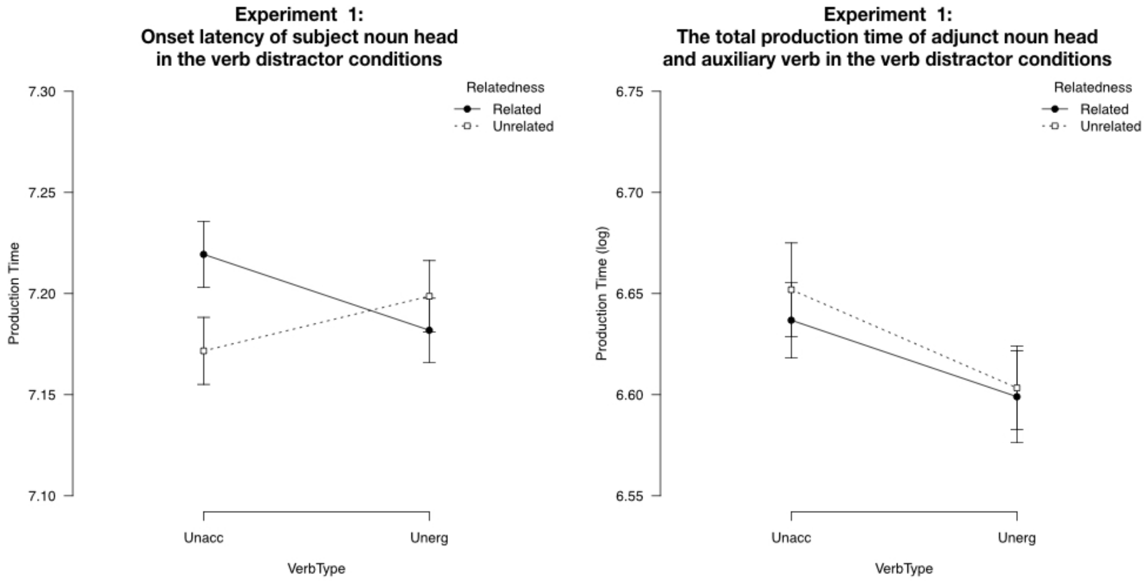

Figure 5. Production time in two regions of interest in the verb distractor conditions in Experiment 1. Error bars represent 95 percent within-subject confidence intervals.

= -0.08, SE = 0.02,t = -4.21, p < .001; Gaussian: $\hat{\beta}$ = -53, SE = 15, t = -3.65, p < .001; Inverse Gaussian: $\hat{\beta}$ = -70, SE = 20, z = -3.56, p < .001). There was no evidence that this effect differs by VerbType, as the interaction between Relatedness and VerbType was not significant in all analyses (ps > .89). The effect of VerbType was not significant either (ps > .72). Thus, under the assumption that the semantic interference effect reflects the lemma retrieval process, we found evidence that speakers *can* engage in the lemma selection of adjunct head noun on a just-in-time basis.

Moreover, the semantic interference effect on adjunct nouns in the second region of interest was stronger than in the preceding regions. Based on an analysis comparing the effect of Relatedness between the second region of interest and the average production time of all preceding regions, there was a significant interaction between Relatedness and Region in all analyses (Gaussian on log-scale: $\hat{\beta}$ = -0.05, SE = 0.02, t = -2.72, p = .006; Gaussian: $\hat{\beta}$ = -35.83, SE = 13.71, t = -2.61, p = .009; Inverse Gaussian: $\hat{\beta}$ = -53, SE = 14, z = -3.86, p < .001). (Note that we did not include the by-subject and by-item random slopes for VerbType and its interaction only in the Inverse-Gaussian analysis because of lack of convergence.) No effect of VerbType or interaction involving VerbType was significant (all ps > .42). The main effect of Relatedness was also significant in all analyses (all $ps < .001$), but we avoid interpreting this effect given the significant interaction effect between Relatedness and Region. The main effect of VerbType was not significant, nor the interaction involving VerbType was significant (all $ps > .5$). The main effect of Region was highly significant in all analyses (all $ps < .001$), again showing only that different regions had different articulation times.

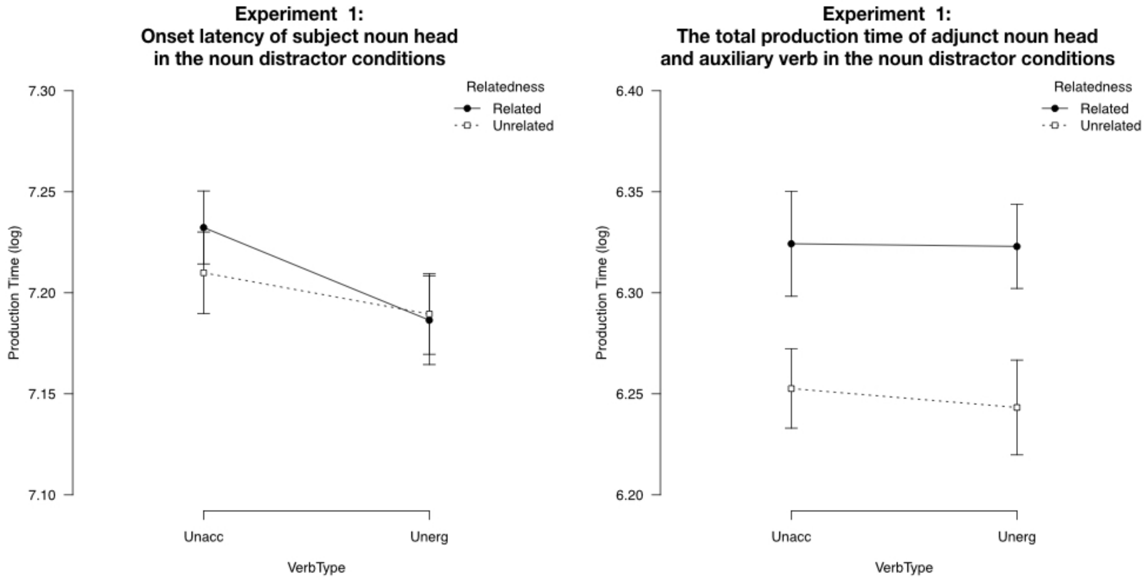

Figure 6. Production time in two regions of interest in the noun distractor conditions in Experiment 1. Error bars represent 95 percent within-subject confidence intervals.

Discussion

Experiment 1 yielded two main findings. First, the semantic interference effect on unaccusative verbs, but not on unergative verbs, was found in the onset latency of the subject head noun, the measure that reflects advance planning. This verb interference effect in the unaccusative conditions weakened or subsided in the later regions of the utterance. Second, the semantic interference effect on adjunct nouns was not found in onset latency of the subject head noun, but found later, in the total production time of the two words preceding the adjunct nouns. This noun interference effect on adjunct nouns was reliably stronger in this second region of interest than in preceding regions. Note that this finding is inconsistent with previous claims that speakers retrieve the nouns of the first phrase of a sentence (Allum & Wheeldon, 2007; Martin & Freedman,

Table 3
Result of mixed effects model analyses on the two regions of interest in the noun distractor conditions in Experiment 1.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.21	0.03	261.71	$< .001^{***}$
Relatedness	-0.01	0.01	0.36	.514
VerbType	-0.03	0.03	-1.24	.215
Relatedness x VerbType	0.03	0.03	1.05	.293
Gaussian				
Intercept	1412	40	35.00	$< .001^{***}$
Relatedness	-17	21	-0.78	.437
VerbType	-45	38	-1.19	.234
Relatedness x VerbType	35	42	0.83	.409
Inverse Gaussian				
Intercept	1632	31	53.14	$< .001^{***}$
Relatedness	-22	26	-0.84	.399
VerbType	-47	37	-1.28	.201
Relatedness x VerbType	-14	26	-0.54	.590
Second region of interest				
Gaussian (log scale)				
Intercept	6.29	0.03	217.86	$< .001^{***}$
Relatedness	-0.08	0.02	-4.21	$< .001^{***}$
VerbType	-0.01	0.04	-0.18	.859
Relatedness x VerbType	-0.01	0.03	-0.14	.890
Gaussian				
Intercept	576	18	31.94	$< .001^{***}$
Relatedness	-53	15	-3.65	$< .001^{***}$
VerbType	-3	29	-0.09	.925
Relatedness x VerbType	1	28	0.05	.961
Inverse Gaussian				
Intercept	646	20	32.69	$< .001^{***}$
Relatedness	-70	20	-3.56	$< .001^{***}$
VerbType	10	28	-0.35	.728
Relatedness x VerbType	1	23	0.05	.961

2001; Smith & Wheeldon, 1999), and is more consistent with the claim that speakers need not retrieve a second noun inside the first phrase (Griffin, 2001) until just before articulation. In all, Experiment 1 suggests that speakers retrieve sentence-final unaccusative verbs before the sentence-medial adjunct noun. This order of lexical planning is directly in line with the hypothesis that the time-course of sentence planning reflects verb-argument dependency relationships at the expense of the linear relationship between words.

However, in Experiment 1, there was no evidence of semantic interference effects on unergative verbs anywhere in the utterance. There are at least two possible reasons for this null result. First, it is possible that our choice of distractor verbs for unergative verbs was not adequate. Although we used latent semantic analysis (Landauer & Dumais, 1997) to match the semantic relatedness of distractors in the unaccusative and the unergative conditions, the cosine-distance obtained from the latent semantic analysis measure is by no means a perfect predictor of the magnitude of semantic interference. Second, it is also possible that the activation of the unergative distractor dissipated before speakers needed to engage in unergative verb retrieval processes. If the activation of distractors dissipates over time (Bloem, van den Boogaard, & La Heij, 2004), it is expected that speakers might show no semantic interference effect on words that they retrieve late in their utterances. The subsequent experiments evaluate these possibilities.

Experiments 2a and 2b

The primary goal of Experiment 2 was to examine why there was no unergative interference effect in Experiment 1. One potential reason was the long interval between the distractor presentation and the hypothesized timing of unergative verb planning. Thus, in Experiments 2a and 2b, we delayed the timing of distractor presentation relative to Experiment 1 by 150 ms (in 2a) and by 450 ms (in 2b). That is, Experiment 2a used an SOA of 0 ms, and Experiment 2b used an SOA of 300 ms. Additionally, Experiment 2a and 2b aimed to replicate the adjunct noun interference effect observed in the second region of interest in Experiment 1.

Note that the unaccusative verb interference effect was not necessarily predicted in Experiment 2a and 2b. The reason is that SOA is known to be a critical factor that modulates semantic interference effects in single-word picture naming studies (Schriefers et al., 1990). Schriefers and colleagues have shown that the semantic interference effect in a single-word picture interference task is most pronounced with an SOA of -150 ms (as in our Experiment 1). If unaccusative verbs are planned before sentence onset, as hypothesized here, it is critical that distractors are presented early enough to elicit semantic interference effects. Thus, given the delay in the distractor presentation in Experiment 2a and 2b, the unaccusative semantic interference effect may not be present.

Methods

Participants. Ninety-nine (forty-eight in Experiment 2a, fifty-one in Experiment 2b) undergraduate students at the University of California, San Diego participated in the experiment for course credits. Two participants in Experiment 2a and four participants in Experiment 2b were replaced due to low accuracy (< 50% in any condition). All participants reported that they learned English as their first language. None participated in Experiment 1. Informed consent was obtained for each participant before the experimental session.

Materials, procedure and analysis. The same materials, procedures, and analyses were used as in Experiment 1, except that the timing of distractor word presentation was delayed by 150 ms (to be a 0 ms SOA) in Experiment 2a, and 450 ms (to be a 300ms SOA) in Experiment 2b.

Results of Experiment 2a (SOA = 0 ms)

13.78% of all data points (635 out of 4608 trials) were excluded from the subsequent analysis as having come from erroneous trials. The word-by-word mean production time, along with the accuracy rate (in %), by each condition is summarized in Table 4. Figure 7 is the difference plot visualizing the semantic interference effect for each word.

Table 4
The by-subject mean production time in Experiment 2a and the accuracy (in percent) by each condition.

