Effects of Semantic Integration on Phrase and Word Ordering Errors in Production

Amy DiBattista (dibattista.a@husky.neu.edu) and Neal J. Pearlmutter (pearlmutter@neu.edu)

Department of Psychology, 125 NI, Northeastern University, 360 Huntington Avenue
Boston, MA 02115 USA

Abstract

Two experiments investigated the locus of the effects of semantic integration (the conceptual relatedness between utterance constituents) on grammatical encoding during language production. In an ordering-error elicitation paradigm, participants produced descriptions of picture stimuli that varied in degree of integration. For both phrase and word ordering errors, integrated stimuli were more error-prone than unintegrated stimuli. The phrase error results support integration effects on phrase-sized units at the functional level. The word error results suggest integration effects on individual lexical items and provide preliminary evidence for positional level integration effects. Implications of the penetration of semantic integration effects through the functional and positional levels are discussed.

Keywords: semantic integration; grammatical encoding; syntactic planning

The process by which lexical items are assigned to serial positions in a spoken utterance can be understood as a translation process from messages to utterances (Garrett, 1975). During this process, the conceptual representation of an intended message, which is nonlinguistic and unconstrained by time, is translated into a linguistic representation that is constrained by time and by a serial order (Bock, 1987).

In Bock and Levelt's (1994) model, language production begins at the message level, at which point a nonlinguistic representation of a speaker's intended meaning is developed. This conceptual information is delivered to the grammatical encoding stage, which is divided into the functional and positional levels. First, at the functional level, *lemmas*, which contain semantic and grammatical class information but no phonological information, are selected and assigned to syntactic roles appropriate to convey the message. Functional-level information is passed to the positional level, at which point *lexemes*, which carry a lexical item's phonological plan, are assigned to sentence frame positions. Next, at the phonological encoding stage, the phonology of the utterance is realized and delivered to the articulation systems.

The division of labor in Bock and Levelt's (1994) model among conceptual, syntactic, and phonological processing allows for investigation of the mechanisms by which information at a higher level influences the level(s) below it. One such inquiry concerns the effects of conceptual properties on syntactic planning, at the grammatical encoding levels. Conceptual properties such as prototypicality (Kelly, Bock, & Keil, 1986; Onishi, Murphy, & Bock, 2008), imageability (Bock & Warren, 1985), and animacy (Bock, Loebell, & Morey, 1992) have been demonstrated to influence assignment of lexical items to syntactic roles or serial positions in utterances.

Semantic integration, defined by Solomon and Pearlmutter (2004) as the degree of conceptual relatedness between utter-

ance constituents to be planned, is another conceptual property suggested to influence syntactic planning. In previous research, subject–verb agreement errors (Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004) and ordering errors (Pearlmutter & Solomon, 2007) were more likely for highly integrated than for less integrated stimuli. Prosodic analyses have demonstrated shorter temporal separation between highly integrated constituents than between less integrated constituents (Gillespie, Pearlmutter, & Shattuck-Hufnagel, 2010). However, the level(s) of grammatical encoding at which these effects arise has not yet been determined. The experiments described below investigate two related inquiries concerning the mechanism of semantic integration. The first is whether integration affects the ordering of full phrases, individual lexical items, or both. The second is how far into the sentence production system integration penetrates, which can be assessed by examining phrase (functional level) and word (potentially positional level) errors.

Semantic Integration and Ordering Errors

Semantic integration (Solomon & Pearlmutter, 2004) arises from the relationship between lexical items, determined by the elements of the phrases in which they occur. It differs from a simple semantic relationship. For example, in the phrase *the ketchup or the mustard*, the nouns *ketchup* and *mustard* are semantically similar. However, the other elements of the phrase, in particular the conjunction *or*, provide no further information about any relationship that may exist between these two nouns. On the other hand, in the phrase *the bracelet made of silver*, *made of* creates a relationship between *bracelet* and *silver*, that of an object and its material.

Semantic integration has been suggested to affect syntax through changes to the timing of planning of utterance constituents (Gillespie et al., 2010; Pearlmutter & Solomon, 2007; Solomon & Pearlmutter, 2004). Lexical items with a higher degree of integration (e.g., the two nouns in *the pizza with the yummy toppings*, which describes an integration-eliciting object–attribute relationship) are planned closer together in time than lexical items with a lower degree of integration (e.g., the two nouns in *the pizza with the tasty beverages*, which describes an accompaniment relationship).

