Semantic Relatedness and Semantic Integration in Subject-Verb Agreement Errors

Darrell Jonathan Penta

Northeastern University

Master's Thesis

Semantic Relatedness and Semantic Integration in Subject-Verb Agreement Errors

Sentence production research concerns itself with understanding the nature of both the linguistic information a speaker brings to the task of converting a thought into speech, and also the specific cognitive mechanisms that make this possible. From the perspective of a speaker, the task itself is essentially effortless under normal circumstances. Although speakers might occasionally "search for the right words" to convey an idea, the underlying processes of speech production are fundamentally rapid, automatic, and inaccessible to conscious thought (Bock, 1995; Bock & Levelt, 1994). Yet the manifest simplicity of speaking belies a set of production operations that are anything but simple: Appropriate words corresponding to parts of the conceptual message that one intends to communicate must be selected and retrieved from memory; the relationship between the words and their respective roles in the message, as well as the interdependencies among the words themselves, have to be coordinated; and a language-specific linear order and form for the words must be derived in a way that is suited to the modality of speech.

A model of sentence production, therefore, is a model of many discrete but connected processes, with near universal consensus that information must be represented in at least a message level, a grammatical level, and a phonological level, each with its own set of constraints relating to what and how information is handled at that level. Differences in opinion about some of the production system's finer points, such as the extent to which the components are autonomous, whether information cascades forward in an incremental or serial fashion, or whether the system permits feedback from lower to higher levels, have led to various iterations of a more or less standard model (cf. Bock, 1995; Bock & Levelt, 1994; Butterworth, 1980; Dell, 1986; Schriefers, Meyer, & Levelt, 1990). Within virtually all models, including the one proposed by Bock and Levelt (1994), production begins with a representation of the speaker's intended meaning being formulated in the message component, or message level. The next stage of processing ultimately involves mapping the message-level representations onto representations containing semantic, grammatical, and morphological information, but there is some dispute about how to characterize the division of labor within the grammatical component.

On the basis of speech error and other data, many have favored distinguishing between the grammatical-level representations that make correspondences with the message level, and those that make correspondences with the phonological level (e.g., Bock & Levelt, 1994; Roelofs, Meyers, & Levelt, 1998). Models consistent with this position generally assume that, within the grammatical level, there is a level of functional processing and a level of positional processing. In the former, lexical concepts, or lemmas, associated with the message are selected, and their grammatical functions (e.g., Subject, Direct Object) are assigned.[1] In the latter, lexical morphemes, or lexemes, and grammatical morphemes are retrieved and then grouped and ordered into hierarchically organized phrasal constituents. In short, these models assert a clear difference between lemmas, which lack morpho-phonological specifications, and lexemes, which lack semantic and syntactic information. On the basis of their own experimental evidence, Ferreira and Humphreys (2001) question whether such a division is necessary or even tenable. They reported that when participants in their study produced exchange errors, for example,

[1] This characterization glosses over the distinctions made in Bock and Levelt (1994) between grammatical functions and grammatical/syntactic relations.

saying *compounded the mix* when originally prompted to say *mixed the compound*, they appropriately adjusted the stress of the misplaced word to reflect its change in grammatical class (i.e., from the noun *COM-pound* to the verb *com-POUND-ed*). The authors took this as evidence against a level of morpho-phonological representation that lacks syntactic information. Caramazza and Miozzo (1998) have argued against the lemma-lexeme division as well, though much of their argument hinges on data from aphasics and written, rather than spoken, language production.

Despite divergent opinions about representations and operations at the grammatical level, there is also strong consensus in the literature that the product of grammatical processing undergoes additional processing at the phonological level, which specifies the phonological structure of the to-be-uttered phrase. Finally, by way of an articulatory component, the representational structure at this level is converted into speech.

Given the complexity of the production process, it is perhaps surprising that cases in which the system fails are the exception rather than the rule. But speech errors, either observed in corpora or elicited experimentally, have long been seen as a valuable source of information about many different aspects of language production. Because some kinds of speech errors display high levels of systematicity (e.g., in word-exchange errors, it is predominantly the case that the exchange will involve words from the same grammatical class) (Garrett, 1975) they can implicate failures at different levels of the system and ultimately aid researchers in narrowing the range of possible designs that a model of production can take (Dell, 1986). The present study is concerned with exploring a particular kind of speech error, namely, subject-verb agreement errors, as a means of learning more about how conceptual, message-level properties, and grammatical properties relate to each other during the production of language.

Grammatical Agreement

Grammatical encoding processes ensure that the constituent elements of a spoken utterance take a grammatical form consistent with the conventions of the language that one speaks. In English, both word order and characteristic grammatical markers on individual words convey non-trivial information about such things as the relationship between the words in an utterance and between the utterance's words and their real-world referents. For example, if a speaker wishes to make a comment about some books, he or she might make *The books* the subject of a to-be-uttered sentence. English stipulates that subjects generally precede verbs, so a verb phrase (VP) such as *are about linguistics* will appear after the subject. English also routinely indicates when more than one of the same thing is being referred to by adding a grammatical marker of plurality at the end of the referent's corresponding noun— in the above example, the phonetic segment represented by the letter "*s*" at the end of the noun *book*. Hence, the noun phrase (NP) *The books* is interpreted to mean more than one book.

Violations of grammatical conventions such as these represent one aspect of language that lends itself to empirical research, since the end-product of grammatical operations is readily observable and measurable. This is certainly the case for subject-verb agreement, where the agreement in question pertains to a match between the grammatical number of the subject and the verb of an utterance. The rules in English surrounding number agreement are relatively straightforward: A subject can either be grammatically singular, in which case it takes a singular verb (e.g., *The guy is*), or grammatically plural, in which case it takes a verb that is inflected, or overtly marked, to show plurality (e.g., *The guys are*); for the majority of English verbs, number marking on main verbs happens only in the present tense. A hallmark syntactic feature of agreement is that, regardless of how much material intervenes, a verb agrees with the subject noun, which may be linearly and hierarchically more distant from the verb than other nouns (e.g., *The guys who stayed up all night playing high-stakes poker are*) (Bock & Miller, 1991). A subject-verb agreement error is defined as a case in which a speaker's utterance contains a violation of the convention of matching subjects and verbs for grammatical number, for example, when a plural subject (e.g., *The guys*) appears with a singular verb (e.g., *is*).

Head-Local Mismatch Effect

In the production literature, agreement errors are traditionally considered to be a good source of information about the syntactic mechanisms involved in sentence processing because number agreement in English, in essence, involves using grammatical information provided by one utterance constituent in order to generate the appropriate form of another constituent (e.g., Barker, Nicol, & Garrett, 2001; Bock & Cutting, 1992; Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Bock & Miller, 1991; Brehm & Bock, 2013; Eberhard, 1999; Franck, Vigliocco, & Nicol, 2002; Gillespie & Pearlmutter, 2011, 2013; Solomon & Pearlmutter, 2004; Thornton & MacDonald, 2003). Among the numerous studies that have sought to experimentally elicit agreement errors, Bock and Miller (1991) was the first to demonstrate the head-local mismatch effect: Across three experiments, participants completed sentences after being presented with preambles in which the head NP (e.g., *The key(s)*) either matched a local NP in number (e.g., *the cabinet(s)*) or did not (e.g., *The key to the cabinets/The keys to the cabinet*).

The authors found that agreement error rates were greater in the mismatch conditions overall but were greatest when a singular head noun was paired with a plural local noun. In nearly all agreement studies since Bock and Miller, this head-local mismatch effect is assessed by measuring the size of the difference in error rates between the mismatch and match conditions. Broadly speaking, the present study is concerned with investigating conceptual or message-level factors that may affect the size of that difference.

Some Conceptual Factors Related to Subject-Verb Agreement

At the same time that subject-verb agreement is a syntactic process, the fact that speech begins with thought and culminates with sound means that all levels of the production system are implicated in agreement to some extent. Of central importance to the present study is research concerned with determining what information from the message level is used in the course of agreement operations at the grammatical level, including the mechanisms proposed to account for how agreement errors occur. A handful of models have been put forth that differ on such things as how connections between the message and grammatical levels are thought to be configured, and where, precisely, agreement computations are believed to take place within the grammatical level. The present experiments are not designed to pit one model against another. Rather, these models are described inasmuch as they relate to the significant findings about conceptual properties' influence on agreement. Among such properties, those related to number meaning have been the focus of many studies.

Notional Versus Grammatical Number of Individual Nouns

On its surface, matching a subject and verb for number sounds like a fairly basic process, and in fact, it is a routine to which few native speakers would give a moment's thought. But from the point of view of the production system, determining whether the subject of a message is plural or singular is a multi-faceted challenge. At the notional level, the quality of singularity or plurality can vary depending upon whether a speaker conceives of a subject noun's referent as a singleton or as a multiple. At the grammatical level, some lexical items have pre-specified number information and so the question of singularity or plurality can vary as a function of the particular noun one chooses for a message. At issue for the processor is the fact that the relationship between message-level and grammatical-level properties— and specifically, the notional and grammatical number of nouns— is not isomorphic, since many types of nouns exhibit non-canonical mappings between notional and grammatical number. For instance, although simple singular count nouns such as *key* are both notionally and grammatically singular, consistently taking a singular verb form such as *is*, bipartite plurals like *scissors* refer to conceptually single entities but are grammatically plural, requiring a plural verb for agreement. Table 1 provides examples of common noun types described with respect to notional and grammatical number. Given these occasional mismatches, properties directly related to number information at the conceptual and grammatical level stand out as potential sources of agreement errors.

Table 1

Notional and Grammatical Number Properties of English Common Nouns, by Category

	Number			
Noun type	Notional	Grammatical	Example	Verb
Singular count	Single	Singular	Key	is *are
Plural count	Multiple	Plural	Keys	*is are
Summation/Bipartite	Single	Plural	Tweezers	*is are
Pluralia Tantum	Multiple	Plural	Suds	*is are
Collective(a)	Single/Multiple	Singular	Herd	is ?are
Mass(b)	Single/Multiple	Singular	Fog	is *are

Note. Asterisks (*) indicate cases in which the given verb would be ungrammatical. Question marks (?) indicate cases in which a grammatical reading is possible. (a)Judgments about the grammaticality of a singular collective noun taking a plural verb vary by noun and by region or dialect. (b)Mass nouns may be marked plural when being used to distinguish between different types of the referent, as in *The teas of Asia are renowned for their delicate flavor*.

Bock, Eberhard, Cutting, Meyer, and Schriefers (2001) offer one account of how mismatches between notional number and grammatical number might translate into agreement errors. An over-simplified example here can illustrate the basics of their marking and morphing model, including the purported source of agreement errors on this model's account: An (American English) speaker intends to communicate about her family. Conceptually, she is thinking of multiple people (i.e., members of her family) and, therefore, from the point of view of the message-level representation, the referent about which she is planning to talk is plural. When production begins, this number information is mapped from the conceptual message to the representation at the level of functional processing in the grammatical component. This results in the subject function being marked for plurality, which consequently provides the number specification for the entire subject NP. As the lexical and inflectional morphology for the plurally-marked NP is realized, however, there is potential for conflict: In American English, it is conventionally the case that *family* is treated as a singular noun, taking a singular verb in agreement. This kind of number information is possibly part of *family*'s lexical specifications. In order to transmit the correct number to the verb for the agreement operation, then, the number of the entire NP must be brought into accord with the number of the subject noun. If it is not adjusted, or if it is adjusted incorrectly, an agreement error may ensue (e.g., *My family IS* vs. **My family ARE*)[2].

Bock et al. (2001) sought to experimentally investigate whether nouns that display a non-canonical mapping between notional and grammatical number could be manipulated to increase mismatch effect sizes. Employing the same sentence completion task as Bock and Miller (1991), Bock and colleagues used NP+PP preambles in which the local noun manipulation included nouns such as those listed in Table 1. They predicted that notionally plural-grammatically plural

[2] The latter would be entirely correct for many native English speakers. The point here is to emphasize the contrast between words that have lexically-specified number information and words that do not.

nouns would increases the rate of agreement errors, but notionally plural-grammatically singular local nouns would not do so. Experiment 1 examined the effects of summation or bipartite plurals (e.g., *tweezers*), which are notionally singular but grammatically plural. On the authors' account, a speaker may have a single entity in mind when forming a message about a bipartite plural noun's referent, but the lexical specification of plurality on the noun should set the number of the subject NP, resulting in a plural rather than a singular verb for agreement. They found that these nouns elicited significantly more errors than singular control nouns, but significantly fewer than simple plural nouns in the mismatch condition. Experiment 2 included *pluralia tantum* nouns (e.g., *suds*), described in Bock et al. as nouns that are generally conceived of as multiples and that are grammatically plural. Similar to the first experiment, these nouns reliably produced more agreement errors than singular controls, but fewer than standard plurals. Experiment 3 presented subjects with collective nouns (e.g., *herd*), which are grammatically singular but which allow for a notionally singular or plural interpretation. These nouns produced error proportions that did not differ from singular local noun controls when they were presented as singular entities (e.g., *herd*), nor from plural local noun controls when they were presented as plural entities (e.g., *herds*).

