Elizabeth Dovenberg
Junior/Senior Honors Thesis
Supervisor: Professor Neal Pearlmutter
17 April 2015

Semantic Relatedness and Its Effect on the Timing of Language Production

Abstract

The timing of language production planning is sensitive to semantic integration, as shown in experiments involving different classes of speech errors (e.g. Solomon & Pearlmutter, 2004; Pearlmutter & Solomon, 2007). Semantically integrated elements are activated with significant overlap in the production planning mechanism, leading to shortened word duration in integrated preambles (Gillespie, Pearlmutter, & Shattuck-Hufnagel, 2014). Semantic relatedness, the connection of two words based on their meanings, differs from integration in that it is contextually independent. The current experiment compared effects of semantic integration and semantic relatedness on word duration measures from a spoken sentence-completion task (Penta, 2014; Experiment 1). Results showed differing effects of integration and relatedness; integration supported previous findings, and effects of relatedness fell into a gradient beginning with significantly delayed word duration and ending with significantly expedited word duration.

The ultimate goal of language production is to generate an utterance to be understood by a conversational partner. Speakers must conceptualize abstract, non-linguistic, non-temporal thoughts, organize grammatical and linguistic knowledge into language-specific patterns, and articulate the formulated message in a comprehendible way. Although speakers carry out this task with relative ease, sentence production is a complex process governed by various grammatical constraints and cognitive resource limitations.

Models of sentence production attempt to explain the processes involved in converting nonlinguistic thoughts into spoken language. It is largely accepted that models of language production include message, grammatical, and phonological components (Bock, 1995; Griffin & Ferreira, 2006; Bock & Levelt, 1994; Dell, 1986; Gillespie, Pearlmutter, & Shattuck-Hufnagel, 2013). Bock and Levelt (1994) introduced a language production model with an initial message-level component, in which speakers generate an abstract, prelinguistic representation of the intended message. The preconceived message proceeds to a grammatical encoding stage that incorporates morphological, grammatical, and semantic information while assigning constituents positions and syntactic roles. There is some disagreement throughout the literature regarding the finer details of the grammatical encoding stage.

Several researchers argue that this stage of language production, in which content words are accessed, is further divided into functional and positional levels. Lemmas selected during the functional level specify semantic information and grammatical function; lexemes selected in the positional level determine the content word's phonological characteristics (Bock & Levelt, 1994; Schriefers, Meyer, & Levelt, 1990; Bock, 1995) Others propose a single grammatical encoding stage that does not include separate lexeme and lemma selection processes; a conceptual message would go through a single phase to determine grammatical role and word position (DiBattista & Pearlmutter, 2011; add more). Speech error studies have shown evidence favoring both single-stage models of grammatical encoding and two-stage models (cf. Schriefers, Meyer, & Levelt, 1990; DiBattista & Pearlmutter, 2011).

Researchers do widely agree that the grammatical component precedes a phonological processing stage, during which the phonological structure of the intended message is organized (Bock, 1995; Bock & Levelt, 1994). The speaker then uses articulatory processes to actually produce the utterance. Although speakers usually construct sentences effectively and efficiently, each stage of language production is affected by processing constraints, resource demands, and cognitive limitations. Errors in sentence production are studied extensively to identify points in production planning and execution processes in which speakers are subjected to the most trouble.

Language Production Errors

Language production research often focuses on speech errors, which are considered to be consequences, rather than failures, of the production system's productiveness and flexibility (Dell, 1986). Mistakes in spoken language are commonly believed to provide productive insight into underlying production planning mechanisms, perhaps more so than successful, errorless speech.

In order to produce a grammatical English sentence, a speaker needs to select, organize, and appropriately convey (among other things) syntactic and morphological information. For example, English predominately uses subject-verb-object word order, and marks specific grammatical features with inflectional morphemes, such as *–s* indicating a plural noun. An utterance like *The boys were playing hockey* suggests that the speaker is talking about multiple people. The speaker must utilize conceptual-level representations (the idea of several boys playing hockey) and apply language-specific grammatical encoding rules (including a plural noun marker and appropriate verb inflection, as well as assigning elements the correct word positions) in order to produce a grammatical English sentence.

Sentence completion and picture description tasks are commonly-used experimental paradigms that allow researchers to examine whether speakers are able to produce sentences with appropriate grammatical inflections and correct serial word order in complex linguistic contexts. Recent research has investigated agreement and exchange errors resulting from manipulations of semantic integration. Agreement and exchange errors are both syntactic in nature and therefore arise in the grammatical encoding stage of language production planning, but they reflect different components of the system: agreement errors, marking of grammatical inflection, and exchange errors, serial word positioning. The following section first describes semantic integration and then reports evidence of integration effects on speech error patterns.

Semantic Integration

Semantic integration, the degree to which pieces of an utterance (e.g., head and local noun) are conceptually related, focuses on the relationship between the head and local noun within a very particular context (Solomon and Pearlmutter, 2004; Penta, 2014). Solomon and Pearlmutter (2004) introduced the term *semantic integration* as they investigated the syntactic planning system in language production. The phrase *the ketchup or the mustard* is considered semantically unintegrated because, although *ketchup* and *mustard* are semantically related to each other (i.e., have similar meanings), the conjunction *or* allows no further relationship to be established between those nouns. On the other hand, *the bracelet made of silver* is a semantically integrated phrase, since *made of* signifies a very tight relationship between *bracelet* and *silver* (DiBattista & Pearlmutter, 2011).

Agreement Errors and Semantic Integration

A typical experimental paradigm, the subject-verb agreement error elicitation task, involves linguistic properties at both the message and grammatical levels, and can demonstrate how they interact during language production. Subject-verb agreement in English is grammatically inflexible in that nouns and verbs must match in number no matter how are apart they appear in a sentence (e.g., *The guys who stayed up all night playing high-stakes poker are*) (Bock & Miller, 1991); more intervening material between the subject and verb strains agreement computation processes (cite).

In experiments investigating subject-verb agreement, sentence preambles are presented to participants, who are instructed to read or repeat the fragments and form sentence completions. Preambles contain complex noun phrases that manipulate the plurality of the head and local (distractor) nouns (e.g., *The key(s) to the cabinet(s)*…(Bock & Miller, 1991)). The most widely observed trend in this research is one of unequal distribution of production errors: a head-local mismatch effect. Subject-verb agreement errors occur much more often in preambles with a singular head noun and plural distractor noun (e.g., *The key to the cabinets *were…*) than the other way around (e.g., *The keys to the cabinet *was…*) (Bock & Miller, 1991; Bock & Cutting, 1992; Bock & Eberhard, 1993; Eberhard, 1997; Barker, Nicol, & Garrett, 2001; Solomon & Pearlmutter, 2004). Experiments using pictorial rather than written or spoken stimuli have also shown the head-local mismatch effect (Gillespie & Pearlmutter, 2011a).