Pearlmutter and Solomon (2007) hypothesized that simultaneity of planning of two constituents would result in ordering errors, such as the exchange error *Although murder is a form of suicide...* (intended: *Although suicide is a form of murder...*; Garrett, 1975). They demonstrated integration effects on ordering error rates in a series of experiments in which participants verbally described grayscale line drawings

Table 1: Correct description and exchange error examples for Experiments 1 and 2.

Response Type	Response
Correct Description	the green spot on the blue apple
Phrase Exchange	the blue apple on the green spot
Noun Exchange	the green apple on the blue spot
Adjective Exchange	the blue spot on the green apple

Figure 1: Picture stimuli used in Experiments 1 and 2. See explanation in the text.

depicting, for example, an apple with a spot on it. Expected responses differed in degree of integration and description preference. Across experiments, ordering errors were more likely for the integrated cases than for the unintegrated cases.

Exchange errors like *the apple on the spot* (intended: *the spot on the apple*) are exchanges of either the phrases *the spot* and *the apple* or the words *spot* and *apple*. Both phrase errors and word errors arise during grammatical encoding, according to Bock and Levelt's (1994) model, so the errors elicited were clearly assignable to the grammatical encoding stage. However, these errors were ambiguous between phrase and word errors. Therefore, it was not possible to determine which specific level of grammatical encoding was affected: the functional level, at which phrase errors and word errors involving lemmas arise; or the positional level, at which word errors involving lexemes arise. The answer to this important question would extend what is currently known about integration and conceptual-property effects on syntactic planning.

The goal of the current experiments was to determine how far into grammatical encoding semantic integration penetrates, by determining if it affects the likelihood of phrase and word ordering errors. Stimuli that could elicit differentiable word and phrase errors were necessary. To this end, we modified Pearlmutter and Solomon's (2007) stimuli by applying color. The words for the colors added adjectives to the expected responses. Ordering errors involving full phrases, nouns, or adjectives were therefore possible.

Table 1 shows examples of an expected correct response for the blue apple/green spot picture, shown in the left panel of Figure 1, and its associated phrase, noun, and adjective exchange errors. The correct description, with the nouns in the order *spot* then *apple*, and the adjectives modifying their intended nouns, would be *the green spot on the blue apple*. A phrase exchange would involve the full phrases *the green spot* and *the blue apple* exchanging. A noun exchange would involve only the nouns *spot* and *apple* swapping, while an adjective exchange would involve only the adjectives *green* and *blue*.

Per Bock and Levelt's (1994) model, phrase errors could be assessed to determine if they varied with level of integration, as a test of whether integration affects the functional level. Word errors under Bock and Levelt's model can arise at the functional level (lemma exchanges) or at the positional level (lexeme exchanges). Integration effects on lemmas would

suggest that integration affects functional-level processing, strengthening evidence from phrase error effects, if found. Integration effects on lexemes would point to integration effects at the positional level. However, word errors elicited in the current design would not be differentiable between lemma or lexeme involvement. Integration effects on word errors, regardless of their lemma–lexeme classification, would suggest the influence of integration on individual lexical items, independent of the other components of their phrases.

Experiment 1

Method

Participants   One hundred undergraduates at Northeastern University participated for credit in an introductory psychology course. Two participants were excluded, one for misunderstanding the instructions, and one for a technical malfunction that prevented completion of the experiment.

Materials and Design   Forty-two stimulus pictures were used, identical to Pearlmutter and Solomon's (2007) picture stimuli, but with color applied. Six of these pictures were practice items, and 36 were experimental items. Each was a line drawing of two common objects, or of an object with an identified attribute. One color was applied per object or attribute, for a total of two colors per picture. Eighteen of the experimental pictures were integrated, and eighteen were unintegrated, as determined by Pearlmutter and Solomon's prior norming. Figure 1 shows examples of an integrated (left panel; a blue apple with a green spot on it) and an unintegrated (right panel; a red shelf above a green sink) picture.

Each picture was describable in at least two ways, determined by Pearlmutter and Solomon's (2007) prior norming. The preferred and unpreferred descriptions determined the preferred and unpreferred prepositions, or linking words, used in describing each picture. The two nouns in each description could occur in two different orders determined by the preferred or unpreferred linking word.