Considering the evidence collectively, the authors concluded that increases in mismatch effect sizes were attributable to the grammatical rather than the notional plurality of the local nouns, consistent with their predictions and with the main premises of the marking and morphing model: Notional number plays some cursory role in agreement to the extent that it is used to guide the selection of appropriate lemmas, and the number information specified at that level of representation is the same given to the entire subject NP up to the point of positional processing.

However, as an utterance is morphologically and inflectionally instantiated, any conflicts in number should be settled in favor of the subject NP's grammatical number. But the authors also discuss an important asymmetry in their data: Since the (notionally singular-) grammatically plural cases like *scissors* and the (notionally plural-) grammatically plural cases like *suds* produced significantly smaller mismatch effect sizes than their standard plural counterparts, the agreement mechanism must be sensitive to more than grammatical information alone. Rather, Bock et al. (2001) propose that the agreement mechanism is differentially sensitive to the morphological information of nouns.

One possibility is that the plural inflection that presumably accompanies regular plural nouns, but that does not accompany invariant grammatical plurals like bipartite and *pluralia tantum* nouns, can increase the likelihood of agreement errors. On an activation-based account, a plural inflection on a local noun can lead to a boost in activation of the plural feature such that the local noun may exert greater influence in establishing number information during the morphing process, allowing a subject NP headed by a singular noun to be erroneously marked as plural. Alternatively, on a retrieval-based account in which both singular and plural forms of nouns are stored in the lexicon (e.g., Seidenberg & McClelland, 1989), competition between the two forms during retrieval might lead to disruptions in the retrieval of an appropriate verb form; when nouns lack a competing form, such as in the case of many invariant plurals like the bipartite *scissors*, the likelihood of errors may be decreased. Bock et al.'s (2001) experiments do not provide direct evidence in favor of either of these explanations, but the authors suggest their research shows the limited capacity of notional number information to influence agreement beyond the scope of providing number information to the lemma-level representation.

Eberhard, Cutting and Bock (2005), however, note cases that challenge this assertion. For example, if the referent of a subject NP is understood to be a singleton in spite of being morphologically marked for plurality, agreement may proceed according to the notional, rather than the grammatical number. This is presumably the case when a subject NP such as *The scrambled eggs at table two,* which is meant to refer metaphorically to the person who ordered the food, not to the food itself, produces a singular verb form (e.g., *...is ready for more coffee*), despite the subject noun being grammatically plural. On balance, then, the evidence suggests that it is not only grammatical number information that matters in agreement.

Conceptual Properties of Noun Phrases

Occasionally, a mismatch between a noun's conceptual number and grammatical number can cause problems for the agreement operation. Yet this is not the only finding related to number meaning that can be brought to bear on the matter of message-level influences on agreement. In fact, number meaning at the level of the noun phrase, rather than at the level of the individual noun, has been successfully manipulated and has been shown to increase mismatch effect sizes. Consider, for example, a phrase such as *The gang on the motorcycles;* although *gang* is a grammatically singular noun, it can reasonably be interpreted in this particular context to refer to the individual members of the gang, each of whom is on his or her own motorcycle. In cases characterized by this sort of distributivity, then, agreement may follow the notional number of the subject noun, resulting in grammatical agreement errors (e.g., *...are raising money for charity*). Bock and Miller (1991) considered cases like this by measuring whether the distributivity of a subject NP with respect to the local NP could potentially increase error rates in the mismatch condition. In the single-token (i.e., non-distributive) condition, single-token NP preambles such as *The key to the cabinets* represented a single instantiation of a head NP that was not distributed across multiple items; conversely, in the multiple-token (i.e., distributive) condition, phrases such as *The label on the bottles* could be interpreted as multiple instances of the token *label* distributed across multiple conceptual items (i.e., *bottles*).

Bock and Miller (1991) found no significant differences in mismatch effect sizes between the single- and multiple-token conditions, which suggested a limited role for conceptual number in agreement. Eberhard (1999) conducted a follow-up study, however, in which the critical question was whether Bock and Miller's experimental preambles were generally too abstract to be conceptually accessible. More specifically, Eberhard suspected that if Bock and Miller's critical items were not sufficiently concrete or easy enough to imagine, a distributed referent's conceptual number may be less accessible and hence, less likely to influence agreement. Creating a new set of stimuli and obtaining imaginability ratings for her items and for items used in previous experiments, Eberhard reported greater mismatch effects for distributed items relative to non-distributed items, at least for items that were highly imageable. The effects of distributivity on subject-verb agreement errors were also confirmed in experiments conducted in languages other than English, including Italian, French, and Dutch (Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996). These findings provide some support to the notion that at least one conceptual feature, distributivity, can predictably elicit a notionally plural reading of a subject NP even when the subject noun is grammatically singular, and that such a reading can manifest as agreement errors.

While research on number meaning has provided a solid theoretical and empirical foundation for proposing an influence of conceptual factors on agreement, additional studies have looked at agreement error rates vis-à-vis conceptual properties of the head and local NPs that are conceivably only indirectly related to notional number. Bock and Miller (1991) addressed the question of whether the animacy of the head or local nouns was a relevant conceptual feature in the agreement process. This question follows from the assumption that the semantic feature of animacy is strongly correlated with subjecthood, and the observation that the subject noun of an NP is ordinarily the locus of grammatical number information used in agreement. The researchers speculated that preambles containing an animate local noun paired with an inanimate head noun might increase the mismatch effect, since cues about animacy (rather than functional role assignment alone) might cause the system to improperly identify the local noun as the appropriate source of number information. Participants in Bock and Miller's Experiment 2 received matched or mismatched head-local pairings in which either the head NP was animate and the local NP inanimate, as in *The nomad(s) of the mountain(s)*, or in which the order was reversed, as in *The mountain(s) of the nomad(s).* The authors reported no significant difference in mismatch effect sizes between the animacy conditions and concluded that animacy of the local NP was not a reliable predictor of agreement errors.

Research by Barker, Nicol, and Garrett (2001), however, increased the number of animacy conditions by adding animacy-matched NP pairings for both animate (e.g., *girl–teacher(s)*) and inanimate nouns (e.g., *blackboard–desk(s)*), and found that inanimate subjects were more likely than animate subjects to increase mismatch effects; that animacy-matched conditions were more error-prone than were their unmatched counterparts; and that the magnitude of the mismatch effect was greatest when an inanimate head noun was paired with an than inanimate subjects overall because the former are both functionally marked as subjects and because their subjecthood is reinforced by their animacy. On the other hand, when both the head and local nouns share the same semantic feature of animacy, this may increase the probability of interference in the system (i.e. because features common to both nouns can be active simultaneously) and lead to increased errors.

In a separate sentence completion task, Thornton and MacDonald (2003) reported larger mismatch effects for sentence preambles in which the local noun was plausible as the subject of a passive sentence. Participants in their experiment were given NP+PP preambles (e.g., *The album(s) by the classical composer(s)*) along with a verb; in one condition, only the head NP was a plausible subject of the verb (e.g., *played,* where only *album* is possible as a thing to be played), while in the other, both the head NP and the local NP made sense as the subject of a different verb (e.g., *praised*, where both *album* and *composer* are possible as a thing to be praised). In this latter condition, mismatch effects were greater as plausible local NPs were more likely to be confused as the source of number information for agreement than when only the head NP was a plausible subject. These findings lend additional support to the claim that non-number related meaning-based properties can reliably influence agreement and increase the magnitude of mismatch effects. The more recent research of Solomon and Pearlmutter (2004) introduces semantic integration—another conceptual property unrelated to notional number— as a factor in agreement errors.

Semantic Integration

Across five sentence completion experiments, Solomon and Pearlmutter (2004) demonstrated robust influences on mismatch effects from semantic integration, which they defined as the degree to which two elements of a to-be-uttered sentence are linked at the message level. On their definition, the relative strength of the linkage varies as a function of how the relationship between the elements is construed within a particular conceptual representation, or message; accordingly, Solomon and Pearlmutter were able to effectively manipulate semantic integration in numerous ways. In Experiment 1, the manipulation was accomplished by varying only the preposition in a standard NP+PP preamble. Participants read items such as those in (1), which shows one critical item preamble in each of its four versions. In addition to the local noun number manipulation involving the singular-singular match (e.g.,1a, 1c) and singular-plural mismatch conditions (e.g., 1b, 1d), a PP headed by the preposition *of* was used for the more integrated condition (e.g., 1a-b), while the less integrated condition featured the preposition *with* (e.g., 1c-d). Critically, in all of Solomon and Pearlmutter's experiments, integration was measured with ratings. A separate set of participants rated semantic integration, and preambles such as (1a-b) were rated as more integrated than preambles such as (1c-d).

(1) a. The drawing of the flower
b. The drawing of the flowers
c. The drawing with the flower
d. The drawing with the flowers

The difference in integration in (1) rests on the relationship between the head and local NPs that is signaled by the preposition: In (1a-b), *of* is used to describe a relationship in which the local noun *flowers* is the representational content of the head noun *drawing.* The two NPs in this case are thus conceptually tightly integrated. By comparison, in the less integrated preambles (1c-d), the preposition *with* signals what would most likely be interpreted as an accompaniment relationship. That is, there is no special connection between the head and local noun other than that they are perhaps close to each other in physical space. Solomon and Pearlmutter (2004) speculated that head and local NPs in more integrated preambles are planned closer together in time, resulting in there being some overlap in activation of their elements (i.e., in a spreading-activation model) such that some information about these elements, including grammatical number information, may simultaneously be available for computation. As in other activation-based models (e.g., Barker et al., 2001; Eberhard, 1997), it is this potential interference from the local noun that translates into agreement errors. The results of their Experiment 1 were consistent with their predictions, as were the results of their Experiments 2 and 3, which used the prepositions *of* or *for*, respectively, to specify a functional relationship between the head and local NPs in the more integrated condition versus *with* in its accompaniment sense in the less integrated condition.

In Solomon and Pearlmutter's (2004) Experiment 4, the head NP and preposition (e.g., *The pizza with*, in (2)) were not changed across the different versions of a critical item, unlike in the previous three experiments; rather, the integration manipulation was achieved solely by altering the content of the local NP. For the more integrated cases (2a-b), the attribute relationship specified by *with* is reinforced by the content of the local NP in that toppings are routinely understood to be a component of a pizza. In the less integrated cases (2c-d), the accompaniment reading of *with* is more plausible, since beverages are not an attribute of any pizza. These differences in integration manifested as differences in error rates, with a greater mismatch effect for the more integrated conditions compared to the less integrated conditions.

(2) a. The pizza with the yummy topping
b. The pizza with the yummy toppings
c. The pizza with the tasty beverage
d. The pizza with the tasty beverages

Other studies have looked at the effects of semantic integration and its influence on the timing of planning outside the context of subject-verb agreement errors. DiBattista and Pearlmutter (2011) further investigated Pearlmutter and Solomon's (2004) hypothesis that the overlapped planning of two utterance constituents could produce ordering errors (e.g., *The apple on the spot*, where what is intended is *The spot on the apple*)*;* participants in their study were more likely to make phrase- and word-ordering errors in an ordering-error elicitation paradigm when the pictures they described contained integrated versus unintegrated items. Gillespie, Pearlmutter, and Shattuck-Hufnagel (2013) examined word durations and prosodic breaks from recorded responses in previous agreement-error elicitation studies (i.e., Gillespie & Pearlmutter, 2011, Exp. 2; Solomon & Pearlmutter, 2004, Exp. 4); they reported that speakers produced shorter word durations for the words linking the head noun and local noun in integrated preambles compared to unintegrated preambles, and that speakers were more likely to use prosodic breaks to separate less integrated items relative to integrated items, providing some direct support for the timing-based account of the effects of semantic integration.