DistractorType	VerbType	Relatedness	Onset	The	octopus	below	the	spoon	is	accuracy %
Verb	Unacc	Related	1234	191	496	407	159	532	268	86
Verb	Unacc	Unrelated	1211	186	503	411	163	528	271	86
Verb	Unerg	Related	1234	183	480	389	155	523	250	87
Verb	Unerg	Unrelated	1206	186	478	384	147	510	255	90
Noun	Unacc	Related	1260	212	500	434	191	550	280	82
Noun	Unacc	Unrelated	1247	217	508	399	154	521	252	88
Noun	Unerg	Related	1222	191	495	423	176	535	252	87
Noun	Unerg	Unrelated	1214	194	481	406	161	518	237	90

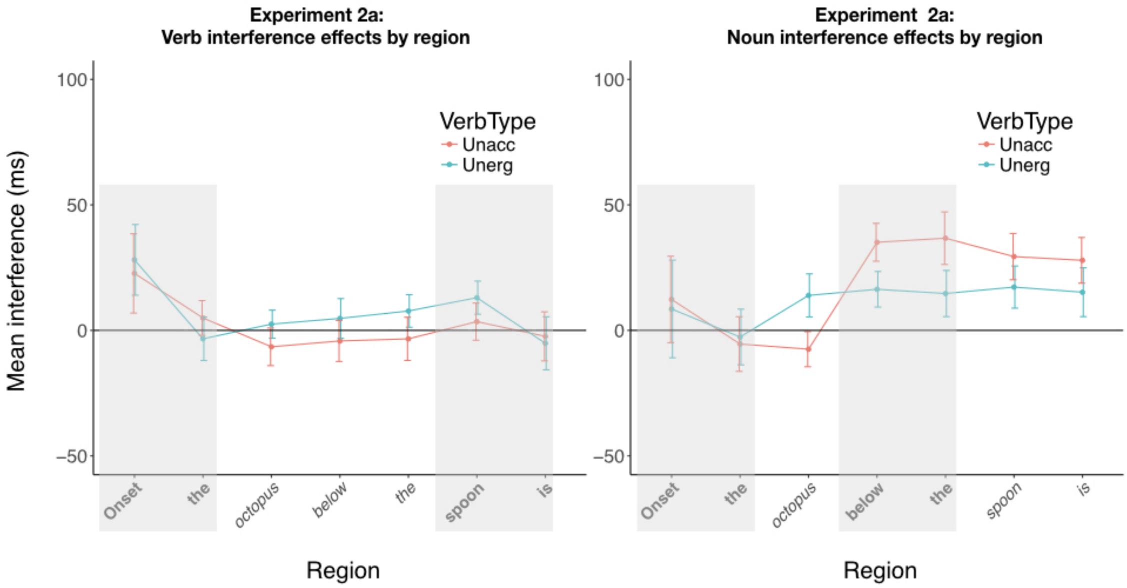

Figure 7. Interference effect (in raw production time) region-by-region by VerbType in the verb distractor condition (left) and the noun distractor condition (right) in Experiment 2a. Error bars represent standard error of the mean. Gray areas represent the regions of interest for each distractor type.

The effect of verb distractors on production time. First, we report the analysis of the verb distractor conditions. The production times in the two regions of interest by each condition are shown in Figure 8. As can be seen in Table 5, this analysis provides little evidence that speakers experienced semantic interference effects before they started to speak the subject noun. The main effect of Relatedness was significant, but only barely in the Inverse Gaussian analysis (Gaussian on log-scale: $\hat{\beta}$ = -0.01, SE = 0.01, t = -0.95, p = .34; Gaussian: $\hat{\beta}$ = -34, SE = 22, t = -1.52, p = .13; Inverse Gaussian: $\hat{\beta}$ = -45, SE = 22, z = -2.01, p = .04). There was no significant interaction between Relatedness or VerbType in any analysis ($ps > .19$). There was also no evidence that the onset latency of the subject noun differed by VerbType: the main effect of verb type was not significant (p = .58). Thus, unlike in Experiment 1, we have little evidence that speakers experienced semantic interference before the onset of the subject noun, or that the size of semantic interference effect was different between unaccusative and unergative verbs. However, note that this is not necessarily in conflict with the results of Experiment 1, because the timing of distractor presentation was different and such a difference is known to influence whether speakers show semantic interference effects in single-word naming tasks (Schriefers et al., 1990).

As can be seen in Table 5, the analysis of the second region of interest revealed no significant main effect of Relatedness, VerbType, or the interaction between Relatedness and VerbType (all $ps > .23$). Thus, Experiment 2a provides no evidence regarding when speakers plan unergative verbs.

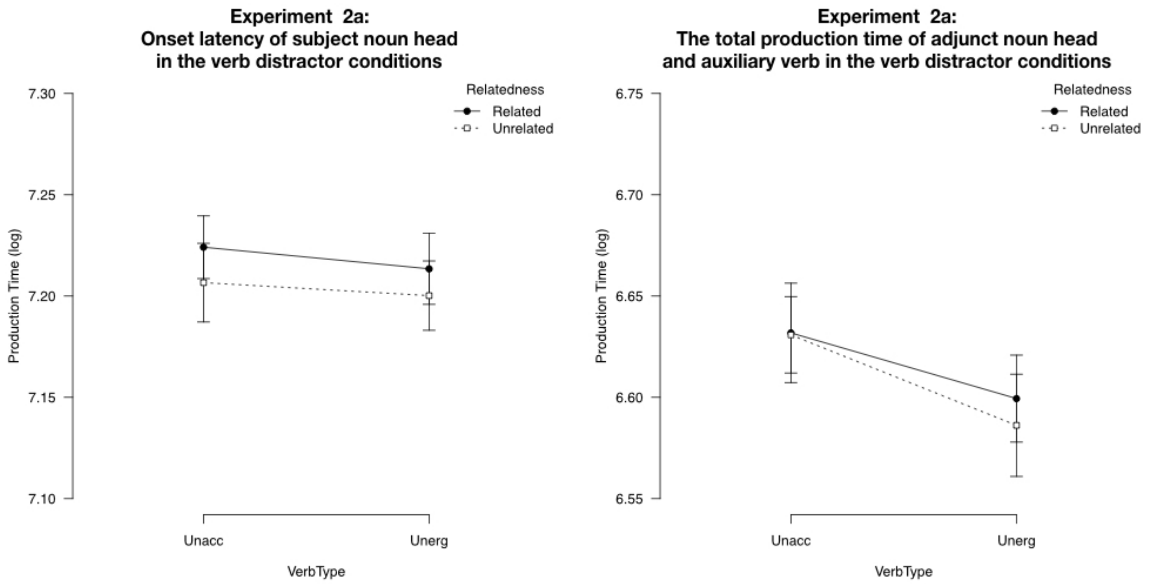

Figure 8. Production time in two regions of interest in the verb distractor conditions in Experiment 2a. Error bars represent 95 percent within-subject confidence intervals.

The effect of noun distractors on production time. Next, we report analyses in the noun distractor conditions. Production times in the two regions of interest by each condition are shown in Figure 9. Analysis on the first region of interest yields little evidence that speakers plan the adjunct noun before the articulation onset of the subject noun. As can be seen in Table 6, speakers were about equally fast in all conditions: the main effect of Relatedness and the interaction between Relatedness and VerbType was not significant (all $ps > .39$). The main effect of VerbType was marginally significant (Gaussian on log-scale: $\hat{\beta}$ = -0.04, SE = 0.02, t = -1.72, p = .08; Gaussian: $\hat{\beta}$ = -52, SE = 29, t = -1.76, p = .08), though not in the Inverse-Gaussian analysis (p = .20).

In contrast, the analysis on the second region of interest revealed that speakers exhibited a semantic interference effect on the adjunct noun as they produced the preposition and the determiner preceding that noun. As can be seen in Table 6 and

Table 5

Results of mixed effects analyses on the two regions of interest in the verb distractor conditions in Experiment 2a.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.21	0.03	228.61	< .001***
Relatedness	-0.01	0.01	-0.95	.340
VerbType	-0.01	0.03	-0.41	.677
Relatedness x VerbType	0.005	0.03	0.17	.86
Gaussian				
Intercept	1422	48	29.47	< .001***
Relatedness	-34	22	-1.52	0.128
VerbType	-9	35	-0.25	0.800
Relatedness x VerbType	8	43	0.18	.856
Inverse Gaussian				
Intercept	1621	33	48.47	< .001***
Relatedness	-45	22	-2.01	.04*
VerbType	-7	29	-0.25	.801
Relatedness x VerbType	32	25	1.29	.199
Second region of interest				
Gaussian (log scale)				
Intercept	6.61	0.03	210.18	< .001***
Relatedness	-0.00	0.01	-0.34	.735
VerbType	-0.04	0.04	-0.78	.434
Relatedness x VerbType	-0.01	0.03	-0.41	.685
Gaussian				
Intercept	788	25	31.04	< .001***
Relatedness	-5	12	-0.39	0.697
VerbType	-30	38	-0.77	0.443
Relatedness x VerbType	-7	24	-0.29	.765
Inverse Gaussian				
Intercept	872	30	28.90	< .001***
Relatedness	4	11	0.33	.744
VerbType	-30	36	-0.81	.416
Relatedness x VerbType	-12	20	-0.57	.568

Figure 9, the main effect of Relatedness was significant in all analyses (Gaussian on log-scale: $\hat{\beta}$ = -0.08, SE = 0.02, t = -4.56, p < 0.001; Gaussian: $\hat{\beta}$ = -55, SE = 13, t = -4.18, p < .001; Inverse-Gaussian: $\hat{\beta}$ = -49, SE = 19, t = -2.59, p < .01). The main effect of VerbType was not significant in any analysis (all ps = .83), suggesting that speakers spend about the same time between two verb types in this region. There was no indication that the effect of relatedness differed by VerbType: The interaction between Relatedness and VerbType was not significant (all ps > .18)

Like in Experiment 1, given reliable noun interference effects in the second region of interest, we conducted an analysis comparing the size of the semantic interference effect between the second region of interest (the total production time of the preposition and the second determiner region) and the average of the preceding regions (the average of the subject noun onset latency and the subject noun). This analysis was carried out to characterize the time-course of semantic interference effects. Note that, as in Experiment 1, we did not include the by-subject or by-item random slopes for VerbType and its interaction with Region because of convergence failure in the inverse-Gaussian analysis. This analysis revealed that there was a significant two-way interaction between Relatedness and Region, suggesting that speakers experienced greater noun interference effects in the second region than in preceding regions.

Results of Experiment 2b (SOA = 300 ms)

11.38% of all data points (568 out of 4992 trials) were identified as erroneous trials, and so were excluded from subsequent analyses. The word-by-word mean production time, along with the accuracy rate (in %) by each condition is summarized in Table 7. Figure 10 is a difference plot showing the semantic interference effect on each word's production time.

The effect of verb distractors on production time. First, we report the analysis of the verb distractor conditions. As can be seen in Table 8 and Figure 11, this analysis provides no evidence that speakers experienced semantic interference effects before they started to speak the subject noun. The main effect of Relatedness,

Table 6
Results of the mixed effects model analyses on the two regions of interest in the noun distractor conditions in Experiment 2a.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.23	0.03	238.01	$< .001^{***}$
Relatedness	-0.01	0.01	-0.42	.673
VerbType	-0.04	0.02	-1.72	.084
Relatedness x VerbType	-0.00	0.03	-0.04	.969
Gaussian				
Intercept	1454	46	31.76	$< .001^{***}$
Relatedness	-11	20	-0.56	0.577
VerbType	-52	29	-1.76	0.078
Relatedness x VerbType	-8	41	-0.20	.838
Inverse Gaussian				
Intercept	1632	31	53.14	$< .001^{***}$
Relatedness	-22	26	-0.84	.399
VerbType	-47	37	-1.28	.201
Relatedness x VerbType	-14	26	-0.54	.590
Second region of interest				
Gaussian (log scale)				
Intercept	6.30	0.03	185.11	$< .001^{***}$
Relatedness	-0.08	0.02	-4.56	$< .001^{***}$
VerbType	-0.01	0.04	-0.18	.857
Relatedness x VerbType	0.04	0.03	1.15	.250
Gaussian				
Intercept	593	21	27.92	$< .001^{***}$
Relatedness	-55	13	-4.18	$< .001^{***}$
VerbType	-4	26	-0.16	0.876
Relatedness x VerbType	33	26	1.27	.204
Inverse Gaussian				
Intercept	692	27	25.38	$< .001^{***}$
Relatedness	-49	19	-2.59	$.009^{**}$
VerbType	-7	26	-0.26	.793
Relatedness x VerbType	39	29	1.34	.182

Table 7

The by-subject mean production time in Experiment 2b, along with the accuracy (in percentage), by each condition

DistractorType	VerbType	Relatedness	Onset	The	octopus	below	the	spoon	is	accuracy %
Verb	Unacc	Related	1145	168	513	388	144	542	266	88
Verb	Unacc	Unrelated	1153	179	511	396	156	534	250	90
Verb	Unerg	Related	1126	166	508	380	146	529	233	89
Verb	Unerg	Unrelated	1125	172	492	376	136	520	230	90
Noun	Unacc	Related	1163	186	520	428	169	551	265	86
Noun	Unacc	Unrelated	1171	181	522	384	154	521	251	88
Noun	Unerg	Related	1155	175	503	415	163	549	237	87
Noun	Unerg	Unrelated	1116	181	498	391	151	527	226	92

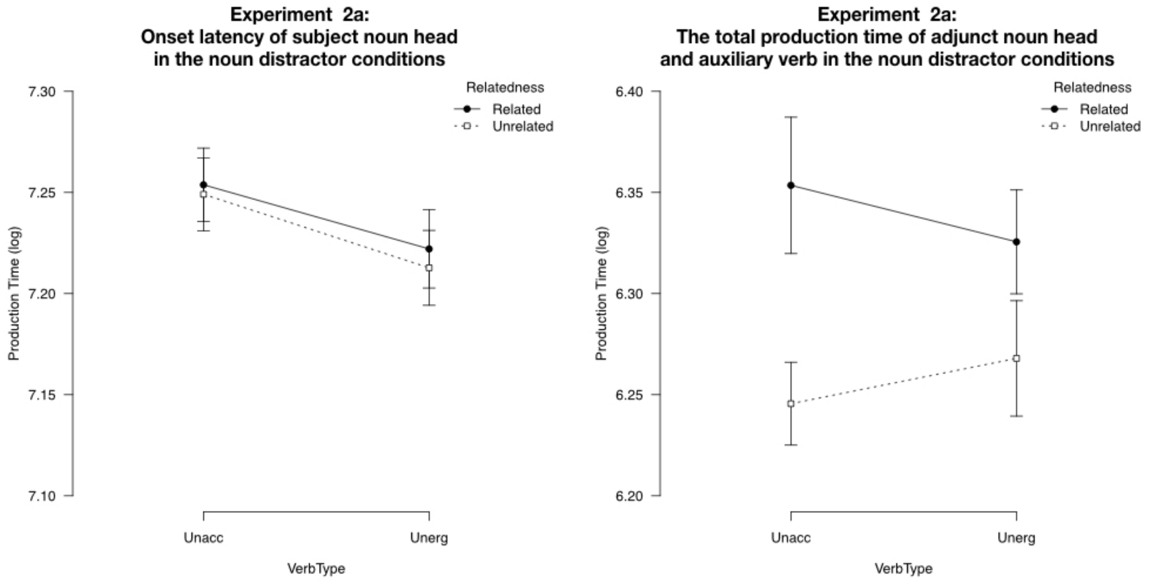

Figure 9. Production time in two regions of interest in the noun distractor conditions in Experiment 2a. Error bars represent 95 percent within-subject confidence intervals.