For the picture of the blue apple with the green spot, the preferred linking word was *on*. This linking word determined the correct order of nouns for this item: *spot* then *apple*. The correct description was therefore *the green spot on the blue apple*. Given the unpreferred linking word *with*, the correct order of the nouns was *apple* then *spot*. The correct unpreferred description of the same picture was thus *the blue apple with the green spot*.

For the picture of the red shelf above the green sink, the

preferred linking word was *above*, which determined the correct noun order to be *shelf* then *sink*. The correct preferred description was thus *the red shelf above the green sink*. The unpreferred linking word was *below*, which determined the correct noun order to be *sink* then *shelf*. The correct unpreferred description was thus *the green sink under the red shelf*.

In addition to its preposition-based descriptions, each item could be described using the conjunction *and* as the linking word. In this case, each noun could appear on either side of the linking word and result in an utterance that made sense, as in *the green spot and the blue apple* or *the blue apple and the green spot*. Each picture could therefore appear in three preference conditions: preferred, unpreferred, and flexible.

Participants completed a familiarization phase before the test phase. Two training lists were generated from the grayscale versions of the practice and experimental items. Each item in the familiarization phase was composed of a picture (e.g., an apple with a spot on it) and the two nouns that named the important parts of each picture (e.g., *apple* and *spot*). The two lists differed only in the left–right presentation order of the two nouns below the picture.

A given participant saw one of the two training lists, six practice items, and one of the 18 test lists, the latter created by crossing the three preference conditions with the six color combinations for each picture.

Procedure   The procedure was identical to that used by Pearlmutter and Solomon (2007), with the exception of the colored stimuli in the test phase and changes to the instructions to reflect the presence of the colored stimuli and the need to use color adjectives in the responses.

The experiment had three parts: a two-part familiarization phase and a test phase. During the first part of the familiarization phase, participants saw grayscale versions of each picture one at a time with the noun labels below. During the second part, they saw the grayscale pictures one at a time without the labels, and were instructed to say the noun labels aloud. The noun labels appeared below the picture 4000 ms later.

During the test phase, the six colored practice items appeared first, one at a time in a fixed order. The 36 colored experimental items appeared after the practice items, one at a time in random order.

Individual test trials began with a fixation cross centered on the computer screen for 1000 ms. The fixation cross was then replaced by a picture, centered where the fixation cross had been. After 1000 ms, an asterisk appeared, centered below the picture. Another 1000 ms after that, the asterisk was replaced by a linking word. Measurement of speech onset time began when the linking word appeared. The participant then produced a description. The picture and the word disappeared simultaneously 3000 ms after the linking word appeared, and the voice key was deactivated.

Apparatus   Stimuli were presented using a PowerMacintosh G3 running PsyScope version 1.2.5 software (Cohen, MacWhinney, Flatt, & Provost, 1993) with a PsyScope but-

ton box. Verbal responses were recorded for later transcription and coding with a Shure SM58 microphone, a Mackie 1202-VLZ Pro mixer/preamplifier, and an Alesis Masterlink ML-9600 compact disc recorder.

Coding   Responses were coded as corrects, ordering errors, or unusable responses. Ordering errors were coded orthogonally to reflect grammatical category involved (nouns, adjectives, phrases, or ambiguous among grammatical categories) and movement type (exchange, anticipation, perseveration, shift, or ambiguous among movement types). Corrects and ordering errors were coded for the presence of filled and unfilled pauses and additions.

Results

Both responses and production latencies were recorded. Only the error results will be presented here, as they address the main experimental question.

Seven participants were excluded because of a high number of individual unusable trials. Four were excluded for very fast production latencies (less than 300 ms). One more was excluded because of a lack of usable responses in the integrated preferred condition. Individual unusable trials from the remaining subjects (14% of the total trials) were then excluded. Data from the remaining 86 participants were used in the error analyses.

Out of the 2671 total usable responses, 2501 (94%) were corrects. Of the 170 ordering errors, 137 (81%) were phrase errors, 8 (5%) were word errors, and the remaining 25 (15%) were ambiguous with regard to grammatical category. Of the error responses, 166 (98%) were exchange errors, and four (2%) were other movement types.

Because there were so few word errors and so few movement error types apart from exchanges, analyses were conducted on phrase exchange errors only. The main analyses were conducted on percentages of error responses, calculated as the number of phrase exchanges divided by the total of phrase exchanges and corrects. Responses were from the preferred and unpreferred conditions only, as phrase exchanges were not possible in the flexible cases, and were included regardless of dysfluencies. Figure 2 shows untransformed error rates for phrase errors by integration and preference.