In summary, three separate sets of experiments have demonstrated the effects of semantic integration on agreement errors, exchange errors, and spoken word durations. Collectively, these findings provide strong motivation for considering how other message-level factors unrelated to number might influence subject-verb agreement. One such factor, semantic relatedness, has received a fair amount of attention in the production literature in general, but has thus far only been featured in one agreement study.

Semantic Relatedness

Semantic relatedness refers to a relationship between two words based on the words' meanings. For example, members of a taxonomic category have been classically treated as semantically related because they tend to have shared features (cf. Bock & Levelt, 1994; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Unlike semantic integration, where the relationship between two elements is determined by the context of the specific message in which they occur, semantic relatedness reflects a relationship in which context is not necessarily relevant. To date, numerous experiments have provided evidence that semantic relatedness can influence language production.

Bock (1986) conducted an experiment using a production-priming paradigm in which participants heard and then verbally repeated a prime word before describing a pictured event aloud. Bock's semantic prime condition contained items that were generally "related in meaning" (p. 577), including hyponyms (e.g., *canine–dog*) and synonyms (e.g., *cop–policeman*). Bock measured the frequency with which the primed target words appeared before the unprimed target words in participants' utterances and found that semantically-primed target words appeared more often as the grammatical subject of active or passive sentences, compared to unprimed target words, which occurred more often as the grammatical object. Based on these findings, Bock concluded that the semantic features of words are capable of influencing the word order and syntactic structure of an utterance.

In a Dutch-language experiment, Vigliocco, Lauer, Damian, and Levelt (2002, Experiment 1) used a procedure in which participants saw, and then translated from English, target words that were presented in the context of either a semantically homogeneous or semantically heterogenous block. Semantic homogeneity was defined in terms of category membership such that the category Animal, for example, included words such as *chicken*, *dog*, and *frog*. The experimenters reported that response latencies were greater when target items appeared in the homogenous versus the heterogeneous block, which they attributed to competition between semantically related words during retrieval.

Most relevant to the present study is Barker, Nicol, and Garrett (2001), which looked at the effect of relatedness in inducing agreement errors in a sentence-completion task. In Barker et al.'s Experiment 2, participants heard preambles for which head NPs and local NPs had either a high or low degree of semantic overlap, where items in the high-semantic overlap condition had "substantial overlap of semantic features (e.g., sailboat/canoe)" (p. 101), including synonyms and members of the same taxonomic category. Barker et al. reported greater mismatch effects for semantically related items relative to unrelated items and articulated the mechanism for these effects in terms of an activation-based framework (e.g., Eberhard, 1997). They reasoned as follows: Agreement errors arise when undue weight is given to the grammatical number feature of a local noun, where the number information should properly be taken from the subject head noun. One means of encouraging the system to give more weight to the local noun is to make it more likely to be considered the subject of a preamble, for example, by imbuing it with qualities that are strongly associated with subjecthood (e.g., Barker et al., 2001), or by making it a plausible candidate for subjecthood (Thornton & MacDonald, 2003).

One open issue, considered in detail in Experiment 2, is how to define semantic relatedness. Most language production experiments, as well as numerous priming experiments (Lucas, 2000), treat semantic relatedness as a matter of feature overlap. Another measure of relatedness is association, which is defined as a regular co-occurrence between two words. While relatedness and association are orthogonal factors, in practice, they commonly co-occur (e.g., *cat* and *dog*; Balota et al., 2006). Moreover, while both relatedness and association seem capable of separately producing priming effects, effect sizes vary widely, as does how the factors are normed and controlled; consequently, scholars have arrived at different conclusions about whether relatedness in the absence of association is capable of priming (cf. Hutchinson, 2003; Lucas, 2000). For Experiment 1, separate measures were obtained for each factor, but we expected the two to correspond.

Experiments 1 and 2 are concerned with exploring semantic relatedness as a factor in increasing mismatch effect sizes. First, we take up the question of how semantic relatedness might contribute to the effects of semantic integration on agreement. The relationship between these factors is unclear in some respects. On the one hand, in all but one of the experiments in Solomon and Pearlmutter (2004), the head and local NPs did not differ within individual items across the semantic integration manipulation, so there was no variation in relatedness across each version. On the other hand, in their Experiment 4, the integration manipulation involved varying the local noun (e.g., *pizza—topping(s)* for the integrated condition vs. *pizza—beverage(s)* for the unintegrated condition). It is possible, therefore, that in Experiment 4, there was a difference in semantic relatedness for the critical items, in addition to a difference in semantic integration. Moreover, while Solomon and Pearlmutter advised the participants who rated their stimuli for integration to ignore any meaning-based relatedness between the noun pairs outside the context of their preambles, relatedness may nevertheless have influenced participants' integration ratings anyway. Furthermore, while Barker et al. (2001) found increased error effects in their semantic overlap conditions relative to their low-overlap conditions, it is unclear whether or how the researchers held semantic integration constant across their critical items. In Experiment 1, we aimed to address these concerns by attempting to manipulate relatedness and integration separately from each other.

Experiment 1

The goal of Experiment 1 was to gain a clearer understanding of the relationship between semantic relatedness and semantic integration as factors contributing to subject-verb agreement errors. It adopted a two (singular vs. plural local noun number) by two (semantically related vs. unrelated) by two (semantically integrated vs. unintegrated) design in a sentence completion task and compared the proportion of errors subjects produced in the singular versus corresponding plural local noun number conditions. Semantic relatedness was manipulated by pairing preamble head NPs with local NPs that were semantically related or unrelated, with norming data serving as the basis for confirming the difference between conditions. A two-way interaction between local noun number and relatedness would replicate Barker et al. (2001) and bolster claims about semantic relatedness as a factor in agreement errors. Semantic integration was manipulated by varying the preposition connecting the head and local NPs: Integrated phrases contained the preposition *for* or *with*, while unintegrated phrases contained *near* or *by*. As with Solomon and Pearlmutter (2004), separate integration ratings were obtained to distinguish more integrated items from less integrated items. An interaction between local noun number and semantic integration would replicate Solomon and Pearlmutter. An interaction between semantic relatedness and integration would clarify some aspects of the relationship between these two factors.

Method

Participants. Two hundred forty-one undergraduate students from Northeastern University participated in this experiment in partial fulfillment of the requirements of an introductory psychology course. Data from three subjects were lost as a result of recording failure, and data from three non-native English speakers were not included in the analyses, resulting in usable data from 235 native English speakers.

Materials and Design. Twenty-four sentence preambles were chosen as the critical stimuli on the basis of norming data, described below. Table 2 provides an example of one critical item in each of its eight versions. Each preamble consisted of a head NP (e.g., *The necklace*) followed by a preposition (*with* or *for* for integrated preambles, *by* or *near* for unintegrated preambles) and a local NP. Local NPs all contained the determiner *the*; followed by an adjective, which was either identical across related and unrelated versions, or else matched for length in characters, syllables and phonemes; followed finally by the local noun. The head NP was always singular, and the eight different versions of an item were created by varying the local NP for related versus unrelated conditions, the local noun number for singular versus plural conditions, and the preposition for integrated versus unintegrated conditions.

Table 2

Experiment 1 Example Stimulus

Semantic integration	Local noun number	Relatedness	
		Related	Unrelated
Integrated	Singular	The necklace with the colorful diamond	The necklace with the colorful feather
	Plural	The necklace with the colorful diamonds	The necklace with the colorful feathers
Unintegrated	Singular	The necklace near the colorful diamond	The necklace near the colorful feather
	Plural	The necklace near the colorful diamonds	The necklace near the colorful feathers

Seventy-two filler preambles were used in addition to the experimental items. Twenty-four fillers were plural head NP+PP preambles that were similar in structure to the critical items. Twelve of the plural head filler items appeared with plural local NPs, and twelve appeared with singular local NPs. The remaining miscellaneous fillers contained a variety of structures, none of which were the same as the critical items' NP+PP structure, and approximately one-third of which did not consist of a subject NP (e.g., *Whenever the lights go out*).

The 24 critical items were chosen from an original list of 45 preambles and were matched for a variety of properties. Table 3 presents the separate means and standard deviations of the relevant properties and conditions for the 24 selected critical items. For all nouns, adjectives, and prepositions used in the preambles, word frequencies taken from the 51-million word SUBTLEX$_{US}$ corpus (Brysbaert & New, 2009), as well as length in characters, phonemes, and syllables were closely matched across conditions. Additionally, stimuli were created so that head and local NPs always matched in animacy; none of the items contained abstract nouns; and all items were designed not to encourage a distributive reading. Items were also designed to be plausible.

Table 3

Experiment 1 Critical Item Frequency and Length

		Length		
Preamble element by condition	Frequency	Characters	Phonemes	Syllables
Preposition				
Integrated	3.77 (0.06)	3.50 (0.51)	3.00 (0)	1.00 (0)
Unintegrated	2.21 (0.59)	3.33 (0.96)	2.67 (0.42)	1.00 (0)
Adjective				
Related	0.75 (0.53)	7.29 (1.65)	6.41 (1.73)	2.33 (0.78)
Unrelated	0.83 (0.57)	7.08 (1.69)	6.25 (1.70)	2.33 (0.78)
Related noun				
Singular	1.09 (0.51)	5.42 (1.64)	4.75 (1.61)	1.58 (0.50)
Plural	0.71 (0.46)	6.54 (1.67)	5.92 (1.56)	1.71 (0.58)
Unrelated Noun				
Singular	0.98 (0.52)	5.96 (1.88)	5.54 (1.77)	2.08 (0.77)
Plural	0.62 (0.54)	7.08 (2.04)	6.58 (1.83)	2.13 (0.79)

Note. Standard deviations are reported in parentheses.
(a) Frequencies are Log10 transformed, calculated as Log10((frequency/million)+1).

Relatedness ratings. Ratings of semantic relatedness for each head-local noun pair were obtained from 96 participants using Mechanical Turk (Buhrmester, Kwang, & Gosling, 2011). Noun pairs (without determiners, adjectives, or prepositional phrases) displayed in capital letters appeared in four versions corresponding to the singular and plural alternations in the related and unrelated conditions, as in (1a-d), as well as in four versions in which the same pairs were presented in reverse order, as in (1e-h). Each version of a noun pair was assigned to a separate list, and each list had three differently randomized presentation orders, resulting in a total of 24 unique lists with 45 noun pairs each.

(1) a. NECKLACE-DIAMOND
b. NECKLACE-DIAMONDS
c. NECKLACE-FEATHER
d. NECKLACE-FEATHERS
e. DIAMOND-NECKALCE
f. DIAMONDS-NECKLACE
g. FEATHER-NECKLACE
h. FEATHERS-NECKLACE

Each participant saw exactly one list. Participants were asked to rate how related pairs of words were on a 1 (*unrelated*) to 7 (*very related*) scale. Examples provided in the instructions indicated that word pairs such as *heavy—big*, *light—dark*, and *cute—little* might be considered very related, whereas a word pair like *rainy—hungry* would be considered unrelated. Participants were instructed to ignore sound and letter similarity and were told that there were no right or wrong answers.

Half of the time, noun pairs were presented with the head noun preceding a local noun, and half of the time the order was reversed. Ten to 12 ratings were obtained for each pair. Differences between ratings for the head-local and local-head pairings were negligible, so only the former were used when choosing the 24 critical items. Table 4 shows the relatedness ratings in the local noun singular and plural conditions for the 24 critical items. As desired, the related noun pairs were rated as substantially more related than the unrelated noun pairs.

Table 4

Experiment 1 Critical Item Mean Relatedness Ratings

	Relatedness	
Local noun number	Related	Unrelated
Singular	6.20 (1.37)	3.02 (1.40)
Plural	6.15 (1.39)	2.95 (1.28)
M	6.18	2.99

Note. Rating scale was 1 (*unrelated*) to 7 (*very related*). Standard deviations are reported in parentheses.

