Squibs and Discussion

ON THE *-OO* SUFFIX OF CAMPBELL'S MONKEYS
Jeremy Kuhn
Sumir Keenan
Kate Arnold
Alban Lemasson

1 Introduction

1.1 Primate Morphology?

Ouattara, Lemasson, and Zuberbühler (2009a,b) make the novel claim that Campbell's monkey alarm calls demonstrate a simple pattern of linguistic morphology. The authors observe that there are at least two distinct alarm calls (called *krak* and *hok*) that are used in two different predatory contexts, and that each may be followed by a low-frequency sound (called *-oo*) that alters the meaning of both calls in predictable ways, allowing contexts with a reduced level of threat. In light of these facts, *-oo* is analyzed as a meaning-bearing, combinatorial unit.

However, the claim that a nonhuman communication system has a combinatorial system (however primitive) is rare in the literature (see section 4 for related patterns) and, indeed, is antithetical to certain

This squib benefited from discussion with Dylan Bumford, Emmanuel Chemla, Didier Demolin, Philippe Schlenker, Dunja Veselinovič, and Charles Yang. The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement N°324115–FRONTSEM (PI: Schlenker) and from the Leverhulme Trust. Research was conducted at Institut d'Etudes Cognitives (ENS), which is supported by grants ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

Authors' affiliations: Jeremy Kuhn: Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, Paris, France

Sumir Keenan: Université de Lyon/Saint-Etienne, Equipe Neuro-Ethologie Sensorielle, ENES/Neuro-PSI, CNRS UMR 9197, Saint-Etienne, France, and School of Psychology and Neuroscience, University of St Andrews, Scotland, UK

Kate Arnold: School of Psychology and Neuroscience, University of St Andrews, Scotland, UK

Alban Lemasson: Université de Rennes 1, Ethologie animale et humaine (UMR 6552–CNRS–Université Caen Normandie), Station Biologique, Paimpont, France

Linguistic Inquiry, Volume 49, Number 1, Winter 2018
169–181

doi: 10.1162/ling_a_00270

claims that structural hierarchy is unique to human language (e.g., Bolhuis et al. 2014). Moreover, it has been noted (Schlenker et al. 2014) that there is redundancy between the apparent semantic contribution of *-oo* and the semantic contribution of a variety of other signal manipulations (e.g., calling rate) that are easiest to explain via noncompositional mechanisms. These facts warrant particular caution when evaluating the pattern as a possible counterexample to generalizations about human language.

Thus, in this squib, we examine the compositional hypothesis further. As counterpoint, we consider a class of more conservative hypotheses in which *-oo* does not itself bear meaning, but instead arises as the side effect of other articulatory processes that noncompositionally affect call meaning. Key to such hypotheses is the premise that *-oo* is articulatorily parasitic on another phonetic process. A major contribution of this squib is thus phonetic: considering the acoustic properties of *-oo*, we conclude that complex calls (*krakoo* and *hokoo*) are produced with two pulses of a single breath-group. Critically, the production of these complex calls requires an additional articulatory gesture and thus an increase in articulatory effort. An increase in articulatory effort would not be expected on an analysis in which *-oo* arises as a phonetic side effect; we accordingly reject these alternative hypotheses, thus strengthening the robustness of the combinatorial analysis.

1.2 Merge as the Putative Defining Feature of Human Language

Bolhuis et al. (2014), following Chomsky (2000), defend the strong hypothesis that the distinguishing feature of human language is the presence of hierarchical syntactic structure. In their words, ''[H]uman language syntax can be characterized via a single operation that takes exactly two (syntactic) elements *a* and *b* and puts them together to form the set {*a*, *b*}.'' This operation, called *Merge* in the Minimalist tradition (Chomsky 2000), allows two elements that are themselves syntactic units to be combined into a complex unit that can serve as the input to another combinatory operation. In human language, this second operation might be a further application of Merge, thus recursively generating structures of arbitrary length.

Of course, the presence of Merge does not guarantee the existence of arbitrarily long sequences; note, for example, that the phrase structure grammar with the terminals {D, N, V} and the rules {S → NP VP, NP → D N, VP → V NP} produces sentences with hierarchical structure, but only generates five-word strings. Relatedly, Rizzi (2016) observes that recursive applications of Merge depend on the presence of a ''temporary workspace,'' short-term memory storage for nonlexical inputs to Merge. Without this workspace, a system can produce binary strings of lexical elements, but cannot store these units for further applications of Merge. For Bolhuis et al. (2014), all nonhuman animal communication systems disallow hierarchy of any depth. As indicated above, the alarm calls of Campbell's monkeys pose a potential counterexample; this is thus the question that we address here.[1]

2 Male Campbell's Monkey Alarm Calls

2.1 Complex Calls

Male Campbell's monkeys (*Cercopithecus campbelli*) produce at least three distinguishable alarm call stems (i.e., calls not followed by *-oo*), called *krak*, *hok*, and *boom*, classifiable both by ear and automatically (Ouattara, Lemasson, and Zuberbühler 2009a,b, Keenan, Lemasson, and Zuberbühler 2013).[2] The *boom* call is unique in several respects (it only appears at the beginning of a call sequence, it involves visible use of superlaryngeal air sacs, it is never suffixed by *-oo*, and it signals the presence of a nonpredatory context); we thus set it aside. Both of the remaining two calls may appear in isolation (''simple calls'': *krak* and *hok*) or followed by the *-oo* suffix (''complex calls'': *krakoo*, *hokoo*). The *-oo* particle never appears in isolation.

Critically, Ouattara, Lemasson, and Zuberbühler (2009a) observe that the addition of *-oo* to a base call alters the meaning in a systematic way, acting to attenuate the force of the call. In their data, *hok* only appears in the presence of eagles (predatory disturbances in the canopy); *hokoo*, too, appears in eagle contexts, but also in contexts of intergroup interaction (nonpredatory disturbances in the canopy). *Krak* only appears in the presence of leopards (predatory disturbances on the ground); *krakoo*, too, appears in leopard contexts, but also in reaction to tree falls, intergroup interaction, and eagles. Schlenker et al. (2014) refine these generalizations with further data. While *hok* is associated with eagles and *krak* with leopards, the association is weaker for *krak* than for *hok*.

Further, on Tiwai Island, which has no leopards, *krak* is used as a general alarm call, including in eagle contexts. For both calls, the complex form is used more widely than the corresponding simple call, including in nonpredatory contexts. These observations motivate an analysis in which the meanings of the complex calls *krakoo* and *hokoo* are compositionally derived. The stem communicates locational information (for Schlenker et al., *hok* indicates an ''upward'' disturbance;

[1] Although Ouattara, Lemasson, and Zuberbühler (2009b) call *-oo* a *suffix* to characterize the call as a minimal meaning-bearing combinatorial unit, this use of terminology should not be interpreted here as committing to any deeper analogy with spoken language, such as postulating *-oo* as a sublexical morpheme or a sentence-final particle. Certainly, either of these phenomena from human language involves syntactic composition; the question here is whether *-oo* does, too.

[2] Ouattara, Lemasson, and Zuberbühler (2009a,b) additionally identify a stem *wak*, but Keenan, Lemasson, and Zuberbühler (2013) provide evidence that it is a variant of the *hok* call.

krak is locationally unspecified); the presence of *-oo* adds information regarding the level of threat.

That *-oo* compositionally modulates threat level is confirmed by the reaction of con- and heterospecifics to natural and artificial stimuli. Ouattara, Lemasson, and Zuberbühler (2009a) report that for Diana monkeys (which associate with Campbell's monkeys), antipredatory behavior occurs only in response to simple calls. This was confirmed experimentally by Coye et al. (2015), who played back recordings of Campbell's *krak* and *krakoo* calls to groups of Diana monkeys. Both male and female Diana monkeys produced more alarm calls in response to *krak* than to *krakoo* sequences. These results held even for calls that were artificially created either by adding *-oo* to *krak* calls or by removing *-oo* from *krakoo* calls.