The analyses were conducted using weighted linear regression on empirical-logit transformed percentages (Barr, 2008), one with subjects as the random factor (t_1), and one with items as the random factor (t_2). Integration, preference, and their interaction were the fixed effects. A sum coding scheme, with unintegrated and preferred as base conditions, was used.

Table 2 shows the weighted linear regression estimates. There was a main effect of integration: More errors occurred in the integrated than in the unintegrated condition. There was a main effect of preference: More errors occurred in the unpreferred than in the preferred condition. The integration $\times$ preference interaction was marginal by participants only.

Corresponding regressions were also conducted with dysfluent responses excluded. The main effects did not change,
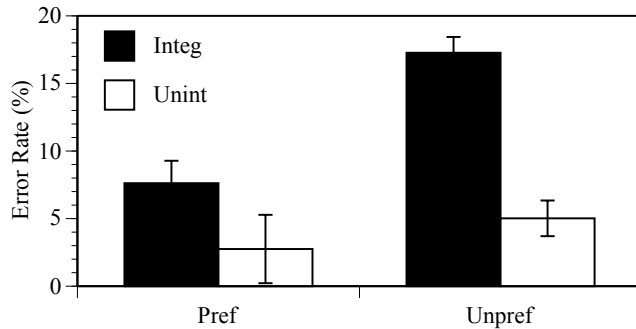
Figure 2: Experiment 1 grand mean phrase error rates by integration and preference condition. Error bars show ± 1 *SEM*, calculated from the analysis by participants.

Table 2: Experiment 1 weighted empirical logit linear regression error rate results.

Effect	By Participants			By Items		
	β	SE	t_1	β	SE	t_2
Sem: Int	.30	.05	5.81**	.55	.08	6.87**
Pref: Unp	.21	.05	4.10**	.30	.07	4.37*
Sem: Int × Pref: Unp	.08	.05	1.56†	.09	.07	1.30

Note. Sem=Integration, Int=Integrated, Pref=Preference, Unp=Unpreferred. **$p < .001$. *$p < .05$. †$p < .10$.

but the interaction was significant by participants. The interaction appeared to result from a very high error rate in the integrated unpreferred condition, which caused a greater difference between the integrated and unintegrated cases in the unpreferred than in the preferred condition.[1]

Discussion

The goal of Experiment 1 was to determine the level(s) of grammatical encoding affected by semantic integration. Integration effects were seen in phrase exchange errors, with errors more likely in the integrated than in the unintegrated condition. As phrase exchanges arise at the functional level, these results show that integration influences grammatical encoding at least this far into the sentence production system. Experiment 1 also replicated Pearlmutter and Solomon's (2007) findings with responses of a more complex syntactic structure (by virtue of the added adjectives).

Experiment 1 did not generate a large enough number of word ordering errors to analyze. Therefore, it was not possible to evaluate integration effects on word errors.

---

[1]Corresponding ANOVAs were conducted on error counts and on untransformed, arcsine-transformed, and empirical-logit-transformed percentages. There were small differences from the regressions in significance levels, but the overall patterns of effects were the same.

The small number of word errors may be a reflection of the nature of ordering errors: Word errors may be more rare than phrase errors in general. Alternatively, the small number of word errors may have been a result of the task, which allowed participants to view the picture for a relatively long time (2000 ms) before the linking word appeared. This long opportunity to plan may have allowed for over-planning of the full phrases as units, thereby driving down the number of word errors. In Experiment 2, we explored this paradigm-specific explanation for the small number of word errors.

Experiment 2

In Experiment 2, we attempted to reduce over-planning of full noun phrases by decreasing the viewing time for the colored version of the picture, with the intention of eliciting an analyzable body of word errors.

Method

Participants   127 Northeastern University students participated in this experiment for credit in an introductory psychology course. One participant was excluded because he was not a native English speaker, and two more participants were excluded due to technical malfunctions.

Materials and Design   The materials and design were identical to Experiment 1.

Procedure   The procedure was identical to Experiment 1 except for the timing of the appearance of the colored picture in the test phase. In the test phase, the grayscale version of the picture appeared first and was replaced by the colored version simultaneously with the appearance of the linking word. The experimental instructions were modified to reflect this.

Coding   Coding was the same as in Experiment 1.