Semantic integration ratings. Ratings of semantic integration were obtained from a different set of 96 participants using Mechanical Turk (Buhrmester et al., 2011). Each of the eight different versions of a given preamble (as in Table 2) was assigned to a separate list, and each of the eight lists list had three differently randomized presentation orders, resulting in a total of 24 unique lists with 45 preambles each. Each preamble was displayed with the head and local noun presented in capital letters, and participants were asked to rate how tightly linked the two nouns were, using a scale of 1 (*not linked*) to 7 (*tightly linked*); participants were also given the example of *The KETCHUP or the MUSTARD* as an unintegrated pair and *the RING made of SILVER* as a highly integrated pair. Participants were instructed to ignore relationships based on the meaning of the words alone, as in *The KETCHUP or the MUSTARD* example, and were told that there were no right or wrong answers. Ten to 12 ratings were obtained for each version of each preamble, resulting in the means presented in Table 5 for the 24 critical items.

Collapsing across local noun number, preambles were rated as more integrated in the related-integrated condition, followed by the related-unintegrated condition, then by the unrelated-integrated condition, and finally, by the unrelated-unintegrated condition. Thus, integrated conditions were rated as more integrated than unintegrated conditions, but the difference in integration between the related and unrelated conditions was even larger. The implications of this confound with respect to interpreting Experiment 1 results are raised in the discussion.

Table 5

Experiment 1 Critical Item Mean Semantic Integration Ratings

		Relatedness	
Semantic integration	Local noun number	Related	Unrelated
Integrated	Singular	5.63 (1.28)	3.64 (1.05)
	Plural	5.76 (1.29)	3.68 (1.07)
	M	5.70	3.66
Unintegrated	Singular	4.78 (1.11)	2.93 (0.85)
	Plural	4.84 (1.13)	2.88 (0.74)
	M	4.81	2.91

Note. Rating scale was 1 (*not linked*) to 7 (*tightly linked*). Standard deviations are reported in parentheses.

Association norms. Word-to-word associations were obtained from 100 Mechanical Turk (Buhrmester et al., 2011) participants to measure the extent to which a preamble's head noun, or the various versions of its local nouns, would elicit one another. (2) lists the five tested nouns from the eight versions of the preamble *The necklace with/near the colorful diamond(s)/ feather(s)*, as in Table 2. For each preamble, the head noun, the related local noun in both its singular and plural form, and the unrelated local noun in its singular and plural form were assigned to one of five separate lists; each list had four randomized presentation orders of its 45 prompt words, resulting in a total of 20 unique lists.

(2)	NECKLACE	*Head noun*
	DIAMOND	*Singular related local noun*
	DIAMONDS	*Plural related local noun*
	FEATHER	*Singular unrelated local noun*
	FEATHERS	*Plural unrelated local noun*

Participants were told that they would see a list of words, and were asked to type the first two words that came to mind after reading each word in the list. The instructions included examples indicating that a word such as *cold* might bring the words *snowy* and *chilly* to mind, and that a word like *friendly* might bring to mind words like *cheerful* and *polite;* participants were also informed that there were no correct or incorrect responses. All prompt words were presented in capital letters. Forty responses were obtained for each noun, and response percentages were calculated as the number of times out of all responses to a given prompt word that a particular local noun was provided in response to a head noun prompt, or that the head noun (in either a singular or plural form) was given in response to a local noun prompt, where both nouns in question were part of the same preamble.

Means were taken of arcsine transformed proportions for all head-to-local (H→L) noun and local-to-head (L→H) noun combinations; however, only the former were used to select critical items. First, the H→L proportions were taken to be the measure of greater importance because participants reading from left to right always encountered the head before the local noun. Second, while any given iteration of a noun pairing may have had a higher proportion in the L→H direction versus the H→L, the differences were small when local noun number was collapsed over, and, more importantly, the overall pattern was consistent with the intended manipulation: Related nouns regularly showed association in both directions, while unrelated nouns showed virtually no association in either direction. Table 6 shows the back-transformed means and standard deviations for each noun combination for the 24 critical items. Head nouns elicited their corresponding related local nouns substantially more than the unrelated local nouns, which were never elicited.

Table 6

Experiment 1 Critical Item Mean Head-to-Local Noun Association Proportions

Local noun number	Relatedness	
	Related	Unrelated
Singular	0.22 (0.30)	0.00
Plural	0.07 (0.17)	0.00

Note. Means of the arcsine transformed proportions of responses were calculated for analyses. The means shown here are back-transformed proportion of local noun responses given in response to a head noun prompt. Standard deviations are shown in parentheses.

Procedure

The procedure was the same as in Solomon and Pearlmutter (2004). All participants were run individually. Preambles were presented visually on a computer screen, and participants were instructed to repeat the preamble aloud, along with an ending, in order to make a complete sentence. Participants received no additional instructions or feedback with respect to the content or structure of the endings they provided. If a participant's rate of speech slowed considerably at any time during the experiment, the experimenter reminded the participant to increase the speed of his or her response.

The preambles were presented on IBM-compatible computers running the MicroExperimental Laboratory (MEL) software package (Schneider, 1988). Participants' responses were recorded onto CD-R for analysis, using a Shure SM58 microphone connected to a Mackie 1202-VLZ Pro mixer/preamp and an Alesis Masterlink ML9600 CD Recorder. The experiment consisted of a total of 96 trials preceded by 5 practice trials.

For each individual trial, a fixation cross appeared on the left edge of the computer screen for 1000ms, followed by a preamble positioned such that its first character was in the same place as the fixation cross. Preambles appeared for 40ms per character or 1000ms, whichever was greater. Participants read the preamble aloud and completed the sentence as quickly as possible. After each preamble was displayed, a blank screen appeared for 3000ms, followed by a message that instructed participants to press the space bar to begin the next trial.

Scoring. Scoring of responses was also identical to Solomon and Pearlmutter (2004). All participant responses were transcribed and assigned to one of four main scoring categories. Correct responses were those in which a participant properly repeated the entire preamble, said the preamble only once, and provided an inflected verb that was correctly marked for number as the first word of the completion. Responses were scored as errors if all of the above conditions were met except that the verb was incorrectly marked for number. Responses were scored as uninflected if all of the criteria for correct responses were met, but the verb was not overtly marked for number. Finally, responses were scored as miscellaneous for trials in which a participant failed to correctly repeat the preamble, a verb did not follow immediately after the preamble, or the criteria for any other category were not met. Trials on which a participant's response included a dysfluency (such as a pause or cough) during or immediately after the preamble were noted as such; in cases where a participant produced a dysfluency but proceeded to complete a sentence in accordance with one of the first three scoring categories, both the dysfluency and the scoring category were noted. Dysfluencies in miscellaneous responses were not counted separately. Individual trials on which a participant made no response were excluded from all analyses.

Results

Across all 5,640 critical trials, there were 2801 correctly-inflected responses, 109 agreement errors, 1241 uninflected responses, 1462 miscellaneous cases, and 27 trials with no response. The percentage of error responses (out of errors plus correct responses) by condition, and the counts of each of the response types, are shown in Table 7.

Table 7

Experiment 1 Error Rates (%) and Response Counts by Condition

Condition	Local noun number	Error Rate	Response Count			
			Error	Correct	Uninflected	Misc
Related						
Integrated	Plural	10.9 (1.6, 2.4)	39 (6)	318 (23)	178 (20)	167
Integrated	Singular	0.0 (0.0, 0.0)	0 (0)	349 (22)	207 (17)	147
Unintegrated	Plural	8.3 (1.4, 1.4)	29 (3)	322 (29)	123 (15)	229
Unintegrated	Singular	0.3 (0.4, 0.3)	1 (0)	376 (34)	133 (6)	192
Unrelated						
Integrated	Plural	6.2 (1.4, 2.0)	22 (5)	332 (29)	149 (16)	196
Integrated	Singular	0.3 (0.4, 0.3)	1 (0)	379 (35)	172 (22)	149
Unintegrated	Plural	4.3 (1.2, 0.9)	16 (3)	352 (34)	134 (17)	199
Unintegrated	Singular	0.3 (0.2, 0.3)	1 (0)	373 (39)	145 (9)	183

Note. Error rates are reported as percentage errors out of errors plus correct responses; in parentheses are standard errors of the mean computed from the analyses by participants and by items. Response counts include dysfluency counts, in parentheses. Misc=Miscellaneous responses.

Analyses were performed on error rates (calculated as the proportion of error responses out of errors plus correct responses), the number of uninflected responses, and the number of miscellaneous responses. For each of these, two separate 2 (local noun number) x 2 (semantic relatedness) x 2 (semantic integration) analyses of variance (ANOVAs) were conducted, one with participants (F_1) and the other with items (F_2) as the random factor. Because error proportions are not normally distributed, separate analyses were conducted using arcsine-transformed error proportions in both by-subject and by-item ANOVAs. Unless otherwise noted, the statistical patterns reported below were identical for analyses on both raw and transformed error proportions; they were also identical when the same analyses were performed on error counts rather than error proportions, and when dysfluency cases were excluded. The statistics reported are for analyses on untransformed error proportions with dysfluency cases included. Figure 1 shows mismatch effects presented as difference scores of error rates for plural local nouns and corresponding singular local nouns in each condition.

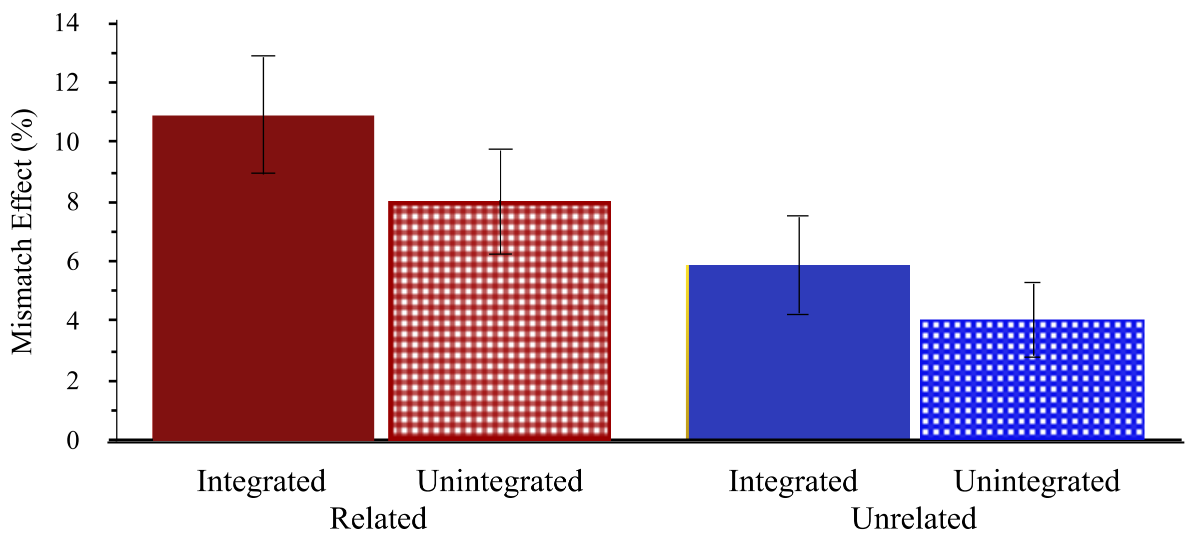

Figure 1. Mismatch effects by condition. Error bars show ± 1 *SEM*, computed from the analyses by participants.

Table 8 shows the results of the ANOVAs on error responses. Participants produced more agreement errors in the plural local noun conditions than in the singular local noun conditions.

Additionally, more agreement errors were made for related preambles versus unrelated preambles. Crucially, there was a two-way interaction between local noun number and relatedness, and the head-local mismatch effect was larger for related cases ($F_1(1,234)=42.80$, $MS_e=220.25$, $p<.001$; $F_2(1,23)=30.27$, $MS_e=31.97$, $p<.001$) than for unrelated cases ($F_1(1,234)=22.51$, $MS_e=152.02$, $p<.001$; $F_2(1,23)=21.40$, $MS_e=12.64$, $p<.001$). There was no main effect of integration, nor a reliable interaction between integration and any other factors.

Table 8

Experiment 1 Error Rate ANOVA Results

Effect	*F1*	*MSe*	*F2*	*MSe*
Local Noun Number	55.95	332.78 ***	31.77	73.84 ***
Relatedness	5.66	188.78 *	5.58	28.85 *
Local Noun Number x Relatedness	6.77	180.99 **	6.15	28.81 *
Integration	0.03	230.12	1.53	36.73
Local Noun Number x Integration	2.14	208.65	2.21	34.00
Relatedness x Integration	0.03	230.12	0.01	17.11
Local Noun Number x Relatedness x Integration	0.08	222.95	0.00	17.39

Note. The statistics reported are for analyses on untransformed error proportions with dysfluency cases included. For *F1* (by-subjects) analyses, all between-groups $df=1$, and within-groups $df=234$; for *F2* (by-items) analyses, all between-groups $df=1$, and within-groups $df=23$.
*$p<.05$, **$p<.01$, ***$p<.001$

Table 9 shows the results of the ANOVAs on uninflected and on miscellaneous responses. Analyses on counts of uninflected items revealed a main effect of local noun number with more uninflected responses for singular items; an effect of integration, with more uninflected responses for integrated items; and a two-way interaction between relatedness and integration, with a larger difference between integrated and unintegrated for related items than for unrelated items.

