Non-sequential lexical planning in sentence production

Shota Momma and Victor S. Ferreira

University of California, San Diego

Author Note

Shota Momma, Department of Psychology, University of California, San Diego

Victor S. Ferreira, Department of Psychology, University of California, San Diego

This research was supported in part by a grant from 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

Intuition suggests and theories propose that speakers plan words in an utterance in roughly the same order as the surface word-order of the utterance, one word at a time. However, grammatical considerations suggest that this may not be how the production system actually plans sentences. Here we argue that abstract argument-predicate dependency controls how a sentence plan unfolds. We show that speakers plan sentences non-sequentially, providing evidence that the syntactic representations of sentence-final verbs are retrieved before certain sentence medial nouns, selectively in specific types of sentences called *unaccusative* sentences (Perlmutter, 1978). We suggest that the abstract argument structure that is not visible on the surface is a critical factor in building an adequate model of sentence production.

Keywords: Sentence production, Argument structure, Advance planning, Working memory, Picture-word interference

Non-sequential lexical planning in sentence production

Introduction

It is intuitive to suppose that speakers plan words in an utterance in roughly the same order as the surface word-order of the utterance, one word at a time. This intuition at least dates back to behaviorism (Watson, 1920), and is captured by many modern models of sentence production, especially those that assume a tight synchronization between the planning and articulation of words (Chang et al., 2006; Christiansen & Chater, 2016; De Smedt, 1990; 1996; V. Ferreira & Dell, 2000; Iwasaki, 2011; Kempen & Hoenkamp, 1987; Levelt, 1989; Slevc, 2011; van Nice & Dietrich, 2003). We call such a view *sequentialism*: the assumption that words in a sentence are planned sequentially from the beginning to the end of an utterance. A different possibility is that abstract argument-predicate dependency controls the time course of sentence planning above and beyond the surface word-order in a sentence. The current study evaluate these two views.

Sequentialism in sentence production

Sequentialism is not based on intuition alone, but is motivated by well-known limitations of human working memory (Christiansen & Chater, 2016; De Smedt, 1996; Kempen & Hoenkamp, 1987; Iwasaki, 2011). Human working or short-term memory is normally considered to be capacity-limited (Miller, 1956) or susceptible to similarity-based interference at some stages of processing (Baddeley & Dale, 1966; Gillund & Shiffrin, 1984 among many others). Hence, holding planned (i.e., retrieved) words in memory while preparing other words for retrieval and production can be cognitively costly, because the planned words interfere with the processing of the other words until they can be spoken. It would be more cognitively economical for speakers to plan and produce sentences one word at a time. This justifies sequentialism because if speakers plan words in a different order than the surface word-order, planned, yet-to-be-articulated words held in working memory could interfere with the retrieval of the other words that need to be spoken first.

In accordance with such memory considerations, many models of sentence production assume some forms of sequentialism. For example, Levelt (1989), based on the model originally proposed by Kempen and Hoenkamp (1987), describes a model of sentence production in which the order of lemma retrieval is roughly in accordance with the linear order of words as spoken. De Smedt (1990, 1996)

formalized this property using the notion of *information cascading* across levels of processing - the notion that information at a higher level of processing can be sent to the next level of processing without delay. Iwasaki (2011) adopted this idea and argued that even the lemmas of sentences in head-final languages can be planned sequentially. These models are generally consistent with the idea proposed by V. Ferreira and Dell (2000). They described a model of sentence production in which speakers produce sentences in accordance with 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 general working memory consideration discussed above: they aim to minimize potential memory cost by maximizing the isomorphism between planning and articulation of words. This results in sequential planning from the beginning to the end of an utterance.

Grammar and sentence production

Human language performance is certainly constrained by memory limitations, but at the same time, natural languages are full of dependency relationships between words that can be arbitrarily far apart. One prominent case of such relationships is the dependency between arguments and verbs (or more precisely, predicates). Argument-verb dependencies often have phonological consequences on arguments. For example, the case marker on the argument can change due to the properties of the verb, or even the presence or absence of overt arguments depends on the argument structure of verbs. Because there are many languages and constructions in which arguments appear before verbs, it seems logically necessary that speakers plan verbs before starting to speak these arguments, at least in some languages and in some constructions. Consistent with such considerations, some prominent production models (e.g., Bock and Levelt, 1994) assume that verb lemmas are necessary to assign grammatical functions to arguments. De Smedt (1996) also described the importance of verbs' lemmas in correctly choosing the clausal type of potentially pre-verbal complement clauses. These claims are often subsumed under the *verb-guidance hypothesis*.

However, the verb-guidance hypothesis has been challenged in the literature (Schriefers et al., 1998; Iwasaki, 2011) mainly because of the memory considerations discussed above, and also because of empirical failures to find the evidence for it. For example, Schriefers et al. (1998) conducted a series of experiments showing that verbs in German verb-final clauses are not planned before speakers start to

produce subject nouns. Based on such evidence, it has been argued that verb lemmas are not necessary to encode the grammatical structures of sentences (Allum & Wheeldon, 2007; Schriefers et al., 1998; Iwasaki, 2011). This verb-independent encoding, sometimes referred to as the *conceptual guidance hypothesis,* serves to reduce speakers' memory cost. If verbs are not needed to encode other sentence arguments, speakers need not hold the yet-to-be-spoken verbs in memory while retrieving intervening sentence material until they can be spoken.

Thus, there is an apparent tension between memory and grammatical demands in sentence production, but Momma et al. (2014, 2015, 2017a) offered a potential reconciliation. They showed that verbs are planned selectively in advance of the production of internal arguments (roughly, non-agentive arguments), but not external arguments (roughly, agentive arguments). Using a paradigm similar to Schriefers et al. (1998), Momma et al. (2015) showed that the verb in Japanese is planned selectively before the speech onset of object-initial (OV) sentences but not subject-initial (SV) sentences. This is furthermore corroborated by two studies of English sentence production. Momma et al. (2017) reported that verbs that take a patient or theme as a sole argument (known as unaccusative verbs, e.g., *fall;* Perlmutter, 1978) are planned in advance of the articulation of the subject noun head, while verbs that take an agent as a sole argument (known as unergative verbs, e.g., *run*) are not. They also reported the same pattern in active and passive production in English (Momma et al., 2014): speakers plan verbs selectively before starting to utter the grammatical subjects of passive sentences. What appears to be systematic across all these studies is that speakers plan the verb in advance selectively before articulating its internal argument (see Hwang & Kaiser, 2014 for seemingly conflicting results, but see also Momma et al., 2015 for discussion).

Interestingly, in linguistic theorizing, it has been argued that internal arguments are the only argument of verb roots (Marantz, 1981; Krattzer, 1996, 2003). Consistent with this theoretical claim, most, if not all, properties of verbs that are irreducible to conceptual structures, including transitivity, selectional restrictions on complement clausal types (Grimshaw, 1990; De smedt, 1996), and the idiosyncratic case marking in languages like Icelandic (Marantz, 1981) only affect the properties of internal arguments. Thus, as long as speakers plan verbs before the production of internal arguments, they can speak grammatical sentences, at least most of the time. At the same time, selective advance planning

is not as costly as blind advance verb planning in terms of memory, especially given the cross-linguistic tendency to put verbs and internal arguments adjacent to each other. Thus, the selective advance verb planning before internal arguments might provide a middle ground for satisfying both grammatical and memory constraints at the same time in sentence production.

However, the verb guidance hypothesis of any form is not easily compatible with sequentialism. If speakers need to plan verbs before producing internal arguments and plan sentences sequentially at the same time, it is predicted that speakers must plan entire sentences before starting to speak when the internal arguments occur sentence initially and verb occurs sentence finally. This is both intuitively and empirically untenable (Christiansen & F. Ferreira, 2005; F. Ferreira & Swets, 2002; Konopka et al., 2012; Wagner et al., 2010; among others). Hence, either the verb guidance hypothesis or sequentialism must be weakened or abandoned. In the current experiments, we examine the time-course of sentence planning in the production of sentences with sentence-initial internal arguments and sentence-final verbs, with additional intervening sentence material. We investigate the relative timing of the planning of sentence-final verbs and sentence-medial nouns. This serves as a strong test for whether speakers plan sentences sequentially, when there is a competing motivation to plan the sentence-final word before starting to articulate internal arguments. If speakers plan sentence final verbs before sentence medial nouns, it offers strong evidence for non-sequentiality in sentence production.

Picture word interference and sentence planning

The timing of retrieving a particular lemma in a sentence can be investigated using what is sometimes called the *extended picture-word interference* (ePWI) paradigm. ePWI is an extension of the widely used picture-word interference paradigm (Lupker, 1976). In a standard PWI paradigm, speakers simply produce a single word as a response to a picture stimulus, and either hear or see a distractor word. Distractor presentation timing varies across studies, but normally are presented temporally close to the picture presentation onset. The classic finding using this paradigm is what is known as *the semantic interference effect*: speakers are slower to name pictures given distractors that are conceptually similar to a target word (Lupker, 1976; Schriefers et al., 1990; Roelofs, 1992; Vigliocco et al., 2002 inter alia). This effect is considered to reflect the lemma retrieval process rather than conceptual preparation processes, because it has been shown that 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 (Meyer, 1996; Schriefers et al., 1998). Traditionally, ePWI is used to test whether a particular word in a sentence is planned 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 plan a particular word in a sentence. Specifically, it is possible to measure the production time of each word of a speakers' utterance and to measure the effect of distractor words on the production duration. We use this word-by-word production time measurement in the current experiments to assess the relative timing of sentence-medial noun planning versus sentence-final verb planning.