Results

Twenty of the participants were excluded because of a high number of unusable trials. Individual unusable trials from the remaining subjects (9% of the total trials) were then excluded. Data from the remaining 104 participants were used in the error analyses.

Out of the 3421 total usable responses, 3205 (94%) were corrects. Of the 216 ordering errors, 140 (65%) were phrase errors, 50 (23%) were word errors, and 26 (12%) were ambiguous between phrase and word errors. Regarding movement type, 207 (96%) of the errors were exchanges, and nine (4%) were other movement types.

Figure 3 shows untransformed error rates, separately for phrase and word errors, as a function of integration and preference. Error rates for phrase errors were computed out of the total of exchanges and corrects, as in Experiment 1. Error rates for word errors were computed out of the total of ordering errors and corrects from all three preference conditions, as word errors were possible in the flexible condition.

Phrase and word errors were analyzed separately using weighted linear regressions as in Experiment 1, except that
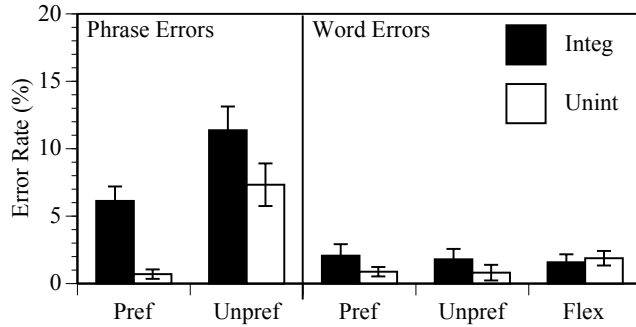
Figure 3: Experiment 2 grand mean phrase error rates (left panel) and word error rates (right panel) by integration and preference condition. Error bars show ± 1 *SEM*, calculated from the analysis by participants.

Table 3: Experiment 2 weighted empirical logit linear regression error rate results for phrase errors.

Effect	By Participants			By Items		
	β	SE	t_1	β	SE	t_2
Sem: Int	.21	.04	4.81**	.53	.09	6.23**
Pref: Unp	.26	.04	5.93**	.50	.08	6.42**
Sem: Int × Pref: Unp	-.07	.04	-1.61*	-.25	.08	-3.26*

Note. Sem=Integration, Int=Integrated, Pref=Preference, Unp=Unpreferred. **$p < .001$. *$p < .05$.

Table 4: Experiment 2 weighted empirical logit linear regression error rate results for word errors.

Effect	By Participants			By Items		
	β	SE	t_1	β	SE	t_2
Sem: Int	.04	.03	1.62*	.12	.10	1.15†
Pref: Prfd	-.02	.04	-.64	.02	.09	.20
Pref: Unp	.05	.04	1.36	-.03	.09	-.33
Sem: Int × Pref: Prfd	.05	.04	1.30†	.15	.09	1.67
Sem: Int × Pref: Unp	.00	.04	.05	.03	.09	.38

Note. Sem=Integration, Int=Integrated, Pref=Preference, Prfd=Preferred, Unp=Unpreferred. *$p < .05$. †$p < .10$.

the flexible condition was the preference factor base for the sum coding in the word error analysis.

Table 3 shows weighted linear regression estimates for phrase errors. Errors were significantly more likely for the integrated than for the unintegrated condition and for the unpreferred than for the preferred condition. The integration × preference interaction was also significant: The difference between the integrated and unintegrated cases was larger in the preferred than in the unpreferred condition.

When dysfluent responses were excluded, the interaction was nonsignificant by participants and marginal by items, but the general patterns of results were the same.

Table 4 shows weighted linear regression estimates for word errors. Errors were significantly more likely in the integrated than in the unintegrated condition. This effect was marginal by items. Only the flexible condition was significantly different (greater) than the grand mean. There were two marginal components of the interaction: the integrated flexible condition was marginally greater than the grand mean by participants and items; and the integrated preferred condition was marginally greater than the grand mean by participants only.

Excluding dysfluent responses, the integration effect was nonsignificant. The flexible condition was greater than the grand mean by participants; this was marginal by items. The unpreferred condition was greater than the grand mean by participants. The preferred condition was marginally greater than the grand mean by items. Some components of the interaction were significant, but differed between the by-participants and by-items analysis. The integrated flexible condition was marginally greater than the grand mean by items only. By participants, the integrated preferred condition was significantly greater, and the integrated unpreferred condition was marginally less, than the grand mean.