2.2 Conjunctive Meaning

Even if complex call meanings are compositional, we should ask whether this composition requires anything beyond conjunction. Notably, even in a system without Merge, if call meanings update an overall information state, the effect is equivalent to the conjunction of the individual calls. On the other hand, any other way of combining meanings requires some kind of function application. Thus, if call combination is found to be nonconjunctive, then syntactic combination is a done deal: the semantic facts alone would be evidence for Merge. On the other hand, if call combination is conjunctive, the need for Merge must be decided on the basis of other facts.

In the case at hand, a conjunctive analysis initially appears not to be viable: of note, as discussed above, Ouattara, Lemasson, and Zuberbühler (2009a,b) show that simple calls occur in a *subset* of the situations where their corresponding complex calls occur. Conjunction can only restrict a meaning; thus, the fact that *-oo broadens* the use of the call suggests that the meaning of *-oo* must be nonrestrictive and thus nonconjunctive. However, Schlenker et al. (2014) show that other, ''pragmatic'' factors complicate the picture. First, they conclude that there is an ''alarm parameter'' that decreases over time. Thus, the reason why *hokoo* appears in all the same situations where *hok* appears is that—after *hok* is repeated for a period of time—the degree of alarm decreases to a sufficiently low level for *hokoo* to be used. Distribution of simple and complex calls supports this hypothesis; in the data from Keenan, Lemasson, and Zuberbühler 2013 (3,344 total calls), in sequences that have both *hok*s and *hokoo*s, an average of 87.5% of *hok*s appear before the majority of cosequential *hokoo*s. (A similar trend holds for *krak*/*krakoo*.) Second, Schlenker et al. propose that there is competition between call types, akin to scalar implicatures in spoken language. This explains why *krak* does not generally appear in situations where *krakoo* would be a more precise call.

In the end, Schlenker et al. (2014) are led to propose that the contribution of *-oo* is restrictive but, for technical reasons, nevertheless not conjunctive (see discussion under their (59)). Their final definition states that for any root *R*, *R-oo* is used for weak *R*-type disturbances.

Although the definition provided by Schlenker et al. (2014) is not technically conjunctive,[3] we consider the situation to be sufficiently unresolved that we cannot conclude on the basis of semantics alone that these calls present a case of syntactic Merge.

2.3 A Single Combinatory Unit

Regardless of whether the semantics is conjunctive or not, facts about timing and distribution nevertheless provide strong evidence that complex calls act as single units that serve as the input for further combinatory processes. In particular, both simple and complex calls are organized into call sequences; in the data from Keenan, Lemasson, and Zuberbühler 2013, sequences have a median of 31 calls and a maximum of 131 calls. In the same data, a pause averaging 4.60 s separates the onset of one call from the onset of the next. Call stems are themselves an average of 0.13 s in duration. In this context, *-oo* shows a strikingly different distribution and timing: *-oo* (average duration 0.093 s) always occurs immediately following a call stem, separated only by a short pause averaging 0.060 s (Ouattara, Lemasson, and Zuberbühler 2009b).

Furthermore, for both simple and complex calls, calls are most commonly found in sequences surrounded by the same call-type. Table 1 provides the O/E (observed over expected frequencies) for each bigram in the data from Keenan, Lemasson, and Zuberbühler 2013 (total counts: *krak*, 479; *hok*, 421; *krakoo*, 1,582; *hokoo*, 862). Values greater than one along the diagonal show that repetition of the same call is more likely than chance for all call types. Naturally, the ''grammar'' that derives these sequences of calls will look dramatically different from the grammars of human languages, and there appears to be no motivation to posit sequence generation via Merge. Nevertheless, even if the system that generates these sequences is a probabilistic model conditioned only on the context of utterance, the difference in cooccurrence frequencies between simple and complex calls can only be stated

Table 1
Observed/Expected frequencies for bigrams in data from Keenan, Lemasson, and Zuberbühler 2013

	____ Krak	____ Hok	____ Krakoo	____ Hokoo
Krak ____	6.42	0.28	0.32	0.03
Hok ____	0.21	4.84	0.07	1.00
Krakoo ____	0.32	0.07	1.80	0.31
Hokoo ____	0.12	0.88	0.36	2.74

[3] The definition of *-oo* is nonconjunctive in the same way that the English adjective *tall* is nonconjunctive, since both must be evaluated with respect to a comparison class; *tall for a six-year-old* is different from *tall for a basketball player*.

by reference to the complex calls *krakoo* and *hokoo* as combinatorial units themselves.

Thus, with respect to both timing and cooccurrence frequency, the complex calls *krakoo* and *hokoo* behave as though they are single calls. In conjunction with the semantic facts motivating decomposition of these calls, we thus have a pattern that appears to exemplify the simplest case of Merge: two units combining to form one complex unit.

2.4 Regarding the ''Holistic'' Hypothesis

At this point, there is nevertheless another, entirely noncompositional hypothesis that is perfectly compatible with the data: namely, that all four forms (*krak*, *hok*, *krakoo*, and *hokoo*) are holistically memorized as atomic units. These four forms could be given exactly the same meanings as those derived by Schlenker et al. (2014), which, in conjunction with Schlenker et al.'s pragmatic analysis, will generate identical results. On this analysis, there is no need for Merge; *-oo* would be no more of a syntactic unit than the *cat* of *catapult.*

In fact, such an analysis can be posited for any system that generates a finite set of forms, be it the four-form inventory of Campbell's monkeys or the set of five-word strings mentioned in section 1.2. In any such case, one cannot falsify a holistic analysis based on form-meaning pairings, as the memorization hypothesis is strictly weaker than the compositional alternative. In the general case, several options can mediate between these hypotheses, but these prove difficult to implement in the case at hand. For example, one can test whether a rule generalizes to a novel form (a *wug* test). For Campbell's monkeys, though, no such data presently exist, owing to the prohibitive difficulty of training a group of monkeys to react to a novel call. More feasibly, one can compare the ''syntactic diversity'' of a set of forms with a model in which compositional parts combine independently and interchangeably. Yang (2013) shows that such a model generates a very close fit to the linguistic systems of human adults and children, and notably does not fit the attested productions of the language-trained chimpanzee Nim Chimpsky. In the case at hand, though, this analytic method is confounded by the small size of the data set (four forms), and by the fact that the contexts that gave rise to calls were often induced by researchers and thus were controlled for frequency.[4]

In light of these challenges, we will not try to put the holistic memorization hypothesis to rest, acknowledging that it is indeed a

[4] Despite these limitations, we can of course still calculate the relevant values. If stems and *-oo* combine independently and interchangeably, we use the data from Keenan, Lemasson, and Zuberbühler 2013 to calculate the expected numbers of calls using the product of the marginal probabilities as *krak*, 555; *hok*, 345; *krakoo*, 1,506; *hokoo*, 938. (For example, the expected count for *krakoo* is (krak + krakoo) × (krakoo + hokoo) / total). These are not far off from the attested values of 479, 421, 1,582, 862, which is consistent with the combinatorial story, though the relevance of this result is mitigated by the issues discussed above.

viable alternative to the compositional theory. On the other hand, we note that what is *lost* on the holistic hypothesis is the semantic connection between *hok* and *hokoo* on the one hand and *krak* and *krakoo* on the other; for example, if each form is memorized independently, there is no principled reason why *hok* and *hokoo* should both relate to aerial disturbances. Thus, our goal here will be to consider a second class of noncompositional hypotheses: namely, that *-oo does* systematically modify the call meaning (thus capturing the relation between simple and complex forms), but that it does so via a noncompositional mechanism. This is spelled out in the following sections.