Solomon and Pearlmutter (2004) conducted the first research investigating influences of semantic integration on subject-verb agreement, hypothesizing that results would support either a serial-based or parallel-activation based system of language production planning. Five experiments utilized a subject-verb agreement error elicitation paradigm in the form of spoken sentence completion tasks. The degree of semantic integration was manipulated by using different prepositions in a NP PP structure, specifically, *of* in tightly integrated stimuli, and *with* in unintegrated stimuli. Head nouns were singular, while local nouns varied in plurality. Mismatch conditions, in which the plurality of the nouns differed (e.g., *the drawing of/with the flowers*), were predicted to produce more subject-verb agreement errors.

Solomon and Pearlmutter (2004) hypothesized that a serial-based system would produce fewer agreement errors in highly integrated *of*-PP phrases; since memory-dependent processes are involved in serial planning systems, moving from a tightly linked NP to a PP would be relatively simple. Parallel-activation systems consider multiple representations simultaneously, hypothetically causing highly integrated phrases to be planned at the same time, or with significant overlap, leading to more agreement errors. In other words, the simultaneity in timing of production planning causes the later noun and its plural marker to be active at the same time as the singular head noun, increasing errors in subject-verb agreement computation due to an influence of the local noun's plurality.

Experiments 1-4 showed clear and uniform affects of semantic integration on subject-verb agreement. More errors occurred in plural local noun conditions, as well as within tightly integrated phrases. Subject-verb agreement errors consistently occurred in those contexts, even after several experimental stimuli adjustments were made: altering the semantic relationship of the highly integrated preposition (Experiment 2); altering the argumenthood status of the PP (*of* was replaced with *for*; Experiment 3); creating preambles with the same preposition but semantically different PPs (*the pizza with the yummy toppings* versus *the pizza with the tasty beverages*; Experiment 4).

Solomon and Pearlmutter (2004) concluded that the degree of semantic integration within a phrase regularly forecasts the likelihood of subject-verb agreement errors, and semantic integration can affect the timing of planning of phrasal elements. Results support a parallel-activation based system of production planning that prepares pieces of tightly integrated phrases with significant overlap.

Exchange Errors and Semantic Integration

Another kind of speech error studied in the production literature is the exchange error. Two elements switch roles in exchange errors, such as *Although murder is a form of suicide* when the intended utterance is *Although suicide is a form of murder* (Garrett, 1975). This type of error can occur at nearly every linguistic level, including sound, morpheme, word and phrase switches (Garrett 1975). Exchange errors involving entire words consistently obey particular rules, albeit with a few exceptions. For example, word switches occur within the same grammatical category (i.e., nouns exchange with other nouns) and within open-class items such as nouns, verbs and adjectives, though are sometimes observed in closed-class items like prepositions (Garrett, 1975; Garrett, 1989). Word exchange errors in production occur between or within the same surface clause or phrase, or spanned across various amounts of intervening material. Example (1) lists examples of various word exchange errors (Garrett, 1975; Pearlmutter & Solomon, 2007).

(1)

a. *I have to fill up the gas with car* (Intended: *car with gas*)

b. *I have had both of enough* (Intended: *enough of both*)

c. *The apple on the spot* (Intended: *spot on the apple*)

d. *Who would else like one?* (Intended: *who else would like one?*)

e. *I want to give my bath a hot back* (Intended: *back a hot bath*)

Although exchange errors likely result from inaccurate allocation of lexical entries to grammatical positions during language production planning, the reason for their occurrence is less straightforward than a speaker simply changing his or her mind during speech production. Word exchange errors are the focus of the discussion here, but exchange errors of all kinds provide evidence for simultaneous activation of various grammatical elements during production planning, supporting predictions made by spreading-activation accounts of language production (Garrett, 1975; Garrett, 1989).

Exchange error research examines phonological and syntactic relationships between words in exchange errors; a few experiments to date have investigated various semantic relations between them. Pearlmutter and Solomon (2007) conducted several experiments investigating the effect of semantic integration on ordering errors, predicting that highly integrated elements (those that are planned with simultaneity in a parallel-activation based production planning system) would increase the risk of ordering errors. Participants were visually presented with and then described simple grayscale drawings of objects varying in semantic integration (e.g., integrated: an apple with a spot on it; unintegrated: a sink with a shelf above it). Higher rates of ordering errors were observed in integrated conditions in all experiments, supporting the hypothesis that closely planned elements cause more disruption in grammatical encoding processes, leading to more word order errors.

Although Pearlmutter and Solomon (2007) demonstrated a straightforward effect of semantic integration on ordering errors, it was unclear whether the exchange errors occurred at the phrase- or word-level. An error like *the apple on the spot* (intended: *the spot on the apple*) could involve the swapping of phrases, *the apple* and *the spot*, or swapping of single words, *apple* and *spot.* DiBattista and Pearlmutter (2011) used a nearly identical experimental design and procedure, but added color to the original Pearlmutter and Solomon (2007) stimuli. Visual presentation of, for example, a blue apple with a green spot on it, could elicit exchange errors of phrases (*the blue apple on the green spot*), nouns (*the green apple on the blue spot*), or adjectives (*the blue spot on the green apple*).

The DiBattista and Pearlmutter (2011) experimental design allowed distinctions to be made regarding whether integration affected entire phrases or individual lexical items, thus suggesting whether those affects occur at the functional and/or positional level of grammatical encoding. Two experiments where participants completed picture-description tasks presented similar results: word exchange errors were more likely when the picture showed highly integrated objects. Ninety-eight percent and 97% of error responses in Experiments 1 and 2, respectively, were exchange errors, supporting the hypothesis that highly integrated elements are activated with significant overlap during production planning processes (Solomon & Pearlmutter, 2007). Overall, results suggested that semantic integration influences at least the functional level of grammatical encoding (as shown by word order errors involving lemmas), and potentially the positional level of language production (DiBattista & Pearlmutter, 2011).