VerbType, or the interaction between Relatedness and VerbType were not significant (all ps > .48).

As can be seen in Table 8, the analysis of the second region of interest revealed no significant main effect of Relatedness, VerbType, or the interaction between Relatedness and VerbType (all ps > .17). Thus, it remains unclear when speakers plan unergative verbs.

The effect of noun distractors on production time. Next, we report the analysis on the noun distractor conditions. The production times in the two regions of interest by each condition are shown in Figure 12. The analysis on the first region of interest provides weak evidence that speakers retrieve the adjunct noun, before the articulation onset of the subject noun. The main effect of Relatedness was significant in the Inverse-Gaussian analysis ($\hat{\beta}$ = -50, SE = 20, t = -2.50, p = .01), though not in the other analysis (ps > .34). However, we consider this effect to be unreliable, because this effect was found to be significant only in the Inverse-Gaussian analysis and was not observed in Experiment 1 or 2a. As can be seen in Table 9, the main effect of Relatedness, and the interaction between Relatedness and VerbType was not significant (all ps > .39). The main effect of VerbType was marginally significant (Gaussian on

Table 8
Results of mixed effects model analyses on the two regions of interest in the verb distractor conditions in Experiment 2b.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.14	0.03	230.37	$< .001^{***}$
Relatedness	0.01	0.01	0.49	.625
VerbType	-0.02	0.03	-0.81	.413
Relatedness x VerbType	-0.02	0.03	-0.62	.534
Gaussian				
Intercept	1328	43	31.21	$< .001^{***}$
Relatedness	12	19	0.65	0.516
VerbType	-24	37	-0.66	0.506
Relatedness x VerbType	-26	39	-0.67	.505
Inverse Gaussian				
Intercept	1513	27	55.51	$< .001^{***}$
Relatedness	38	20	1.91	.056
VerbType	-30	43	-0.71	.480
Relatedness x VerbType	-24	34	-0.70	.485
Second region of interest				
Gaussian (log scale)				
Intercept	6.61	0.03	254.22	$< .001^{***}$
Relatedness	-0.01	0.01	-1.23	.219
VerbType	-0.04	0.04	-0.98	.326
Relatedness x VerbType	-0.01	0.02	-0.41	.683
Gaussian				
Intercept	781	20	39.08	$< .001^{***}$
Relatedness	-9	11	-0.87	0.383
VerbType	-35	30	-1.17	0.240
Relatedness x VerbType	-3	22	-0.13	.895
Inverse Gaussian				
Intercept	849	35	24.54	$< .001^{***}$
Relatedness	-24	17	-1.37	.172
VerbType	-33	36	-0.926	.354
Relatedness x VerbType	-3	23	-0.11	.913

Table 9

Result of mixed effects model analyses on the two regions of interest in the noun distractor conditions in Experiment 2b.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.15	0.03	249.80	$< .001^{***}$
Relatedness	-0.01	0.01	-0.65	.515
VerbType	-0.03	0.02	-1.19	.234
Relatedness x VerbType	-0.02	0.03	-0.66	.507
Gaussian				
Intercept	1344	39	34.77	$< .001^{***}$
Relatedness	-20	21	-0.96	0.337
VerbType	-30	26	-1.12	0.262
Relatedness x VerbType	-36	42	-0.86	.391
Inverse Gaussian				
Intercept	1508	28	54.18	$< .001^{***}$
Relatedness	-50	20	2.50	.012*
VerbType	-36	31	-1.16	.247
Relatedness x VerbType	-62	39	-1.59	.113
Second region of interest				
Gaussian (log scale)				
Intercept	6.28	0.03	203.60	$< .001^{***}$
Relatedness	-0.08	0.02	-4.78	$<.001^{***}$
VerbType	-0.02	0.04	-0.49	.626
Relatedness x VerbType	0.03	0.04	0.72	.470
Gaussian				
Intercept	574	19	30.38	$< .001^{***}$
Relatedness	-58	13	-4.60	$<. 001^{***}$
VerbType	-10	25	-0.39	.695
Relatedness x VerbType	-16	24	0.67	.502
Inverse Gaussian				
Intercept	649	20	32.03	$< .001^{***}$
Relatedness	-90	15	-6.03	$< .001^{***}$
VerbType	-8	22	-0.34	.731
Relatedness x VerbType	24	22	1.09	.275

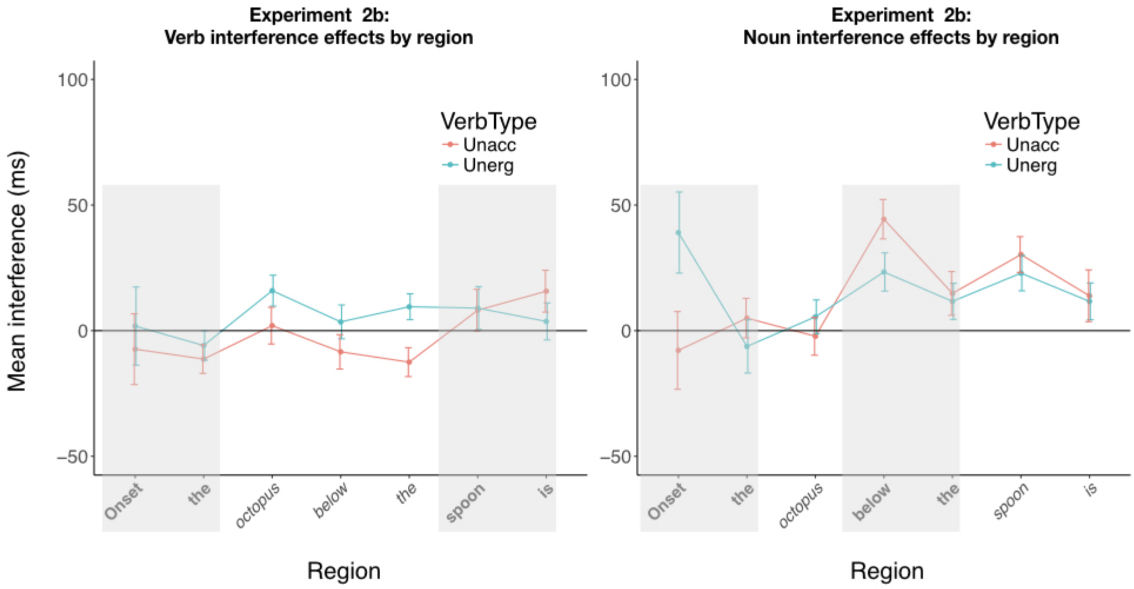

Figure 10. Interference effect (in raw production time) on region-by-region production time by VerbType in the verb distractor condition (left) and the noun distractor condition (right) in Experiment 2b. Error bars represent standard error the of mean. Gray areas represent the regions of interest for each distractor type.

log-scale: $\hat{\beta}$ = -0.04, SE = 0.02, t = -1.72, p = .08; Gaussian: $\hat{\beta}$ = -52, SE = 29, t = -1.76, p = .08), though not in the Inverse-Gaussian analysis (p = .20).

In contrast, the analysis on the second region of interest revealed that speakers experienced semantic interference effect on the adjunct noun as they produced the preposition and the determiner. As can be seen in Table 6 and Figure 9, the main effect of Relatedness was significant in all analysis (Gaussian on log-scale: $\hat{\beta}$ = -0.08, SE = 0.02, t = -4.56, p < .001; Gaussian: $\hat{\beta}$ = -55, SE = 13, t = -4.18, p < 0.001; Inverse-Gaussian: $\hat{\beta}$ = -49, SE = 19,t = 2.59, p < 0.01). The main effect of VerbType was not significant in any analysis (all ps > .83), suggesting that speakers spend about the same time between two verb types in this region. There was no indication that the effect of relatedness differed by VerbType: The interaction between Relatedness and VerbType was not significant in any analysis (all ps > .18)

Given reliable noun interference effects in the second region of interest, we conducted an analysis comparing the size of the semantic interference effect between the second region of interest (the total production time of the preposition and the second determiner region) and the average of the preceding regions (the average of the subject

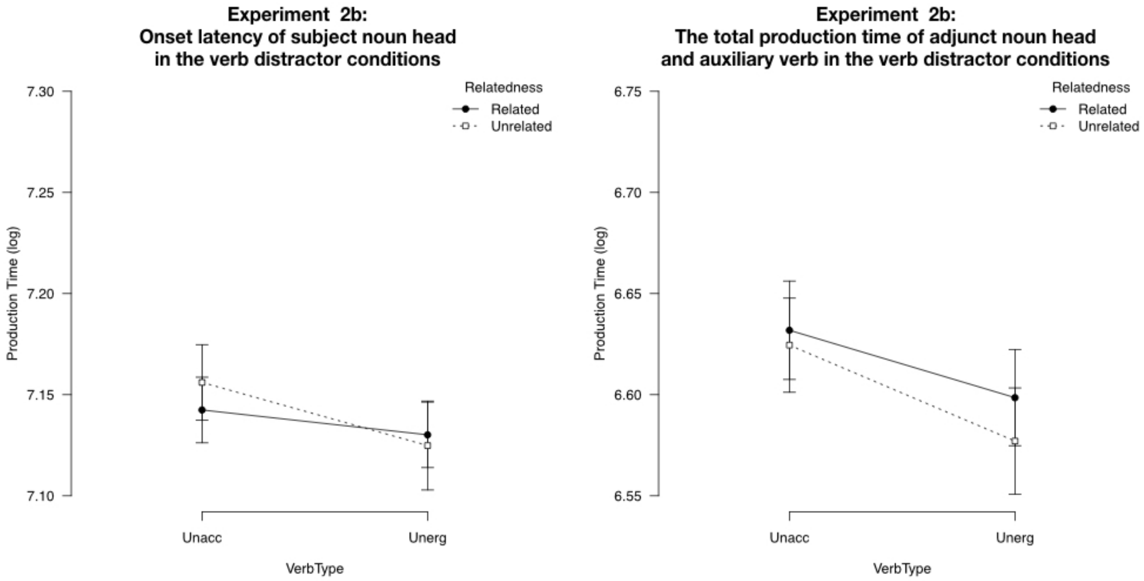

Figure 11. Production time in two regions of interest in the verb distractor conditions in Experiment 2b. Error bars represent 95 percent within-subject confidence intervals.

noun onset latency and the subject noun). This analysis was carried out to characterize the time-course of semantic interference effects. Note that, like in Experiment 1, we did not include the by-subject and by-item random slopes, only in the Inverse-Gaussian analysis, for VerbType and its interaction with Region in the random effects structure because of convergence failure. This analysis revealed that there was a significant two-way interaction between Relatedness and Region, suggesting that speakers experienced greater noun interference effects in the second region than the average of the preceding regions.

Discussion

Experiment 2a and 2b yielded two main results. First, Experiments 2a and 2b both independently replicated the temporal pattern of semantic interference effects on adjunct nouns: the adjunct noun interference effect was not found in the subject noun onset (with the exception of one significant result that is likely spurious), but was found in the total production time of the preposition and the second determiner preceding that noun. This result is a replication of Experiment 1 and suggests that speakers retrieve the adjunct nouns on a just-in-time basis.