Current study

In the current study, we report three experiments that are only minimally different from each other. In all three experiments, we used the extended version of the ePWI paradigm to investigate the relative timing with which speakers plan the sentence-final verb and sentence-medial nouns (see the method sections below for more details), building on the previous study by Momma et al. (2017a). This is to test whether the relative order in which speakers retrieve lemmas correspond to the surface word-order of an utterance. Given pictures as in Figure 1 and the instruction to describe the entity indicated by the red arrow, speakers consistently say sentences like (1) and (2).

(1) *The octopus below the spoon is boiling.* [Unaccusative sentence]

(2) *The octopus below the spoon is swimming.* [Unergative sentence]

To measure the relative timing at which speakers retrieve the relevant lemmas, we presented visual distractor words that were conceptually related to the noun head inside a subject modifying adjunct (e.g., *knife*) or 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 lemma. Experiments 1, 2, and 3 have essentially the same structure, except that the timing at which distractors are presented is different (150 ms before the picture onset in Experiment 1, at the same time as picture onset in Experiment 2, and 300 ms after the picture onset in Experiment 3).

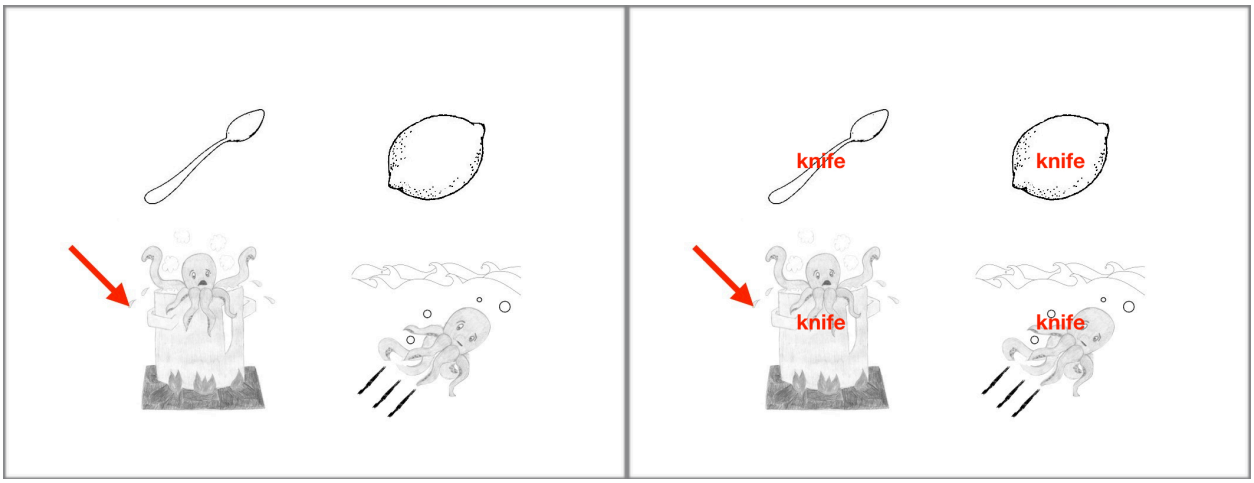

Figure 1. Two example picture stimuli used in all experiments reported here, with (right) and without (left) distractor words superimposed. This picture elicited *The octopus below the spoon is boiling*. When the red arrow points to the octopus in the bottom-right corner, it elicited *The octopus below the lemon is swimming*. In half the stimulus lists, the position of two actions are switched. Across different picture stimuli, the action pictures corresponding to unaccusative verbs appeared both in the right and left side of the screen roughly equally often.

Experiment 1

The goal of Experiment 1 was to test whether the lemma retrieval of unaccusative verbs (e.g., *boil*) selectively precedes the lemma retrieval of the noun inside an intervening adjunct phrase (e.g., *spoon*). Based on a previous study that showed that unaccusative verbs are selectively planned before their subject nouns are articulated (Momma et al., 2017a), it was predicted that the timing of any verb interference effect should temporally precede the timing of any adjunct-noun interference effect. In contrast, based on the same previous study, unergative verbs were predicted to be retrieved after the subject noun phrase production, one or two words before speakers need to produce them (Momma et al., 2017b). Thus, we predicted that the timing of the verb interference effect in unergative sentences temporally follows the timing of the noun interference effect.

Method

Participants. Forty-eight undergraduate students at University of California, San Diego participated in the experiment for course credit. Two participants were excluded because their first

language was not English, according to self-report. Another participant was also excluded because of low accuracy rate (< 50% in at least one of the conditions). All remaining forty-five participants reported that they learned English as their first language. Informed consent was obtained for 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 1. Twelve of the event pictures corresponded to unaccusative sentences (e.g., *the dog is spinning*). The remaining twelve event pictures corresponded to unergative sentences (e.g., *the dog is barking*). The unaccusative and unergative verbs were chosen based on transitivity alternation tests, and were confirmed using additional tests that are known to correlate with unaccusativity. The tests included the transitivity test, the pronominal modifier formation test, and the *-er* affixation tests (see https://shotam.github.io/Stimuli/nlPWI_test.pdf for the list of verbs with the tests used to classify them). Each of the twenty-four complex pictures contained two events (e.g., *spinning* and *barking* events) sharing the same event participant (e.g., *dog*), and two object pictures (e.g., *apple* and *spoon*). In half of the pictures, event pictures were placed on the top-half of the display. In other half of the pictures, object pictures were placed on the bottom-half of the display. The overall frequency of verbs was not significantly different between unergative (M = 8.61; SD = 1.68) and unaccusative conditions (M = 8.89; SD = 1.00) based on the CELEX corpus; these were not different, according to a two-tailed independent sample t-test assuming equal variance ($p > .5$).

For each picture, the related distractor words were first chosen based on intuition, and their relatedness was verified using the cosine similarity measure in Latent Semantic Analysis (Landauer & Dumais, 1997). For verbs, we specifically chose distractors in such a way that (a) the relatedness between the subject noun and the verb distractor does not differ between unaccusative and unergative conditions, (b) the relatedness between 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 measure does not differ systematically 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 means 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 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, not surprisingly, 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 judgement 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 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, not surprisingly, statistically reliable ($t(23) = 6.65$, $p < .0001$).

For each picture, the positioning of the two object entities (e.g, *apple* and *spoon*) 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 dog above the apple/spoon*) were identical 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 which action 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 completely even because of the three participants excluded from the analysis). The entire set target sentences are available online (https://shotam.github.io/Stimuli/nlPWI_stimuli.pdf).

Design. We manipulated three independent variables. First, we manipulated the type of verbs used in an utterance (*VerbType*: unaccusative vs. unergative), the type of the distractors (*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*) design. As discussed above, speakers uttered sentences that contained either an unergative verb (e.g., *the octopus above the spoon is swimming*) or an unaccusative verb (e.g., *the octopus above the spoon is boiling*) while seeing a noun distractor that is related (e.g., *knife*) or unrelated (e.g., *apple*), or verb distractor that is related (e.g., *melt/run*) or unrelated (e.g., *fall*/*smile*). No distractor-target pair started with the same syllable or rhymed with each other.

In accordance with Momma et al. (2017a), we used the 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 corresponding event and object descriptions using a picture booklet containing all experimental picture components (but never the full picture scene like in Figure 1). Following this familiarization session, participants practiced describing the complex pictures. In this practice session, participants were instructed to first find a red arrow in the picture, and describe which one of the two participants (indicated by the red arrow) is doing what action. After one practice trail for each picture, participants proceeded to the experimental session.

Each experimental trial started with the fixation of 500 ms. After the fixation, a complex picture as in Figure 1 was presented with a brief click sound that was later used to identify the onset of the picture in the audio recording to measure timing. The entire experiment session was audio recorded, and the audio recording was than transcribed by undergraduate research assistants.

Analysis. The transcriptions and the audio file corresponding to each individual trial was aligned using an automatic text-to-speech forced alignment algorithm (*P2FA*, Yuan & Liberman, 2008). From the output of the text-to-speech alignment algorithm, two types of information were extracted. First, the onset of the subject noun head relative to the picture onset was extracted. Consistent with the previous study investigating the timing of verb planning (Momma et al., 2017a), we used the onset of the subject noun head instead of its determiner, because the determiner can be said without looking at the picture and hence is not in itself necessarily informative about the grammatical encoding process. Second, the durations of each word (except for the first determiner) were extracted.

Any trials containing overt hesitations (e.g., *um*), errors, or with a production onset of more than 5000 ms or production time of 1500 ms in any region were excluded from the analysis (13.2% of the

data). For each participant, noun onset latencies and durations more than three standard deviations away from the participant's mean were excluded (< 2% for all measurements). The remaining data points were log-transformed and submitted to statistical analyses. All categorical variables were deviation coded (i.e., -0.5 vs. 0.5). We analyzed the effect of distractor relatedness and verb type in each distractor target condition separately, because the three-way interaction was not of theoretical interest. All the analyses reported in this article used the maximal random effects structure (Barr et al., 2013) unless otherwise noted. Also, all p-values are derived using *lmerTest* package (Kuznetsova et al., 2015) unless otherwise noted.

Results

The mean raw production time for each region in each condition is summarized in Table 1. All the mixed effects analyses conducted on the regions of interest are reported in Table 2. A difference plot that visualizes the interference effects is shown in Figure 2.

Condition	Onset + The	octopus	below	the	spoon	is
Verb distractor						
Unacc. Rel.	1485 [15]	506 [6]	403 [6]	138 [4]	538 [5]	262 [7]
Unacc. Unrel.	1409 [13]	524 [7]	408 [4]	139 [5]	537 [5]	267 [8]
Unerg. Rel	1433 [15]	496 [7]	395 [6]	144 [5]	530 [5]	235 [6]
Unerg. Unrel	1463 [13]	508 [5]	394 [6]	140 [5]	524 [5]	236 [6]
Noun distractor						
Unacc. Rel.	1488 [19]	518 [6]	422 [7]	157 [6]	556 [7]	268 [8]
Unacc. Unrel.	1460 [24]	518 [6]	405 [5]	136 [4]	552 [7]	253 [8]
Unerg. Rel	1431 [25]	501 [6]	424 [7]	158 [6]	558 [8]	245 [6]
Unerg. Unrel	1422 [21]	506 [6]	399 [5]	142 [6]	533 [4]	222 [8]

Table 1. The mean region-by-region raw production time along with standard errors in square brackets in Experiment 1.