Semantic integration has been shown to disrupt the production planning mechanism, resulting in subject-verb agreement errors (specifically, the head-local mismatch) and ordering errors (specifically, exchange errors). Recent evidence has also demonstrated the influence of integration on word durations in sentence completion tasks. Gillespie, Pearlmutter, and Shattuck-Hufnagel (2013) conducted the first experiments illustrating, using prosodic analyses, that words between highly integrated sentential elements were produced more quickly than words between unintegrated items. Prosody is able to reflect ongoing language planning processes; for example, the availability of sentential elements greatly influences prosodic boundaries, with more predictable words uttered more quickly than unpredictable ones (cite).

Their two experiments used sentence completion tasks to investigate whether semantic integration influences timing of syntactic planning. Temporal durations of words in between highly integrated elements were predicted to be shorter than those in unintegrated conditions. Stimuli in Experiment 2 were originally from Solomon & Pearlmutter (2004, Experiment 4), and included head NP PP local NP sentence preambles. Word durations were coded using ToBI and Praat software, and linear mixed-effect models that included speech rate, phonological environment, and accessibility/predictability control predictors were used to predict word duration measures.

Results from Experiment 2 showed effects of semantic integration on temporal separation of words. Words in between highly integrated elements were shorter in duration (i.e., said faster) than in unintegrated conditions. Semantic integration effects were only significant in word positions in between the head and local nouns, suggesting that effects are present when highly integrated elements are simultaneously activated in the production planning system.

Overall, results in Gillespie et al.'s Experiment 2 matched the predictions and results from Solomon & Pearlmutter (2004), providing further evidence for the influence of semantic integration on the timing of (syntactic) production planning.

Semantic Relatedness

Semantic relatedness concerns the connection of two words based on their meanings. Two words' meanings can be related to synonymy (e.g., *package–parcel, trash–garbage*), taxonomic category (e.g., *dog–cat, bread–cake*), part-whole relation (e.g., *button–shirt, page-book*), or contextual relation (e.g., *bread–butter, dog–bone)*, among other associations (e.g., Meyer & Schvaneveldt, 1971; Penta, 2014). Semantic relatedness differs from semantic integration in that the former considers two words in isolation, detached from any specific utterance, while the latter is dependent on linguistic context (Penta, 2014).

Priming paradigms are common experimental methods to examine effects of semantic relation on language production processes. Evidence has shown that semantic relatedness can impact word retrieval (Vigliocco, Lauer, Damian, & Levelt, 2002), syntactic organization, and sentence structure (Bock, 1986). Two studies to date have used sentence-completion tasks to analyze influences of semantic relatedness on agreement computation.

Barker, Nicol, and Garrett (2001) used three sentence-completion tasks to investigate effects of semantic overlap (relatedness), noun animacy, and plausibility on subject-verb number agreement in production. Their Experiment 2 manipulated the meaning relationships of nouns within a complex NP, and examined whether high semantic overlap increased the likelihood of agreement errors. Semantic overlap was high in preambles like *The canoe by the sailboat(s)* and low in preambles like *The canoe by the cabin(s)* (Barker et al., 2001). Results showed that subject-verb agreement errors occurred almost twice as often in utterances with high semantic overlap, suggesting that number agreement processes in language production planning are sensitive to not only syntactic information, but also semantic features of preamble elements (Barker, Nicol, & Garrett, 2001; Eberhard, 1997).

Penta (2014) conducted two experiments that independently tinkered semantic integration and semantic relatedness in order to examine the relationship between them. Experiment 1 included a sentence-completion task in which local noun number, semantic relatedness, and semantic integration were all manipulated. Relatedness was manipulated by using head and local nouns that were either very similar in meaning or semantically unrelated, and semantic integration was determined by individual prepositions, with integrated phrases linking head and local NPs with *for* or *with*, and unintegrated utterances, with *near* or *by*. Previous norming studies provided the baseline ratings for related/unrelated and integrated/unintegrated conditions (see Penta, 2014 for descriptions of integration, relatedness, and association norming procedures). Sentence preambles were of the form head NP PP local NP. Across conditions, all nouns, adjectives, and prepositions in the preambles were matched for frequency, length (in characters, phonemes, and syllables), plausibility, and animacy. An example of a trial set is seen in (2) below.

(2)

a. The necklace with the colorful diamond(s) (Related – Integrated)

b. The necklace near the colorful diamond(s) (Related – Unintegrated)

c. The necklace with the colorful feather(s) (Unrelated – Integrated)

d. The necklace near the colorful feather(s) (Unrelated – Unintegrated)

Stimuli were presented on a computer screen to subjects, who were asked to read aloud and complete each sentence as quickly as possible, although no strict time constraints were enforced. Subjects' agreement error rates were calculated and compared across the singular and plural local noun number conditions. The experimental procedure and scoring system were identical to that in Solomon and Pearlmutter (2004).

Participants produced more subject-verb agreement errors in conditions where the head noun was singular and local noun was plural, providing further evidence for the mismatch effect described earlier. There was no mismatch effect observed in integrated conditions, not replicating previous research (e.g., Solomon & Pearlmutter 2004). More agreement errors were seen in related than unrelated conditions, similar to results found by Barker et al. (2001).

An issue in experiments that manipulate semantic links between words is that defining meaning relationships is not straightforward (Penta, 2014). Semantic relation is typically specified in terms of shared characteristics (e.g., words belonging to the same taxonomic category) or of the degree of association between two elements; words are associated if the presence of one increases the likelihood of observing the other (cite). Another issue in this line of research is that semantic relationships often overlap with associative relationships: *cat* and *dog* belong to the same taxonomic category, but the presence of one (*cat*) quickly activates the other (*dog*). Since both semantic relation and association influence and activate these nouns, and probably speed up production of the second-retrieved word (cite), experiment results may not be able to tell apart effects of semantic overlap and effects of association; this type of problem occurs in priming studies in particular (cite).

Past research involving subject-verb agreement errors (Gillespie & Pearlmutter, 2011b; Solomon and Pearlmutter, 2004), phrase and word exchange errors (DiBattista & Pearlmutter, 2011), and ordering errors (Solomon, 2004; Pearlmutter & Solomon, 2007) demonstrated that highly integrated elements caused more difficulties in language production planning than unintegrated elements. Research has also shown the influence of semantic integration on word durations and the timing of production planning (Gillespie et al., 2013). Experiments manipulating semantic relatedness illustrated the increased likelihood of subject-verb agreement errors in utterances containing elements with very similar meanings (Barker et al., 2001; Penta, 2014).