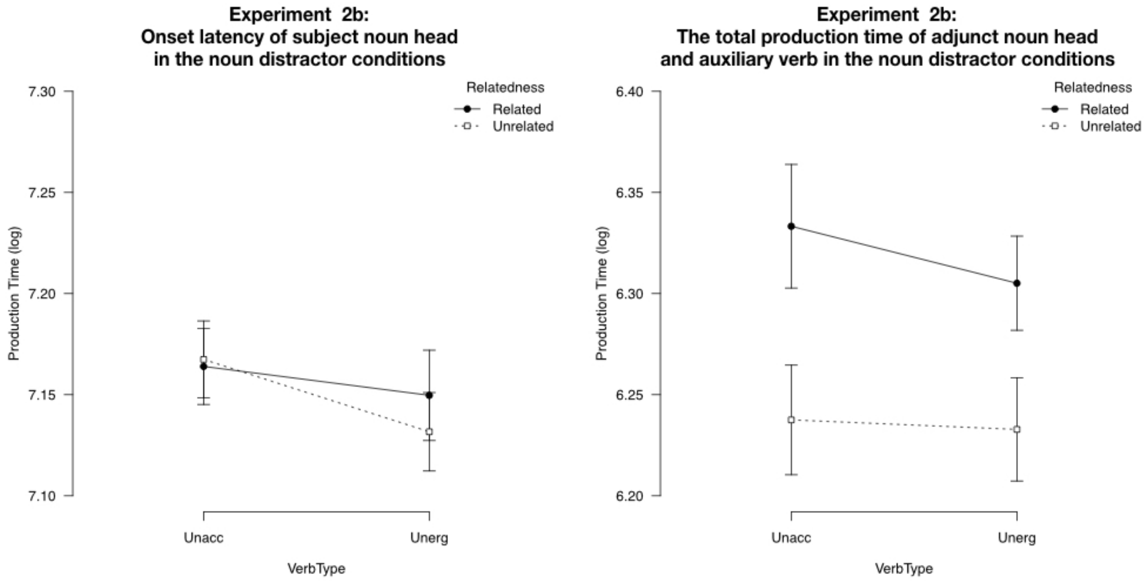

Figure 12. Production time in two regions of interest in the noun distractor conditions in Experiment 2b. Error bars represent 95 percent within-subject confidence intervals.

Second, as in Experiment 1, we again failed to find semantic inference effects on unergative verbs in any region (though there were numerical differences in the predicted direction in several production time measures, as can be seen in Table 4). Thus, there remain two possible reasons for the absence of unergative verb interference effects. It is possible that the lack of unergative verb interference was due to the temporal interval between the distractor presentation and the timing of unergative verb retrieval. Under this explanation, the distractor presented with the SOA of 0 ms or 300 ms was still not late enough to interfere with the unergative verb retrieval process that was hypothesized to happen late in an utterance. This interpretation is consistent with the hypothesis that speakers plan unergative verbs on a just-in-time basis. It is, however, also possible that the absence of the effect was because the selected distractors were not sufficiently related to the unergative verbs to-be-produced. We continue to evaluate these two possibilities in the subsequent experiments.

Experiment 3

So far, the unaccusative verb interference effect on the subject noun onset latency was only observed when the distractor words were presented 150 milliseconds before the picture, that is, only in Experiment 1. This result is not entirely surprising given previous studies that show that the semantic interference effect is sensitive to SOA manipulations (Schriefers et al., 1990). However, to be more confident that the difference was indeed due to the SOA difference, we tested whether the unaccusative semantic interference effect in Experiment 1 is replicable, but potentially only when the distractor was presented early (i.e., with an SOA of -150 ms). Thus, in Experiment 3, we used the same pictures and the same set of verb distractors (with the same paring between pictures and distractors) as in Experiments 1, 2a and 2b but manipulated SOA as a within-subject factor. The primary goal was to examine whether the critical effect (Relatedness x VerbType interaction on a subject noun onset latency) replicates, potentially only with the SOA of -150 ms.

Experiment 3 focused mainly on verb planning, so we used verb distractors only. This decision was made because the effects associated with noun planning were unambiguously present three times in Experiments 1, 2a, and 2b.

Methods

Participants. Sixty undergraduate students at the University of California, San Diego participated in the experiment for course credit. Two participants were replaced because of their poor accuracy rate (< 50 percent in at least one of the conditions). All participants reported that they learned English as their first language. None participated in Experiment 1, 2a or 2b. Informed consent was obtained for each participant before the experimental session.

Materials, Procedure and Analysis. The same materials, procedures, and analyses were used as in Experiments 1, 2a, and 2b, except that all noun distractors were replaced with verb distractors. This means that there was no DistractorType factor, and all verb distractors were used twice in the related condition and twice in the unrelated condition. Also, Experiment 3 manipulated SOA as a within-subject factor (-150 ms vs. 0 ms).

Also, we analyzed each level of the SOA factor separately, for both theoretical and practical reasons. First, our central claims do not hinge on whether relatedness effects vary reliably across different SOAs. Second, even if we failed to find the three-way interaction between Relatedness, VerbType and SOAs, it could merely reflect lack of statistical power to detect the three-way interaction. Third, fitting Inverse Gaussian distribution on the overall data involves convergence problems that are difficult to resolve.

Results

12.07% of all data points (695 out of 5760 trials) were excluded from subsequent analyses as erroneous trials. The word-by-word mean production time, along with the accuracy rate (in %) in each condition are summarized in Table 10. Figure 13 is the difference plot showing the semantic interference effect for each word.

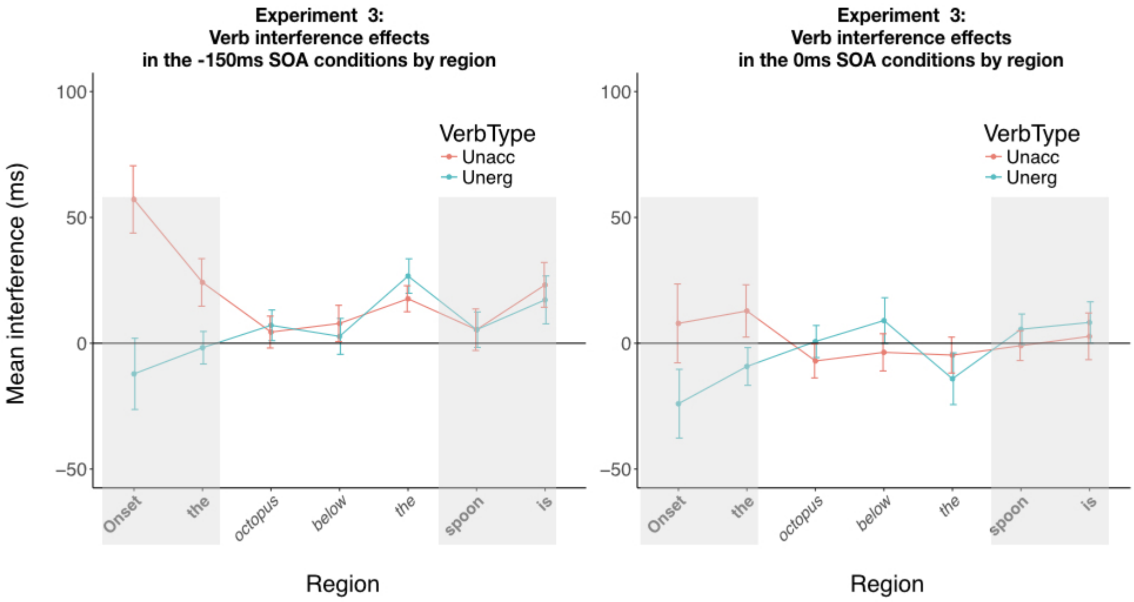

Figure 13. Interference effect (in raw production time) by each region by VerbType in the -150 ms SOA condition (left) and in the 0 ms SOA condition (right) in Experiment 3. Error bars represent standard errors. The regions of interest are shaded in gray.

In the -150 SOA conditions, speakers were around 81 milliseconds slower to start speaking the subject noun given related distractors in the unaccusative condition. In comparison, speakers were 14 milliseconds faster to start speaking the subject head noun given related distractors in the unergative condition. This pattern is shown in Figure 14. As can be seen in Table 10, there was a significant two-way interaction

Table 10

The by-subject mean production time by each condition by each region in Experiment 3.

SOA	VerbType	Relatedness	Onset	The	octopus	below	the	spoon	is	accuracy %
-150ms	Unacc	Related	1167	204	535	409	164	554	288	87
-150ms	Unacc	Unrelated	1110	180	530	401	146	549	265	88
-150ms	Unerg	Related	1101	183	506	400	172	538	275	88
-150ms	Unerg	Unrelated	1113	185	499	398	146	533	257	90
0ms	Unacc	Related	1156	207	525	406	156	545	268	86
0ms	Unacc	Unrelated	1149	194	532	410	161	546	265	89
0ms	Unerg	Related	1118	188	516	414	159	542	273	87
0ms	Unerg	Unrelated	1142	197	515	405	174	536	265	89

between Relatedness and VerbType in the -150 SOA condition in all analyses (Gaussian on log-scale: $\hat{\beta} = 0.07$, $SE = 0.03$, $t = 1.97$, $p = .049$; Gaussian: $\hat{\beta} = 111$, $SE = 49$, $t = 2.27$, $p = .023$; Inverse-Gaussian: $\hat{\beta} = 139$, $SE = 23$, $z = 5.97$, $p < .001$). As can be seen in Table 12, no other effects were significant, except the main effect of Relatedness in the Inverse-Gaussian analysis ($p = .032$).

Like in Experiment 1, we conducted the analysis that compares the size of the unaccusative interference effect between the first region of interest and the average of the following regions in the -150 SOA condition. This analysis suggests that speakers exhibited a greater interference effect in the first region of interest than the average of the following regions. There was a significant two-way interaction between Relatedness and Region in the unaccusative condition, though the effect was not significant in the Gaussian analysis on log scale (Gaussian on log-scale: $\hat{\beta} = 0.02$, $SE = 0.02$, $t = 1.00$, $p = .32$; Gaussian: $\hat{\beta} = 72$, $SE = 33$, $t = 2.17$, $p = .046$; Inverse-Gaussian: $\hat{\beta} = 52$, $SE = 19$, $z = 2.71$, $p < .01$). The main effects of Relatedness and Region were also both significant in all analysis (all $ps < .05$), but they are not of theoretical interest. This result suggests that the interference effect on unaccusative verbs observed in the first region of interest weakened or subsided in the following regions.

In the 0 ms SOA conditions, speakers were 26 milliseconds slower to start speaking the subject noun given the related distractors in the unaccusative conditions. In contrast, they were 33ms faster in the unergative conditions. This pattern is shown in Figure 15. As can be seen in Table 12, the two-way interaction between Relatedness and VerbType was significant, but only in the Inverse-Gaussian analysis (Gaussian on log-scale: $\hat{\beta} = 0.03$, $SE = 0.03$, $t = 0.88$, $p = .379$; Gaussian: $\hat{\beta} = 52$, $SE = 45$, $t = 1.16$, $p = .245$; Inverse-Gaussian: $\hat{\beta} = 89$, $SE = 25$, $z = 3.51$, $p < .001$). No other effect was significant in either SOA conditions ($ps > .218$).

The analysis comparing the size of interference effect between the first region of interest and the average of the following regions in the 0 ms SOA condition was conducted, but this analysis did not reveal the interaction between Relatedness and Region (all $ps > 0.5$). The main effects of Region were also significant in all analyses (p

$<$ .001) showing only that the regions varied in overall duration.

In the second region of interest, we again failed to find a reliable semantic interference effect on unergative verbs. As can be seen in Table 11, there was a marginally significant effect of Relatedness in the unergative condition. Thus, it remains unclear when speakers retrieve unergative verbs, and it also remains unclear whether the failure to obtain the semantic interference effect on unergative verbs is due to the long time interval between the distractor processing and unergative verb retrievals or due to insufficiently related distractors.

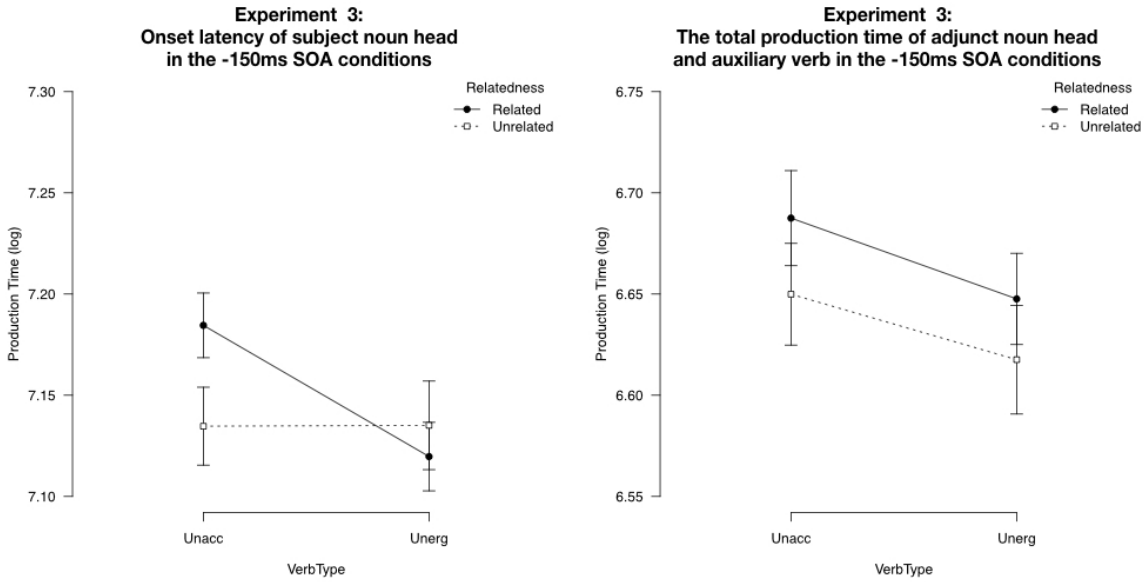

Figure 14. Production time in two regions of interest in the -150 ms SOA conditions in Experiment 3. Error bars represent 95 percent within-subject confidence intervals.