3 Noncompositional Modification?

3.1 What Does It Mean to Bear Meaning?

The analysis of compositionality in section 2.1 is predicated on the assumption that *-oo* itself bears a meaning. For human language, we can say that a morpheme bears a certain meaning if it makes a stable semantic contribution in all contexts; semantic judgments can be gathered from intuitions of native speakers. For primates, conclusions must be drawn from indirect evidence; as we have already shown, this can include both the context of use and the response of conspecifics or heterospecifics to recordings of the signal in question. Together, these demonstrate that a certain proximate factor is responsible for the signal, and that other animals can interpret the signal in order to react appropriately.

However, these diagnostics cannot determine whether the segment itself bears the meaning or whether the meaning is inferred indirectly. To illustrate this point, we can look to cases of ‘‘paralinguistic’’ meaning in human speech. Consider, for example, [+excited], a nonconcatenative modification of the intensity, pitch range, and speed of an acoustic signal, which combines productively with any utterance and adds the (presupposed) semantic content that the speaker is excited. As with monkey alarm calls, this meaning can be deduced from the context of use (heightened emotional state) and from the reactions of conspecifics to the signal manipulation (‘‘Calm down!’’). Intuitively, though, this inference is quite different in origin from the semantic contribution of combinatorial morphemes; whereas morphemes bear meaning themselves, the paralinguistic modification results from the way that the context (the emotional state) directly affects articulation. (In light of section 2.2, it bears noting that the meaning of [+excited] is semantically conjunctive.)

In human communication more generally, the phonetic properties of speech have been shown to vary with respect to communicational and situational demands (Picheny, Durlach, and Braida 1986). Lindblom (1990) describes principles governing these phonetic adaptations in terms of trade-offs: ‘‘hyperarticulated’’ speech is used to facilitate perception in contexts in which communication is harder or more important (e.g., slow and clear speech in a loud environment); otherwise, when perceptual demands are less severe, speech defaults to an articulatorily easier form. Exactly analogous kinds of patterns have been shown to hold for nonhuman communication; for example, Candiotti, Zuberbühler, and Masson (2012b) show that female Diana monkey contact calls display greater interindividual acoustic distinctiveness in dark environments (where caller identification relies on sound) than in bright environments.

3.2 Noncompositional Modification for Campbell's Monkeys

There are independent reasons to think that threat level affects the form of Campbell's monkey alarm calls in a noncompositional manner. As noted above, the presence of *-oo* is associated with contexts with decreased levels of threat. Additionally, though, the level of threat influences Campbell's monkey calls in other ways. First, Lemasson et al. (2010) show that low-threat contexts are correlated with a slower calling rate. Second, Keenan, Lemasson, and Zuberbühler (2013) show that both *hok* and *krak* calls can be divided into subtypes; the less phonetically stereotyped version of each form is correlated with low-threat contexts.

In both of these cases, compositional analyses are difficult to implement. For call rate, the modification applies to a global property of a call sequence. For call distinctiveness, the modification is most easily stated in terms of the phonetic similarity among multiple lexical items. Neither of these situations is conducive to an explanation in terms of local composition. On the other hand, both the variable calling rate and the acoustic variance can be given a simple noncompositional explanation based on environmental-level factors. On a sequence level, increased calling rate may track emotional state, and it increases the redundancy of a signal. On a call level, increased distinctiveness between call types maximizes discriminability, so reduces the chance of communicating the wrong signal in high-threat contexts, where ambiguity can be fatal (Cheney and Seyfarth 1990; see also Arnold and Zuberbühler 2013). Notably, these patterns fit in neatly with the trade-offs discussed in section 3.1; high-threat environments, where communication is more important, induce signals that are perceptually clearer, but that are articulatorily harder to produce.

Given that the meaning contributed by *-oo* may also be expressed through noncompositional mechanisms, we may well ask whether *-oo* itself should be analyzed in noncompositional terms.[5] On such a hypothesis, *krakoo* and *hokoo* are simply phonetic variants of *krak* and *hok.* Since *krakoo* and *hokoo* would then be syntactically atomic, there would be no need for Merge; the hypothesis thus presents a more

[5] Importantly, though, this redundancy with noncompositional mechanisms does not *necessitate* a noncompositional analysis for *-oo.* After all, even in human language, discourse particles and expressives may express content that can equally well be communicated noncompositionally. For example, the emotive content of *fucking* in the sentence *I'm going to the fucking store* will often be redundant with the semantic content communicated noncompositionally by the tone of voice in which the sentence is uttered.

conservative alternative to the compositional analysis of Ouattara, Lemasson, and Zuberbühler (2009a,b) and Schlenker et al. (2014).

If this is indeed the case, we would then expect *-oo* to adhere to the same principles of communication as the other noncompositional indicators of threat level. In particular, as noted earlier, *high*-threat contexts induce clearer and faster signals, at the cost of greater articulatory effort. If the presence of *-oo* in *low*-threat environments arises from the same principles, we make two predictions: first, if *-oo* alters the signal perceptually, it should do so in the opposite direction—a slower or less clear signal; second, being the unmarked form, calls with *-oo* should be articulatorily easier to produce than calls without it. In what follows, we will show that these predictions are not borne out: *-oo* has no effect on perceptual properties, and in fact requires *increased* articulatory effort. These results provide grounds to reject the articulatory hypothesis, thus providing support for a morphological analysis.

3.3 Perceptual Effects of -oo

In principle, *-oo* could affect temporal properties of a call sequence; for example, the time it takes to enunciate *-oo* could have the direct effect of slowing down the calling rate. However, this hypothesis is implausible given the durations involved. The shortest average time between calls reported by Lemasson et al. (2010) is roughly 2 s (in visual eagle scenarios) and ranges up to about 6 s. The smallest significant difference between threats of different levels is approximately 1 s. In contrast, the average length of the *-oo* suffix is less than 0.1 s (Keenan, Lemasson, and Zuberbühler 2013). Thus, the amount of time that it takes to enunciate *-oo* is sufficiently small that its addition alone would not alter the call rate enough to have an effect on the meaning.

Alternatively, *-oo* could affect distinguishability via an acoustic effect on the call stem. Just as coarticulation of an English vowel with a following nasal results in a reduced vowel space (Wright 1986), if Campbell's monkey calls include an *-oo* suffix, then overlap of articulatory gestures could plausibly result in a diminished formant space. However, this possibility, too, is not borne out. Keenan, Lemasson, and Zuberbühler (2013) show that the semantic effect of acoustic subtype can be dissociated from the presence of the *-oo* suffix: holding stem subtype constant, both *krak* variants are used more frequently in response to direct observation of a predator; *krakoo* forms are used more frequently in response to another monkey's predator call. This hypothesis is further falsified by the playback experiments of Coye et al. (2015), in which Diana monkeys showed differential behavior to *krak* and *krakoo*, even when these stimuli were artificially constructed from the stems of the other call.

In summary, *-oo* tracks the threat level of the context, independent of any effect on the call sequence or call stem. There is thus no evidence that the presence of *-oo* affects other perceptual properties of call sequences.

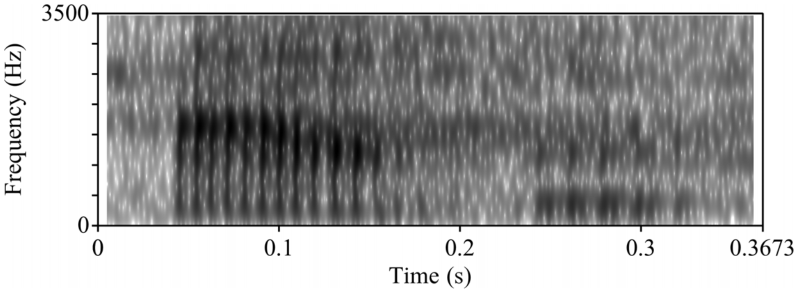

Figure 1
Suspended phonation between *krak* stem and *-oo*. (Recording by KA on Tiwai Island.)