Experiment

The current experiment investigates how the degree of overlap in language production planning and changes in the timing of production are affected by semantic integration and semantic relatedness, separately or in conjunction with each other. If semantic relatedness influences word durations similarly to semantic integration, intervening material in semantically related utterances will be shorter in duration compared to words in semantically unrelated utterances (Gillespie et al., 2013). The magnitude of the potential interaction between integration and relatedness is unknown.

Method

The current experiment analyzed the original data from Penta (2014), measuring word durations in utterances from his Experiment 1. Thirty-two of the most fluent participants, those with the fewest disfluencies in sentence completions, were randomly chosen for coding. Each subject produced 24 sentence completions for a total of 768 preamble trials. Items in which subjects did not complete the sentence or produced errors within the preamble were not coded and were excluded from analysis. No individual experimental item was excluded more than three times in any given condition across all participants. The final analyses included 698 trials from the original data set.

One trained student used Praat software to hand-code boundaries of all words in each preamble, as well as the following three words of the sentence completion. Each preamble and completion was labeled according to part of speech and the exact word uttered (see Table 1).

Table 1: Coding key across conditions

Word	Label
The	D1
necklace	N1
with	P1
the	D2
colorful	A1
diamonds	N2
was	C1
very	C2
expensive	C3

Note: D = determiner; N = noun; P = preposition;
A = adjective; C = completion word.

TextGrid files outputted from Praat were read into R (cite version?), where mixed-effect models predicted word duration measure as a function of various control predictors and random and fixed effects. The variables included in these models were largely similar to those in the Gillespie et al. (2013) regression models. Fixed effects included semantic integration, semantic relatedness, association, and plausibility. Random effects included subject and item intercepts and random slopes to account for by-subject and by-item variability.

Control predictors, or nuisance factors, included length of the current word (*w*) in characters, phonemes, and syllables, lengths in phonemes of the preamble's first noun (N1), first preposition (P1), first adjective (A1), and second noun (N2), and frequencies of the current word (*w*), previous word (*w* - 1) and following word (*w* + 1). Three non-residualized regression models were used to predict word duration:

1) Integration-only model: Semantic integration + control predictors
2) Relatedness-only model: Semantic relatedness + control predictors
3) Combined model: Semantic integration + semantic relatedness + control predictors

The three above models were run for each part of speech in the preamble: D1, N1, P1, D2, A1, and N2. Table 2 shows the variables considered in each non-residualized regression model.

Table 2: Control predictors for each word position model

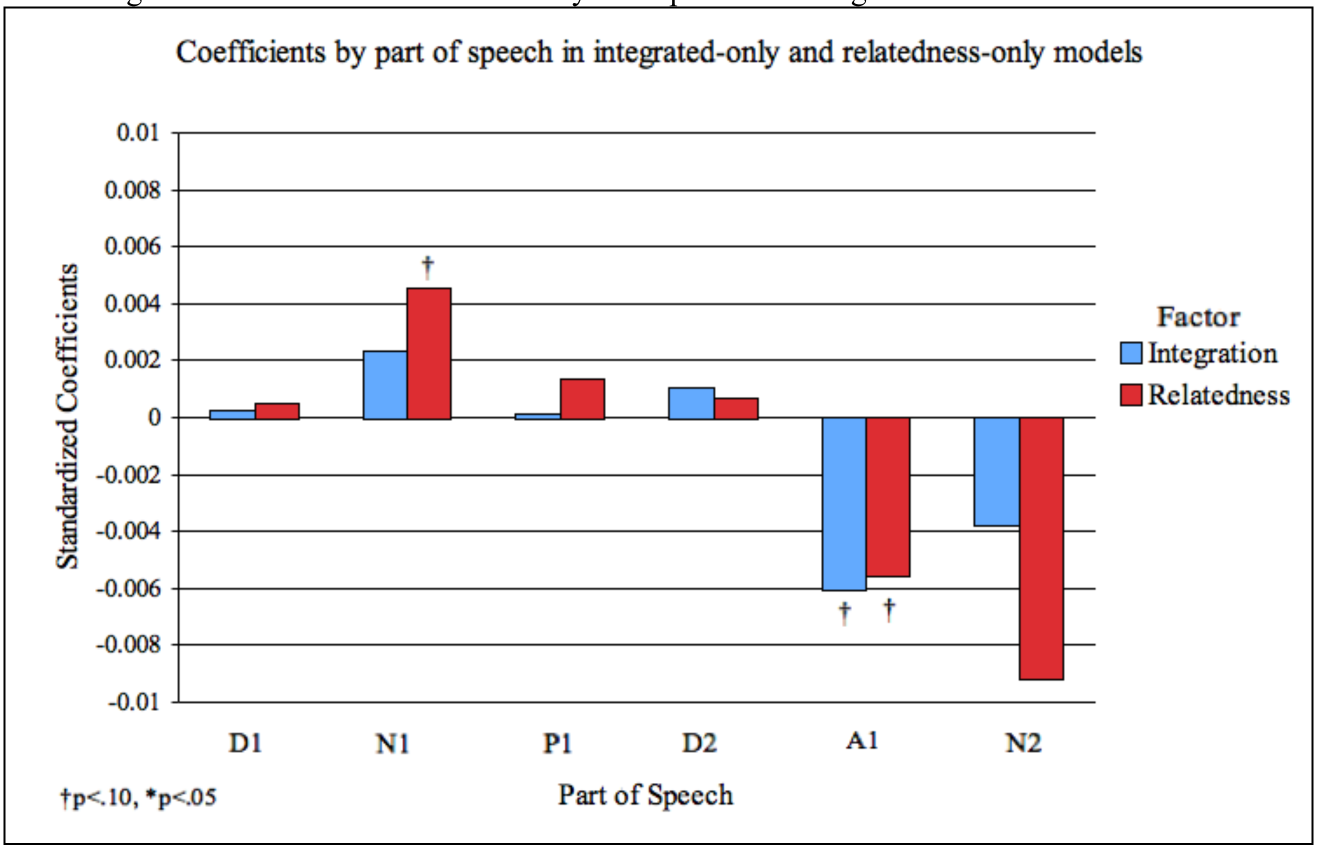
Control Predictor	Word Position					
	D1	N1	P1	D2	A1	N2
Length char (*w*)		✔	✔		✔	✔
Length syll (*w*)		✔			✔	✔
Length (N1)	✔	✔	✔	✔	✔	✔
Length (P1)	✔	✔	✔	✔	✔	✔
Length (A1)	✔	✔	✔	✔	✔	✔
Length (N2)	✔	✔	✔	✔	✔	✔
Freq (*w*)		✔	✔		✔	✔
Freq (*w* - 1)			✔	✔		✔
Freq (*w* + 1)	✔	✔		✔	✔	✔

Note: D1 = first determiner; N1 = first noun; P1 = first preposition; D2 = second determiner; A1 = first adjective; N2 = second noun; *w* = current word; *w* – 1 = previous word; *w* + 1 = following word; Length = length in phonemes; Length char = length in characters; Length syll = length in syllables; Freq = frequency.