Discussion

Experiment 3 replicated the pattern of verb interference effects observed in Experiment 1. In the -150 SOA condition, speakers exhibited the verb interference effect in the subject noun onset selectively in the unaccusative condition. This result reinforces the conclusion that speakers retrieve unaccusative verbs selectively in advance of the subject head noun articulation onset. Within 0 ms SOA, the critical interaction effect (that indicates selective interference effects in the unaccusative condition) was only reliable in the Inverse-Gaussian analysis in the 0 ms SOA condition. Thus, the

Table 11

Results of mixed effects model analyses on the first region of interest in the negative SOA conditions in Experiment 3.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.15	0.02	310.88	< .001***
Relatedness	-0.02	0.02	-1.04	.300
VerbType	-0.03	0.03	-1.10	.270
Relatedness x VerbType	0.07	0.03	1.97	.049*
Gaussian				
Intercept	1328	31	42.61	< .001***
Relatedness	-26	24	-1.11	.269
VerbType	-38	36	-1.06	.289
Relatedness x VerbType	111	49	2.28	.023*
Inverse Gaussian				
Intercept	1449	25	57.32	< .001***
Relatedness	-49	23	-2.14	.032 *
VerbType	-37	28	-1.333	.183
Relatedness x VerbType	139	23	5.97	< .001***
Second region of interest				
Gaussian (log scale)				
Intercept	6.65	0.03	235.08	< .001***
Relatedness	-0.03	0.01	-2.24	.025*
VerbType	-0.04	0.04	-0.88	.380
Relatedness x VerbType	0.01	0.02	0.42	.675
Gaussian				
Intercept	821	24	34.69	< .001***
Relatedness	-21	11	-1.00	.047*
VerbType	-26	32	-0.81	.417
Relatedness x VerbType	8	20	0.37	.71
Inverse Gaussian				
Intercept	929	24	39.23	< .001***
Relatedness	-14	18	-0.82	.412
VerbType	-22	23	-0.94	.345
Relatedness x VerbType	30	25	1.20	.231

Table 12
Results of mixed effects model analyses on the first region of interest in the 0 ms SOA conditions in Experiment 3.

	Estimate	SE	t/z	p
First region of interest				
Gaussian (log scale)				
Intercept	7.16	0.02	311.22	< .001***
Relatedness	-0.00	0.02	-0.09	.921
VerbType	-0.02	0.03	-0.87	.380
Relatedness x VerbType	0.03	0.03	0.880	.379
Gaussian				
Intercept	1350	33	41.54	< .001***
Relatedness	-2	22	-0.08	.935
VerbType	-35	37	-0.94	.348
Relatedness x VerbType	52	45	1.16	.245
Inverse Gaussian				
Intercept	1461	25	59.00	< .001***
Relatedness	-25	20	1.23	.218
VerbType	-28	26	-1.057	.291
Relatedness x VerbType	89	25	3.51	< .001***
Second region of interest				
Gaussian (log scale)				
Intercept	6.65	0.03	248.18	< .001***
Relatedness	-0.01	0.01	-0.73	.467
VerbType	-0.01	0.04	-0.21	.830
Relatedness x VerbType	-0.01	0.02	0.51	.610
Gaussian				
Intercept	817	22	36.45	< .001***
Relatedness	-8	11	-0.77	.442
VerbType	-6	29	-0.21	.835
Relatedness x VerbType	-10	21	-0.48	.633
Inverse Gaussian				
Intercept	910	24	37.60	< .001***
Relatedness	-21	16	1.34	.181
VerbType	-3	25	-0.126	.900
Relatedness x VerbType	-7	22	-0.347	< .729

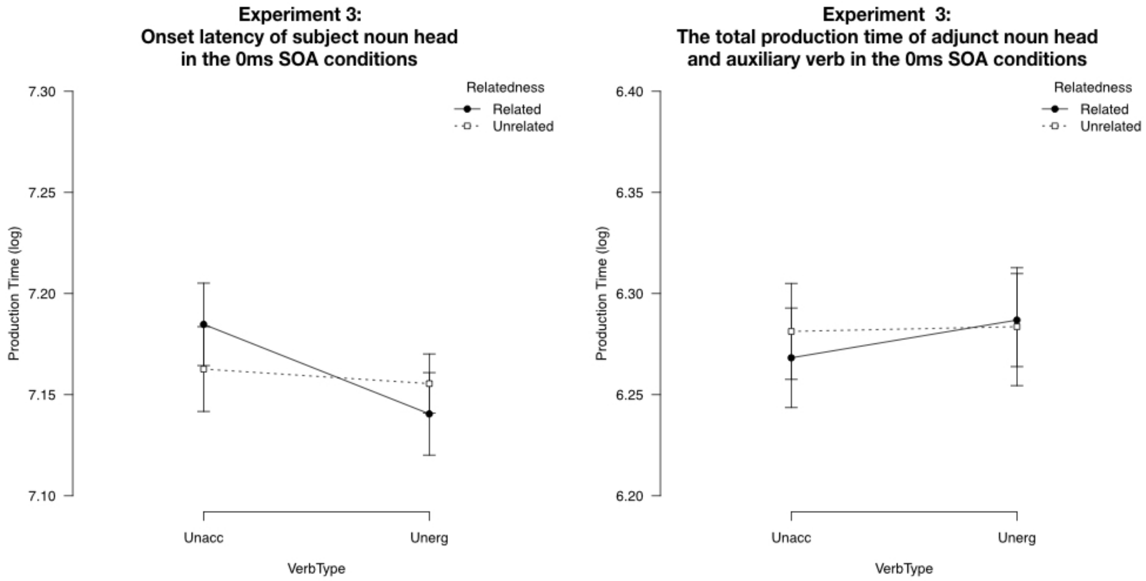

Figure 15. Production time in two regions of interest in the 0 ms SOA conditions in Experiment 3. Error bars represent 95 percent within-subject confidence intervals.

absence of the verb interference effect in the unaccusative conditions in Experiment 2a and 2b is likely because distractor words were presented too late to reliably affect the unaccusative verb retrieval process, which by the current hypothesis occurs early in the sentence planning process.

Finally, we again failed to observe any semantic interference effect on unergative verbs. Thus, Experiment 3 is not informative about when speakers plan unergative verbs. We continue to investigate this in Experiment 4, but next we use different means to assess the timing of unergative verb planning.

Event codability effects in Experiment 1-3

In Experiment 1-3, we relied on the timing of semantic interference effects to make inferences about the timing of verb retrieval. The effectiveness of this method, however, was limited by the possibility that the activation of distractor words may dissipate over time. If this is the case, semantic interference effects are not suitable for probing lemma retrieval processes that occur late in an utterance. Certainly, we reliably found semantic interference effects on adjunct head nouns, so the activation of distractor words is high enough to cause interference effects sentence-medially. But the onset of the sentence-final unergative verbs was about 1 second later than the onset of the adjunct head nouns in Experiment 1-3. Thus, it is possible that the lack of verb interference effect in the unergative conditions in Experiment 1-3 may be because the activation of verb distractors dissipated before speakers retrieve unergative verb lemmas. This problem may explain why we failed to obtain a reliable unergative interference effect. However, this explanation is not conclusive.

In the picture naming literature, there is another effect that is considered to reflect lemma retrieval: *codability.* It has been firmly established that name agreement is a strong predictor of response latency in single-word picture naming (see Alario et al., 2004; Snodgrass & Vanderwart, 1980 among many others). Codability is often operationalized using a quantity that expresses response variability for a certain picture stimulus (Snodgrass & Vanderwart, 1980), sometimes known as the *H-statistic.* Mathematically, the H-statistic can be expressed as:

$$H = \sum_{i=1}^{k} p_i \; log_2 \; (1/p_i)$$

In this equation, k is the number of different names given to each picture and p_i is the proportion of subjects giving each name. Note that pictures with a higher H statistic are less codable. For the ease of interpretation, we define codability score as the sign-flipped version of the H statistic (i.e., codability $= -H$).

Previous research has established that pictures with greater codability (i.e., pictures with less response variability) are named faster in a single-word picture-naming task (Snodgrass & Vanderwart, 1980), even with a pre-experiment familiarization phase (Alario et al., 2004), and in phrase-level or sentence-level production tasks (Griffin, 2001). If the codability effect reflects lemma retrieval rather than conceptual preparation processes as argued previously (Alario et al., 2004; Griffin, 2001; Johnson, 1992; Johnson, Paivio, & Clark, 1996), the timing of codability effects in an utterance is informative about when the lemma retrieval of that word occurs in an utterance.

Based on the hypothesis that unaccusative verbs are retrieved in advance of the production of their arguments, but adjunct nouns and unergative verbs are retrieved on a just-in-time basis, the following predictions about the effect of event codability on production time can be made. First, speakers should be faster to start speaking the subject noun (corresponding to the production time of the first region of interest used in Experiment 1-3) in response to pictures with higher event codability, but only when saying sentences with unaccusative verbs. Second, speakers should be faster to produce the adjunct noun and auxiliary verb (corresponding to the production time of the second region of interest in the verb distractor conditions in Experiment 1-3) in response to the picture with higher event codability, but only when saying sentences with unergative verbs.

Codability norms. We obtained three codability scores for each stimulus picture in Experiment 1-3, by conducting online post-hoc norming tests and by using the database provided by Szekely et al. (2004).

First, an event codability score for each picture was obtained via an online post-hoc norming test. Fifty participants from Amazon Mechanical Turk typed a verb that best describes each event picture used in Experiment 1-3. Their responses were then re-coded into citation form, and then the codability score for each event picture was computed, according to the formula presented above. The event codability score was marginally significantly different between unaccusative and unergative conditions ($t(22) = 1.79$, $p = .09$). This (marginal) difference could not easily be avoided, because drawing pictures that reliably elicit unaccusative verbs with animate subjects is difficult. However, based on a re-analysis of Experiments 1 and 3, this difference did not explain the critical pattern of selective verb interference effects in the first region of interest. Including the event codability score (henceforce EventCodability) and its interaction with Relatedness did not improve the fit of the model in which the critical interaction effects were found (all $ps > .87$). In another analysis where Relatedness was replaced with EventCodability, there was no significant interaction between EventCodability and VerbType either (all $ps > .87$). These results suggest that the (marginally significant) difference in the event codability between unaccusative and unergative conditions is unlikely to be the cause of the selective verb interference effects in Experiment 1 and 3. Thus, we consider it unlikely that the evidence for the selective

Table 13
The mean codability scores of picture components corresponding to subject nouns (participant codability), adjunct nouns (object codability), and verbs (event codability), in the unaccusative and unergative conditions.

VerbType	Participant codability	Object codability	Event codability
Unaccusative	0.86	0.37	2.05
Unergative	0.88	0.38	1.45

verb planning in the unaccusative condition was due to the difference in event codability between the pictures that elicited each type of description.

Second, we also obtained the codability score of each event participant (corresponding to subject nouns) for each event picture, by conducting online post-hoc norming tests. Fifty participants from Amazon Mechanical Turk typed a noun that best describes each event participant picture used in Experiment 1-3. As can be seen in Table 13, this score was closely matched between the unaccusative and unergative conditions (reflecting that, for example, the octopus is as likely to be named *octopus* when it is shown swimming as when it is shown boiling). This codability score was used as a control predictor in the models reported below.

Finally, we used norms obtained by Szekely et al. (2004), for the codability scores of each object picture (corresponding to the adjunct nouns). Thus, we did not conduct a norming study to obtain the codability score for the object corresponding to the adjunct nouns. As can be seen in Table 13, this score was closely matched between the unaccusative and unergative conditions (by design, this score should be identical if the data were perfectly balanced, as nouns were counterbalanced across verb type). This codability score was used as a control predictor in the models reported below. Table 13 summarizes the mean codability scores for event participant, object, and event depicted in the pictures used in the unaccusative and unergative conditions.

The timing of event codability effects. All participants from Experiments 1-3 were included in the analysis. That is, the current analysis contained the data from all 219 speakers. A mixed effects model that assessed the effect of the codability scores of the event picture (Event Codability) was built, separately for the each of the two regions of interest used in the verb distractor conditions in Experiment 1-3.

In all codability analyses, all experimental factors introduced so far (Relatedness, VerbType, DistractorType, SOA), as well as their interactions were included as control predictors (i.e., as fixed effects, with no by-subject or by-item random slopes, see Barr et al., 2013). The factor Experiment (Experiment 1 vs. 2a vs. 2b vs. 3) was initially included but removed because it caused the model to be rank-deficient (but keeping it did not change the statistical results). Additionally, factors that are known to affect the processing difficulty of each of the content words (subject noun log frequency, participant codability, object codability, adjunct noun log frequency, and verb log frequency), as well as their interaction with VerbType were also included as control predictors. Frequency was obtained from the SUBTLEX corpus (Brysbaert & New, 2009). The random effects structure of this model was maximal with respect to the effects of interest (i.e., with respect to the event codability factors and any interaction factors involving them). For the analysis of the effect of event codability, we used a model comparison approach to test the statistical significance of factors, because all the models we report below involve many different factors that are known to be to some extent collinear (e.g., frequency and codability are correlated to some degree). The model comparison approach using the maximum likelihood ratio test is relatively robust against potential multi-collinearity issues (Jaeger, 2010).