3.4 The Articulatory Production of -oo

Finally, we turn to the articulation of *-oo*. As discussed above, if *-oo* gives rise to meaning indirectly, we would expect it to be associated with a decrease in articulatory effort. This is not borne out. To analyze the articulation of *-oo*, two acoustic facts are of particular import. First, the *-oo* suffix is always separated from the stem by a brief pause, averaging 0.060 s in length (Ouattara, Lemasson, and Zuberbühler 2009b). Second, *-oo* is characterized by a low-pitch band with no higher frequency bands. Figure 1 provides an example spectogram of *krakoo*.

Several sources could account for the pause in phonation: (a) a laryngeal gesture that allows a moment of voicelessness during continued airflow; (b) stopping the airflow by obstruction (as for plosives in human speech); (c) stopping the airflow by a pulmonary gesture. Are any of these explanations compatible with decreased articulatory effort? Possibilities (a) and (b) are not; both require an an additional articulatory gesture that would not naturally occur otherwise. Possibility (c), on the other hand, could result from decreased articulatory effort if the *-oo* suffix is produced via inspiration. As with a human hiccup, the moment at which airflow changes direction would be accompanied by a brief pause of phonation, and continued phonation during inspiration could plausibly result from reduced attention to articulation. The hypothesis that some guenon vocalizations may be produced via inspiration has also been suggested by Riede and Zuberbühler (2003) for Diana monkeys. However, this hypothesis does not square with the acoustic facts. Eklund (2008), in a review of ingressives in both human and animal sound production, characterizes ingressives as acoustically noisier and less cyclic than their egressive counterparts. Here, *-oo* is not notably noisier than the stem, and it has a cyclic period that is visibly distinguishable in both the waveform and the spectrogram. Moreover, if the pause is produced entirely with the source of phonation (with no additional articulatory gesture), then the formants produced by inspiration should be identical to those produced by expiration, since the vocal tract filtering the call remains the same. This is not the case; the spectral bands change dramatically from stem to *-oo*. We conclude that *-oo* is not produced via ingression; the call stem and the *-oo* suffix are produced as part of a single breath-group. Importantly, if the stem and *-oo* are two pulses of a single breath-group, the pause in airflow between the two must result from an additional articulatory gesture; this additional gesture requires an increase in articulatory effort. This fact is at odds with any theory in which *-oo* is an articulatory side effect.

4 Discussion

Whether and how animal calls ''bear meaning'' has been discussed in both the biological literature (Cheney and Seyfarth 1990) and the philosophical literature (Grice 1957, Quine 1973). The present squib extends this discussion to a possible case of hierarchical composition within a nonhuman primate (Ouattara, Lemasson, and Zuberbühler 2009a,b). The topic is of particular importance to recent claims (e.g., Bolhuis et al. 2014) that hierarchical structure is the defining characteristic of human language. We investigated the hypothesis that *-oo* produced by Campbell's monkeys is a combinatorial, meaning-bearing unit, using as counterpoint the hypothesis that the use of *-oo* arises indirectly from articulatory mechanisms. In this spirit, we discussed both the contextual factors that influence call articulation and the articulation of *-oo* itself. Ultimately, we were able to reject a class of hypotheses in which *-oo* is a side effect of articulation, thus bolstering the hypothesis that *-oo* itself carries semantic content.

The argumentation developed here is useful for further investigations into the evolution of hierarchical compositionality in human language, especially as more repertoires of acoustically complex calls are being described for nonhuman primates: see Bouchet et al. 2010 on the *uh* unit in mangabeys; Candiotti, Zuberbühler, and Lemasson 2012a on the A-calls of female Diana monkeys; Bouchet, Blois-Heulin, and Lemasson 2012 on the *i* unit in De Brazza's monkeys; Arnold and Zuberbühler 2006 on the *pyow-hack* sequences of putty-nosed monkeys; and Bene et al. 2012 on various calls among colobus monkeys. While these patterns are a far cry from the complex combinatory processes of human language, detailed examination of them—especially informed by modern linguistic theory—promises to offer insight into the evolution of syntactic and semantic composition in natural language.

References

Arnold, Kate, and Klaus Zuberbühler. 2006. Semantic combinations in primate calls. *Nature* 441:303. doi:10.1038/441303a.

Arnold, Kate, and Klaus Zuberbühler. 2013. Female putty-nosed monkeys use experimentally altered contextual information to disambiguate the cause of male alarm calls. *PLoS ONE* 8.e65660. doi:0.1371/journal.pone.0065660.

Bene, Jean-Claude, Karim Ouattara, Eloi Anderson Bitty, and Inza Kone. 2012. Combination calls in olive colobus monkeys (*Pro-*

colobus verus) in Taï National Park, Côte d'Ivoire. *Journal of Asian Scientific Research* 2:466–477.

Bolhuis, Johan J., Ian Tattersall, Noam Chomsky, and Robert C. Berwick. 2014. How could language have evolved? *PLoS Biology* 12.e1001934. doi:10.1371/journal.pbio.1001934.

Bouchet, Hélène, Catherine Blois-Heulin, and Alban Lemasson. 2012. Age- and sex-specific patterns of vocal behavior in De Brazza's monkeys (*Cercopithecus neglectus*). *American Journal of Primatology* 74:12–28. doi:10.1002/ajp.21002.

Bouchet, Hélène, Anne-Sophie Pellier, Catherine Blois-Heulin, and Alban Lemasson. 2010. Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): A multi-level acoustic analysis. *American Journal of Primatology* 72:360–375. doi:10.1002/ajp.20791.

Candiotti, Agnès, Klaus Zuberbühler, and Alban Lemasson. 2012a. Context-related call combinations in female Diana monkeys. *Animal Cognition* 15:327–339. doi:10.1007/s10071-011-0456-8.

Candiotti, Agnès, Klaus Zuberbühler, and Alban Lemasson. 2012b. Convergence and divergence in Diana monkey vocalizations. *Biology Letters* 8:382–385. doi:10.1098/rsbl.2011.1182.

Cheney, Dorothy, and Richard Seyfarth. 1990. *How monkeys see the world: Inside the mind of another species.* Chicago: University of Chicago Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on Minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Coye, Camille, Karim Ouattara, Klaus Zuberbühler, and Alban Lemasson. 2015. Suffixation influences receivers' behaviour in nonhuman primates. *Proceedings of the Royal Society B* 282(1807). 20150265. doi:10.1098/rspb.2015.0265.

Eklund, Robert. 2008. Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association* 38:235–324. doi:10.1017/S0025100308003563.

Grice, H. P. 1957. Meaning. *The Philosophical Review* 66:377–388.

Keenan, Sumir, Alban Lemasson, and Klaus Zuberbühler. 2013. Graded or discrete? A quantitative analysis of Campbell's monkey alarm calls. *Animal Behavior* 85:109–118. doi:10.1016/j.anbehav.2012.10.014.

Lemasson, Alban, Karim Ouattara, Hélène Bouchet, and Klaus Zuberbühler. 2010. Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften* 97:1023–1027. doi:10.1007/s00114-010-0715-6.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H & H theory. In *Speech production and speech modelling*,

ed. by William J. Hardcastle and Alain Marchal, 403–439. Dordrecht: Kluwer.
Ouattara, Karim, Alban Lemasson, and Klaus Zuberbühler. 2009a. Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences* 106:22026–22031. doi:10.1073/pnas.0908118106.
Ouattara, Karim, Alban Lemasson, and Klaus Zuberbühler. 2009b. Campbell's monkeys use affixation to alter call meaning. *PLoS ONE* 4.e7808. doi:10.1371/journal.pone.0007808.
Picheny, M. A., N. I. Durlach, and L. D. Braida. 1986. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research* 29:434–446. doi:10.1044/jshr.2904.434.
Quine, W. V. O. 1973. On the reasons for the indeterminacy of translation. *Journal of Philosophy* 12:178–183.
Riede, Tobias, and Klaus Zuberbühler. 2003. The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *Journal of the Acoustical Society of America* 114:1132–1142. doi:10.1121/1.1580812.
Rizzi, Luigi. 2016. Monkey morpho-syntax and Merge-based systems. *Theoretical Linguistics* 42:139–145. doi:10.1515/tl-2016-0006.
Schlenker, Philippe, Emmanuel Chemla, Kate Arnold, Alban Lemasson, Karim Ouattara, Sumir Keenan, Claudia Stephan, Robin Ryder, and Klaus Zuberbühler. 2014. Monkey semantics: Two 'dialects' of Campbell's monkey alarm calls. *Linguistics and Philosophy* 37:439–501. doi:10.1007/s10988-014-9155-7.
Wright, James T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In *Experimental phonology*, ed. by John J. Ohala and Jeri J. Jaeger, 45–67. Orlando, FL: Academic Press.
Yang, Charles. 2013. Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences* 110:6324–6327. doi: 10.1073/pnas.1216803110.