Results

Figure 1 shows standardized coefficients at each word position predicted from non-residualized integration-only and relatedness-only models. That is, each integration coefficient was calculated from a model containing integration as the only fixed effect, while the opposite is true for relatedness coefficients.

Figure 1. Standardized coefficients by word position in single-fixed-effect models.



Note: D1 = first determiner; N1 = first noun; P1 = first preposition; D2 = second determiner; A1 = first adjective; N2 = second noun

The sections below describe results at individual word positions. Unless indicated otherwise, results are reported from the combined model, which included both semantic integration and semantic relatedness as fixed effects. Control predictor coefficients (estimates, standard errors, and p-values) from the combined model are provided in tables in each section.

Determiner 1 (D1). At the preamble's first determiner, no effects of integration or relatedness on duration were observed. Semantic relatedness negligibly increased D1 duration.

Integration minimally decreased D1 duration in the combined model, but in the integration-only model, integration slightly increased D1 length (Estimate=0.0002548; Standard error=0.0012303; p=0.836017). None of these effects approached significance.

Table 3: Coefficient Values for D1 Combined Model

Factor	Estimate	Standard Error	p-value	Significance
Intercept	0.0565977	0.0143235	0.00075	***
Relatedness	0.0007278	0.0018363	0.69199	
Integration	-0.0003142	0.0018883	0.86788	
Plausibility	-0.0004843	0.0013548	0.72088	
N1 Length	-0.0080667	0.0027398	0.00815	**
P1 Length	-0.0014498	0.0012704	0.25421	
A1 Length	0.0007966	0.002515	0.75403	
N2 Length	-0.0007099	0.0012148	0.55915	
Frequency (N1)	0.0117456	0.0051889	0.03484	*

Note: *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Noun 1 (N1). Semantic relatedness was associated with slower N1 duration, reaching marginal significance in the relatedness-only model (Estimate=0.0045268; Standard error= 0.002362; p=0.0557) and significance in the combined model (p=0.0235). Semantic integration did not have a significant effect on N1 duration in the combined model. The integration-only model showed a non-significant pattern in the opposite direction, with integration correlated with faster N1 production (Estimate=0.0023164; Standard error=0.0026062; p=0.3744).

Table 4: Coefficient Values for N1 Combined Model

Factor	Estimate	Standard Error	p-value	Significance
Intercept	0.382664	0.044031	0.000000008	***
Relatedness	0.010557	0.00465	0.0235	*
Integration	-0.007709	0.00512	0.1327	
Plausibility	-0.00468	0.00265	0.0779	†
Frequency (N1)	-0.034597	0.014288	0.0257	*

N1 Length (char)	0.173908	0.063186	0.0125	*
N1 Length (phon)	-0.006633	0.024507	0.7895	
N1 Length (syll)	0.021072	0.010665	0.0632	†
P1 Length	0.00223	0.002486	0.3699	
A1 Length	-0.005149	0.007052	0.471	
N2 Length	-0.002807	0.002389	0.2405	
Frequency (P1)	0.001376	0.003397	0.6855	

Note: *** p<0.001, ** p<0.01, * p<0.05, † p<0.1; char = characters; phon = phonemes; syll = syllables.

Preposition 1 (P1). Semantic relatedness was correlated with increased P1 duration, with a marginally significant effect in the combined model and a non-significant effect in the same direction in the relatedness-only model (Estimate=0.0013674; Standard error=0.0013910; p=0.3259). Semantic integration was associated with decreased P1 duration in the combined model, but effects did not reach significance. The integration-only model also predicted slowed P1 production, but the effect was negligible, approaching zero (Estimate=0.0001019; Standard error=0.001529; p=0.94689).

Table 5: Coefficient Values for P1 Combined Model

Factor	Estimate	Standard Error	p-value	Significance
Intercept	0.2803354	0.0217407	< 0.0000000000000002	***
Relatedness	0.00516	0.0027802	0.0639	†
Integration	-0.0048085	0.0030538	0.11584	
Plausibility	-0.0011184	0.0015984	0.48438	
Frequency (P1)	-0.020598	0.0024089	< 0.0000000000000002	***
P1 Length (char)	-0.1142672	0.0267118	0.0000227	***
P1 Length (phon)	0.0440342	0.0164348	0.00758	**
N1 Length	0.0008583	0.0031203	0.7861	
A1 Length	-0.0046751	0.0028748	0.11569	
N2 Length	-0.0008325	0.0013938	0.55052	

Frequency (N1)	-0.0057426	0.0059286	0.34394	

Note: *** p<0.001, ** p<0.01, * p<0.05, † p<0.1; char = characters; phon = phonemes.

Determiner 2 (D2). No significant effects of relatedness or integration were observed at the second determiner's position. Semantic relatedness was minimally associated with faster D2 duration in the combined model, but with slower duration in the relatedness-only model (Estimate=0.0006649; Standard error=0.0011484; p=0.56284). Semantic integration showed negligible patterns in both models, with slower D2 duration in integrated conditions.

Table 6: Coefficient Values for D2 Combined Model

Factor	Estimate	Standard Error	p-value	Significance
Intercept	0.1080462	0.01204	1.33E-15	***
Relatedness	-0.0006797	0.002281	0.76581	
Integration	0.001717	0.0025189	0.49568	
Plausibility	-0.0016421	0.0013083	0.20992	
N1 Length	-0.001259	0.0027303	0.64879	
P1 Length	-0.002026	0.0012102	0.0946	†
A1 Length	0.0080589	0.002414	0.00267	**
N2 Length	-0.0006096	0.0012464	0.62498	
Frequency (P1)	-0.0046259	0.001666	0.00565	**
Frequency (A1)	0.0037879	0.0034008	0.27001	

Note: *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Adjective 1 (A1). Semantic relatedness and integration showed similar effects on A1 duration, with all models correlated with quicker A1 production. Marginally significant effects were observed in the integration-only model (Estimate=-0.00599; Standard error=0.003269; p=0.067364) and relatedness-only model (Estimate=-0.005548; Standard error=0.003262; p=0.089526). The combined model predicted a similar pattern but integration and relatedness effects did not reach significance.