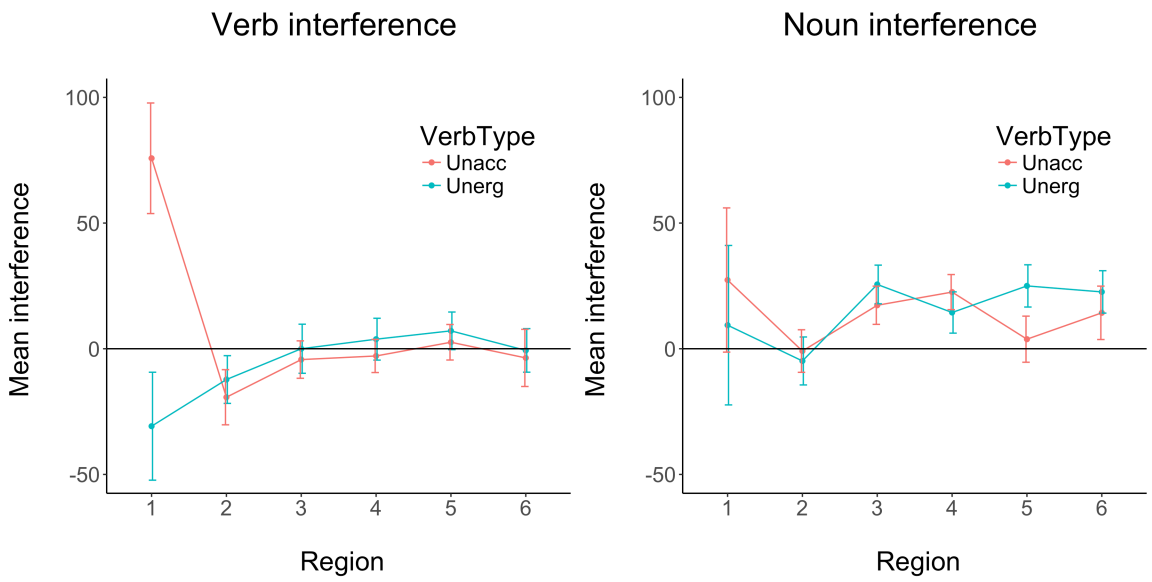

Figure 2. Interference effect (in raw production time) by region by VerbType in the verb distractor condition (left) and in the noun distractor condition (right) in Experiment 1. The error bars represent standard error of the means.

Term	β	SE	|t|	p
		Verb distractor conditions		
Subject noun onset				
Intercept	7.23	0.03	230.32	< 0.0001
Relatedness	0.01	0.01	0.93	0.34
VerbType	0.00	0.02	0.16	0.87
Rel. x VerbType	-0.06	0.02	2.45	0.02 *
Noun inside the adjunct				
Intercept	6.24	0.03	245.95	< 0.0001
Relatedness	0.00	0.01	0.07	0.95
VerbType	-0.02	0.04	0.64	0.52
Rel. x VerbType	0.01	0.02	0.49	0.63
Auxiliary verb				
Intercept	5.35	0.04	128.82	< 0.0001
Relatedness	-0.01	0.02	0.54	0.59
VerbType	-0.09	0.05	1.71	0.09 .
Rel. x VerbType	0.03	0.05	0.72	0.47
		Noun distractor conditions		
Subject noun onset				

Intercept	7.23	0.03	238.76	< 0.0001
Relatedness	0.00	0.01	0.58	0.56
VerbType	-0.03	0.02	1.52	0.13
Rel. x VerbType	-0.02	0.03	0.84	0.40
Preposition				
Intercept	6.24	0.03	222.35	< 0.0001
Relatedness	0.06	0.01	4.00	0.0003 ***
VerbType	-0.01	0.03	0.33	0.74
Rel. x VerbType	0.02	0.03	0.80	0.43
Determiner				
Intercept	4.82	0.03	151.6	< 0.0001
Relatedness	0.07	0.03	2.95	0.005 **
VerbType	0.00	0.03	0.28	0.78
Rel. x VerbType	-0.05	0.05	1.07	0.29

Table 2. Linear mixed effects model estimates in regions of interest in Experiment 1.

Verb distractor conditions. For statistical analyses in the verb distractor conditions, we planned to analyze (a) the subject noun onset latency and (b) the production time of the noun inside the adjunct phrase and the auxiliary verb. This is based on the hypothesis that (a) unaccusative verbs are planned before the subject noun production and (b) unergative verbs are planned immediately before they need to produced (i.e., during the production of neighboring words that precede the unergative verbs). We also report uncorrected statistical results for other regions as well, but readers should note that the analyses of other regions should be interpreted more conservatively.

Subject noun onset latency. Related verb distractors selectively slowed down subject noun onset in the sentence with unaccusative verbs, but not with unergative verbs. This is supported by a significant interaction between VerbType and Relatedness in the mixed effects model with maximal random effects structures. Planned subset analyses revealed that Relatedness was significant ($\beta = 0.04$, SE = 0.02, $|t| = 2.57$, $p = .02$) in the unaccusative conditions. In contrast, in the unergative conditions, the main effect of Relatedness was not significant ($\beta = -0.02$, SE = 0.02,$|t| = 1.18$, $p = .25$). This suggests that verb distractors selectively slowed down the subject noun onset in the unaccusative sentences, but not unergative sentences.

Production time of the noun inside adjuncts. The production time of the noun inside adjuncts did not change reliably across the conditions. No statistically reliable effects were found in this region (all ps > .5). Planned subset analyses by VerbType also revealed no statistically reliable effects of Relatedness (Unaccusative: β = 0.00, SE = 0.02, |t| = 0.22, p =.83, Unergative: β = 0.01, SE = 0.02, |t| = 0.41, p = .69).

Auxiliary verb production time. The production time of the auxiliary verb production time did not change reliably across the conditions. No reliable effects were found in this region (p > .1), except the main effect of VerbType that was marginally significant (p = .09). Planned subset analyses by VerbType also revealed no statistically reliable effects of Relatedness (Unaccusative: β = 0.02, SE = 0.04, |t| = 0.69, p = .49, Unergative: β = 0.00, SE = 0.03, |t| = 0.01, p = .99).

Production time in other regions. Other regions were also analyzed in a post-hoc fashion, using the same statistical analysis procedure as above. We did not find any reliable effects without correction for multiple comparisons (all uncorrected ps > .1).

Noun distractor conditions. For statistical analyses in the noun distractor conditions, we planned to analyze (a) the subject noun onset latency and (b) the production time of the preposition and determiner for the adjunct noun. This is to test whether speakers plan the adjunct noun head before unaccusative verbs, or plan the noun head right before they need to say it (i.e., a word or two before the actual production). Since we did not expect any difference between the unergative and unaccusative conditions with respect to the timing of noun planning, we did not conduct planned subset analyses by VerbType (unlike in the verb distractor conditions).

Subject noun onset latency. The subject noun onset latency did not change reliably across the conditions. No reliable effects were found in this region (ps > .5).

Preposition production time. The production time of the preposition was increased in both the unaccusative and unergative conditions. This is supported by the significant main effect of Relatedness (β = 0.06, SE = 0.01, |t| = 4.00, p = .0003), with no significant interaction with VerbType (p > .4).

Determiner production time. The production time of the determiner was increased in both the unaccusative and unergative conditions. The main effect of Relatedness was significant (p = .005), suggesting that the distractors related to the target noun inside adjuncts slowed down the production time of the determiner for the noun inside adjuncts.

Production time in other regions. We compared three remaining regions in post-hoc analyses. There was a significant main effect of Relatedness at the noun inside adjuncts and the auxiliary verb production time (β = -0.03, SE = 0.01, uncorrected p = .03, β = -0.07, SE = 0.03, uncorrected p = .01, respectively). However, those effects cannot reflect the planning process of the nouns inside adjuncts. The main effect of VerbType was also marginally significant in the auxiliary production time (β = -0.08, SE = 0.05, uncorrected p = .07).

Discussion

In Experiment 1, we observed that speakers are slower to start articulating the subject noun when presented with related verb distractors, but only in sentences with unaccusative verbs. This is a conceptual replication of Momma et al. (2017a). This unaccusative verb interference effect temporally preceded the interference effect on the noun inside the subject-modifying adjuncts. This temporal pattern of interference suggests that speakers plan unaccusative verbs but not unergative verbs before the noun inside adjuncts. This is a case of non-sequential sentence planning, and it suggests that the timing at which speakers plan a certain word in a sentence is controlled by the argument structure of a sentence above and beyond surface word order.

However, Experiment 1 failed to obtain an interference effect for unergative verbs anywhere in the utterance. This null effect might raise the possibility that the order of lemma retrieval is actually influenced by the order in which speakers see the picture elements, rather than argument structures. That is, it may be that speakers plan words in the order that they view the corresponding information in the stimulus picture. Under this explanation, the null effect of verb interference in the unergative conditions is either due to the failure of experimental design (e.g., poor choice of related distractors) or insufficient statistical power.

As discussed above, we hypothesized from previous studies that unergative verbs are retrieved later, roughly when speakers utter words just preceding the target word. A possible reason why we did not obtain an interference effect might be that the unergative verbs are planned too late to be interfered with by the distractors. Recall that the distractor words were presented 150 ms before the picture presentation, and speakers took on average about 2.5 seconds to start saying the noun inside adjuncts (the roughly estimated point in the utterance where they are expected to start retrieving unergative verbs). If the

activation of distractor words decays sufficiently before speakers start retrieving unergative verbs, it is expected that verb interference effects are weak and not easily observable. Thus, we conducted essentially the same experiments, with delayed distractor presentation, in Experiments 2 and 3.