Miscellaneous responses showed an effect of local noun number, with more miscellaneous responses for plural items; an effect of integration, with integrated items generating fewer miscellaneous responses; and a two-way interaction between integration and relatedness (significant by subjects only), with the difference between the related and unrelated conditions greater for unintegrated items than for integrated items.

Table 9

Experiment 1 Uninflected and Miscellaneous Response ANOVA Results

Effect	*F1*	*MS*	*F2*	*MSe*
Uninflected Responses				
Local Noun Number	6.53	0.43 *	5.12	5.41 *
Relatedness	2.19	0.41	0.45	19.33
Local Noun Number x Relatedness	0.03	0.46	0.03	3.84
Integration	38.38	0.40 ***	11.60	13.13 *
Local Noun Number x Integration	1.01	0.51	1.42	3.54
Relatedness x Integration	9.78	0.41 **	4.54	8.69 *
Local Noun Number x Relatedness x Integration	0.06	0.45	0.48	5.27
Miscellaneous Responses				
Local Noun Number	16.70	0.46 ***	5.83	12.86 *
Relatedness	0.07	0.51	0.02	15.76
Local Noun Number x Relatedness	0.04	0.54	0.02	11.00
Integration	22.07	0.50 ***	8.72	12.38 *
Local Noun Number x Integration	0.18	0.58	0.31	3.27
Relatedness x Integration	5.55	0.47 *	2.45	10.42
Local Noun Number x Relatedness x Integration	2.39	0.51	4.02	2.99

Note. The statistics reported are for analyses on counts of both uninflected responses (with dysfluency cases included) and miscellaneous responses. For *F1* (by-subjects) analyses, all between-groups *df*=1, and within-groups *df*=234; for *F2* (by-items) analysis, all between-groups *df*=1, and within-groups *df*=23.
*p<.05, **p<.01, ***p<.001

Discussion

Experiment 1 was designed to assess the separate effects on subject-verb agreement of semantic integration and semantic relatedness, two non-number-related message-level factors that have each been shown to increase mismatch effect sizes. Norming data revealed a confound between these factors: While integrated preambles were rated as more integrated than unintegrated preambles, the difference was greater for related items compared to unrelated items.

There are at least two important findings from Experiment 1. First, participants made more subject-verb agreement errors when computing preambles containing plural nouns versus those containing singular local nouns. This replicates the often-reported mismatch effect, as in Bock and Miller (1991) and Solomon and Pearlmutter (2004). Second, the size of the mismatch effect was greater for related items relative to unrelated items. This is consistent with the results of Barker et al. (2001), who reported larger mismatch effect sizes for preambles containing head-local noun pairs that contained a high degree of semantic overlap.

There was no significant interaction between semantic integration and local noun number, and semantic integration did not influence mismatch effects in Experiment 1. These results do not replicate previously reported results (e.g., Solomon & Pearlmutter, 2004). However, the manipulation of integration was weaker in Experiment 1 than in prior experiments, and the numerical pattern across all combinations of relatedness and integration matched the pattern of integration ratings. Additional regression analyses are planned, which should offer some clarification of the present results and should make it possible to better separate integration and relatedness.

Given the clear influence of relatedness, however, the goal of Experiment 2 was to examine this factor in greater detail. Specifically, we investigated whether different types of semantic relationship and associative relationships are differentially capable of increasing the mismatch effect size.

Experiment 2

The manipulation of semantic relatedness in Experiment 1 resulted in greater mismatch effects for related items relative to unrelated items, providing additional support for the argument that meaning-based properties unrelated to conceptual number can have a measurable effect on agreement. However, there are numerous different kinds of semantic relationships, and many ways of manipulating relatedness experimentally. For example, in Barker et al. (2001), preambles in the semantically related condition contained noun pairings in which the nouns shared semantic features, representing synonyms or members of the same taxonomic category (henceforth, “category coordinates”). In contrast, a range of different semantic relationships were included in Experiment 1. For the purposes of Experiment 2, we sought to test the effects on agreement of specific semantic and associative relationship types by creating sentence preambles in which these relations were manipulated independently of each other. This meant first identifying relationships that could be suitably manipulated.

Semantic Relationships

There is a long history within priming research of manipulating different types of semantic relationships to various effects on both language production and comprehension (Fischler, 1977; Lucas, 2001; Meyer & Schvaneveldt, 1971; Moss, Ostrin, Tyler, & Marslen-

Wilson, 1995; Shetlon & Martin, 1992). A perennial concern raised in the literature, however, is that there are rarely hard and fast boundaries for separating out relationship types, and researchers occasionally take issue with the way a given relationship has been operationalized in previous work (e.g., Fishler, 1977; Hutchinson, 2000). An unavoidable issue in this respect is that, even if it were possible to create an agreed-upon list of definitional criteria for the different relationships, any pair of related words may satisfy criteria in more than one semantic category. Another issue is that semantically related words can also be associatively related. Since priming effects have been separately attributed to both semantic and associative relations, the potential confound between them imposes certain constraints with respect to choosing and manipulating relationship types. Among the candidate relationship types that were considered for Experiment 2, the two described below were the most appealing on a number of grounds.

Category Coordinates. The first and most compelling reason for featuring category coordinate relations in Experiment 2 is that it enables us to compare our results with the results of Barker et al. (2001). While only two of the 24 preambles used in Experiment 1 contained nouns that were normatively coordinates,[3] in Barker et al., the head and local noun pairings were ostensibly all category coordinates and other relations based on feature-overlap.[4] Moreover, effects of category coordinates have been established in the broader language production literature, showing both inhibitory effects (Hantsch, Jescheniak, & Schriefers, 2005; Schriefers,

---

[3] The pairs *table–chair* (Furniture) and *pencil–eraser* (Writing/Office Supplies) were the only pairs included in the Van Overschelde, Rawson, and Dunlosky (2004) update of the Battig and Montague (1969) norms.

[4] According to the authors, their study included synonyms, superordinate- and subordinate-level category members as well. If other types of relations involving feature overlap were included in, the authors fail to note as much, and efforts to obtain their stimuli have not been successful.

Meyer, &Levelt, 1990) and facilitation (Bock, 1986; Huttenlocher & Kubicek, 1983;), depending upon the experimental design. Lastly, category coordinates can be manipulated while avoiding an overlapping associative relationship. While there is conflicting evidence about how effective unassociated coordinates are in producing priming effects (cf. Hutchinson, 2002; Lucas, 2000; Moss et al., 1995), the fact that they can be isolated from associative relationships mades them ideal for Experiment 2.

Attributes. In contrast to category coordinates, attribute relations are both poorly defined and underrepresented in the literature. Under the name "essential attributes," a particular version of these relationships was featured in a single priming study conducted by Lucas (2001). Examples of essential attribute pairs provided by Lucas, with parenthetical information reflecting her view of the nature of the relationship between the prime and target, include: *sugar* (contains) *calories*, *trees* (have) *roots*, and *water* (is made of) *hydrogen.* In Lucas's experiments, both a lexical decision task and a separate naming task were used in which targets followed the presentation of their prime either immediately or after 200ms; significant priming effects were found only in the lexical decision task, and only when the target appeared 200ms after the prime.

Despite a lack of strong evidence of priming effects for attribute relations, there are a number of reasons for including these in Experiment 2. First, unlike many other semantic relationships, it is possible to impose a stringent in-or-out criteria when deciding what noun pairs belong in the attribute relation category, and this is accomplished by loosening one of Lucas's (2001) criteria and strengthening another. First, the question of what constitutes the essence of a thing, such that an individual might incorporate that information into the meaning representation of a concept, is far from a settled matter (cf. Medin & Ortony, 1989; Sloman & Malt, 2003).

Creating items based solely upon this criterion, therefore, seemed somewhat unreasonable for the purposes of Experiment 2. Rather, we defined an attribute relationship as one in which a word represents a salient property, attribute, or feature of another. Importantly, while we approached the creation of candidate attribute pairs with an emphasis on the property being centrally important to the whole, we did not feel it was necessary to justify pairings on an essentialist explanation. Examples from Experiment 2 include *jar—lid, rifle—trigger, book—page.*

Another reason for including attribute relations is that these kinds of relations are inherently semantic in nature, which is to say that the conceptual connection between the words provides a strong basis for an associative relationship between them as well. It is not because attribute pairings happen to co-occur, for example, that makes them related, but it is precisely because they are related that they do co-occur. Lucas (2001) noted this as a problem in attempting to create pairs of unassociated essential attributes, and Hutchinson's (2002) review of her study confirms that some of her critical item pairs were also associated, albeit weakly. Our own norming data also confirm that attributes are generally associated with the thing they describe. In short, whereas it is possible to remove associations between otherwise semantically related category coordinate items, attribute relations are both semantically and associatively related. This kind of relationship, therefore, offers a handy point of contrast to category coordinates.

Associative Relationships

Associations are relationships wherein one word commonly co-occurs with another, whether the reason for the co-occurrence is semantically based or not. Because associated items are also often semantically related, however, researchers manipulating relatedness for priming studies may inadvertently vary the strength of association between words, making it difficult to assess the source of the reported effects (Balota et al., 2006). On the basis of a meta-analysis, Hutchinson (2003) argues that unassociated semantically related primes fail to obtain reliable effects, and proposes that association strength is the key predictor of priming ability. Other studies suggest that semantic relationships can induce priming effects without associative relationships (cf. Fischler, 1977; Meyer & Schvaneveldt, 1971). Including associates that are not semantically related in Experiment 2 affords us the opportunity to make comparisons between these items and the associated attribute pairs, which explicitly include a semantic component, and also to the category coordinate pairs, which lack any association.

Overview

In Experiment 2, we investigated whether different semantic and associative relationships can differentially affect mismatch effect sizes, using the same type of NP+PP preambles and sentence completion task as in Experiment 1. An example of a category coordinate critical item preamble in its four versions is given in Table 10. As in Experiment 1, the head noun (e.g., *saw*) did not vary across conditions, and there was a local noun number manipulation (singular vs. plural). The category coordinate manipulation involved varying the local noun so that it was either a coordinate of the head noun (e.g., *wrench(es)*) in the related condition, or a non-coordinate control (e.g., *rag(s)*) in the unrelated condition.

Table 10

Experiment 2 Category Coordinate Example Stimulus

	Local noun number	
Relationship	Singular	Plural
Related	The saw by the oily wrench	The saw by the oily wrenches
Unrelated	The saw by the oily rag	The saw by the oily rags

For the purposes of Experiment 2, category coordinate critical items were created such that the nouns pairs in each preamble were taken from published category norms (Van Overschelde et al., 2004) and were selected to minimize association. Coordinates were always at the same taxonomic level and included no synonyms (e.g., *spider—bee*; *saw—wrench*). Association was measured in the same type of norming procedures as in Experiment 1. Lucas's (2001) conclusion that relatedness in the absence of association is sufficient to prime predicts a larger mismatch effect for the related items relative to unrelated items; conversely, if Hutchinson (2002) is right, and association is the critical component, we would expect no difference in mismatch effects across relatedness conditions. between the conditions. An interaction would be consistent with Barker et al.'s (2001) findings.

An example of a preamble with the semantic-associative relationship manipulation is provided in Table 11. Across all versions of a preamble, the head noun was not varied (e.g., *jar*) and local noun number (singular vs. plural) was manipulated. For the attribute relationship condition, the local noun (e.g., *lid(s)*) represented an associated attribute of the head noun. At the associate level, the local noun (e.g., *cookie(s)*), selected from Nelson, McEvoy, and Schreiber's (1998) published association norms, was associated but not otherwise strongly semantically related to the head noun. The unrelated versions featured a local noun (e.g., *bagel(s)*) that was neither an attribute nor an associate of the head noun.