Unaccusativity in Sentence Production

Shota Momma
University of California, San Diego
L. Robert Slevc
University of Maryland
Colin Phillips
University of Maryland

Linguistic analyses suggest that there are two types of intransitive verbs: unaccusatives, whose sole argument is a patient or theme (e.g., *fall*), and unergatives, whose sole argument is an agent (e.g., *jump*).[1] Past psycholinguistic experiments suggest that this distinction affects how sentences are processed: for example, it modulates both compre-

[1] Some have claimed that unaccusative verbs that can participate in a transitive alternation are not truly unaccusative verbs, suggesting that the subject of those alternating verbs does not undergo movement and instead is base-generated in the subject position (Haegeman 1994). We acknowledge that some evidence suggests that this distinction may have some processing consequence. However, the evidence is equivocal at best (see Friedmann et al. 2008). Hence, we adopt here the more common view that both alternating and nonalternating verbs are unaccusative verbs (e.g., Perlmutter 1978).

Linguistic Inquiry, Volume 49, Number 1, Winter 2018
181–194

doi: 10.1162/ling_a_00271

hension processes (Bever and Sanz 1997, Friedmann et al. 2008) and production processes (Kegl 1995, Kim 2006, M. Lee and Thompson 2004, J. Lee and Thompson 2011, McAllister et al. 2009). Given this body of evidence, it is reasonable to assume, as we do here, that this distinction is directly relevant to psycholinguistic theorizing. However, especially in production, exactly *how* this distinction affects processing is unknown, beyond the suggestion that unaccusatives somehow involve more complex processing than unergatives (see M. Lee and Thompson 2011). Here we examine how real-time planning processes in production differ for unaccusatives and unergatives. We build on previous studies on lookahead effects in sentence planning that show that verbs are planned before a deep object is uttered but not before a deep subject is uttered (Momma, Slevc, and Phillips 2015, 2016). (We use terms like *deep subject* in a theory-neutral fashion, with no intended commitment to a specific syntactic encoding.) This line of research sheds light on the broader issue of how the theory of argument structure relates to sentence production.

1 Unaccusativity and the Timing of Verb Planning in Sentence Production

The Unaccusative Hypothesis claims that the subject of an unaccusative verb originates as the object of the verb (e.g., Burzio 1986, Perlmutter 1978). Supporting this hypothesis, a range of linguistic phenomena, including *ne*-cliticization and auxiliary selection in Italian (Burzio 1986), English resultatives (Levin and Rappaport Hovav 1995), and possessor datives in Hebrew (Borer and Grodzinsky 1986), suggest that the subjects of unaccusative verbs behave like objects. Reflecting this object-like nature of unaccusative subjects, in transformational theories such as Government-Binding Theory (Chomsky 1981) unaccusative subjects are considered to be base-generated in the object position and moved to the subject position (e.g., Burzio 1986).

Recent studies on the time course of sentence planning suggest that speakers plan verbs (specifically, verbs' lemma representations; Kempen and Huijbers 1983; see Levelt, Roelofs, and Meyer 1999 for a detailed review) before they articulate a deep object, but not before they articulate a deep subject.[2] These studies support an intermediate position between production models that assume that verbs must be planned before all arguments (e.g., Bock and Levelt 1994) and models that assume that no advance verb planning is needed (e.g., Schriefers, Teruel, and Meinshausen 1998). Specifically, we have shown (Momma, Slevc, and

[2] Most prominent models of sentence production (e.g., Garrett 1975, Bock and Levelt 1994, and their more recent variants) divide structure building into two processes: functional (i.e., assignment of grammatical function) and positional (i.e., assignment of hierarchical and/or linear position). Although one might ask which of these processes is responsible for advance verb planning before deep objects, the current findings do not distinguish these alternatives.

Phillips 2016) that verbs are planned before the object noun is uttered but not before the subject noun is uttered in Japanese active sentences. Similarly, we have shown that verbs are planned before subject nouns are uttered in passive but not in active sentences in English (Momma, Slevc, and Phillips 2015). These studies together suggest that verbs are planned before a deep object, regardless of case marking/grammatical function or whether a noncanonical word order is involved. This finding makes an interesting prediction about the production of intransitive sentences. If unaccusative subjects are deep objects, unlike unergative subjects, then unaccusative sentences but not unergative sentences should require advance planning of the verb before the subject noun is articulated. If this prediction is correct, it would show that the subject of unaccusative sentences is processed like a deep object in sentence production, and that split intransitivity directly affects the time course of speaking.

One way to study the timing of verb planning in sentence production is to use the extended picture-word interference paradigm (Meyer 1996, Momma, Slevc, and Phillips 2015, 2016, Schriefers, Teruel, and Meinshausen 1998). In an extended picture-word interference experiment on verb planning, participants describe pictures depicting an action/event in sentential form. At the same time as they see each picture, or slightly before/after, they also see or hear a distractor word. This distractor word is sometimes semantically related to the target verb, which could cause interference in verb processing.[3] This interference can delay verb-related computation, specifically lemma retrieval, which surfaces as a delay in production. (Interference is always measured by comparison with an unrelated distractor word.) The critical question is when this interference effect is observed. If it delays the onset of the subject noun, one can infer that the verb's lemma is planned (i.e., retrieved in advance) before the subject noun is uttered. This pattern would demonstrate that some computation involving the verb's lexical representation is performed before the subject noun is uttered. On the other hand, if an interference effect is observed after the onset of the subject noun, one can infer that the verb is planned after the subject noun is sent for articulation, suggesting that no computation involving the verb's lexical representation is performed before the subject noun is uttered. Therefore, the timing of verb-related interference is informative about what kinds of computations are involved in the production of verbs' arguments. In the current study, we used extended picture-word interference to specifically examine the timing of verb planning in unaccusative and unergative sentences.

[3] By *verb*, we specifically refer to the verb root. We remain agnostic about whether the inflectional component of a verb is planned together with the verb root or not.

2 Experiment

2.1 Participants

Twenty-four native speakers of English participated for either class credit or monetary compensation.

2.2 Materials and Design

Twenty-four pictures of events were selected. Half corresponded to unergative verbs (e.g., *sleep*), and half corresponded to unaccusative verbs (e.g., *float*); see the example in figure 1. The participants of the events corresponding to the unergative verbs were all animate, but half of the participants of the events corresponding to the unaccusative sentences were inanimate. This imbalance in the number of animate subjects in unergative vs. unaccusative conditions was due to the practical difficulty of drawing a picture in which animate participants undergo the action denoted by certain unaccusative verbs (e.g., *melt*). The six animate participants in the unaccusative pictures were exactly matched to the six animate participants in the unergative pictures. This identical subset of nouns was used to test whether any difference between unaccusative and unergative conditions could be attributed solely to the difference in verbs. A full list of target sentences and distractors can be found in appendix 1 (available online at http://www.mitpressjournals.org/doi/suppl/10.1162/ling_a_00271). Also, for each verb, we applied five different tests for unaccusativity, and the results of these tests are reported in appendix 2 (available online at the same address).