Experiment 2

Experiment 2 repeated Experiment 1, except that the distractor words were presented 150 ms later, at the same time as the picture onset in Experiment 2.

Methods

Participants. Forty-eight undergraduate students at University of California, San Diego participated in the experiment for course credit. None participated in Experiment 1. Three of the participants were excluded because of low accuracy rate (< 50% in at least one of the conditions). All the remaining forty-five participants reported that they learned English as their first language. Informed consent was obtained for each participant before the experimental session.

Materials, Design, Procedure and Analysis. Materials, design, procedure, and analysis were the same as in Experiment 1, except that the distractor onset was synchronized with the picture onset (i.e., the picture-distractor SOA = 0 ms).

Results

The mean raw production time for each region in each condition is summarized in Table 3. All the mixed effects analyses conducted on the regions of interest are reported in Table 4. A difference plot showing interference effects is shown in Figure 3.

Condition	Onset + The	octopus	below	the	spoon	is
Verb distractor						
Unacc. Rel.	1442 [13]	494 [5]	402 [5]	148 [4]	524 [5]	257 [7]
Unacc. Unrel.	1408 [12]	499 [4]	404 [4]	146 [5]	516 [5]	246 [5]
Unerg. Rel	1426 [12]	480 [4]	380 [5]	150 [5]	518 [4]	238 [6]
Unerg. Unrel	1405 [13]	476 [4]	380 [4]	137 [5]	499 [6]	231 [6]
Noun distractor						
Unacc. Rel.	1479 [14]	490 [6]	422 [6]	166 [7]	532 [7]	257 [7]
Unacc. Unrel.	1463 [13]	504 [5]	393 [5]	145 [4]	507 [6]	237 [7]
Unerg. Rel	1432 [15]	496 [6]	415 [5]	158 [5]	521 [7]	236 [8]
Unerg. Unrel	1421 [14]	481 [6]	393 [6]	152 [5]	505 [6]	221 [6]

Table 3. The mean region-by-region raw production time along with standard errors in square brackets, in Experiment 2.

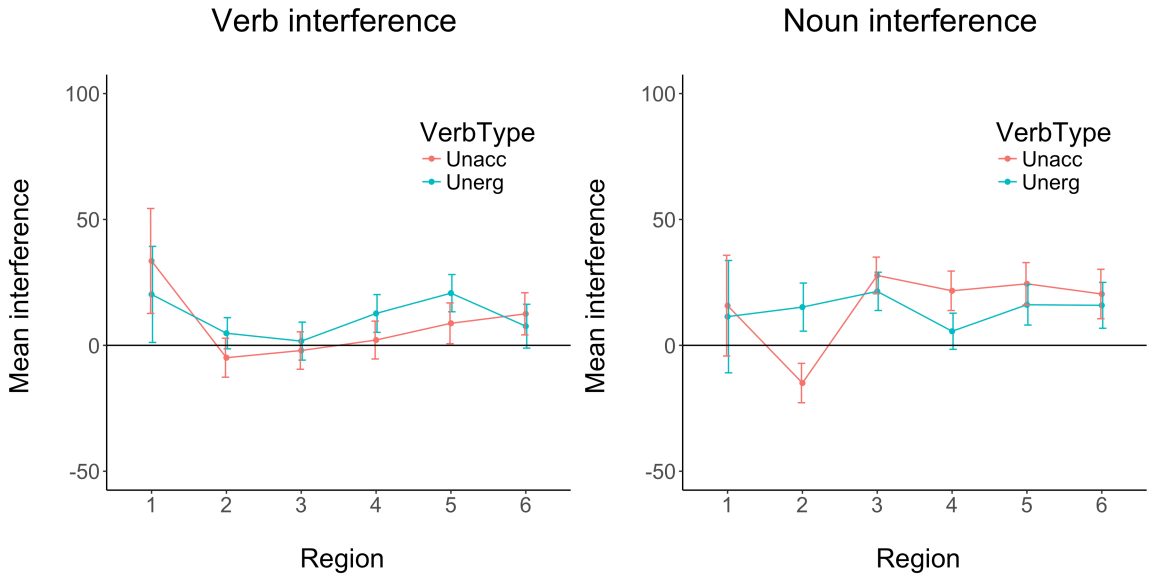

Figure 3. Interference effect (in raw production time) by region by VerbType inthe verb distractor condition (left) and the noun distractor condition (right) in Experiment 2. The error bars represent standard error of the means.

Term	β	SE	|t|	p
		Verb distractor conditions		
Subject noun onset				
Intercept	7.21	0.03	231.73	< 0.0001
Relatedness	-0.01	0.01	1.19	0.24
VerbType	-0.01	0.02	0.69	0.49
Rel. x VerbType	0.01	0.02	0.31	0.76
Noun inside the adjunct				
Intercept	6.20	0.03	225.01	< 0.0001
Relatedness	0.02	0.01	1.93	0.06 .
VerbType	-0.01	0.04	0.49	0.63
Rel. x VerbType	0.02	0.02	1.11	0.27
Auxiliary verb				
Intercept	5.34	0.04	119.06	< 0.0001
Relatedness	0.03	0.03	1.10	0.28
VerbType	-0.07	0.06	1.25	0.22
Rel. x VerbType	-0.02	0.05	0.42	0.68

Noun distractor conditions				
Subject noun onset				
Intercept	7.23	0.03	234.04	< 0.0001
Relatedness	0.01	0.01	0.46	0.56
VerbType	-0.04	0.02	2.11	0.13
Rel. x VerbType	0.01	0.03	0.45	0.40
Preposition				
Intercept	5.94	0.03	187.98	< 0.0001
Relatedness	0.06	0.01	4.42	0.0001 ***
VerbType	-0.01	0.03	0.34	0.74
Rel. x VerbType	-0.01	0.03	0.27	0.79
Determiner				
Intercept	4.86	0.04	136.28	< 0.0001
Relatedness	0.05	0.03	1.58	0.12
VerbType	-0.01	0.04	0.18	0.86
Rel. x VerbType	-0.03	0.07	0.51	0.61

Table 4. Linear mixed effects model estimates in regions of interest in Experiment 2.

Verb distractor conditions. The same regions of interest were chosen as in Experiment 1.

Subject noun onset latency. Verb distractors did not delay the subject noun onset latency reliably. No reliable effects were found in this region (*p*s > 0.2). The planned subset analyses by VerbType revealed no reliable difference in the unaccusative or unergative conditions (Unaccusative: β = 0.02,SE = 0.02, |t| =1.03, *p* =.31, Unergative: β = 0.01, SE = 0.02,|t| =0.60, *p* =.55). Note however that the numerical pattern is the same as in Experiment 1.

Production time of the noun inside adjuncts. As predicted, the verb distractor selectively disrupted unergative verb planning. Supporting this, the planned subset analyses by VerbType revealed that Relatedness did not reliably affect the production time of this region in unaccusative conditions (β = 0.01, SE = 0.01, |t| = 0.66, *p* = .51), but does reliably affect the production time in unergative conditions (β = 0.03, SE = 0.01, |t| = 2.19, *p* = .03). The interaction between Relatedness and VerbType, however, was not significant in the mixed effects model.

Auxiliary verb production time. The auxiliary verb production time was unaffected by the distractor relatedness. No reliable effects were found in this region (*p*s > 0.2). The planned subset

analyses by VerbType revealed no reliable difference in the unaccusative or unergative conditions (Unaccusative: β = 0.04,SE = 0.04,|t| =0.98, *p* =.34, Unergative: β = 0.01,SE = 0.03,|t| =0.46, *p* =.65).

Production time in other regions. Other regions are also analyzed, using the same statistical analysis procedure as in the regions of interest. We did not find any reliable effects even without correction for multiple comparisons (uncorrected *p*s > .1). As in Experiment 1, we did not conduct subset analyses by VerbType, because we did not expect the noun planning timing to differ between the unaccusative and unergative conditions.

Noun distractor conditions. The same regions of interest were analyzed as in Experiment 1.

Subject noun onset latency. The subject noun onset latency was unaffected by the distractor relatedness. The main effect of VerbType was significant (*p* < .05). No other effects were significant (*p*s > .1).

Preposition production time. The main effect of relatedness was highly significant (*p* = .0001), suggesting that the distractors related to the target noun inside adjuncts slowed down the production time of prepositions. No other effects were found to be statistically significant.

Determiner production time. The determiner production time was unaffected by the distractor relatedness. No effect reached significance in this region. However, the effect of Relatedness showed a numerical trend in a way that is consistent with Experiment 1.

Other regions. As in Experiment 1, we compared three other regions in the post-hoc manner. There was a significant effect of Relatedness on production time of the noun inside adjuncts (β = 0.04, SE = 0.01, |t| =3.86, uncorrected *p* = .0001) and the auxiliary verb (β = 0.07, SE = 0.02, |t| =3.08, uncorrected *p* = .002). No effects in any other regions were significant (uncorrected *p*s > .1).

Discussion

In Experiment 2, the pattern of the interference effects in the noun distractor conditions is consistent the results of Experiment 1. This suggests that the noun interference effects are robust across experiments regardless of SOA.

The pattern of the verb interference effects are different from Experiment 1. Namely, in the unaccusative condition, the interference effect on the subject noun onset latency was not significant in Experiment 2. This might be due to the fact that the distractor verbs are presented slightly later in

Experiment 2: the timing of unaccusative verb retrieval and distractor word processing may have less overlap in Experiment 2 than in Experiment 1. Also, in the unergative condition, the reliable effect of unergative verb distractor is observed in the duration of the noun inside the adjunct phrase in the planned comparison test. This suggests the unergative verbs are planned as speakers incrementally produce the sentence medial materials in unergative sentences.