Table 7: Coefficient Values for A1 Combined Model

Factor	Estimate	Standard Error	p-value	Significance
Intercept	0.500871	0.040998	< 0.0000000000000002	***
Relatedness	-0.002506	0.00478	0.600236	
Integration	-0.004164	0.004793	0.385345	
Plausibility	0.006468	0.00361	0.07367	†
Frequency (A1)	-0.060436	0.012277	0.00000415	***
A1 Length (char)	-0.048581	0.04406	0.2724	
A1 Length (phon)	0.107239	0.027072	0.000213	***
A1 Length (syll)	0.00176	0.019994	0.9303	
N1 Length	0.022417	0.012677	0.091983	†
P1 Length	-0.004335	0.003244	0.181855	
N2 Length	-0.004887	0.004015	0.223988	
Frequency (N2)	-0.002719	0.004949	0.582907	

Note: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, † $p<0.1$; char = characters; phon = phonemes; syll = syllables.

Noun 2 (N2). Semantic relatedness was associated with decreased N2 duration, with marginally significant effects in the combined model and same-direction non-significant effects in the relatedness-only model (Estimate=-0.009171; Standard error=0.005771; p-value=0.11252). Semantic integration in the integration-only model was also correlated with decreased duration, though effects did not reach significance (Estimate=-0.003740; Standard error=0.005889; p=0.52565). Integration in the combined model showed an opposite but non-significant association with increased N2 duration.

Table 8: Coefficient Values for N2 Combined Model

Factor	Estimate	Standard Error	p-value	Significance
Intercept	0.475121	0.051243	1.78E-15	***
Relatedness	-0.014626	0.008728	0.09428	†
Integration	0.007437	0.008897	0.40351	
Plausibility	-0.013911	0.00645	0.03151	*

Frequency (N2)	-0.009304	0.009047	0.30427	
N2 Length (char)	0.102802	0.054054	0.05865	†
N2 Length (phon)	0.116972	0.025985	9.84899E-06	***
N2 Length (syll)	-0.040046	0.012286	0.00119	**
N1 Length	0.003812	0.010926	0.73052	
P1 Length	0.006305	0.005836	0.28044	
A1 Length	0.006606	0.009969	0.51404	
Frequency (A1)	-0.009202	0.014543	0.5306	
Frequency (C1)	0.007976	0.004145	0.05476	†

Note: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, † $p<0.1$; char = characters; phon = phonemes; syll = syllables.

Discussion

Gillespie and colleagues (2013) reported that highly integrated conditions (sentence preambles in which head and local nouns are tightly conceptually linked) were associated with significantly shorter word durations at the preamble's first preposition (P1), second determiner (D2), and first adjective (A1). That is, the intervening material between head and local nouns was produced more quickly when the nouns were semantically integrated than when the nouns were unintegrated (Gillespie et al., 2013).

Overall, effects of semantic integration on word duration in the current experiment generally replicated findings in Gillespie et al. (2013). At D1 position, both studies showed non-significant interactions, with duration negligibly longer in integrated than unintegrated conditions. A significant effect at this position would be very unusual, so these results are not surprising. Gillespie and colleagues (2013) showed a non-significant effect of integration on the first noun (N1), with shorter N1 durations in integrated conditions. In the current experiment, a similar non-significant pattern was observed in the combined model, which included both semantic integration and semantic relatedness as fixed effects.

At the preamble's first preposition (P1), both studies reported shorter durations in integrated conditions, though effects were only significant in Gillespie et al. (2013); the range of magnitudes of interaction may be related to differences in experimental stimuli between the two studies, namely, the individual preposition. Preambles in Gillespie et al. (2013) contained the same preposition, *with*, in all items across all conditions. Penta (2014) manipulated integration by using different prepositions (in integrated conditions, *for* and *with*; in unintegrated conditions, *by* and *near*). Prepositions assume varying relationships between the nouns within the preamble, which may have influenced the magnitude of interaction. The preposition *with* may signify an attribute relationship (*the necklace with the colorful diamonds*), an accompaniment relationship (*the sweater with the clean skirts*; Gillespie et al., 2013), or a part-whole relationship (*the bicycle with the warped tires*). Possible relationships indicated by the preposition *for* can be beneficiary (*the bodyguard for the nervous actresses*) or part-whole (*the banana for the popular sundae*). Prepositions *by* and *near* commonly signify location relationships (*the dam by the musty factory; the church near the giant driveway*). The prepositional relations above are nowhere near exhaustive of all possibilities (e.g., Srikumar & Roth, 2013; Table 1), but the fact that multiple potential relationships do exist and vary for each preposition is noteworthy. The differences in magnitude of integration effects observed in Gillespie et al. (2013) and Penta (2014) may be explained by the latter experiment using multiple prepositions within sentence preambles, and thus a greater variety of potential meaning relationships between the head and local nouns, while the former was limited to the possible relationships signified by *with*.

Duration at the second determiner (D2) was significantly shorter in integrated conditions than in unintegrated conditions in Gillespie et al. (2013), but the current experiment observed a non-significant effect in the opposite direction: D2 duration increased in integrated preambles. The difference between current results and those found previously may be due to carry-over effects from processing the preposition**.**

Semantic integration showed similar non-significant effects on the second noun (N2) duration in Gillespie et al. (2013) and in the integration-only model in the current experiment, with shorter N2 duration in integrated conditions.

Word duration measures sometimes patterned in the same direction in both studies (i.e., both speeding or both slowing production), but only reached significance in Gillespie et al. (2013). This mismatch in significance occurred at the P1 and A1 word positions. In other cases, duration measures patterned in the same direction in both studies, but only in certain regression models: N1 and P1 duration patterns were similar to Gillespie et al. (2013) in the combined model but not in the integration-only model; N2 duration patterns were similar to Gillespie et al. (2013) in the integration-only model but not in the combined model. Varying degrees of integration manipulations may be able to explain the different magnitudes of interaction.