For each picture, a semantically related distractor verb was selected from the set of target verbs for the other pictures. These distrac-

Figure 1
Example pictures for unergative sentences (left: *The doctor is sleeping*) and unaccusative sentences (middle: *The doctor is floating*), and example picture with superimposed distractor (right: either a related distractor (e.g., *drown*), an unrelated distractor (e.g., *burn*), or (as shown here) xxxx).

tors always corresponded to one of the other target verbs to maximize the chance of obtaining an interference effect (Roelofs 1992). Semantic relatedness was estimated on the basis of the cosine distance measure from latent semantic analysis (Landauer and Dumais 1997). This measure reflects how close two words are in a multidimensional semantic space with values ranging from 0 to 1. As a point of reference, the clearly related pair *cat* and *dog* receives a value of .36, while the pair *cat* and *desk* receives a value of .01. The average cosine distance between each target verb and its related distractor verb was .31 for the unergative verbs and .35 for the unaccusative verbs, values that did not differ significantly ($p > .5$). Each related distractor word was re-paired with a picture from the same verb type to create the unrelated picture-distractor pairs. The average cosine distance between each target verb and its unrelated distractor verb was .08 for the unergative verbs and .14 for the unaccusative verbs. Unsurprisingly, there was a significant difference in the cosine distance between related and unrelated pairs, both with unergative verb pairs ($p < .001$) and with unaccusative verb pairs ($p < .001$). Importantly, the mean cosine distance between the related and unrelated pairs differed by .23 for unergative verbs and by .21 for unaccusative verbs, so the relatedness manipulation was comparable for the two verb types.

In sum, the current study had a 2×2 within-subjects design, with Verb Type (unergative vs. unaccusative) as a between-items factor and Relatedness (related vs. unrelated) as a within-items factor. There were 24 filler trials where distractors were replaced with ‘‘xxxx’’. In total, there were 72 trials, and participants saw each picture three times: once with a related distractor, once with an unrelated distractor, and once as a filler with xxxx. Note that this number of repetitions is many fewer than in some previous picture-word interference studies (e.g., Schriefers, Meyer, and Levelt 1990).

2.3 Procedure

Participants were first familiarized with the pictures and the target sentences corresponding to each picture, until they felt comfortable with each picture and sentence. This familiarization session was used in order to increase the accuracy and reaction time stability of their production, and it is a standard procedure in picture-naming studies (e.g., Schriefers, Teruel, and Meinshausen 1998).

The experimental session directly followed this familiarization session. Participants were instructed to ignore the written distractor word (in red font) on top of the picture and to describe the picture in sentential form (in present progressive) as soon as they could, except when they saw xxxx as a distractor. When they saw xxxx, they were instructed to not describe the picture and instead press the spacebar. This prevented them from visually ignoring the distractor, thereby ensuring that the distractor words were processed at least to the extent that they could be distinguished from xxxx. On each experimental trial, the participant saw a fixation cross at the center of the screen for 750 ms. Then, a distractor verb (equally often related or unrelated to the target verb) or xxxx appeared at the center of the screen in red font for 500 ms. A short time (150 ms) following the appearance of the distractor, a picture from the studied set appeared on the screen for 1,500 ms. A blank screen, shown for 2,000 ms, separated the trials. Two values were measured manually using Praat (Boersma and Weenink 2015): the time from picture onset to utterance onset, and the duration of the subject noun head and following auxiliary verb *is*. These measures were log-transformed and submitted to statistical analysis.

2.4 Results

The results for onset latencies are summarized in table 1 and figure 2 (left) and the results for duration in table 2 and figure 2 (right). A mixed-effects model with maximal random effects structure in the sense of Barr et al. 2013 was constructed. For the model of subject noun duration, the number of syllables of the noun was included as a predictor.

The model of onset latency revealed a main effect of Relatedness ($\beta = -0.07$, $SE = 0.02$, $|t| = 3.42$, $p < .01$), but no main effect of Verb Type. The interaction between Verb Type and Relatedness was significant ($\beta = 0.06$, $SE = 0.03$, $|t| = 2.08$, $p < .05$). Following planned comparisons revealed that distractor relatedness affected onset latency in unaccusative sentences ($t1(23) = 4.56$, adj. $p < .001$; $t2(11) = 2.92$, adj. $p < .05$) but not in unergative sentences ($t1(23) = 0.11$; adj. $p = .9$; $t2(11) = 0.71$, adj. $p > .9$).

In contrast, the duration measure showed a significant interaction in the opposite direction ($\beta = -0.05$, $SE = 0.02$, $|t| = 2.58$, $p < .01$), with no main effect of Verb Type or Relatedness. The number of syllables in the preverbal noun was also significantly related to duration ($\beta = 0.12$, $SE = 0.03$, $|t| = 4.24$, $p < .001$), but did not interact with Verb Type. Planned comparisons revealed that participants lengthened the utterance of the subject in unergative sentences ($t1(23) = 2.85$, adj. $p < .01$; $t2(11) = 3.18$, adj. $p < .05$) but not in unaccusative sentences ($t1(23) = -0.6$, adj. $p > .9$; $t2(11) = -0.88$, adj. $p > .8$).

Table 1
Mean speech onset latency for each condition, with within-subject standard errors (calculated according to Morey 2008) in square brackets

	Mean latency		Onset interference effect
	Related	Unrelated	(Related-Unrelated)
Unaccusative	1,161 [11]	1,083 [15]	78
Unergative	1,176 [18]	1,171 [16]	5

Table 2
Mean subject noun + auxiliary verb articulation duration for each condition, with within-subject standard errors (calculated according to Morey 2008) in square brackets

	Mean duration		Duration interference effect
	Related	Unrelated	(Related-Unrelated)
Unaccusative	548 [6]	551 [4]	−3
Unergative	584 [7]	553 [4]	31

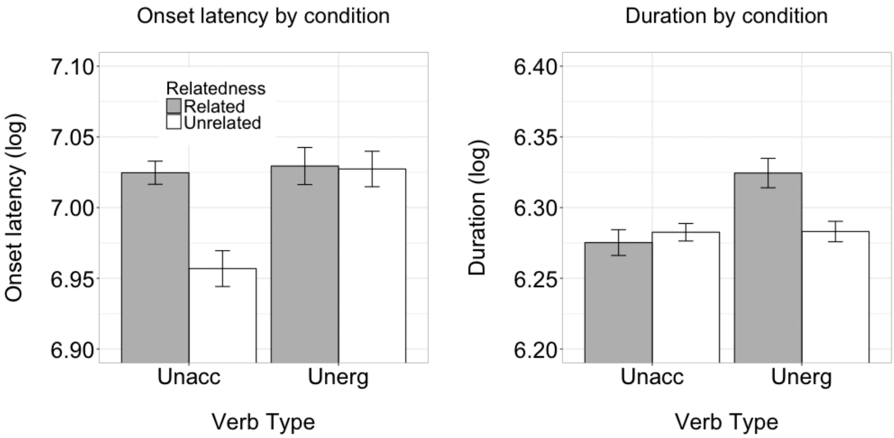

Figure 2
Mean speech onset latency (left) and duration (right) by condition (log-transformed), with associated within-subject standard errors

To ensure that these effects were not due to idiosyncratic differences between items,[4] a secondary analysis examined the subset of 12 items that elicited exactly the same set of animate subject nouns in the unergative and unaccusative conditions. This yielded the same qualitative pattern of results: a greater onset interference effect for unaccusatives than unergatives (60 ms vs. 21 ms), and a greater duration interference effect for unergatives than unaccusatives (23 ms vs. −14 ms).

[4] We conducted this analysis because animacy is sometimes argued to be used for assigning grammatical functions (e.g., Branigan, Pickering, and Tanaka 2008). To the extent that advance verb planning occurs because of grammatical function assignment, it is possible that animacy might change the time course of verb planning (e.g., subject function assignment might require verbs only when the noun head is inanimate).