Discussion

In Experiment 2, integration effects on phrase errors were found, replicating Experiment 1's findings and further supporting integration effects at the functional level. Experiment 2 expands on Experiment 1 by demonstrating integration effects on word error rates. The word-error effects were in the same direction as the phrase-error effects: more errors in the integrated than in the unintegrated condition.

These word errors may be lexeme misorderings (assignment of lexical items to incorrect slots in the sentence frame), and so are potentially assignable to the positional level. The integration effects for these errors therefore represent preliminary evidence that integration penetrates grammatical encoding past the functional level, to the positional level, and thus that integration may influence both syntactic role assignment and serial position assignment.

These word errors may also be lemma misorderings (assignment of single lexical items to incorrect syntactic roles). If this is the case, then this experiment provides further support for functional-level integration effects, but not for positional-level effects. Establishing integration effects on functional-level word errors would contribute to our knowledge about semantic integration in a different way. Phrase-error-only effects would have suggested that integration cannot affect an independent lexical item (e.g., a noun) without affecting the other components of the surrounding phrase (e.g., the determiner and adjective). The fact that effects sur-

faced for word errors suggests that integration can affect independent lexical items. Two nouns or adjectives in an utterance can overlap temporally—and exchange—without involving the other components of the phrase.

General Discussion

The experiments reported in this paper investigated semantic integration, with the goal of determining how far into the sentence production system it can penetrate. Taken together, Experiments 1 and 2 suggest that integration affects ordering of phrase-level constituents and individual lexical items. These experiments present strong support for functional-level integration effects, along with preliminary evidence for positional-level integration effects.

Assumptions about the sentence production system determine to what extent a conceptual relationship like semantic integration would be expected to affect the functional and positional levels. As information in Bock and Levelt's (1994) model flows top-down, and each level is solely affected by the level above it, integration effects at the functional level are a reasonable outcome, whereas effects at the positional level are more surprising.

An alternative to consider is a more liberal view of information flow through the system. Vigliocco and Hartsuiker (2002) reviewed research strongly supporting maximal information flow from one level to the next. In the maximal input case, it may be possible for conceptual information delivered to the functional level to be sent further to the positional level, rather than remaining encapsulated at the functional level. The positional level could then be affected by integration, resulting in simultaneity of lexeme planning.

Yet another alternative is single-stage grammatical encoding, as posited by Ferreira and Humphreys (2001). During single-stage grammatical encoding, lexical items would be organized according to syntactic category information, specified for lexical and morphological information, and assigned to serial positions. If grammatical encoding is a single stage, conceptual information would need only flow one level down to affect syntactic role assignment and serial order position. This fits well with our results, and would eliminate the need to differentiate between lemma and lexeme involvement.

The word errors found in these experiments are not firmly attributable to the positional level, as it is not possible to determine if they involved lemmas or lexemes. Further research is needed to confidently claim integration effects on positional-level processing. Future experiments aim to determine semantic integration's full scope with effects on ordering errors more definitely attributable to the positional level, such as stranding errors or shifts; and to explore integration effects on non-error production.

Acknowledgments

References

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Bock, J. K. (1987). Co-ordinating words and syntax in speech plans. In A. W. Ellis (Ed.), *Progress in the psychology of language*. London: Lawrence Erlbaum.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego: Academic Press.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*, 150–171.

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

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.

Ferreira, V. S., & Humphreys, K. R. (2001). Syntactic influences on lexical and morphological processing in language production. *Journal of Memory and Language*, *44*, 52–80.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation*. New York: Academic Press.

Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject–verb agreement production. *Cognition*, *118*, 377–397.

Gillespie, M., Pearlmutter, N. J., & Shattuck-Hufnagel, S. (2010). Prosodic phrasing reflects planning processes in sentence production. Talk presented at the 23rd Annual CUNY Conference on Human Sentence Processing, New York, NY.

Kelly, M. H., Bock, J. K., & Keil, F. C. (1986). Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory and Language*, *25*, 59–74.

Onishi, K. H., Murphy, G. L., & Bock, K. (2008). Prototypicality in sentence production. *Cognitive Psychology*, *56*, 103–141.

Pearlmutter, N. J., & Solomon, E. S. (2007). Semantic integration and competition versus incrementality in planning complex noun phrases. Talk presented at the 20th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

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

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound and syntax in sentence production. *Psychological Bulletin*, *128*, 442–474.