Semantic integration manipulations in the current experiment (which used preambles from Penta, 2014, Experiment 1) were much weaker than semantic integration manipulations in Gillespie et al. (2013), which used preambles from Solomon & Pearlmutter (2004, Experiment 4). Both groups conducted norming studies in which participants rated the degree of two nouns' conceptual relationship on a seven-point scale, with 1 indicating a loosely integrated or unintegrated pair of nouns and 7 indicating a tightly integrated pair of nouns. Comparisons between mean semantic integration ratings from Gillespie et al. (2013) and Penta (2104) show that the difference between integrated and unintegrated elements is much smaller in the latter, in both singular local noun and plural local noun conditions. That is, the difference between semantically integrated items and unintegrated items was more striking in Gillespie et al.'s

(2013) stimuli than in the preambles used in the current experiment. Weaker integration manipulations may be able to explain the differing degrees of interaction, in which duration measures patterned in the same direction but only reached significance in Gillespie et al.'s (2013) second experiment.

Effects of semantic relatedness, if similar to those of semantic integration, should show shorter word durations (faster production) in related conditions and longer word durations in unrelated conditions. This is the first experiment investigating spoken word duration as a function of relatedness; as such, predictions of semantic relatedness and integration similarly affecting word duration, and therefore affecting the timing of language production planning, were based on neither prior evidence nor substantiated claims. As results show, semantic relatedness generally influenced word duration differently than semantic integration. Significant differences are observed at word positions N1 (first noun), P1 (first preposition), and N2 (second noun). Possible reasons for the contrasting influences are discussed below.

Integration and relatedness effects differed slightly in word positions D1 and D2, but no effects from any regression model in any condition reached significance, with all standardized coefficients approaching zero in determiner positions. Significant effects of integration and relatedness were not expected at D1, so this result is not surprising. Gillespie et al. (2013) found significant integration effects at D2, with the determiner produced more quickly in integrated preambles. An effect of semantic relatedness at D2, however, is not expected here, since intervening material between head and local nouns had no influence on the degree of relatedness. The intervening material does influence the degree of semantic integration, since integration is linguistically and contextually dependent (Solomon & Pearlmutter, 2004). Although the determiner remained constant throughout all preambles and did not individually affect integration ratings, faster production of D2 in Gillespie et al. (2013) could be explained by the large influence of its surrounding environment (P1 and A1) on integration.

Semantic integration and relatedness affected first noun (N1) durations in opposite directions: though effects were non-significant, integrated conditions in the integrated-only model were associated with shorter N1 duration. In related conditions, N1 was produced significantly slower than in unrelated conditions. It is unclear why N1 production time increased in related preambles, but the slowed articulation could be explained with regard to previous priming literature (cite). As the language production planning mechanism recognizes or articulates a word (the 'target'), other lexical items associated with the target are highly activated, leading to their faster retrieval (cite). For example, saying (or thinking of) the target word *baseball* spikes activation in words highly associated with it (e.g., *bat, glove, grass, home run, basketball)* and allows the speaker to retrieve those words more quickly. Participants perhaps took longer to produce N1 because their production planning mechanism was busy activating elements associated with the target. This explanation is plausible, since N2 production was significantly faster in related conditions. The language production planning system took more time to produce N1, but in doing so sped N2 retrieval and production later in the preamble.

Different influences of integration and relatedness are also observed at the first preposition (P1). In the combined regression model, P1 duration was shorter in integrated conditions (though with non-significant effects), and longer in related conditions (with marginally significant effects). The opposite pattern may be due to a few factors, independently or combined: overlap effects from longer N1 duration, or the preposition's role in determining integration but not relatedness. The first noun (N1) was produced significantly slower in related preambles, so carry-over effects from lexical activation of N2 may have caused P1 to be produced slower in related conditions.

In the preambles in this experiment and others (e.g. Solomon & Pearlmutter, 2004; Gillespie et al., 2013), P1 defines the degree of integration: *the necklace with the colorful diamonds/feathers* is integrated only because the preposition *with* implies a relationship in which the head noun (*necklace*) is composed of the local noun (*diamonds/feathers*). As discussed earlier, prepositions vary in the relationships they define (e.g., attribution, accompaniment, location). Relatedness, however, is not concerned with P1 in the same way; the degree of relatedness is determined solely by the head and local nouns (*the necklace with/near the colorful diamonds* is related regardless of the preposition used). In short, relatedness does not care what preposition is included in the preamble, but integration does. Thus, effects of integration at P1 similar to those found in Gillespie et al. (2013) should not necessarily be expected in the current experiment, which focuses on effects of semantic relatedness on duration measures.

At A1, integration and relatedness influenced word duration in the same direction and magnitude: durations were shorter in integrated and related conditions. This pattern is somewhat surprising after observing very different patterns of integration and relatedness on N1 and P1 durations. If P1 was indeed produced more slowly in related conditions due to carry-over effects from N1, that overlapping difficulty had likely subsided by the time participants were articulating A1. The pattern observed at A1 follows the above explanation of longer N1 duration ultimately leading to shorter N2 duration; A1 falls in between these nouns and is perhaps produced more quickly to facilitate speeded N2 retrieval.

N2 production was significantly faster in related conditions, contradicting effects of integration. As mentioned, this facilitation is likely due to spiked activation of N2 at the time of

N1 articulation. Priming literature shows clear patterns of faster retrieval and production of words that are highly associated with a target word. Because N1 and N2 are semantically related, they are also (probably) associated (cite). Semantically integrated conditions led to longer N2 duration, but integration does not care about N2; conversely, semantic relatedness is dependent on the relationship between N1 and N2.

A few notes regarding data analysis processes deserve remark. Association was excluded in all analyses due to uneven and unclear distribution. Penta (2014, Experiment 1) did not intentionally manipulate associative relationships between head and local nouns, so differences in effects of association were probably driven by a few outlying instances of very strong association measures, with the majority of experimental stimuli being loosely associated or unassociated. However, his Experiment 2 did involve manipulations of semantic and associative relationships to investigate mismatch effects resulting from different meaning relationships, specifically, category coordinates and attribute relations (Penta, 2014).

In the initial analyses, no regression models were residualized, meaning that all control predictors (also called ‘nuisance factors’) were included in every model (refer to Table 2). After analyses were completed, regressions that included only local nuisance factors for each word position (D1, N1, P1, D2, A1, N2) were run for all models (integration-only, relatedness-only, combined). Local nuisance factors are control predictors that affect the current word’s immediate surrounding environment. For example, the N1 regression model with only local nuisance factors considered the following predictors: integration, plausibility, N1 frequency, N1 length in characters, phonemes, and syllables, P1 length, and P1 frequency. Control factors that did not occur in the direct environment, such as N2 length and frequency, were excluded.