A potential concern is that any difference in semantic relatedness between the target nouns and distractor verbs across conditions might have confounded the results. For instance, if the target noun *doctor* and the related verb distractor *drown* (unaccusative) are more closely related than the target noun *doctor* and the unrelated verb distractor *burn* (unergative), the effects reported above could reflect an interference effect on noun retrieval rather than on verb retrieval. To address this concern, we conducted a post-hoc analysis on relatedness measures between target nouns and distractor verbs, using cosine distance values derived from latent semantic analysis (Landauer and Dumais 1997). In the unaccusative conditions, the mean cosine distance (*SEM*) was 0.14 (0.03) in the related condition and 0.10 (0.05) in the unrelated condition ($p > .5$). In the unergative conditions, the mean cosine distance was 0.13 (0.03) in the related condition and 0.09 (0.03) in the unrelated condition ($p > .3$). Thus, in both cases, cosine similarity between verb distractors and nouns was 0.04 higher in the related than the unrelated condition. This small difference in relatedness was not statistically reliable ($p > .3$) and, more importantly, was identical for the unaccusative and unergative verb conditions. Furthermore, when these cosine distance values were incorporated into the mixed-effects models for the onset data reported above, semantic relatedness did not interact with Relatedness, Verb Type, or the Relatedness by Verb Type interaction (all ps $> .35$). Thus, lexical relatedness between the target nouns and distractor verbs, at least as reflected in latent semantic analysis, cannot account for the key finding of our study.

A similar potential concern is that semantic relatedness between target nouns (e.g., *doctor*) and target verbs (e.g., *float*/*sleep*) might create some unintended interference, and, if this differs across verb types, might lead to the contrasting pattern of interference reported above. To address this concern, we measured the relatedness (again via cosine distances using latent semantic analysis) between target nouns and target verbs. Target nouns and verbs were somewhat related: the average cosine distance was 0.25 ($SEM = 0.06$) in the unaccusative conditions and 0.20 ($SEM = 0.06$) in the unergative conditions; however, this difference was far from significant ($p > .5$). Additionally, when target noun/verb relatedness was incorporated into the mixed-effect models reported above, it did not interact with the effect of Relatedness between distractor and target verbs ($p > .5$). Thus, we find no evidence that the pattern of semantic interference reported here is attributable to semantic relationships between target nouns and target verbs or between target nouns and distractor verbs.

3 Discussion

Given past evidence that verbs must be planned before a deep object is uttered but not before a deep subject is uttered, the current study examined the processing consequences of the Unaccusative Hypothesis. As predicted, verb interference was found before subject onset in unaccusative sentences, but during subject articulation in unergative

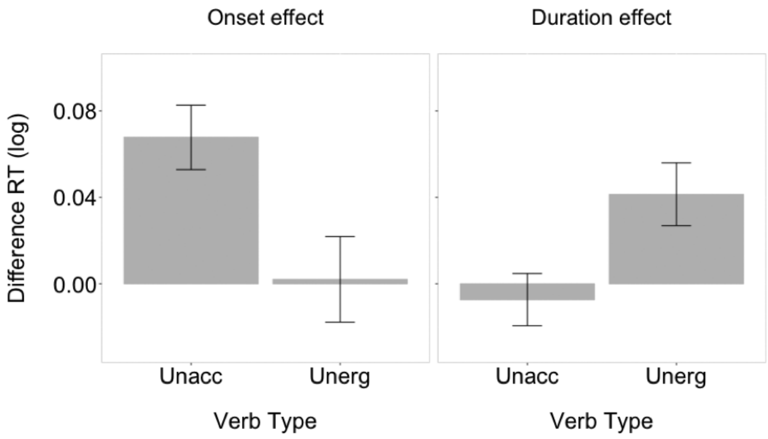

Figure 3
Effect of verb associates (related-unrelated) on subject NP onset latency (left) and duration (right). (RT = response times, in ms)

sentences (figure 3). This suggests that verbs are indeed planned before the utterance of subject nouns in unaccusative sentences, but during the utterance of subject nouns in unergative sentences. Thus, we conclude that the unaccusative/unergative distinction is realized by the producer, and that this distinction is reflected in the selective advance planning of verbs in unaccusative sentences.

The selective advance verb planning in unaccusative sentences is naturally explained if the subject of unaccusative sentences is like an object at some level of representation. This is because advance verb planning has been selectively found before direct object nouns in Japanese sentences and before the subject in English passive sentences, but not before the subject in Japanese sentences or before the subject in English active sentences (Momma, Slevc, and Phillips 2015, 2016).

3.1 Relation to Previous Production Research on Unaccusativity

Unaccusativity in sentence production has been mainly studied in the context of agrammatic aphasia (Kegl 1995, J. Lee and Thompson 2011, M. Lee and Thompson 2004, McAllister et al. 2009, Thompson 2003). These studies tested whether producing unaccusative as compared to unergative sentences leads to increased difficulty (and hence increased error rates). In general, unaccusative sentences do seem to be more difficult than unergative sentences for people with agrammatic aphasia in both natural speech (Kegl 1995) and elicited speech (M. Lee and Thompson 2004), although results are not entirely consistent across speech elicitation methodologies (J. Lee and Thompson 2011). Notably, however, little work has investigated the production of unaccusative vs. unergative sentences by healthy speakers.[5] One exception is a study by Kim (2006), who found that unaccusative sentences prime passive sentences. This suggests that some representation or process is shared between unaccusative and passive sentences, such as a shared movement operation. The current study goes beyond these earlier findings in that it tells specifically how unaccusative and unergative sentences are processed differently in speaking.

Some might argue that the current results do not necessarily show that verbs *must* be planned before unaccusative subject nouns are uttered. Some studies find that perceptual factors modulate structural choice (e.g., passives are more likely to be produced when patient entities are visually cued; Gleitman et al. 2007), at least when the depicted events are difficult to name (van de Velde, Meyer, and Konopka 2014). This type of data is often cited as support for *linearly incremental* production (e.g., Bock and Ferreira 2014), in which there is little or no planning beyond the first constituent or word at sentence onset. By this view, verb planning might not be necessary to produce passive and unaccusative sentence subjects. However, evidence that visual cuing affects structural choice does not necessarily imply an absence of advance planning beyond the first noun phrase. Indeed, Gleitman et al. (2007) also showed that passive utterances took longer to initiate than active utterances, even with perceptual priming of the patient/theme entity. This suggests that speakers do not simply start articulating patient nouns when they are perceptually more available. Instead, even if the first word of the sentence is determined on the basis of its perceptual availability, some additional computations are performed before it is articulated. These computations might well involve advance planning of the verb to establish some syntactic/semantic relation for its internal argument, as proposed here. Thus, the existing evidence is compatible with the strong interpretation of the current results and, until it is disproven, we maintain this strong, more readily falsifiable hypothesis that verbs must be planned before internal arguments can be uttered.

[5] J. Lee and Thompson (2011) also investigated healthy participants' production of unaccusative verbs with a task in which participants produced unergative (*The black dog is barking*) vs. unaccusative (*The black tube is floating*) sentences by rearranging written words on a screen. Using eye-tracking, they found that healthy participants fixated on the noun more than the adjective when producing the noun in unaccusative sentences but not in unergative sentences. They used this finding to argue that unaccusative sentences are processed more sequentially. We do not think, however, that this result is informative about how unaccusative sentences are processed differently from unergative sentences: the observed contrast is not motivated by a model of production, since the unergative sentences showed numerically the same pattern and no interaction analysis was reported, and since the written-word-rearranging task does not engage normal production processes.