As visualized in the partial regression plots in Figure 16, speakers were *slower* to start speaking the subject noun head in response to the pictures with *more* codable events in sentences with unaccusative verbs, but not in sentences with unergative verbs. There was a significant interaction between EventCodability and VerbType ($\chi^2(1) = 5.85$, $p = .016$) in the first region of interest. The planned subset analyses revealed that speakers were slower to start speaking in response to the pictures with more codable events in the unaccusative conditions ($\chi^2(1) = 7.01$, $p = .008$), but not in the unergative conditions ($\chi^2(1) = 0.25$, $p = .610$). That is, we observed a reverse codability effect selectively in the unaccusative sentences. This result still suggests that speakers plan some aspects of events before starting to speak the subject noun selectively in the unaccusative conditions. However, the effect was in the opposite direction than predicted. We return to this issue below.

In the second region of interest, as event codability increased, speakers were *faster* to speak in the unergative conditions, but not in the unaccusative condition. There was a significant interaction between EventCodability and VerbType in this region ($\chi^2(1) = 8.36$, $p = .004$). The planned subset analysis also revealed that speakers were faster to speak in response to the more codable pictures in the unergative conditions ($\chi^2(1) = 15.84$, $p < .001$), but not in the unaccusative conditions ($\chi^2(1) = 0.29$, $p = .588$). This pattern suggests that speakers retrieve unergative verbs, but not unaccusative verbs on a just-in-time basis. This codability effect is in the predicted direction.

A remaining mystery is why we observed a *reverse*-codability effect in the unaccusative conditions. We suspect that this might be because of the strong positive correlation between the event codability of unaccusative pictures and how strongly pictures were associated with unaccusative verbs. For example, when speakers did not use *bounce* to describe a picture of a penguin bouncing, they often used unergative verbs like *jump*. Thus, it could be that as the codability score for an unaccusative-eliciting picture decreases (i.e., as speakers use a wider variety of verbs to describe a picture), so too does the degree to which that picture elicits an unaccusative verb at all (because the wider varieties of verbs is more likely to include more unergative verbs). That is, the likelihood that the pictured event is even conceived of as an unaccusative type event. As such, because only unaccusative verb planning (by hypothesis) happens before sentence onset, eliciting fewer unaccusative productions overall should allow faster onset times.

We tested this hypothesis by quantifying how strongly each picture is associated with unaccusative verbs. In the norming data, we counted how many unaccusative verbs participants in the norming studies produced, and calculated the proportion of unaccusative verbs for each picture. This unaccusativity score for each picture was positively correlated with the event codability scores ($r = .59$, $t(10) = 2.31$, $p = 0.04$). In other words, more codable pictures in the unaccusative conditions were more consistently described with unaccusative verbs by speakers who were not pre-trained to produce a certain words for target pictures. Under the current hypothesis in which speakers need to plan unaccusative verbs before their subjects, speakers may initiate verb retrieval in advance when the sole argument is non-agentive (because unaccusative subjects are by definition non-agentive). Thus, when speakers recognized the event participant as non-agentive more consistently, they may perform an additional process of retrieving verbs before starting to speak its subject noun more consistently. This increased consistency in engaging in advance unaccusative verb retrieval may be the reason that speakers were slower to start speaking in response to the more codable event pictures in the unaccusative conditions.

Consistent with this interpretation, when the production time in the first region of interest was residualized against how strongly a picture is associated with unaccusative verbs (or equivalently, how likely the event participant is recognized as non-agentive, estimated from the norming data described above), EventCodability was no longer a significant predictor of the (residualized) production time ($\chi^2(1) = 0.065$, $p = .80$). Thus, a reasonable interpretation of the reverse codability effect is that speakers were slower to start speaking in response to the event pictures that are more strongly associated with unaccusative verbs. This effect may have masked any effect of event codability, and the two effects are not readily isolable due to collinearity. Regardless of whether this interpretation is correct, the reverse event codability effect still shows that speakers were affected by event properties before starting to speak the subject noun, selectively in the unaccusative conditions. Thus, the reverse codability effect is broadly consistent with the hypothesis that speakers retrieve unaccusative verbs but not unergative verbs before starting to speak subject nouns.

Together, the timing of codability effects suggests that speakers plan unergative verbs on a just-in-time basis. It also suggests, given our interpretation of the reverse codability effect, that speakers retrieve unaccusative verbs before starting to speak the subject noun.

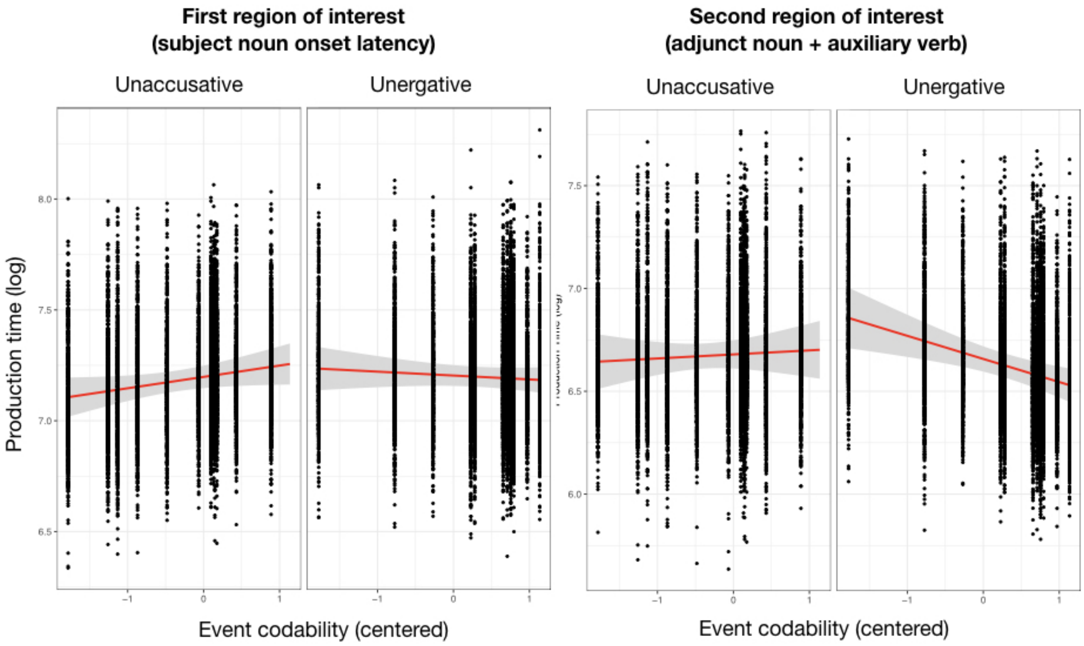

Figure 16. Partial regression plots visualizing the production time in the two regions of interest as a function of event codability in the unaccusative and the unergative conditions.

Experiment 4

Experiments 1-3 established first, that speakers exhibit a semantic interference effect and a *reverse* event codability effect before they start speaking the subject noun of unaccusative sentences, but not unergative sentences. Second, speakers exhibit unergative codability effects right before they speak unergative verbs. We offered a potential explanation of why the unergative interference effect was not reliably observed: the weak unergative interference may be due to the temporal interval between the timing of distractor presentation and the timing of unergative lemma retrieval processes. But it remains possible that the unergative distractors used in Experiment 1-3 were insufficiently related to the target unergative verbs. The goal of Experiment 4 was to examine this possibility by removing the intervening material between the subject noun phrase and the verbs.

(3) The octopus is boiling [unaccusative]

(4) The octopus is swimming [unergative]

Unlike in Experiment 1- 3, the target sentences did not contain adjunct modifiers, but they were otherwise identical to the target sentences in Experiment 1-3. If the absence of the reliable effect of unergative verb interference was due to the long time interval between the distractor presentation and unergative verb retrieval, a robust semantic interference effect from unergative verb distractors in Experiment 4 is predicted. On the other hand, if the inconsistency of the unergative verb interference effects was merely due to the unergative verb distractors being insufficiently related to those unergative verbs, the unergative verb interference effect should not be observed in Experiment 4.

Methods

Participants. Fifty undergraduate students at the University of California, San Diego, participated in the experiment for course credit. All participants reported that they learned English as their first language. None participated in Experiment 1-3. Informed consent was obtained for each participant before the experimental session.

Materials. The same set of event pictures used in Experiment 1-3 were used in Experiment 4, except that the pictures for the adjunct nouns (e.g., *lemon* and *spoon*) were removed (to simplify the task). The same set of verb distractors were used. The pairing between verb distractors and event pictures were also the same as in Experiment 1-3.

Procedure. Participants underwent a similar familiarization as in Experiment 1-3, where they studied the entire set of event pictures with target sentences and then practiced describing each picture once. Following this familiarization phase, they performed the similar sentence description task with distractor words superimposed on each picture (i.e., the ePWI task). Stimulus presentation parameters (e.g., when fixation cross, the stimulus picture appear and disappear, etc.) were identical to Experiment 1-3. The distractors appeared 150 milliseconds before the presentation of the pictures. (i.e., the SOA was -150).

Because the primary purpose of this experiment was to test whether the related distractors were equally effective between the unaccusative and unergative conditions,

Table 14
The mean onset latency of verbs and accuracy rate by each condition in Experiment 4

VerbType	Relatedness	Verb onset latency	Accuracy %
Unaccusative	Related	2076	95
Unaccusative	Unrelated	1980	97
Unergative	Related	2030	98
Unergative	Unrelated	1928	99

we focused on the onset latency of the verb, rather than on the production time of pre-defined regions of interest. This decision was made because previous studies showed that verb interference effects in unergative sentences could appear later than sentence onset (Momma et al., 2018).

Analysis. The onset latency of verb production of each trial was estimated using the text-to-speech alignment technique used in Experiment 1-3. The same procedures were used as in Experiment 1-3, except that we initially focused exclusively on the verb onset latency in Experiment 4 for the reasons discussed above. For brevity, we only report the results of Gaussian analysis on log-transformed production time data, but the statistical pattern did not change in Gaussian analysis or Inverse-Gaussian analysis.

Results

5.97% of all trials (149 out of 2496 trials) were identified as erroneous and thus were excluded from subsequent analyses. Table 14 summarizes the verb onset latency, along with the accuracy rate in each condition. Speakers were on average 99 milliseconds slower to start speaking verbs given the related verb distractors. There was no evidence that this semantic interference effect was different between the unaccusative and the unergative conditions. There was a significant main effect of Relatedness ($\hat{\beta}$ = -0.05; SE = 0.01, t = -4.93, $p < .001$), with no interaction between Relatedness and VerbType (p = .79). There was also no reliable effect of VerbType (p = .49).

The sentence onset latency along with region-by-region production time in each condition is summarized in Table 15, and the region-by-region semantic interference

Table 15
The by-subject mean production time by each condition by each region in Experiment 4.

VerbType	Condition	Onset	The	octopus	is
Unaccusative	Related	1164	166	481	249
Unaccusative	Unrelated	1104	156	484	237
Unergative	Related	1139	175	470	232
Unergative	Unrelated	1063	154	468	228

effect is visualized in Figure 17.

Discussion

Experiment 4 showed that speakers showed roughly the same amount of verb interference in the unaccusative and unergative conditions when verbs were positioned closer to sentence onset. This result suggests that the lack of unergative interference in Experiments 1-3 was unlikely to be due to distractors being insufficiently related to the unergative verbs.

However, visual inspection of Figure 17, as well as the post-hoc analysis on the subject noun onset latency, suggest that the temporal pattern of unergative verb interference in Experiment 4 is inconsistent with the previous study by Momma et al. (2018). Namely, Momma et al. (2018) observed that the unergative verb interference effect was found in the total production time of the subject noun and the auxiliary verb, not in the onset latency measure. In contrast, the current study showed that speakers experienced the semantic interference effect on unergative verbs (as well as unaccusative verbs) in the onset latency measure (the simple effect of Relatedness in the unergative conditions: $\hat{\beta}$ = -0.06, SE = 0.02, t = -3.51, p = .005).

This discrepancy may be explained if speakers can be flexible in terms of how many words they retrieve before starting to speak. It is likely that many factors modulate whether speakers retrieve unergative verbs in advance of speaking their subject head nouns. Among them are working memory load (Wagner, Jescheniak, & Schriefers, 2010), time pressure, task difficulty (F. Ferreira & Swets, 2002), recent experience

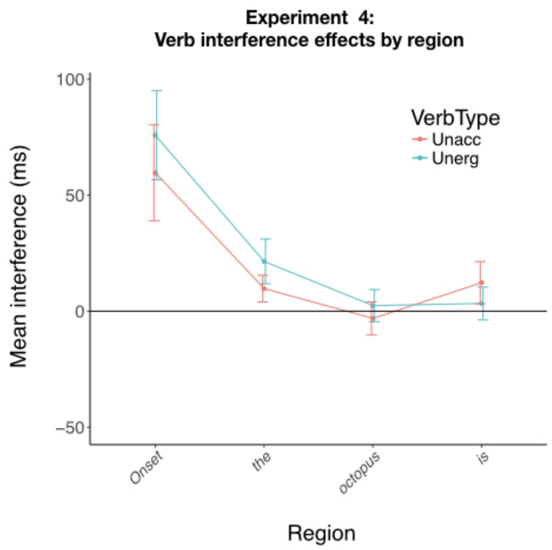

Figure 17. Interference effect (in raw production time) on region-by-region production time by VerbType in Experiment 4. Error bars represent standard error of the mean.