Experiment 3

In Experiment 3, we repeated the same experiment as in Experiment 1 and 2, except that the distractor words were presented 300 ms after the picture onset in Experiment 3.

Methods

Participants. Thirty-two[1] undergraduate students at University of California, San Diego participated in the experiment for course credits. None participated in Experiment 1 or 2. All participants reported that they learned English as their first language. Informed consent was obtained for each participant before the experimental session.

Materials, Design, Procedure and Analysis. Materials, design, procedure, and analysis were the same as in Experiment 1 and 2, except that the distractor onset was 300 ms after the picture onset (i.e., picture-distractor SOA 300ms).

Results

The mean raw production time for each region in each condition is summarized in Table 5. All the mixed effects analysis conducted on the regions of interest was reported in Table 6. A difference plot showing interference effects is shown in Figure 4.

Condition	Onset + The	octopus	below	the	spoon	is
Verb distractor						
Unacc. Rel.	1331 [14]	511 [6]	392 [5]	131 [4]	534 [7]	257 [6]
Unacc. Unrel.	1336 [20]	526 [6]	403 [4]	145 [4]	534 [7]	248 [7]
Unerg. Rel	1312 [15]	514 [6]	386 [6]	136 [5]	535 [8]	235 [7]
Unerg. Unrel	1304 [23]	506 [7]	391 [7]	133 [3]	519 [8]	232 [6]

[1] We stopped running more participants because the academic quarter ended, and because 32 is a multiple of 8, which corresponds to the number of stimulus lists we used.

Noun distractor						
Unacc. Rel.	1366 [17]	530 [7]	437 [8]	160 [8]	556 [8]	273 [10]
Unacc. Unrel.	1376 [14]	528 [7]	393 [7]	136 [7]	527 [7]	243 [7]
Unerg. Rel	1329 [20]	518 [8]	427 [8]	151 [5]	552 [8]	236 [8]
Unerg. Unrel	1297 [20]	510 [6]	401 [6]	142 [5]	539 [6]	226 [9]

Table 5. The mean region-by-region raw production time along with standard errors in square brackets, in Experiment 3.

Term	β	SE	|t|	p
		Verb distractor conditions		
Subject noun onset				
Intercept	7.13	0.04	183	< 0.0001
Relatedness	0.00	0.01	0.19	0.84
VerbType	-0.02	0.03	0.95	0.35
Rel. x VerbType	0.01	0.03	0.42	0.68
Noun inside the adjunct				
Intercept	6.24	0.03	223	< 0.0001
Relatedness	0.01	0.02	0.75	0.46
VerbType	-0.02	0.04	0.58	0.57
Rel. x VerbType	0.02	0.03	0.70	0.48
Auxiliary verb				
Intercept	5.34	0.05	109.58	< 0.0001
Relatedness	0.03	0.03	0.9	0.37
VerbType	-0.05	0.04	1.12	0.27
Rel. x VerbType	0.02	0.05	0.39	0.70
		Noun distractor conditions		
Subject noun onset				
Intercept	7.15	0.04	192.17	< 0.0001
Relatedness	0.00	0.02	0.15	0.88
VerbType	-0.04	0.02	2.18	0.03 *
Rel. x VerbType	0.02	0.03	0.69	0.49
Preposition				
Intercept	5.96	0.04	147.08	< 0.0001
Relatedness	0.08	0.02	4.48	< 0.0001 ***
VerbType	-0.01	0.03	0.39	0.70
Rel. x VerbType	-0.04	0.03	1.17	0.25

Determiner				
Intercept	4.84	0.04	126.53	< 0.0001
Relatedness	0.06	0.03	1.79	0.08 .
VerbType	0.00	0.04	0.07	0.94
Rel. x VerbType	0.06	0.05	1.19	0.24

Table 6. Linear mixed effects model estimates in regions of interest in Experiment 3.

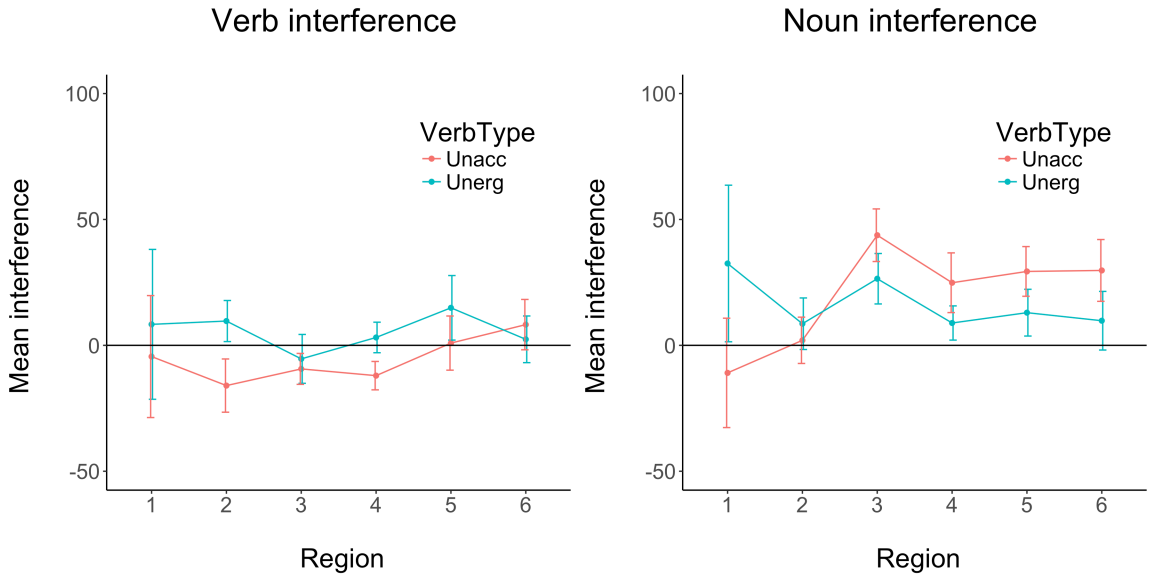

Figure 4. Interference effect (in raw production time) by region by VerbType inthe verb distractor condition (left) and the noun distractor condition (right) in Experiment 3. The error bars represent standard error of the means.

Verb distractor conditions. The same regions of interest were chosen as in Experiments 1 and 2.

Subject noun onset latency. No effects were found to be significant in this region (*p*s < 0.3). The planned subset analyses by VerbType revealed no reliable effects of Relatedness (Unaccusative: β = 0.00, SE =0.02, |t| = 0.20, *p* =.85, Unergative: β = 0.00, SE =0.02, |t| = 0.44, *p* = .66).

Production time of the noun inside adjuncts. No effects were found to be significant in this region (*p*s < .2). The planned subset analyses by VerbType revealed no reliable effects of Relatedness (Unaccusative: β = 0.00, SE =0.02, |t| = 0.14, *p* =.88, Unergative: β = 0.02, SE =0.02, |t| = 0.89, *p* = .38). Note however that the numerical pattern is the same as in Experiments 1 and 2.

Auxiliary verb production time. No effects were found to be significant in this region (*p*s < .2). The planned subset analyses by VerbType revealed no reliable effects of Relatedness (Unaccusative: β = 0.04, SE =0.04, |t| = 1.01, *p* =.32, Unergative: β = 0.01, SE =0.04, |t| = 0.36, *p* = .72).

Other regions. Other regions are also analyzed, using the same statistical analysis procedure as in the regions of interest. We did not find any reliable effects even without correction for multiple comparisons (uncorrected *p*s > .1).

Noun distractor conditions. The same regions of interest were analyzed as in Experiments 1 and 2.

Subject noun onset latency. The main effect of VerbType was significant (*p* < .05). No other effects were significant.

Preposition production time. The preposition production time was slowed down by the related distractors. Supporting this, the main effect of relatedness was significant (*p* < .0001). No other effects were found to be statistically significant.

Determiner production time. No effects reached significance in this region. However, the effect of Relatedness showed marginally significant effect (*p* = 0.08) that is consistent with Experiment 1 and 2.

Other regions. Just like in Experiment 1 and 2, we compared three other regions in the post-hoc manner. There was a significant effect of Relatedness on the production time of the noun inside subject-modifying adjuncts and the auxiliary verb (β = 0.04, SE = 0.01, |t| = 2.94, uncorrected *p* = 0.003, β = 0.06, SE = 0.03, |t| = 2.16, uncorrected *p* = 0.04, respectively). No effects in any other regions were significant (all uncorrected *p*s > 0.1).

Discussion

The pattern of interference in the noun distractor conditions is consistent with the results of Experiments 1 and 2. However, the verb distractors did not cause significance interference effects anywhere in the sentence as in Experiment 1, though the numerical pattern of the production time of the noun inside adjuncts for the unergative condition were consistent with Experiments 1 and 2. In order to assess the overall pattern across three experiments reported here, we next report the statistical analyses on the datasets that combined data across experiments.

Cross-experiments analyses

Here we report statistical analyses on interference effects observed in all three experiments, using vincentile analyses and peak onset latency analysis.

Unaccusative verb interference

In three experiments, we found diverging statistical patterns of unaccusative verb interference effects in the subject noun onset latency. The null effect of the unaccusative verb distractors in Experiment 3 was expected, because distractor words that are presented 300 ms after picture onset have been shown to not affect speech onset latency of single word production in previous PWI studies any more than noise (Schriefers et al., 1990). If unaccusative verbs are the first word that is planned in the current experiment, it is expected that the distractor that was presented 300 ms after the picture onset does not affect unaccusative verb planning. However, based on this absence of interference effect in Experiment 2, one may argue that the selective verb interference effects in unaccusative sentences are not very robust.