3.2 Unaccusativity, Argument Structure, and Sentence Production

The question remains: why is the verb selectively planned before a deep object is uttered? One reasonable explanation is that some computation that is needed to retrieve the lemma of an internal argument noun, or to assign some syntactic/semantic status to deep object nouns after lemma retrieval, depends on the lexical representation of the verb. Although we cannot rule out a contribution from noncanonical thematic-to-syntactic mappings in the current study, we have found advance verb planning before canonical objects in Japanese (Momma, Slevc, and Phillips 2016), which is unlikely to be due to computations involved in establishing a noncanonical mapping between thematic roles and grammatical functions. Also, given that advance verb planning was found in unaccusative sentences in the current study and in English passive sentences (Momma, Slevc, and Phillips 2015), the current effect is not likely due to computations related to accusative case assignment.

The remaining candidates include (a) phrase structure building for the underlying object position and/or (b) assignment of internal argument roles. Both candidates have to do with the role of the argument structure of verbs in sentence production, at either the syntactic or the semantic level. The first possibility relates to the claim that the phrase structure for a VP depends on the lexical properties of the verb (i.e., subcategorization) that are not deducible from the conceptual representation alone (Grimshaw 1990).[6] Under this view, a nonlinguistic message-level representation is sufficient for incorporating external arguments into the sentence structure, but lexical representations of verbs are necessary for incorporating internal arguments into the structure. The second possibility is that the difference in thematic roles between external and internal arguments is the key contrast. This account is in line with Kratzer's (1996) linguistic analysis of argument-predicate relationships, according to which only the internal arguments are true arguments of verbs. Under this view, the assignment of an argument role to the object might selectively require selecting a specific verb, while the assignment of an agent/causer argument role to an external argument might be done independently from the lemma of the verb head. In other words, the nonlinguistic message is sufficient to assign a thematic role to the external argument without the involvement of the verb lemma. In contrast, the verb lemma is necessary to assign thematic roles to the internal arguments.

Future studies should aim to distinguish between these possibilities. For instance, as a reviewer suggests, there are some classes of unergative verbs—namely, *internally caused verbs*—that take a theme

[6] We assume that conceptual representation is equivalent to nonlinguistic representation of a message, which is assumed to be the starting point of production in many production models (e.g., Bock and Levelt 1994).

as their sole argument (e.g., *glow* and *sparkle*). To the extent that these verbs are actually unergative syntactically, comparing the verb planning of theme unergative and unaccusative sentences could reveal whether the effects reported here are due to thematic or syntactic differences. More broadly, the current paradigm can be extended to other classes of verbs that have been the center of attention in argument structure research, such as psych verbs (Belletti and Rizzi 1988), verbs that take instrumental subjects, and other types of unaccusative verbs that do not alternate (Levin and Rappaport Hovav 1995). Given these prospects, we hope that the current study opens up opportunities for developing a line of research that will inform how theories of argument structure relate to theories of sentence production.

References

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Belletti, Adriana, and Luigi Rizzi. 1988. Psych verbs and θ-theory. *Natural Language and Linguistic Theory* 6:291–352.

Bever, Thomas, and Montserrat Sanz. 1997. Empty categories access their antecedents during comprehension: Unaccusatives in Spanish. *Linguistic Inquiry* 28:69–91.

Bock, Kathryn, and Victor S. Ferreira. 2014. Syntactically speaking. In *The Oxford handbook of language production*, ed. by Matthew Goldrick, Victor Ferreira, and Michele Miozzo, 21–46. New York: Oxford University Press.

Bock, Kathryn, and Willem J. M. Levelt. 1994. Language production: Grammatical encoding. In *Handbook of psycholinguistics*, ed. by Morton Gernsbacher, 945–984. Orlando, FL: Academic Press.

Boersma, Paul, and David Weenink. 2015. *Praat: Doing phonetics by computer* (Computer program). Available at http://www.praat.org/.

Borer, Hagit, and Yosef Grodzinsky. 1986. Syntactic cliticization and lexical cliticization: The case of Hebrew dative clitics. In *The syntax of pronominal clitics: Syntax and semantics 19*, ed. by Hagit Borer, 175–215. New York: Academic Press.

Branigan, Holly P., Martin J. Pickering, and Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118:172–189.

Burzio, Luigi. 1986. *Italian syntax.* Dordrecht: Reidel.

Chomsky, Noam. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Friedmann, Naama, Gina Taranto, Lewis P. Shapiro, and David Swinney. 2008. The leaf fell (the leaf): The online processing of unaccusatives. *Linguistic Inquiry* 39:355–377.

Garrett, Merrill. 1975. Syntactic process in sentence production. In *Psychology of learning and motivation.* Vol. 9, *Advances in research and theory*, ed. by Gordon Bower, 133–177. New York: Academic Press.

Gleitman, Lila R., David January, Rebecca Nappa, and John C. Trueswell. 2007. On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language* 57:544–569.

Grimshaw, Jane. 1990. *Argument structure.* Cambridge, MA: MIT Press.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory.* Oxford: Blackwell.

Kegl, Judy. 1995. Levels of representation and units of access relevant to agrammatism. *Brain and Language* 50:151–200.

Kempen, Gerard, and Pieter Huijbers. 1983. The lexicalization process in sentence production and naming: Indirect election of words. *Cognition* 14:185–209.

Kim, Christina. 2006. Structural and thematic information in sentence production. In *NELS 37*, ed. by Emily Elfner and Martin Walkow, 1:59–72. Amherst: University of Massachusetts, Graduate Linguistic Student Association.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. by Johan Rooryck and Laurie Zaring, 109–137. Berlin: Springer.

Landauer, Thomas K., and Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104:211–240.

Lee, Jiyeon, and Cynthia K. Thompson. 2011. Real-time production of unergative and unaccusative sentences in normal and agrammatic speakers: An eyetracking study. *Aphasiology* 25: 813–825.

Lee, Miseon, and Cynthia K. Thompson. 2004. Agrammatic aphasic production and comprehension of unaccusative verbs in sentence contexts. *Journal of Neurolinguistics* 17:315–330.

Levelt, Willem J. M., Ardi Roelofs, and Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22:1–38.

Levin, Beth, and Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax–lexical semantics interface.* Cambridge, MA: MIT Press.

McAllister, Tara, Asaf Bachrach, Gloria Waters, Jennifer Michaud, and David Caplan. 2009. Production and comprehension of unaccusatives in aphasia. *Aphasiology* 23:989–1004.

Meyer, Antje S. 1996. Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language* 35:477–496.

Momma, Shota, L. Robert Slevc, and Colin Phillips. 2015. The timing of verb planning in English active and passive sentence production. Poster presented at the 28th annual CUNY Conference on Human Sentence Processing. Los Angeles, CA.

Momma, Shota, L. Robert Slevc, and Colin Phillips. 2016. The timing of verb planning in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42:813–824.

Morey, Richard. 2008. Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology* 4:61–64.

Perlmutter, David. 1978. Impersonal passives and the Unaccusative Hypothesis. In *Papers from the Fifth Annual Meeting of the Berkeley Linguistics Society*, ed. by Jeri Jaeger, 157–189. Berkeley: University of California, Berkeley Linguistics Society.

Roelofs, Ardi. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42:107–142.

Schriefers, Herbert, Antje S. Meyer, and Willem J. M. Levelt. 1990. Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* 29:86–102.

Schriefers, Herbert, Encarna Teruel, and Raik-Michael Meinshausen. 1998. Producing simple sentences: Results from picture-word interference experiments. *Journal of Memory and Language* 39:609–632.

Thompson, Cynthia K. 2003. Unaccusative verb production in agrammatic aphasia: The argument structure complexity hypothesis. *Journal of Neurolinguistics* 16:151–167.

van de Velde, Maartje, Antje S. Meyer, and Agnieszka E. Konopka. 2014. Message formulation and structural assembly: Describing ‘‘easy’’ and ‘‘hard’’ events with preferred and dispreferred syntactic structures. *Journal of Memory and Language* 71: 124–144.