(Konopka, 2012), the phonological length of the first word of an utterance (Griffin, 2001) and potentially numerous other factors (e.g., how careful speakers want to be in what they say). The difference between Experiment 4 and Momma et al. (2018) may be due to many differences in how the tasks were set up, the properties of picture stimuli or target sentences, and so forth. For example, one clear difference between Momma et al. (2018) and Experiment 4 is that only Momma et al. (2018) involved a type of go-no-go task, in that participants were asked to suppress their speech conditionally in some trials. This additional task might have increased the task difficulty in Momma et al. (2018), potentially reducing how many words speakers retrieve in advance.

Importantly, however, the flexibility in whether to retrieve verbs in advance is selective to unergative verbs. When unergative verbs appear relatively far downstream, or, as in Experiment 1-3, when the experimental task is relatively complex, speakers seem to retrieve unergative verbs late, as suggested by the cross-experiment analysis of codability effects in Experiments 1-3. In comparison, when unergative verbs appear closer to the sentence onset and the task is relatively simple (as in Experiment 4), speakers seem to retrieve unergative verbs before sentence onset. Critically, these cross-experiments differences in when speakers retrieve verbs selectively influence unergative verb production, but not unaccusative verb production. This difference is congruent with the claim that speakers retrieve unaccusative verbs before their subject noun to establish the argument-verb dependency, which exists regardless of how complex the task is and how far downstream the verb appears in linear distance.

General Discussion

We reported five extended picture-word interference studies to study the time-course of lexical retrieval in sentences headed by two different types of intransitive verbs: unaccusative and unergative verbs. Experiment 1 suggested that speakers retrieved sentence-final unaccusative verbs before the onset of sentence-initial subject nouns, but retrieved sentence-medial adjunct nouns on a just-in-time basis. Experiments 2a and 2b both supported the conclusion of Experiment 1 that speakers retrieved sentence-medial adjunct nouns on a just-in-time basis. Experiment 3 replicated Experiment 1, again showing the unaccusative verb interference effects in the subject noun onset. Experiment 3 also showed that this interference effect could be more reliably obtained when the distractor is presented sufficiently early. Experiment 4 showed that the absence of the reliable semantic interference effect on unergative verbs in Experiment 1-3 was unlikely to be due the distractor verbs being insufficiently related to the unergative verbs, but instead likely because the timing of distractor presentation is temporally too far from the timing of unergative lemma retrieval processes. Experiment 4 also showed that unergative verbs *can* be retrieved in advance of the sentence onset when they appear closer to the sentence onset. Furthermore, the temporal analyses of event codability effects in Experiment 1-3 offered converging evidence for the claim that speakers retrieve sentence-final unaccusative verbs at the beginning of the sentence, but retrieve unergative verbs on a just-in-time basis. This time-course of lexical retrieval is congruent with the prediction of the hypothesis that the time-course of sentence planning reflects argument structures above and beyond surface word-order.

The role of word-order and argument structure in speaking

In order to speak grammatically, speakers need to be able to map conceptual representations onto structural representations, according to the grammar of their language. To achieve this mapping systematically, speakers need to access information stored with verbs, because verbs' argument structures determine which event participants map onto which grammatical roles in a manner that is not derivable from conceptual structure alone (Grimshaw, 1990). For this reason, some previous models of grammatical encoding (e.g., Bock & Levelt, 1994; De Smedt, 1996; F. Ferreira, 2000) assume verb guidance, the hypothesis that speakers retrieve verbs early in sentence planning and use them to encode the rest of the sentence grammatically. Models that assume verb-guidance normally assume that both subject and object arguments need verb guidance for their grammatical encoding (Bock & Levelt, 2002; F. Ferreira, 2000). However, the current results, along with the results from previous studies (Momma et al., 2016, 2018), suggest that verb guidance is plausible but selective to internal arguments. Speakers selectively retrieve verbs before the articulation of theme or patient arguments, but not agent arguments. This pattern is naturally congruent with the linguistic hypothesis that internal arguments are the only true arguments of verb roots (Kratzer, 1996, 2003).

But the problem with any verb guidance hypothesis is that verbs can appear arbitrarily far from their internal arguments in linear distance. If speakers obligatorily retrieve words sequentially in a 'beginning to end' manner, they must either (a) give up retrieving verbs in advance when they appear further away in a sentence, and instead speak with uncertainty about clausal structure, or (b) plan the entire portion of sentences up to the verbs (see *the head principle* by Martin & Freedman, 2001 for a related view). However, once the assumption that lexical retrieval processes obligatorily occur sequentially is abandoned, speakers can retrieve only the words that are necessary for grammatical encoding. The current results show that speakers are in fact capable of retrieving the verb before starting to articulate their internal arguments, without first retrieving linearly intervening words (here, the noun phrase inside subject-modifying adjuncts).

Why are verbs needed to encode their internal arguments grammatically? There are several possible hypotheses (see Momma et al., 2018 for discussion), and we focus on four here.

First, it is possible that speakers retrieve unaccusative verbs in advance to encode semantic dependencies between internal arguments and verbs. Under this account, the relational meaning of internal arguments (theme or patient arguments) is verb-dependent, but the relational meaning of external arguments (agent arguments) is verb-independent. Kratzer (Kratzer, 2003, p.4) captures the basic intuition behind this claim by, "Themes lack the conceptual independence of agents. Theme arguments seem to be tightly linked to their verbs. Agents are different. Actions seem to have agents independently of how we describe them." Thus, agent arguments are relatively independent of verbs, but theme or patient arguments are dependent on verbs, in the interface between syntax and semantics, or in the semantic representation.

If this representational difference in argument-verb relation is transparently reflected in sentence production, it may be possible to explain why speakers retrieve verbs before speaking their internal arguments but not before speaking their external arguments. Specifically, internal arguments require verb roots to obtain their relational meaning in a sentence, but external arguments do not need a verb root to obtain relational meaning (agent roles). Critically, the semantic interference effect affects the retrieval of verb roots, because verb roots carry the verb meaning that the related (verb) distractor is similar to, and thus their retrieval is susceptible to semantic interference.

The second possibility is that speakers retrieve unaccusative verbs in advance to encode the syntactic dependency between internal arguments and verbs. Under this account, speakers need to retrieve verbs in advance to integrate the internal arguments to sentence structure, because the internal structure of verb phrases is dependent on verbs' argument structure. In linguistics, it has been long observed that the phrase structure of a verb phrase depends on verbs' subcategorization (e.g., Chomsky, 1965). In some theories of syntax, the subject arguments of unaccusative verbs are their objects underlyingly and thus are a part of verb phrases (Perlmutter, 1968, 1978). On this account, in order to syntactically encode the subjects of unaccusative verbs, it is necessary to retrieve verbs. This may be the cause of advance verb retrieval in sentences with unaccusative verbs.

The third possibility is that speakers retrieve unaccusative verbs in tandem with internal arguments because an internal argument and its verb, but not an external argument and its verb, form a semantic unit. Under this account, unaccusative verbs and their subject argument form integrated units in semantic representation, and parts of a single semantic unit is planned in tandem. This account may be closely related to the notion of *semantic integratability* (Solomon & Pearlmutter, 2004), which is argued to influence the scope of planning in sentence production.

The fourth possibility is that speakers retrieve unaccusative verbs in tandem with internal arguments because internal arguments and verbs form a syntactic constituent. If this is the case, unaccusative verbs and their subject arguments form verb phrases in the underlying syntactic structure. Combined with the idea that speakers retrieve lemmas in advance within a first phrase (Smith & Wheeldon, 1999), it can be argued that speakers retrieve both the derived subject and verb in tandem, before starting to speak unaccusative sentences.

Thus, the current results can be explained in terms of semantic dependency, syntactic dependency, semantic constituency, or syntactic constituency. Based on the current data, we cannot distinguish between these explanations. However, one way to empirically distinguish dependency-based and constituency-based accounts is to test whether speakers retrieve object arguments before producing verbs. Under the dependency-based hypothesis, speakers should not need to retrieve object nouns in advance, because verbs are not dependent on their object arguments. On the other hand, under the constituency-based hypothesis, it can be predicted that speakers need to retrieve the object noun before saying the verb. The precise reasons that speakers plan verbs before their internal arguments remain inconclusive, but the above describes two explicit hypotheses and some empirical ways to test them. However, regardless of which account is correct, the current results suggest that how speakers plan utterances reflects the abstract relationship between arguments and their verbs, over and above surface word-order.

Is sentence production not incremental?

The current results may at first seem to be inconsistent with the widely accepted view that sentence production is incremental. But this depends on what it means for the sentence production system to be incremental. There are at least three distinct definitions of incrementality in the literature. First, at the most general level of description, incrementality refers to the idea that speakers can (but not must) interleave the planning and articulation of a single sentence (Bock & Ferreira, 2013; F. Ferreira & Swets, 2002; Levelt, 1989). This version of incrementality is not controversial in the literature and is not in conflict with the current results. In fact, the present findings support it by showing that speakers do not (need to) retrieve the adjunct noun and unergative verbs until immediately before they need to speak them.

Second, incrementality can refer to the idea that the retrieved words and planned structures are immediately integrated to the overall representation of a sentence (Momma & Phillips, 2018). This version of incrementality is also compatible with the current results: the unaccusative verbs that are retrieved in advance may be immediately integrated with the intermediate representation of a sentence without delay. Note that this notion of incrementality is parallel to the idea of incrementality in syntactic parsing (Demberg, Keller, & Koller, 2013; Momma & Phillips, 2018; Sturt & Lombardo, 2005). In parsing, incrementality refers to the property of the parser that immediately integrates the input representation to the overall representation of a sentence so that the syntactic representation is connected throughout a comprehension process.

Finally, incrementality can refer to the idea that planning and articulation of a particular element of an utterance synchronizes as much as possible (Christiansen & Chater, 2016; De Smedt, 1996; F. Ferreira, 2000; Van Nice & Dietrich, 2003). The current results are not readily compatible with this version of incrementality, in the sense that the retrieval and articulation of unaccusative verbs was shown to be systematically de-synchronized. This version of incrementality is what gives rise to the sequential retrieval assumption, as we discussed in the introduction, and is primarily motivated by working memory considerations. A natural concern, then, is that advance unaccusative verb retrieval processes may not be cognitively economical. Thus, we next address how the current results relate to working memory in sentence production.

Working memory and sentence planning

As discussed in the introduction, a primary motivation for sequential word retrieval is to minimize memory cost during sentence production. Contrary to this memory consideration, the current results suggest that speakers retrieve unaccusative verbs that appear far downstream in the early stages of sentence planning. If speakers keep the retrieved verb in working memory until it can be spoken, the process of producing unaccusative sentences (and also passive sentences, according to the previous results by Momma et al., 2015) may be costly. Also, if the current claims were to apply cross-linguistically, it is predicted that verb-final languages, all else being equal, are harder to speak than non-verb-final languages.

However, whether these predictions follow from the current claim depends on how memory cost is defined in sentence production. There are at least two reasons to postulate that retrieving verbs in advance might not be costly in terms of working memory. First, as long as verbs that are retrieved in advance are immediately integrated into the overall structure of a sentence (in line with the second notion of incrementality discussed above), verbs become a part of a syntactic chunk, that is, a constituent. Just like chunking alleviates memory cost in other domains of cognition (Miller, 1956), it is possible that speakers may hold arbitrarily many words as long as they can form a single connected constituent structure. Under this view, a single syntactic chunk needs to be held in working memory throughout the course of producing a sentence, so sentence production is of course not cost free. But as a result of chunking, the cost of producing a sentence does not increase proportionally to the number of words. Thus, it is not clear if holding multiple words in memory is necessarily as costly in terms of working memory.

Second, verbs may not be a source of interference for other elements of a sentence. It is well-known in the working memory literature that the similarity between relevant items largely determines working memory cost. This is true for both capacity-based views (Baddeley & Hitch, 1974; Just & Carpenter, 1992; Miller, 1956)) and interference-based views (Nairne, 1990). Generally speaking, when some item or items are held in working memory, maintenance and retrieval of other similar items can be harder. The view that similarity is a major factor affecting processing cost is widely accepted in psycholinguistics, in particular, sentence comprehension (Lewis & Vasishth, 2005; Van Nice & Dietrich, 2003). Critically, it has been suggested that lemmas belonging to the different syntactic categories (e.g., nouns vs. verbs) may not interfere with each other in sentence production, even when they are conceptually similar to each other (Dell et al., 2008; Momma, Buffinton, Slevc, & Phillips, submitted). For example, Momma et al. (submitted) showed that speakers are slower to say a word in a sentence (e.g., *singing*) when given a conceptually similar distractor word in immediate memory (*whistling*), but only when the conceptually similar distractor is perceived as in the same syntactic category as the target. For example, when both the to-be-produced and the distractor words are nouns, as in *her skillful singing/whistling* or are verbs as in *she is skillfully singing/whistling*, they interfered with each other. Critically, when both the to-be-produced and the distractor words did not share the same syntactic category, they did not interfere with each other. Given this category specificity of retrieval interference, buffering verbs as nouns are processed (or vice versa) may not be cognitively costly. If similarity-based retrieval interference is a primary determinant of processing cost only when items match in syntactic category, non-sequential lexical retrieval processes may not be costly, as long as only one item of a certain category is held in memory. In other words, speakers may process sentences in 'one-word-of-each-category-at-a-time' fashion. For example, retrieving a sentence-final verb first may not incur processing cost on the retrieval and production of linearly preceding nouns. Thus, given what we know about memory cost, the (selective) advance verb planning mechanism we described here may not be particularly costly. One prediction of this view is that relative clause modifying subject nouns of unaccusative verbs (and also passive verbs) are predicted to be difficult to produce, because under the current view, unaccusative verbs and passive verbs need to be held in working memory while the verbs of subject-modifying relative clause is retrieved. This prediction can be tested experimentally, and also using corpus studies. For example, relative clauses modifying subject nouns may be predicted to be rarer in unaccusative than in unergative sentences, under the linking hypothesis that more difficult structures are less likely to be produced. Future studies may test these predictions.