Subtle differences in predictions from models with all nuisance factors and models with only local nuisance factors should be mentioned. The combined regression model, which simultaneously considered integration and relatedness as fixed effects, is discussed first. The combined model with all nuisance factors showed a marginal effect of relatedness at N2, with faster N2 production in semantically related conditions. In the model containing only local nuisance factors, the marginal effect at N2 became significant in the same direction. In other words, removing non-local control predictors increased the significance of the effect of relatedness at N2.

A few differences in non-significant effects deserve mention here. At A1, integration and relatedness patterned in the same direction in both models. In the model with only local nuisance factors, the magnitude of difference between integration and relatedness effects was more pronounced. That is, in the regression with only local nuisance factors, integration sped duration more and relatedness sped duration less. Similarly, effects of integration at N2, although non-significantly increasing word duration in both models, were more pronounced in the model with local nuisance factors.

The integration-only and relatedness-only models also showed faint differences between regression models with all or with only local nuisance factors. In the model with all nuisance factors, marginally significant effects were seen: relatedness increased duration of N1, and relatedness and integration both decreased duration of A1. In the regression model containing only local nuisance factors, effects of relatedness at N1 and effects of integration at A1 became significant. Effects of relatedness at A1 lost significance. It should also be noted that integration effects at N2 (though non-significantly decreasing duration in both models) was somewhat less pronounced in the model with only local nuisance factors.

Taken together, models including all or only local nuisance factors showed similar predictions of word durations. The directions of effects of integration and relatedness were preserved in all models, though in some cases the magnitude of effect fluctuated. There are no clear theoretical reasons for preferring one regression model to the other, but by comparing models that consider all nuisance factors to those that consider only local nuisance factors, it can be observed that non-local control predictors account for a small portion of variance at nearly all word positions.

Overall, effects of semantic relatedness on word duration fall on a gradient: word durations are longer in semantically related conditions in the beginning of the preamble (D1, N1, P1, D2), and become shorter in the later part of the preamble (A1, N2). Although not all effects reach significance, they consistently follow the structure of the gradient. Effects of semantic relatedness are not restricted to the intervening material between the head and local nouns, as semantic integration effects have been (Gillespie et al., 2013). Semantic relatedness influences word duration differently than semantic integration, falling into a gradient that begins with significantly delayed production and ends with significantly expedited production.

General Discussion

The purpose of this experiment was to investigate effects of semantic integration and semantic relatedness on the timing of language production planning, specifically using measures of word duration. Data from Penta's (2014) Experiment 1, a spoken sentence-completion task manipulating integration and relatedness, were coded and analyzed using non-residualized multiple-regression models. This experiment was the first investigation of effects semantic relatedness on word duration; as such, it was unclear whether relatedness and integration would influence the timing of language production separately or in conjunction with one another, and the magnitude of potential interactions was unknown. It was hypothesized that semantic relatedness would show similar effects on duration measures as integration, with intervening material in semantically related utterances being shorter in duration compared to words in semantically unrelated utterances (Gillespie et al., 2013). However, different influences of integration and relatedness were observed. Semantic integration showed largely similar patterns to previous literature, with words in between two semantically related nouns being produced more quickly than when in between unintegrated nouns (with the exception of the preamble's second determiner, which showed non-significant effects nearing zero). Evidence of semantic integration effects in this study provides further support for a parallel-activation based system of language production (e.g. Solomon & Pearlmutter, 2004, Gillespie et al., 2013).

Semantic relatedness influenced word duration along a gradient scale, with words near the beginning of sentence preambles (D1, N1, P1, D2) being produced more slowly than words at the end of the preamble (A1, N2). As participants proceeded through the preambles, word durations became decreasingly slower and increasingly faster; that is, the magnitude of effects of semantic relatedness on word duration differed at each word position, but consistently followed the gradient.

Taken together, results of this experiment demonstrate different influences of semantic integration and semantic relatedness on word duration. As this experiment is the first to investigate effects of semantic relatedness on the timing of language planning and production, additional research is essential to support the findings discussed here. Semantic integration and relatedness both influence speech planning processes, and further investigation into the relationship between the two will hopefully provide fruitful insights into inner workings of the mechanism that times, plans, and executes language production.

Acknowledgments

I would like to thank Neal Pearlmutter and Darrell Penta for helping with data analyses, and Maureen Gillespie for providing Praat instructions and analyses tips. I greatly appreciate comments and suggestions from Neal Pearlmutter, Joanne Miller, and Darrell Penta on earlier drafts of this work. This experiment uses original data from Penta's (2014) Experiment 1 and uses coding and analysis techniques from Gillespie et al. (2013). This project is supported by the Northeastern University Honors Program and Department of Psychology.

References

Barker, J., Nicol, J., & Garrett, M. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research*, *30*(1), 91-114.

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

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of Perception and Cognition* (pp. 181-216). Orlando, Fla: Academic Press.

Bock, K., & Cutting, C. J. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language, 31*, 99-127.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, *8*(1), 57-99.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945-984). San Diego, CA: Academic.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45-93.

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

DiBattista, A., & Pearlmutter, N. J. (2011). Effects of semantic integration on phrase and word ordering errors in production. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2295-2300). Austin, TX: Cognitive Science Society.

Eberhard, K. M. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language, 36*, 147-164.

Ferreira, F., Engelhardt, P. E., Traxler, M., & Gernsbacher, M. (2006). Syntax and production. *Handbook of Psycholinguistics*, 61-91.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 133-177). London: Academic Press.

Garrett, M. F. (1989). 3 Processes in language production. *Linguistics: The Cambridge Survey: Volume 3, Language: Psychological and Biological Aspects*,*3*, 69.

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. *Handbook of Psycholinguistics*, *2*, 21-59.

Gillespie, M., & Pearlmutter, N. J. (2011a). Effects of semantic integration and advance planning on grammatical encoding in sentence production. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1625-1630). Austin, TX: Cognitive Science Society.

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

Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject–verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 511-528.

Gillespie, M., Pearlmutter, N. J., & Shattuck-Hufnagel, S. (2013). *Prosody reflects semantic integration in sentence production*. Manuscript in revision.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

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 Language Processing, La Jolla, CA.

Penta, D. J. (2014). *Semantic relatedness and semantic integration in subject-verb agreement errors*. Unpublished master's thesis, Northeastern University, Boston, Massachusetts.

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

Srikumar, V., & Roth, D. (2013). Modeling semantic relations expressed by prepositions. *Transactions of the Association for Computational Linguistics*, *1*, 231-242.

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

Vigliocco, G., Lauer, M., Damian, M. F., & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 46-58.