We thus need to assess the consistency between Experiment 1 and 2 by combining the datasets from these experiments. However, considering only mean differences can be misleading. As shown by Momma et al. (2015), production time is highly skewed, just like reaction times in general, and the interference effects in the ePWI task are likely not the effects on the central tendency. We thus conducted a vincentile analysis (Lago et al., 2015; Staub, 2010, Vincent, 1912, among others) on the verb distractor conditions of the combined dataset of Experiment 1 and 2. Since each experiment involved only 12 trials per condition per participant, we used only 4 bins instead of 10 (see Lago et al., 2015). According to previous studies that are similar to the current one (e.g., Momma et al., 2015), we expected that the current interference effects reside in the tail of the production time distribution, that is, in slow responses.

As can be seen in Figure 2, the interference effect was most prominent in the last vincentile only in the unaccusative conditions. In contrast, there is no evidence for verb interference effects in unergative conditions in any of the experiments in any vincentile. To assess this pattern statistically, a mixed effects model with maximal random effects structures for experimental factors (but without the slope for *VincentileRank*) was constructed. Confirming the visual impression, there was a significant 3-way interaction between *Relatedness*, *VerbType*, and *VincentileRank*, based on the result of maximum likelihood ratio test between models with and without the three-way interaction term ($\chi 2$ (3) = 12.82, *p*

= .005). in the combined dataset of Experiment 1 and 2. This three-way interaction suggests that the magnitude of the verb interference effects are bigger for slower responses, selectively in the unaccusative conditions. Thus, the selectivity of the unaccusative verb interference effects on the subject noun onset latency is robust.

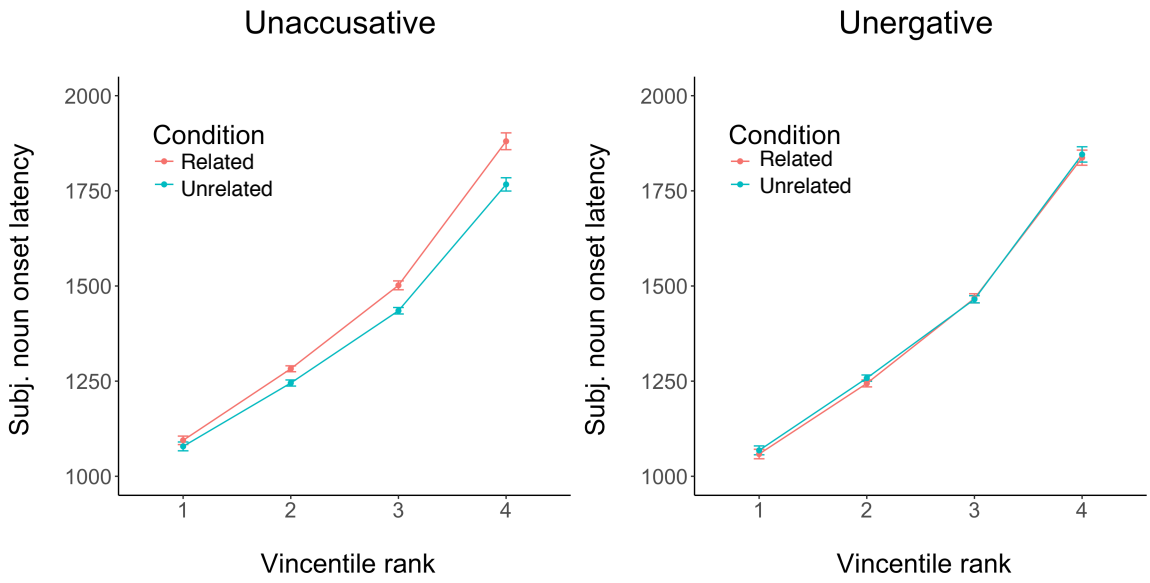

Figure 5. Vincentile plot showing verb interference effects in the unaccusative (left) and unergative (right) conditions, based on the combined dataset of Experiment 1 and 2. The error bars represent standard errors.

Unergative verb interference

We found the unergative verb interference effect in the production time of the noun inside adjuncts only in Experiment 2. However, the numerical pattern was consistent across all experiments. To assess whether the effect in Experiment 2 was spurious or real, we conducted the combined analysis across all three experiments. Since the interference effect is likely not an effect on the central tendency of the production time distribution (see above), we conducted a vincentile analysis again. The vincentile plot is shown in Figure 3. Visual inspection reveals that, in the unergative conditions, the interference effects are most prominent at the right tail of the distribution. In comparison, this pattern is not apparent in the unaccusative conditions. Statistically, in a mixed effects model where *VincentileRank* is treated as a predictor, the two-way interaction between *Relatedness* and *VincentileRank* was significant, based on model comparison with and without the relevant interaction term ($\chi^2(10) = 19.76$, $p = .03$), though there

was no three-way interaction between *Relatedness*, *VincentileRank* and *VerbType* ($\chi^2(3) = 1.40$, $p = .7$). We further analyzed the unergative and unaccusative conditions separately, and found that the two-way interaction is significant in the unergative conditions ($\chi^2(3) = 10.79$, $p = .01$) but not in unaccusative conditions ($\chi^2(3) = 4.1576$, $p = .24$). This pattern suggests that the interference effect is most prominent on slower trials, selectively in unergative sentences.

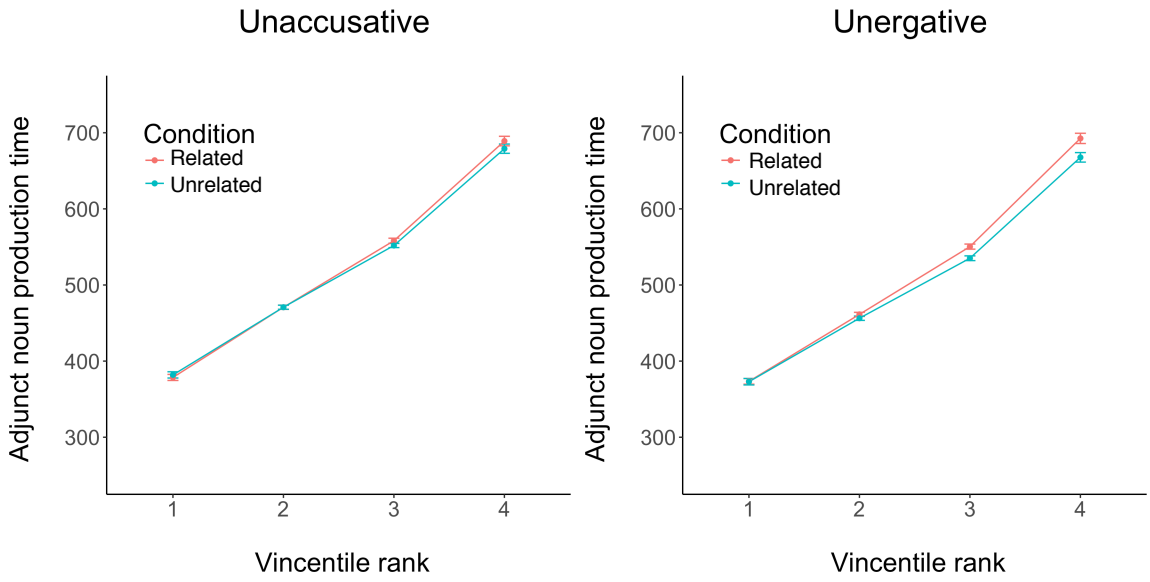

Figure 6. Vincentile plot showing the verb interference effects in the unaccusative (left) and unergative (right) conditions, based on the combined data of Experiment 1, 2 and 3.

Peak interference latency analysis

Combining all the datasets obtained in three experiments, we compared the timing at which speakers show the maximum amount of interference for each distractor type (noun vs. verb). In this comparison, we first converted each measurement (subject noun onset latency and production time of each following word) into z-scores for each measure within each participant. Following this z-score transformation, we identified which region in a sentence showed the largest related - unrelated difference (i.e., the interference effect in z-score). This is done for verb and noun distractor conditions separately. We focused on the subject noun onset latency (Region 1) and the production time of the subject noun head (Region 2), preposition (Region 3), and the determiner in the noun inside the subject-modifying adjuncts (Region 4). We did not include the production time of later regions, because they cannot reflect the noun planning processes (nouns are already spoken at region 5). We then treated the position of peak

interference as an ordinal dependent variable, and examined whether participants overall showed peak verb interference in an earlier region than the peak noun interference. We predicted that speakers, when speaking unaccusative sentences but not when speaking unergative sentences, should show peak verb interference earlier than peak noun interference. This prediction is derived from our main hypothesis that unaccusative verbs are selectively planned earlier than sentence-medial nouns inside subject-modifying adjuncts.

The mean peak interference position across conditions is summarized in Table 7. As can been seen, in the unaccusative conditions, the peak of the verb interference effect was earlier than the peak of the noun interference effect. The opposite patten was found in the unergative conditions. The peak of verb interference effect was later than the peak of noun interference effect.

VerbType	Distractor Type	
	Verb	Noun
Unacc.	2.13 [0.26]	3.13 [0.22]
Unerg.	2.75 [0.27]	2.58 [0.21]

Table 7: Mean position of peak interference effect across condition across all three experiments. Unit = regions.

These data were statistically analyzed using a cumulative link model (Christensen, 2015), a statistical model suitable for analyzing ordinal data. This analysis revealed that the main effect of DistractorType was significant. (β= -1.51, SE = 0.55, |z| = 2.71, p = .007). The main effect of VerbType was not significant, though it was trending towards significance (β = -0.83, SE = 0.51, |z| = 1.63, p = .10). Critically, the interaction between *DistractorType* and *VerbType* was significant (β = 1.89, SE = 0.77, |z| = 2.47, p = .01). In order to better understand this interaction effect, we conducted a planned subset analyses for unaccusative and unergative conditions. This subset analysis suggests that the position of peak verb interference effect is earlier than peak noun interference effect in unaccusative conditions (β = -1.42, SE = 0.56, |z| = 2.53, p = .01). In comparison, the position of peak verb interference effect was not statistically different from peak noun interference effect in the unergative conditions (β = 0.45, SE = 0.53, |z| = 0.85, p = .39).