Flexibility in planning scope

From the early days of modern sentence production research, it has been noted that how much speakers plan at a time, that is, *scope of planning*, may be flexible (Levelt, 1989), and this insight has been verified experimentally (F. Ferreira & Swets, 2002; Griffin, 2001; Konopka, 2012; Wagner et al., 2010). Under this view, a natural question in the current context is whether speakers *must* retrieve verbs before starting to speak their internal arguments.

Phenomenologically, it is implausible that speakers *must* retrieve verbs before their internal arguments in all circumstances. For example, speakers can name an object that happens to come to mind for whatever reasons in a phrasal format (e.g., *the computer*) without having any idea about what the continuation of the sentence is. Speakers should have no problem determining the continuation of this sentence using unaccusative verbs (e.g., *the computer fell from the table*) *after* having uttered the noun phrase. Of course, this is a phenomenological observation that needs to be interpreted cautiously, but according to this intuition, we consider it unlikely that speakers *must* retrieve a verb before its internal arguments under all circumstances.

When, then, do speakers need to plan verbs before their internal arguments? One possibility is that speakers must retrieve a verb to speak a noun phrase realizing its relational meaning in a sentence, not as a fragment that can be repaired into an internal argument. When speakers produce a particular noun phrase as a fragment and later fix it into a full sentence, they need not retrieve verbs in advance to starting to speak their internal arguments. In contrast, when they speak internal arguments as internal arguments, speakers may need to retrieve verbs before their internal arguments. One empirical way to distinguish these two 'modes' of speaking would be to measure the fluency of speech. When speakers say a noun phrase as a fragment, not as an internal argument, there should be a measurable pause (or filler, like *um*) before being able to produce the verb. Thus, it is possible that speakers must retrieve verbs before starting to speak their internal arguments *in fluent speech*.

As far as we are aware, no existing evidence contradicts the strong hypothesis that speakers must plan verb lemmas before speaking internal arguments as internal arguments. Also, note that the model that allows flexible verb planning is theoretically less constrained, and therefore is harder to be falsified. Therefore, until proven otherwise, we adopt the more readily falsifiable hypothesis that speakers *must* plan verbs before their internal arguments when they speak the internal arguments as internal arguments.

Methodological remarks

The current study combines the word-by-word production time measurement (using an automatic forced alignment algorithm) with an interference paradigm to track the real-time cost associated with specific sub-processes in sentence production. Certainly, other methodologies, such as visual-world eye-tracking during speaking (e.g., Griffin & Bock, 2000 among others) also allows measurement of what speakers process in real-time, but the current method nicely complements them. The advantage of the current method is that it allows researchers to investigate the processing cost of sentence parts that do not directly correspond to an easily definable region in a picture, such as verbs (though see, e.g., Hwang & Kaiser, 2014 for an attempt). Most of the time these types of sentence parts encode abstract relational information, and arguably how speakers encode relational information is the critical but missing part of sentence production theories. It is possible to imagine extending the current paradigm to probe the timing of even more abstract processing, such as the processing of tense, empty categories, functional heads, and so forth, by carefully choosing the right kinds of distractors. This sort of investigation is not easily possible with existing methods, and thus the current method opens up new opportunities for investigating previously under-investigated aspects of how sentence planning unfolds over time in speaking. Of course, this methodology has some limitations; for example, as the current results suggest, the interference effect that occurs late in a sentence may not be reliably detected.

Concerning the semantic interference effects observed in the onset latency measures, one consistent pattern was that Inverse-Gaussian mixed models revealed more reliable effects than ordinal Gaussian mixed models. This difference is likely because the distribution of onset latency measures is highly skewed, and the semantic interference on onset latency measures reside in slow trials. As can be seen in Figure 18, the unaccusative interference effects found in Experiment 1 and 3 were most prominent in slow trials.

Given this nature of interference effects, it is likely that log-transformation (or Inverse-transformation) 'shrinks' the effects at the slower trials. We did find significant effects either way, but the degree of significance was different, both in Experiment 1 and 3. Thus, it is important to be cautious when applying a transformation to production time data to satisfy the normality assumption of linear mixed effects models, especially when there is an *a priori* reason to believe that the effect resides in slow trials, like in the current study.

Conclusion

Since the seminal work by Garrett (1975), a major goal of sentence production research has been to understand how speakers translate conceptual representations into

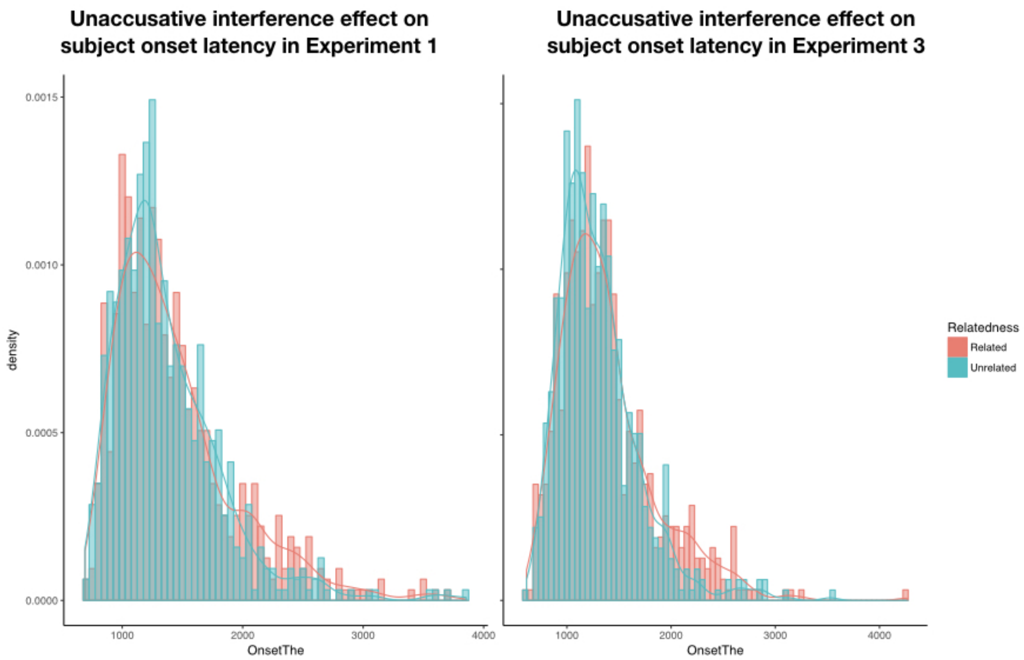

Figure 18. The density plots visualizing the distribution of subject noun onset latency in the unaccusative verb distractor conditions by distractor relatedness in Experiment 1 and 3

syntactic representations. Verbs' argument structures are a critical part of a semantic-syntax or conceptual-grammatical interface. Thus, they are likely to play a key role in this translation process. The current studies show that speakers retrieve words in sentences non-sequentially and that the order of lemma retrieval reflects verb-argument dependencies over and above surface word-order. Therefore, developing an adequate model of sentence production requires both incorporating theories of abstract argument structures and abandoning the default simplifying assumption that the time-course of lexical planning is transparently reflected in the surface word-order of a sentence.

References

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155.

Alexiadou, A., Anagnostopoulou, E., & Everaert, M. (2004). *The unaccusativity puzzle: Explorations of the syntax-lexicon interface* (Vol. 5). Oxford University Press on Demand.

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791.

Baddeley, A. D., & Dale, C., Harold. (1966). The effect of semantic similarity on retroactive interference in long-and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *5*(5), 417–420.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bloem, I., van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, *51*(2), 307–323.

Bock, K., & Ferreira, V. S. (2013). Syntactically speaking. In *The oxford handbook of language production.*

Bock, K., & Levelt, W. (2002). Language production. *Psycholinguistics: Critical concepts in psychology*, *5*, 405.

Bock, K., & Levelt, W. J. (1994). Language production: Grammatical encoding. In *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior research methods*, *41*(4), 977–990.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review*, *113*(2), 234.

Chomsky, N. (1965). Aspects of the theory of syntax.

Chomsky, N. (1986a). *Barriers* (Vol. 13). MIT press.

Chomsky, N. (1986b). *Lectures on government and binding: The pisa lectures* (No. 9). Walter de Gruyter.

Chomsky, N., & Keyser, S. J. (1982). *Some concepts and consequences of the theory of government and binding*. MIT press.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (odawa). *Cognition*, *98*(2), 105–135.

Collina, S., Tabossi, P., & De Simone, F. (2013). Word production and the picture-word interference paradigm: the role of learning. *Journal of psycholinguistic research*, *42*(5), 461–473.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283–321.

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*(4), 583–608.

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, *39*(4), 1025–1066.

De Smedt, K. (1996). Computational models of incremental grammatical encoding. In K. d. S. A. Dijkstra (Ed.), *Computational psycholinguistics: Ai and connectionist models of human language processing* (pp. 279–307). London UK: Taylor and

Francis.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. *Aspects of language production*, 291–330.

Ferreira, F., & Swets, B. (2002). How incremental is language production? evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive psychology*, *40*(4), 296–340.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151–167.

Garrett, M. (1975). The analysis of sentence production1. In *Psychology of learning and motivation* (Vol. 9, pp. 133–177). Elsevier.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, *91*(1), 1.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, *27*(1), 1–40.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274–279.

Grimshaw, J. (1990). *Argument structure*. the MIT Press.

Haegeman, L. (1991). *Introduction to government and binding theory* (Vol. 2). Blackwell Oxford.

Hwang, H., & Kaiser, E. (2014). The role of the verb in grammatical function assignment in english and korean. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1363.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–144.

Iwasaki, N. (2010). Incremental sentence production: Observations from elicited speech errors in japanese. In *Processing and producing head-final structures* (pp. 131–151). Springer.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, *61*(1), 23–62.

Johnson, C. J. (1992). Cognitive components of naming in children: Effects of referential uncertainty and stimulus realism. *Journal of Experimental child psychology*, *53*(1), 24–44.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*(1), 113.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122.

Kaplan, R. M., Bresnan, J., et al. (1982). Lexical-functional grammar: A formal system for grammatical representation. *Formal Issues in Lexical-Functional Grammar*(47), 29–130.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive science*, *11*(2), 201–258.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185–209.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162.

Kratzer, A. (1996). Severing the external argument from its verb. In *Phrase structure and the lexicon* (pp. 109–137). Springer.

Kratzer, A. (2003). The event argument and the semantics of verbs. *Ms., University of Massachusetts, Amherst*.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

Levelt, W. J. (1989). Speaking: From intention to articulation. *A Bradford book; ACL-MIT Press series in natural-language processing*.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1–38.

Levin, B., & Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface* (Vol. 26). MIT press.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375–419.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485–495.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, *89*(5), 483.

Marantz, A. (1981). *On the nature of grammatical relations* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Martin, R. C., & Freedman, M. L. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory*, *9*(4-6), 261–280.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of memory and Language*, *35*(4), 477–496.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Momma, S., Buffinton, J., Slevc, L. R., & Phillips, C. (submitted). Syntactic category constrains lexical competition in speaking.

Momma, S., & Phillips, C. (2018). The relationship between parsing and generation. *Annual Review of Linguistics*, *4*.

Momma, S., Slevc, L., & Phillips, C. (2015). The timing of verb planning in active and passive sentence production. In *Poster presented at the 28th annual cuny conference on human sentence processing, los angeles, ca, march 19–21.*

Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 813.

Momma, S., Slevc, L. R., & Phillips, C. (2018). Unaccusativity in sentence production. *Linguistic Inquiry*, *49*(1), 181–194.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269.

Perlmutter, D. M. (1968). *Deep and surface structure constraints in syntax.* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. In *annual meeting of the berkeley linguistics society* (Vol. 4, pp. 157–190).

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107–142.

Sag, I. A., & Pollard, C. J. (1987). *Head-driven phrase structure grammar.* Morgan Kaufmann.

Schnur, T. T. (2011). Phonological planning during sentence production: Beyond the verb. *Frontiers in psychology*, *2*, 319.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, *29*(1), 86–102.

Schriefers, H., Teruel, E., & Meinshausen, R.-M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language*, *39*(4), 609–632.

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of experimental psychology: Learning, memory, and cognition*, *37*(6), 1503.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*(3), 205–246.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, *49*(1), 1–46.

Steedman, M. (2000). The syntactic process.

Sturt, P., & Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, *29*(2), 291–305.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... others (2004). A new on-line resource for psycholinguistic studies. *Journal of memory and language*, *51*(2), 247–250.

Van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from german picture descriptions. *Linguistics*, *41*(5; ISSU 387), 825–850.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive psychology*, *48*(4), 422–488.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423.

Yamashita, H., & Chang, F. (2001). "long before short" preference in the production of a head-final language. *Cognition*, *81*(2), B45–B55.

Yuan, J., & Liberman, M. (2008). Speaker identification on the scotus corpus. *Journal of the Acoustical Society of America*, *123*(5), 3878.