Table 11

Experiment 2 Semantic-Associative Relationship Example Stimulus

	Local noun number	
Relationship	Singular	Plural
Attribute	The jar near the sticky lid	The jar near the sticky lids
Associate	The jar near the sticky cookie	The jar near the sticky cookies
Unrelated	The jar near the stale bagel	The jar near the stale bagels

If attribute relations have an influence on agreement over and above the influence of associative relations on agreement, we would expect mismatch effects to be greater for attributes than for associates. Alternatively, if association alone is the critical factor, both attribute and associates will presumably generate significantly larger effect sizes than the unrelated control items.

Method

Participants. One hundred thirty-three undergraduate students from Northeastern University participated in this experiment for course credit as described in Experiment 1. Data from two subjects were lost due to recording failure, and data from one non-native speaker were excluded from analyses. Additionally, one subject's data were excluded on the basis of too few usable trials.

Materials and Design. Two different sets of preambles were created for Experiment 2. From an original list of 46 preambles, 24 preambles were selected that featured a two-level manipulation of category coordinate status (i.e., related vs. unrelated) crossed with local noun number. From a separate list of 43 candidate preambles, an additional 24 preambles were selected that crossed the local noun number manipulation with a three-level manipulation of semantic-associative relationship status: associated attributes vs. associates vs. unrelated). Each of the 48 critical item preambles consisted of a singular head NP (e.g., *The saw*) followed by a locative preposition (e.g., *by, near*) and a local NP. Local NPs all contained the determiner *the,* followed by an adjective, which was either identical across each version of an item or else matched for length in characters, syllables and phonemes, followed finally by the local noun. The four versions of each category coordinate preamble and the six versions of each association preamble were created by varying the local NP (i.e., either related and unrelated for category items; either attribute, associate, or unrelated for attributes items) and the local noun number (singular vs. plural). In addition to the experimental items, 120 filler preambles were used. Thirty-six of these were plural head NPs containing a local noun number manipulation, and the remainder were miscellaneous fillers such as those used in the first experiment.

For all nouns, adjectives, and prepositions used across all conditions within each set of preambles, word frequencies (Brysbaert & New, 2009), as well as length in characters, phonemes, and syllables, were matched as closely as possible. For each of these properties, the separate means and standard deviations of the 24 category coordinate items are presented in Table 12, and for association items, in Table 13. Items were created so that head and local NPs always matched in animacy; none of the items contained abstract nouns; and all items were designed not to encourage a distributive reading. Items were also designed to be plausible.

Table 12

Experiment 2 Category Coordinate Critical Item Frequency and Length

		Length		
Preamble element by condition	Frequency(a)	Characters	Phonemes	Syllables
Preposition	1.95 (0.71)	4.21 (1.53)	4.04 (1.37)	1.38 (0.48)
Adjective				
Related	0.95 (0.72)	6.13 (1.79)	5.21 (1.50)	1.92 (0.57)
Unrelated	0.95 (0.72)	6.13 (1.79)	5.21 (1.50)	1.92 (0.57)
Related noun				
Singular	1.01 (0.52)	5.54 (1.35)	5.25 (1.48)	1.96 (0.68)
Plural	0.46 (0.36)	6.71 (1.31)	6.38 (1.41)	2.08 (0.57)
Unrelated Noun				
Singular	0.81 (0.43)	6.21 (1.73)	5.79 (1.68)	1.88 (0.60)
Plural	0.40 (0.32)	7.29 (1.77)	6.92 (1.75)	1.96 (0.61)

Note. Standard deviations are reported in parentheses.
(a)Frequencies are Log10 transformed, calculated as Log10((frequency/million)+1).

Table 13

Experiment 2 Semantic-Associative Relationship Critical Item Frequency and Length

		Length		
Preamble element by condition	Frequency(a)	Characters	Phonemes	Syllables
Preposition	2.15 (0.83)	3.92 (1.61)	3.88 (1.33)	1.25 (0.44)
Adjective				
Attribute	0.92 (0.63)	6.04 (1.65)	5.21 (1.50)	1.75 (0.61)
Associate	0.90 (0.57)	6.04 (1.43)	5.17 (1.49)	1.75 (0.61)
Unrelated	1.02 (0.66)	6.04 (1.57)	5.33 (1.55)	1.75 (0.61)
Attribute noun				
Singular	0.82 (0.52)	4.96 (1.55)	4.21 (1.22)	1.42 (0.58)
Plural	0.55 (0.49)	5.92 (1.50)	5.29 (1.27)	1.42 (0.58)
Associate noun				
Singular	1.15 (0.54)	5.33 (1.83)	4.63 (1.84)	1.63 (0.82)
Plural	0.61 (0.50)	6.38 (1.86)	5.63 (1.86)	1.67 (0.82)
Unrelated Noun				
Singular	0.97 (0.48)	5.33 (1.66)	4.79 (1.44)	1.46 (0.51)
Plural	0.56 (0.40)	6.38 (1.69)	5.79 (1.44)	1.46 (0.51)

Note. Standard deviations are reported in parentheses.
(a)Frequencies are Log10 transformed, calculated as Log10((frequency/million)+1).

Relatedness ratings. Ratings of semantic relatedness were obtained following the same procedures as in Experiment 1. Two separate surveys were used because the decision to add an unrelated level to the semantic-associative relationship preambles was made after the first set of preamble candidates was created. Consequently, survey one featured noun pairs from 46 category coordinate preamble candidates and 43 semantic-associative relationship preamble candidates in which the local noun was either the attribute or the associate. Responses were collected from 96 participants using Mechanical Turk (Buhrmester et al., 2011). Noun pairs (without determiners, adjectives, or prepositional phrases) displayed in capital letters appeared in four versions corresponding to the singular and plural alternations in the related and unrelated conditions for category coordinate items, as in (3a-d), as well as in four versions in which the same pairs were presented in reverse order, as in (3e-h). For all orderings of semantic-associative relationship pairs, the local noun was either the attribute noun in its singular or plural form, as in (4a-b, g-h), or the associate noun in either its singular or plural form, as in (4c-d, i-j). Each version of a noun pair was assigned to a separate list, and each list had three randomized presentation orders, resulting in a total of 24 unique lists with 89 items total.

(3) a. SAW-WRENCH
b. SAW-WRENCHES
c. SAW-RAG
d. SAW-RAGS
e. WRENCH-SAW
f. WRENCHES-SAW
g. RAG-SAW
h. RAGS-SAW

(4) a. JAR-LID
b. JAR-LIDS
c. JAR-COOKIE
d. JAR-COOKIES
e. JAR-BAGEL
f. JAR-BAGLES
g. LID-JAR
h. LIDS-JAR
i. COOKIE- JAR
j. COOKIES- JAR
k. BAGEL-JAR
l. BAGELS-JAR

Each participant saw exactly one list. Half of the time, noun pairs were presented with the head noun preceding a local noun, and half of the time the order was reversed. Ten to 12 ratings were obtained for each pair. Differences between ratings for the head-local and local-head pairings were negligible, so only the former were used when choosing the critical items. Table 14 shows the relatedness ratings in the local noun singular and plural conditions for the 24 category coordinate critical items, which indicated that the related noun pairs were rated as more related than the unrelated noun pairs. Ratings for semantic-associative relationship pairings are presented in conjunction with those obtained in the second survey, described below.

Table 14

Experiment 2 Category Coordinate Critical Item Mean Relatedness Ratings

	Status	
Local noun number	Related	Unrelated
Singular	4.21 (0.86)	2.29 (0.87)
Plural	4.29 (0.78)	2.32 (0.79)
M	4.25	2.31

Note. Rating scale was 1 (*unrelated*) to 7 (*very related*). Standard deviations are reported in parentheses.

A second relatedness survey, using the same procedures as described above for the first relatedness survey, was created in order to obtain relatedness ratings for noun pairs from the unrelated versions of the semantic-associative critical items; for noun pairs from the attribute versions semantic-associative, which were used as a control for the survey; and also for a subset of the critical items used in Solomon and Pearlmutter (2004). Ratings of the latter were collected for the purposes of future analyses and are not discussed here. A separate group of 96 Mechanical Turk (Buhrmester et al., 2011) participants provided ratings for pairs in which the head noun either preceded or was followed by a singular or plural attribute noun, as described above, or one of the two versions of the unrelated noun, as in (4e-f, k-l). As with the first survey, each version of a noun pair was assigned to a separate list, and each of the eight lists had three randomized presentation orders, resulting in a total of 24 unique presentation lists. Each list contained 88 items in total.

Each participant saw only one list. Ten to 12 ratings were obtained for each noun pair. Because the head-attribute local noun pairs appeared on both surveys, a total of 20-24 ratings were obtained for each version of these items and 10-12 ratings were obtained for each version of associate local nouns as well as for the unrelated nouns. Table 15 shows the mean ratings in the local noun singular and plural conditions for the 24 critical items. Attribute pairs were rated as the most related, followed by associate pairs, followed by unrelated pairs.

Table 15

Experiment 2 Semantic-Associative Relationship Critical Item Mean Relatedness Ratings

	Relationship		
Local noun number	Attribute	Associate	Unrelated
Singular	5.84 (0.40)	4.72 (0.91)	3.28 (0.91)
Plural	5.98 (0.44)	4.61 (0.93)	2.96 (1.43)
M	5.91	4.66	3.12

Note. Rating scale was 1 (*unrelated*) to 7 (*very related*). Standard Deviations are reported in parentheses.

Semantic integration ratings. Given the pattern of relatedness and integration ratings in Experiment 1, we had reason to suspect that the related items in Experiment 2 design would generally be rated as more integrated than their unrelated counterparts. Despite the fact that integration was not manipulated in Experiment 2, separate semantic integration ratings were obtained to try to minimize differences in integration between item versions. As with relatedness norming, it was necessary to use two separate surveys to obtain integration ratings because the unrelated versions of the semantic-associative relationship items had not yet been created at the start of norming procedures. Responses for the first survey were obtained from a separate group of 48 participants using Mechanical Turk (Buhrmester et al., 2011). The first survey contained: (1) each of the category coordinate versions shown in Table 10; (2) the attribute and associate versions of semantic-associative relationship preambles in both their singular and plural versions, shown in the top two rows of Table 11; and (3) for the 24 Experiment 1 critical items such as that shown in Table 4, the integrated version of preambles containing their unrelated and related singular and plural local nouns. Inclusion of these latter items was to ensure that participants saw an appropriate range of semantic integration. Each of the four versions of these preambles was assigned to a separate list and each of the four lists had three randomized presentation orders, resulting in a total of 12 unique lists containing 113 items.

Each participant saw only one list. Ten to 12 ratings were obtained for each version of each preamble, resulting in the means presented in Table 16 for the 24 category coordinate items. As indicated, the difference in integration between related and unrelated items was minimal for these items.

Table 16

Experiment 2 Category Coordinate Critical Item Mean Semantic Integration Ratings

	Status	
Local noun number	Related	Unrelated
Singular	3.23 (0.58)	2.74 (0.64)
Plural	3.11 (0.51)	2.62 (0.60)
M	3.17	2.68

Note. Rating scales were 1 (*not linked*) to 7 (*tightly linked*). Standard deviations are reported in parentheses.

Using a separate survey, and following the same procedures as described above for the first semantic integration survey, we obtained integration ratings for: (1) the attribute and unrelated versions of the 24 semantic-associative relationship items in both singular and plural versions, as in the first and last rows of Table 11; and (2) 58 preambles representing singular and plural local noun alternatives of a subset of the critical items used in Solomon and Pearlmutter (2004). These latter items were included to ensure that participants saw surveys with a range of integration, and also, for the purposes of future analyses, which are not discussed here. Forty-eight Mechanical Turk participants contributed ratings to the second survey. The four versions of the semantic-associative preamble critical items were each assigned to a separate list, and each list had three random presentation orders, for a total of 12 unique presentation lists. Of the 82 items on each list, 24 were the various semantic-associative relationship versions. Because the attribute preambles appeared in both surveys, a total of 20-24 ratings were obtained for each version of these items, while 10-12 ratings were obtained for each version of the unrelated preambles. Table

17 shows the mean integration ratings for the 24 critical items. Attribute items were rated slightly higher than associate items, while unrelated items had a lower average rating.

Table 17

Experiment 2 Semantic-Associative Relationship Critical Item Mean Semantic Integration Ratings

	Relationship		
Local noun number	Attribute	Associate	Unrelated
Singular	4.64 (0.42)	4.10 (0.76)	2.88 (0.88)
Plural	4.62 (0.35)	4.22 (0.72)	3.00 (0.60)
M	4.64	4.16	2.94

Note. Rating scales were 1 (*not linked*) to 7 (*tightly linked*). Standard deviations are reported in parentheses.