Furthermore, we computed the position of peak interference for each item, using the same z-score used above. This is analogous to F2 analysis in ANOVA often performed in psycholinguistics, and is useful for assessing whether the effect of interest is consistent across items. This analysis revealed the same pattern as the analysis above. The main effect of *DistractorType* was marginally significant (β = -0.98, SE = 0.51, |z| = 1.93, p = .05), and the main effects *VerbType* was significant (β = -1.62, SE = 0.56, |z| = 2.89, p = .003). Critically, the interaction between *DistractorType* and *VerbType* was significant (β = 2.45, SE = 0.78, |z| = 3.12, p = .002). The planned subset analyses suggest that the position of the peak verb interference effect is earlier than the peak noun interference effect in unaccusative conditions (β = -1.54, SE = 0.57, |z| = 2.70, p = .007), but not in in the unergative conditions (β = 0.82, SE = 0.54, |z| = 1.52, p = .13).

Discussion

The combined analyses yielded three important results. First, the combined vincentile analyses of Experiment 1 and 2 suggest that the unaccusative-unergative contrast in the interference effect on the subject noun onset latency is robust, and is most prominent at the right-tail of the subject noun onset latency distribution. When the verb distractors are presented early enough, speakers delayed the production of the subject noun selectively in unaccusative sentences. This is what is expected based on the hypothesis that verbs are planned selectively before the production of internal argument nouns.

Second, the combined analyses across all three experiments revealed that the unergative interference effects on the production time of the noun inside subject-modifying adjuncts are statistically detectable, though weak. Though it is not entirely clear that the verb interference effect in this region is selective to the unergative sentences, due to the lack of the three way interaction, the relatedness effect seems to be present only in the unergative conditions: the two-way interaction between *Relatedness* and *VincentileRank* was found only in unergative sentences. This suggests that the unergative verbs are selectively retrieved at a later time in the sentence production process, after speakers started saying the noun inside the subject-modifying adjunct phrase.

Finally, the peak latency analysis offered direct evidence that the relative timing of unaccusative verb planning, but not unergative verb planning, is earlier than the planning of the noun inside subject-modifying adjuncts. In conjunction, the across-experiment analyses strongly suggest that the sentence-

final unaccusative verbs, but not sentence-final unergative verbs, are selectively planned before the sentence-medial noun inside subject-modifying adjuncts.

In combination, these analyses all point to the same conclusion: sentence-final unaccusative verbs are planned before the sentence-medial noun inside the subject-modifying adjuncts, which in turn is planned before sentence-final unergative verbs.

General Discussion

The results of the three experiments revealed the following. Experiments 1 and 2 revealed that the verb interference effect was found in the subject noun onset latency, selectively in unaccusative sentences. Experiments 1, 2, and 3 in combination revealed that the verb interference effect was found in the production time of the noun inside adjuncts selectively in unergative sentences (though this effect is weaker). In all experiments, the adjunct noun interference effect was consistently found in the production time of the preposition and also in later regions. The averaged interference effects across all experiments are visualized in Figure 7.

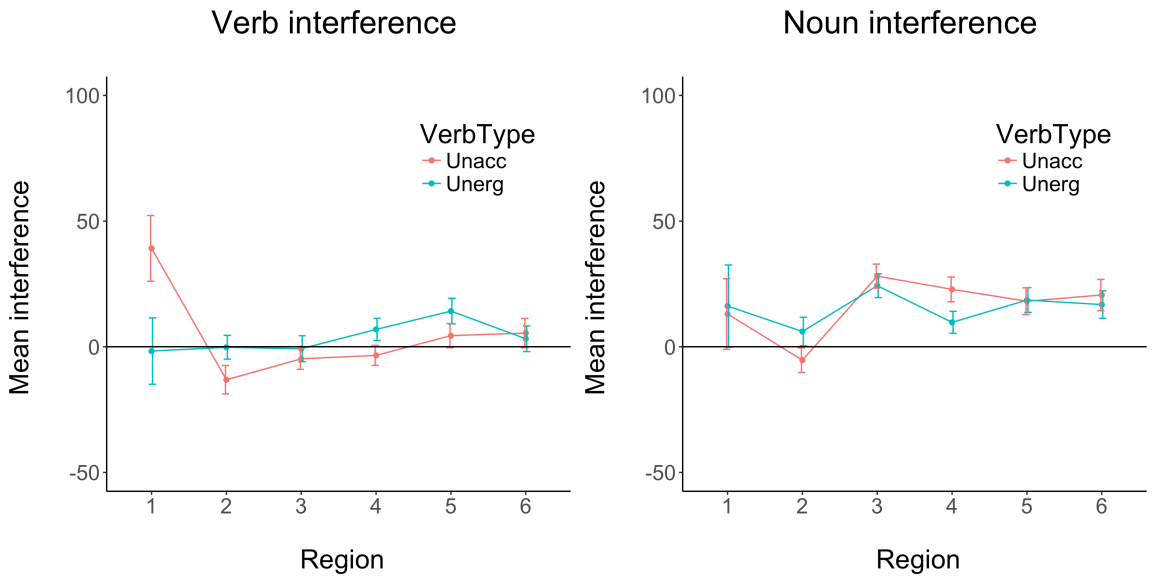

Figure 7. Interference effect (in raw production time) by region by VerbType in the verb distractor condition (left) and the noun distractor condition (right) across all experiments. The error bars represent standard error of the means.

Additionally, the peak interference latency analysis with the combined dataset confirmed that the verb interference effect preceded the noun interference effect, selectively in unaccusative sentences. In

combination, the current experiments suggest that the timing of unaccusative verb planning but not unergative verb planning is temporally earlier than the timing of planning the noun inside subject-modifying adjunct. This is a clear case of non-sequential planning that demonstrates the impact of the abstract argument structure on the time course of sentence planning.

The current study showed that there is a clear discrepancy between how sentence planning unfolds and how the words in a sentence are ordered. This discrepancy has two main theoretical implications, one related to the architecture of grammatical encoding system and the relationship between sentence planning and memory, and the other related to the relationship between abstract linguistic structure and sentence planning.

Architecture of sentence production mechanisms

The main results of the current study are not easily compatible with models of grammatical encoding that reduce sentence production to a sequence of word productions. For example, models that take a simple recurrent neural network architecture as a generator of sentence, such as Chang et al. (2006), is consistent with such a view. Aside from connectionist models, models that are often labeled as *radically incremental* assume that every word is produced immediately upon retrieval (Van Nice et al., 2003; Christiansen & Chater, 2016), and hence reduce the production process to a chain of word productions. Less radically incremental models also tend to assume a limited role of abstract argument structures in planning processes, especially in verb-final constructions or verb final languages (e.g., Kempen & Hoenkamp, 1987; De Smedt, 1990, Iwasaki, 2011). What is shared among many of these models is that the representation of sentence structure is taken as an 'emergent property' of the dynamics of word production processes in sequence, rather than as an active controller of word production dynamics. In other words, sentence structures are the result, not the cause, of a particular word sequence of a sentence. The current results can be taken as a direct case against a strict version of such an idea. The current results instead suggest that abstract argument structures that are not realized in the surface structure of a sentence control what processes take place when in sentence production. We thus argue that the non-surface linguistic structures play important roles in building adequate sentence production models.

Memory and sentence planning

It is important to consider the concern that advance verb planning is costly in terms of memory. The current proposal suggests that speakers plan at the outset a verb that is articulated far downstream in a sentence. If speakers keep the retrieved verb in working memory until it can be spoken, the process of producing unaccusative sentences may be costly. However, the depth of this concern depends on how memory cost is defined in sentence production. It is well-known in the working memory literature that working memory cost is largely determined by the similarity between relevant items. This is true for both capacity-based views (Baddley & Hitch, 1974; Just & Carpenter, 1992; among many others) 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 is 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 Dyke & McElree, 2006, among many others). 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 et al., 2016). For example, Momma et al. (2016) 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 interfere with each other. Critically, when both the to-be-produced and the distractor words do not share the same syntactic category, they do not interfere with each other. Given this category specificity of retrieval interference, buffering verbs as nouns are processed (or vice verse) may not be cognitively costly. If conceptual similarity-based retrieval interference is a primary determinant of processing cost only when items match in syntactic category, non-sequential planning may not be as 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, planning 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

particularly costly. This may be one of the reasons why speakers are able to speak relatively fluently yet grammatically determined planning mandates.

Argument structure and (cross-linguistic) sentence production

The current results also provide additional evidence for the relevance of the linguistic distinction between (alternating) unaccusative and unergative verbs in sentence production. More generally, the current results provide support for the idea that internal arguments selectively require verbs' lemmas to be produced (Momma et al., 2014; 2015; 2017a). As stated in the introduction, Momma and colleagues previously found that verbs are selectively retrieved before the production of object nouns in Japanese, and the subject nouns of passive sentences and unaccusative sentences in English. This advance verb planning is shown to be selective; the evidence for advance verb planning was not found before the production of the subject noun in Japanese active sentences or in English active sentences (regardless of the transitivity). The generalization that cuts across all these studies and the current study is that verbs' lemmas are retrieved selectively in advance of the production of internal arguments. Notably, this generalization is naturally consistent with the linguistic analysis of Kratzer (1996; 2003). Kratzer provided the linguistic analysis in which verbs' internal arguments (logical objects) are true arguments of the verb, but verbs' external arguments are not. Kratzer offered rather extensive linguistic arguments substantiating this claim, which are not reviewed here. Briefly, the idea is that the root of the verb does not assign external argument roles, but internal argument roles, by comparison, are assigned by verb roots themselves. This captures the varieties of linguistic contrast between subjects and objects, including the current findings.

Furthermore, this proposal has the virtue of potentially capturing two cross-linguistic patterns, though both of them are speculative at this point. First, it might explain the distribution of idiosyncratic cases, like the one found in Icelandic. Idiosyncratic cases are morphological cases that are determined by non-predictable properties of verbs. Since idiosyncratic case marking affects the phonological shape of the arguments, it seems logically necessary that speakers plan verbs before they can start speaking arguments that are potentially idiosyncratically case-marked. Interestingly, idiosyncratic cases are sometimes considered to be found only in internal arguments and not external arguments (Marantz, 1981) cross-linguistically. Thus, if the production system plans verbs consistently in advance of internal

arguments (for the independent reasons described above), speakers can ensure that idiosyncratic or systematic cases are correct for internal arguments. This might explain why ideosyncratic cases are largely restricted to verbs' internal arguments.

Second, the current proposal may capture one of the factors that contributes to the distribution of word-orders in the worlds' languages. For simple transitive sentences, there are six logically possible word-order among subjects (S), objects (O) and verbs (V): SOV, SVO, VSO, VOS, OVS, OSV. However, there is clear unevenness in the distribution (Tomlin, 1982). SOV and SVO languages are both very common (roughly 45% and 42% of worlds' languages, respectively), but VSO, VOS, and OVS languages are rare (9%, 3%, 1%, respectively), and OSV languages are extremely rare (below 1%). We argued above that advance verb planning may not be as costly as it is sometimes assumed, but it would not be surprising if speakers still prefer to release planned verbs as soon as possible. For example, when a relative clause is involved in a subject constituent, advance verb planning is predicted to be costly because a relative clause contains another verb, and thus causes within-category similarity-based interference with the planned verb. If the preference to release verbs as soon as possible does exist, it is predicted that speakers should disprefer separating objects and verbs by subjects, selectively when objects precede the verb[2]. This dispreference might penalize OSV languages, skewing the distribution of word-order in worlds' languages. Of course, we do not claim that this is the only factor that makes OSV languages particularly rare, but it might be one of the many contributing factors. For example, there may be a general tendency to not separate O and V because they form a syntactic or semantic unit (verb phrase or predicate). There also may be a general tendency to put subjects before objects, potentially because grammatical functions that are higher in the grammatical hierarchy (Keenen & Comrie, 1977) might be assigned to more accessible words, which tend to be spoken earlier in a sentence (Bock, 1980 ; Branigan et al., 2008; Christiensen & F. Ferreira, 2005; McDonald, Bock & Kelley, 1993; Slevc, 2011). These constrains may conspire to make OSV languages particularly rare.

Locus of the contrast

[2] This argument assumes that objects need not be planned in advance before the verb. We have some evidence that objects need not be planned before the production of verbs (Momma et al. 2013).

At which processing stage does the currently observed unaccusative-unergative contrast arise? Many influential models of sentence production assume that there are two distinct stages in grammatical encoding processes (Bock & Levelt, 1994; Bock & V. Ferreira, 2014; Garrett, 1975; V. Ferreira & Slevc, 2007). In these models, grammatical functions (subject, object, verb, etc.) are assigned to the retrieved lemmas in the first stage, often referred to as the *functional level* of processing. Constituent structures and linear orders are determined at the second stage, often referred to as the *positional level* of processing.

In this architecture, the observed unaccusative-unergative contrast must arise before the functional level, that is, the message level or the interface between the message level and the functional level. This is because unergative and unaccusative subjects do not differ from each other in terms of grammatical function or linear position. Unless we assume the dominance-only sub-stage of the positional level processing (see Pickering & Ferreira, 2008 for a review of counter-evidence), the unaccusative-unergative distinction must be realized at a processing stage earlier than the functional stage of processing. One possibility is the transition from the message level representation to the functional level representation. We might call this *thematic-syntactic interface*. This thematic-syntactic interface somehow encodes the asymmetry between agentive vs. non-agentive arguments. When an argument is agentive, the subject function can be assigned without verb lemmas, because agent arguments are somehow linked to the subject function in some privileged fashion, or because the functional head that is independent from the verb root is sufficient. In contrast, when an argument is non-agentive, verbs' argument structures are necessary to map the argument to a specific grammatical function. This computational necessity may explain the timing contrast between unaccusative and unergative verbs observed in the current experiments.

Potential concerns

As briefly discussed above, it may be that the unaccusative verb interference effects that were found earlier than the noun inside the adjunct is an artifact of our experimental design, namely, how entities are organized in the picture stimulus. Speakers in the current experiments most likely looked at the action picture first given what they have to say first, and action pictures contains the visual information corresponding to the verb. This order of visual information intake might be the factor determining the order in which lemmas are retrieved. However, this does not explain why unergative verb

interference was absent at the subject noun onset, and why it is found later in the sentence instead. If the order of lemma retrieval is simply mirroring the order of visual information intake, this is unexpected. Thus, we argue that the current findings are unlikely to be due to be the how visual information is organized in the picture stimuli.

Methodological contributions

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 the specific sub-processes in sentence production, just like self-paced reading methods can track real-time costs in sentence comprehension. Certainly, visual-world eye-tracking during speaking (e.g., Griffin & Bock, 2000) also allows measurement of what speakers processing in real-time, but the current method nicely complements it. The virtue of the current method is that it allows researchers to locate the processing cost of sentence parts that do not directly correspond to an easily definable region in a picture, such as verbs. Most of the time these types of sentence parts encode relational information, and arguably how speakers encode relational information is the critical but missing part of sentence production theory. We can imagine extending the current paradigm to probe the timing of even more abstract processing, such as the processing of tense, functional heads, and so forth, by carefully choosing the right kinds of distractors. This is not easily possible with existing methods, and thus the current method opens up new opportunities for investigating previously uninvestigated aspects of how sentence planning unfolds over time in speaking.

Conclusion

A critical factor that determines the structure and meaning of a sentence is its predicates' argument structure, which is not necessarily realized in the surface linear structure of a sentence. The current studies demonstrated that the argument structure of predicates directly impacts how sentence plans unfold, in a way that is consistent with independently motivated linguistic theories. Based on the finding that the timing of sentence-final verb planning temporally precedes the timing of a sentence-medial noun, we argue that the hierarchical relationship encoded in verbs' argument structure guides sentence planning. Thus, incorporating grammatical representations like argument structures that are not visible on the surface structure is critical in developing an adequate model of sentence production.

References

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, H. C. (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. *Psychology of Learning and Motivation*, 8, 47-89.

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.

Bock, J. K. & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds), *The Oxford Handbook of Language Production* (pp. 21-46). New York: Oxford University Press.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 467-484.

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

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118(2), 172-189.

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

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.

De Smedt, K. (1990). IPF: An incremental parallel formulator. *Current Research in Natural Language Generation* (pp. 167-192).

De Smedt, K. (1996). Computional models of incremental grammatical encoding. In A. Dijkstra & K. de Smedt (Eds.). *Computational Psycholinguistics: AI and Connectionist Models of Human Language Processing* (pp. 279-307). London: Taylor & Francis.

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.

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.

Ferreira VS, Slevc LR. Grammatical encoding. In: Gareth Gaskell M, (Ed.). *The Oxford Handbook of Psycholinguistics*. Oxford University Press; 2007. pp. 453–469.

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation*, 9, 133-177.

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

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

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

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.

Iwasaki, N. (2010). Incremental sentence production: Observations from elicited speech errors in Japanese. In Yamashita, H., Hirose, Y., & Packard, J. (Eds.). *Processing and producing head-final structures* (pp. 131-151). Springer, Netherlands

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

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1), 63-99.

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

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

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: Johan Rooryck, Lauri Zaring (Eds.). *Phrase Structure and the Lexicon*, (pp. 109-137). Dordrecht: Kluwer.

Kratzer, A. (2003): *The event argument and the semantics of verbs*. Unpublished Manuscript. University of Massachusetts at Amherst. [Downloadable at http://semanticsarchive.net/ Archive/GU1NWM4Z/]

Lago, S., Shalom, D., Sigman, M., Lau, E., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149.

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.

Lashley, K. S. (1951). The problem of serial order in behavior. In *Cerebral mechanisms in behavior* (pp. 112-136).

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: 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.

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

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

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25(2), 188-230.

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., Slevc, L. R., & Phillips, C. (2013) *Advance selection of verbs in head-final language production*. Poster presented at the 26th annual CUNY Conference on Human Sentence Processing, Columbia, SC. March 21-23.

Momma, S., Slevc, L. R., & Phillips, C. (2014) The timing of verb retrieval in English passive and active sentences. *Proceedings of Mental Architecture for Processing and Learning of Language*, Tokyo, Japan.

Momma, S., Slevc, L. R., & Phillips, C. (2015) The timing of verb planning in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 42, pp. 813-824.

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

Momma, S., Slevc, L. R., Buffinton, J. & Phillips, C. (2016). *Similar words compete, but only when they're from the same category*. Poster presented at the 29th annual CUNY Conference on Human Sentence Processing, Gainesville, FL. March 3-5.

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

Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. In *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society* (pp. 157–190). Berkeley, CA: University of California, Berkeley Linguistics Society.

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

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), 1503-1514.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, 114(3), 447-454.

Tomlin, R. S. (1982). *Basic Word Order (RLE Linguistics B: Grammar): Functional Principles* (Vol. 13). Routledge.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language,* 55(2), 157-166.

Van Nice, K. Y. & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics*, 41, 825–849.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, 85(3), 61-69.

Vincent, S. B. (1912). The function of the viborissae in the behavior of the white rat. *Behavioral Monographs*, 1, No. 5.

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.

Watson, J. B. (1920). Is thinking merely action of language mechanisms?(V.). *British Journal of Psychology*, 11(1), 87-104.

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