Association norms. Two separate surveys were designed to obtain word-to-word associations for Experiment 2 critical items because the unrelated versions of semantic-associative relationship preambles items had not yet been created at the start of norming procedures. For the first survey, 100 Mechanical Turk participants gave two words for each head or local noun prompt. (5) gives an example of the five tested nouns from the category coordinate preamble, *The saw by the oily wrench/wrenches/rag/rags*, as in Table 10. Likewise, five nouns from each attribute or associate preamble from the semantic-associative relationship items were tested. Each of the five nouns was assigned to a separate list; each list had four randomized presentation orders, resulting in a total of 20 unique lists with 89 items each.

(5)	SAW	*Head noun*
	WRENCH	*Singular related local noun*
	WRENCHES	*Plural related local noun*
	RAG	*Singular unrelated local noun*
	RAGS	*Plural unrelated local noun*

Forty responses were obtained for each noun, and response percentages were calculated as the number of times out of all responses to a given prompt word that a particular local noun was provided in response to a head noun prompt, or that the head noun (in either a singular or plural form) was given in response to a local noun prompt, where both nouns in question were part of the same preamble. Means were taken of arcsine transformed proportions for all head-to-local (H→L) noun and local-to-head (L→H) noun combinations; however, as in Experiment 1, only the former were used to select critical items.

Category coordinate preambles were designed such that the head noun was not expected to elicit either the related or unrelated local noun, and in fact, none of the local nouns was ever elicited in response to its corresponding head noun. Data from the semantic-associative relationship preambles are discussed in the context of the second survey, in which a separate group of 100 Mechanical Turk participants provided responses for the five nouns derived from each version of an attribute or unrelated association preamble. In addition to these items, a subset of critical items from the experiments in Solomon and Pearlmutter (2004) were included to ensure that participants were exposed to a sufficient range of integration, and so that additional analyses could be conducted in the future. The ratings for these items are not discussed here. As in the first survey, the five tested nouns from the different versions of the attribute and unrelated preambles were also assigned to five separate lists, and each list had four randomized presentation orders, resulting in a total of 20 unique lists. The lists contained 89 items, 24 of which were various versions of the semantic-associative preambles.

Forty responses were obtained for each noun, and the same set of calculations described for the first survey was performed on the data from the second survey. Table 18 shows the back-transformed means and standard deviations for each head-to-local noun combination for the 24 semantic-associative critical items. The results indicate that attribute nouns were elicited more than associates, and both were elicited more than unrelated local nouns, which were never elicited.

Table 18

Experiment 2 Semantic-Associative Relationship Critical Item Mean Head-to-Local Noun Association Proportions

	Relationship		
Local noun number	Attribute	Associate	Unrelated
Singular	0.16 (0.20)	0.09 (0.12)	0.00
Plural	0.08 (0.15)	0.04 (0.13)	0.00

Note. Means of the arcsine transformed proportions of responses were calculated for analyses. The means shown here are back-transformed proportion of local noun responses given in response to a head noun prompt. Standard deviations are shown in parentheses.

Procedure and Scoring. The procedure and scoring were exactly the same as in Experiment 1.

Results

Across all 6,241 critical trials, there were 3698 correctly-inflected responses, 118 agreement errors, 921 uninflected responses, 1438 miscellaneous cases, and 66 trials with no response. The percentage of error responses (out of errors plus correct responses) and the numbers of each of the other response types by condition are shown in Table 19.

Table 19

Experiment 2 Error Rates (%) and Response Counts by Condition

			Response count			
Condition	Local noun number	Error rate	Error	Correct	Uninflected	Misc
Related	Plural	5.7 (1.2, 1.4)	26 (3)	426 (31)	124 (8)	196
Related	Singular	0.0 (0.0, 0.0)	0 (0)	448 (30)	126 (11)	194
Unrelated	Plural	6.2 (1.2, 1.8)	26 (0)	391 (23)	138 (13)	216
Unrelated	Singular	0.6 (0.3, 0.3)	3 (2)	466 (30)	127 (22)	175
Attribute	Plural	5.4 (1.5, 1.5)	17 (3)	298 (19)	71 (5)	129
Attribute	Singular	0.0 (0.0, 0.0)	0 (0)	319 (16)	78 (5)	119
Associate	Plural	10.1 (1.7, 1.9)	34 (4)	303 (26)	59 (3)	120
Associate	Singular	0.5 (0.3, 0.3)	2 (0)	361 (25)	63 (5)	89
Unrelated	Plural	2.2 (1.1, 0.9)	8 (0)	348 (17)	57 (4)	103
Unrelated	Singular	0.6 (0.4, 0.4)	2 (0)	338 (17)	78 (9)	97

Note. Error rates reported as percentage errors out of errors plus correct responses; in parentheses are standard errors of the mean computed from the analyses by participants and by items. Response counts include dysfluency counts in parentheses. Misc=Miscellaneous responses.

As in Experiment 1, analyses were performed on error rates, calculated as error responses out of errors plus correct responses, the number of uninflected responses, and the number of miscellaneous responses for category coordinates and association items. Category coordinates were subjected to two separate 2 (local noun number) x 2 (category category status related vs. unrelated) ANOVAs, one with participants (F_1) and the other with items (F_2) as the random factor. Semantic-associative items were subjected to two separate 2 (local noun number) x 3 (attribute vs. associate vs. unrelated) ANOVAs, one with participants (F_1) and the other with items (F_2) as the random factor. Analyses were conducted using arcsine transformed error proportions in both by-subject and by-item ANOVAs. Except where noted, the statistical patterns reported below were identical for analyses on both raw and transformed error proportions; they were also identical when the same analyses were performed on error counts rather than error proportions, and when dysfluency cases were excluded. The statistics reported are for analyses on untransformed error proportions with dysfluency cases included. Figure 2 shows mismatch effects presented as difference scores of error rates for plural local nouns and corresponding singular local noun conditions.

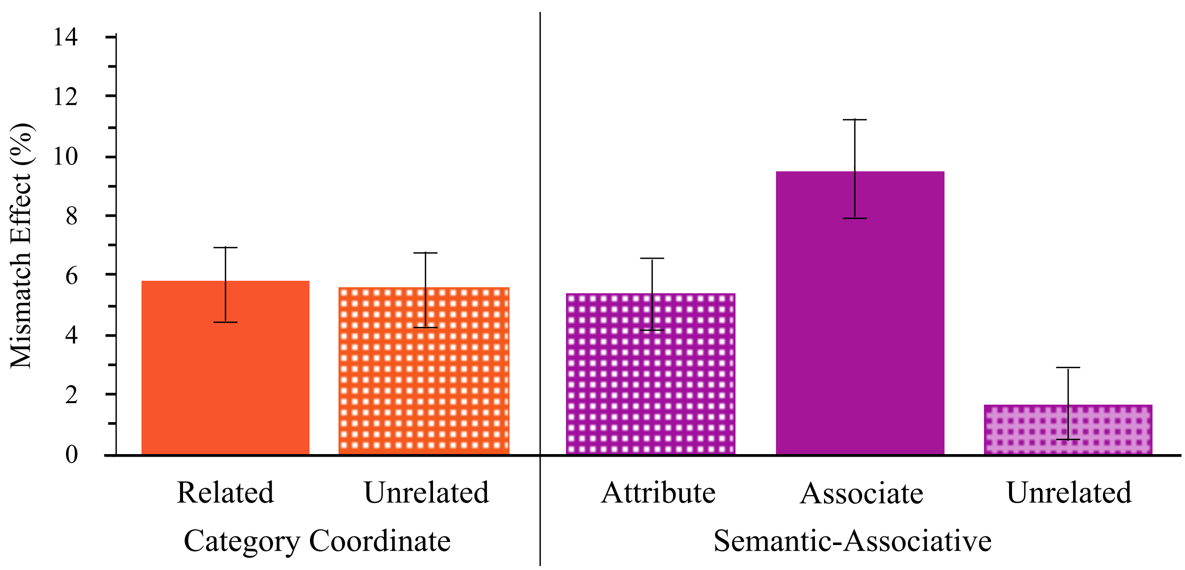

Figure 2. Mismatch effects by condition. Error bars show ± 1 *SEM*, computed from the analyses by participants.

Table 20 shows the results of the ANOVAs on category coordinate error responses. The analyses revealed the expected main effect of local noun number, with participants making more agreement errors for preambles containing plural local nouns than for their singular counterparts. There was no main effect of relatedness, nor any hint of a two-way interaction between relatedness and local noun number. Table 21 shows the results of the ANOVAs on category coordinate uninflected and miscellaneous responses. None of these analyses revealed any significant main effects or interactions.

Table 20

Experiment 2 Error Rate ANOVA Results for Category Coordinates

Effect	*F1*	*MSe*	*F2*	*MSe*
Local Noun Number	31.63	122.10 ***	14.19	46.23 **
Related	0.47	74.74	0.09	19.74
Local Noun Number x Related	0.01	66.33	0.14	20.50

Note. The statistics reported are for analyses on untransformed error proportions with dysfluency cases included. For F1 (by-subjects) analyses, all between-groups *df*=1, and within-groups *df*=128; for by-items analysis, all between-groups *df*=1, and within-groups *df*=23.
**p<.01, ***p<.001

Table 21

Experiment 2 Uninflected and Miscellaneous Response ANOVA Results for Category Coordinates

Effect	*F1*	*MSe*	*F2*	*MSe*
Uninflected Responses				
Local Noun Number	0.30	0.53	0.28	3.04
Relatedness	0.65	0.67	0.36	6.45
Local Noun Number x Relatedness	0.45	0.73	0.31	5.74
Miscellaneous Responses				
Local Noun Number	2.67	1.34	2.60	7.41
Relatedness	0.00	0.78	0.00	6.94
Local Noun Number x Relatedness	2.69	1.10	2.24	7.08

Note. The statistics reported are for analyses on counts of both uninflected responses (with dysfluency cases included) and miscellaneous responses. For by-subjects analysis, all between-groups *df*=1, and within-groups *df*=128; for by-items analysis, all between-groups *df*=1, and within-groups *df*=23.

Table 22 shows the results of the ANOVAs on semantic-associative relationship error responses. The analyses revealed a main effect of local noun number, as well as a clear effect of relatedness. The interaction pattern was examined with a series of 2 x 2 ANOVAs including as factors local noun number and the various possible pairings of the three levels of relatedness. The details of the local noun number main effects are not described here.

Table 22

Experiment 2 Error Rate ANOVA Results for Semantic-Associative Relationships

Effect	*F1*	*MSe*	*F2*	*MSe*
Local Noun Number(a)	31.85	184.64 ***	28.32	40.87 ***
Relatedness	7.41	95.56 ***	7.18	20.53 ***
Local Noun Number x Relatedness	6.25	114.32 **	7.14	21.76 **

Note. The statistics reported are for analyses on untransformed error proportions with dysfluency cases included. Except as noted, for *F1* (by-subjects analyses), all between-groups *df*=2, and within-groups *df*=256; for *F2* (by-items) analyses, all between-groups *df*=2, and within-groups *df*=46.
(a)For *F1*(by-subjects), between-groups df=1, and within-groups *df*=128; for *F2* (by-items) analyses, between-groups *df=1*, and within-groups *df=23.*
**$p<.01$, ***$p<.001$

Comparing attributes and associates, there was a main effect of relationship, with associates generating more errors than attributes ($F_1(1,128)=4.81$, $MS_e=121.83$, $p<.05$; $F_2(1,23)=4.68$, $MS_e=22.02$, $p<.05$). There was also a marginally significant interaction between relatedness and local noun number ($F_1(1,128)=2.95$, $MS_e=123.57$, $p<.10$; $F_2(1,23)=2.15$, $MS_e=27.44$, $p>.10$), the mismatch effect being greater for associates ($F_1(1,128)=26.17$, $MS_e=193.01$, $p<.001$; $F_2(1,23)=20.92$, $MS_e=47.75$, $p<.001$) than for attributes ($F_1(1,128)=14.14$, $MS_e=137.53$, $p<.001$; $F_2(1,23)=15.87$, $MS_e=26.97$, $p<.001$).

Associates also generated more agreement errors than unrelated items ($F_1(1,128)=12.72$, $MS_e=107.95$, $p<.001$; $F_2(1,23)=11.88$, $MS_e=24.51$, $p<.05$), and there was a local noun number by associate interaction ($F_1(1,128)=11.84$, $MS_e=120.55$, $p<.05$; $F_2(1,23)=14.88$, $MS_e=20.76$, $p=<.05$)

where the head-local mismatch effect size was greater for associated preambles relative to unrelated preambles ($F_1(1,128)=3.76$, $MS_e=82.75$, $p<.10$; $F_2(1,23)=4.54$, $MS_e=9.85$, $p<.05$).

A comparison of attribute and unrelated items revealed a marginal main effect of relationship as related items generated more errors than unrelated items ($F_1(1,128)=2.89$, $MS_e=56.89$, $p<0.10$; $F_2(1,23)=3.17$, $MS_e=15.07$, $p<0.10$. There was also an interaction that was significant by items and marginally significant by subjects ($F_1(1,128)=3.54$,$MS_e=98.85$, $p<.10$; $F_2(1,23)=5.74$, $MS_e= 17.07$, $p <.05$), with larger mismatch effects for attributes relative to unrelated items.

The results of the ANOVAs on uninflected and miscellaneous responses are presented in Table 23. Analysis of uninflected responses revealed an effect of local noun number that was significant by items and marginally significant by subjects, with fewer uninflected responses for plural items than for singular items.

Analysis on counts of miscellaneous responses revealed a main effect of local noun number, significant by subjects and marginally significant by items, with fewer miscellaneous responses for singular versus plural items. There was also a main effect of relatedness, significant by subjects but not items, with more miscellaneous responses for related versus unrelated items.

Table 23

Experiment 2 Uninflected and Miscellaneous Responses ANOVA Results for Semantic-Associative Relationships

Effect	*F1*	*MS*	*F2*	*MS*
Uninflected Responses				
Local Noun Number	3.46	0.38 †	4.83	1.47*
Relatedness	1.48	0.48	1.15	3.29
Local Noun Number x Relatedness	0.73	0.44	0.66	2.60
Miscellaneous Responses				
Local Noun Number	4.14	0.69 *	3.67	4.18†
Relatedness	4.04	0.63 *	2.27	5.96
Local Noun Number x Relatedness	1.33	0.53	0.99	3.92

Note. The statistics reported are for analyses on counts of both uninflected responses (with dysfluency cases included) and miscellaneous responses. For *F1*(by-subjects) analyses, all between-groups *df*=2, and within-groups *df*=256; for *F2* (by-items) analyses, all between-groups *df*=2, and within-groups *df*=46.
†p<.10, *p<.05

Discussion

The purpose of Experiment 2 was to investigate how differences in semantic and associative relationship type might manifest as differences in mismatch effect sizes. Barker et al. (2001) used category coordinates and noun pairs with feature overlap in their critical items, and they reported larger mismatch effects for related items versus unrelated items. The mismatch effects for our category coordinate items and their unrelated controls did not differ.

One possible explanation for this outcome is that our manipulation may not have been strong enough. Our norming data indicated that related items were rated as more related than unrelated items, while being very closely matched on integration, and completely matched on association. However, it is possible that the difference in relatedness between the related and unrelated versions was not large enough to see measurably different influences on agreement. Also, while related noun pairs were selected from published norms (Van Overschelde et al., 2004) to represent typical members of the same taxonomic category, we did not obtain separate ratings to measure the extent of category coordinate status for related head and local nouns. It is possible that the underlying semantic relationship of the critical item noun pairs may have gone unnoticed– for example, if they were not readily perceived as representative members of a given taxonomic category.

The failure to increase error effects may also be attributable to the fact that category coordinate critical items in Experiment 2 were also unassociated. Again, it is an open question in the priming literature as to whether coordinates also need to be associated in order to produce significant effects, with Lucas (20001) contending in her meta-analysis that they do not, and Hutchinson (2002) contending that they do. The results of Experiment 2 are consistent with the latter claim, applied to effects on agreement.

Our results also failed to replicate Barker et al. (2001). However, without more information on their specific stimuli, we can only speculate as to what might have contributed to their findings. First, the possibility that their critical item noun pairs were associatively related in addition to being semantically related cannot be ruled out. Second, their critical items may have involved other types of semantic relationships. Finally, their relatedness factor might have been confounded with additional factors, such as semantic integration.

Our analyses also showed a clear effect of association, with attributes and associates both generating mismatch effect sizes greater than the unrelated preambles. In Experiment 1, relatedness and association were confounded, and it was not possible to distinguish the effects of one factor on agreement from the other. Experiment 2, on the other hand, demonstrates that the mismatch effects are the result of association, and not semantic relatedness. Associated versions of the Experiment 2 preambles contained noun pairs for which no strong semantic relationship was present; attribute pairings were both semantically and associatively related, but there were not additive effects of attribute status above and beyond the effects of association.

General Discussion

Experiments 1 and 2 were designed to investigate the effects of message-level properties on agreement during language production, and specifically, properties unrelated to number meaning. The first experiment noted similarities between two factors that have both separately been shown to increase mismatch effect sizes. One factor, semantic integration, is operationally defined as a tight conceptual relationship between elements of a to-be-uttered message, with the context within which those elements appear being critically important in creating integration (e.g., Solomon & Pearlmutter, 2004). The second factor, semantic relatedness, is defined as a context-independent relationship between words based on semantic properties of the words (e.g., Barker et al., 2001).

Experiment 1 showed that, while higher semantic integration ratings corresponded to numerically larger mismatch effects, only the relatedness manipulation reliably increased mismatch effects. Experiment 2 pursued the question of relatedness's effects on agreement by separately manipulating different semantic and associative relationships. We found that the semantically related but unassociated category coordinates did not increase mismatch effects compared to unrelated controls. Attribute items, which were semantically and associatively related, did not produce effects beyond their corresponding associate controls, which lacked a semantic relationship, but both associates and attributes proceeded greater mismatch effects than the unrelated preambles.

Both experiments provide some insight into the differences in effects on agreement of semantic integration, semantic relatedness, and association, but to better understand the relationship among these factors, additional investigation is needed. Specifically, we are currently in the process of preparing our data so that they can be submitted to a series of logit regressions. A logit regression is a linear regression on the "log odds" transform of the probability of an error at each point along the regression line. Analyses using logit regressions of this kind will enable us to consider the relative contribution of each our predictor variables, as well as other factors such as animacy and plausibility, both alone in an various combinations; we expect these analyses to be particularly informative with respect to two issues.

First, the norming data from Experiment 1 clearly showed that integration ratings were influenced by relatedness, with more related noun pairs yielding higher rated integration. For example, the related-unintegrated preambles were rated as more integrated than unrelated-integrated preambles. While we were nevertheless able to manipulate integration independently, the manipulation itself was weaker than we had originally intended. Hence, we may have failed to detect an effect of integration due to a lack of power, and a follow up study seeking to examine relatedness and integration would need to attempt a stronger manipulation. Experiment 2's norming showed similar (albeit weaker) evidence of the confound: Related category coordinate preambles were rated as slightly more integrated than their unrelated counterparts; and, both attributes and associates were rated as slightly more integrated than their unrelated controls, with a small difference in integration between the former favoring attributes over associates. The logit regressions should provide information about the influence of integration independent of relatedness and other factors.

Second, while differences in relatedness, integration, and association closely matched the pattern of agreement errors in Experiment 1, in Experiment 2, this was not the case. With respect to category coordinates, although related items were rated as more related than unrelated items, there was no different in mismatch effects. For the semantic-associative relationships, attributes were rated as more related and were also more associated. This was entirely consistent with the fact that attributes were selected to be both semantically and associatively related. The largest mismatch effects were not generated by attributes, however, but by associates. In this case, the logit regressions might help explain why these patterns were obtained.

The results of suggest that, at least for these experiments, association between the head and local noun is most important predictor of mismatch effects. As was noted in the introduction of Experiment 2, associations are based on probabilities of co-occurrence, and while there is no question that many semantically related items are also associated, the reverse is not necessarily true. Consequently, a better way of reexamining the previously reported effects of semantic integration would be to consider the role that association, rather than semantic relatedness, might have played in Solomon and Pearlmutter's (2004) Experiment 4. We have begun collecting data to this end, it appears to be the case that, with few exceptions, the majority of critical items in that experiment did not contain associated head-local noun pairs. This implies that no special pre-existing relationship between the elements of a message is required for integration to produce its effects; this provides some insight into what semantic integration *is not,* but clearly, only more research will be able to say what it *is*.

Finally, a separate empirical question that we have begun pursuing is whether the timing based mechanism proposed by Solomon and Pearlmutter (2004) to account for the effects of integration on agreement is applicable with respect to the effects reported in the present experiments. Following the methods of Gillespie et al. (2013), our lab has selected 32 recordings of Experiment 1 participants on the basis of their having the greatest number of correct, error, or uninflected responses, as well as the fewest dysfluencies, for critical items; these are currently being analyzed using Praat software (Boersma & Weenink, 2014) to collect information about spoken word and pause durations. As with Gillespie et al., if the duration of words intervening between related head and local nouns is shorter than for words between unrelated head and local nouns, this would be taken as direct evidence for semantic or associative relationships increasing the rate at which sentence elements are planned.

References

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 285-375). doi:10.1016/B978-012369374-7/50010-9

Barker, J., Nicol, J., & Garrett, M. (2001). Semantic factors in the production of number agreement. *Journal of psycholinguistic research, 30*, 91-114. doi: 10.1023/a: 1005208308278

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of experimental psychology: Learning, memory, and cognition, 12*, 575-586. doi: 10.1037/0278-7393.12.4.575

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. doi: 10.1016/0749-596x(92)90007-k

Bock, J. K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive psychology, 43*, 83-128. doi:http://dx.doi.org/10.1006/cogp.2001.0753

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.

Bock, J. K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of memory and language, 40*, 330-346. doi:http://dx.doi.org/10.1006/jmla.1998.2616

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.84, retrieved 26 August 2014 from http://www.praat.org/

Brehm, L., & Bock, J. K. (2013). What counts in grammatical number agreement? *Cognition, 128*, 149-169. doi:10.1016/j.cognition.2013.03.009

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods,* 41, 977-990.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science,* 6, 3-5.

Butterworth, B. (1980). Some constraints on models of language production. In B. Butterworth (Ed.) *Language Production: Volume 1: Speech and Talk* (pp. 423-459). London: Academic Press.

Caramazza, A., & Miozzo, M. (1998). More is not always better: A response to Roelofs, Meyer, and Levelt. *Cognition, 69*, 231-241.

Collins, A.M., & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321. doi: 10.1037/0033-295x.93.3.283

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of verbal learning and verbal behavior, 20*, 611-629. doi: 10.1016/s0022-5371(81)90202-4

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.

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in english. *Journal of memory and language, 41*, 560-578. doi:http://dx.doi.org/10.1006/jmla.1999.2662

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological review, 112*, 531-559. doi:10.1037/0033-295x.112.3.531

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.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & cognition, 5*, 335-339. doi:10.3758/BF03197580

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and cognitive processes, 17*, 371-404. doi:10.1080/01690960143000254

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.

Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject–verb agreement production. *Cognition, 118*, 377-397. doi:10.1016/j.cognition.2010.10.008

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. doi:10.1037/a0029005

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

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2005). Semantic competition between hierarchically related words during speech planning. *Memory & cognition, 33*, 984-1000. doi:10.3758/BF03193207

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic bulletin & review, 10*, 781-813. doi:10.3758/BF03196544

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic bulletin & review, 7*, 618.

Lucas, M. (2001). Essential and perceptual attributes of words in reflective and on-line processing. *Journal of psycholinguistic research, 30*, 605-625. doi: 10.1023/a:1014283106728

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). New York: Cambridge University Press.

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. doi:10.1037/h0031564

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of experimental psychology: Learning, memory, and cognition*, *21*, 863-883. doi: 10.1037/0278-7393.21.4.863

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

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.

Roelofs, A., Meyer, A. S., & Levelt, W. J. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition, 69*, 219-230.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology,* 8, 382-439.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior research methods, instruments, & computers,* 20, 206-217.

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. doi: 10.1016/0749-596x(90)90011-n

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of experimental psychology: Learning, memory, and cognition, 18*, 1191-1210. doi: 10.1037/0278-7393.18.6.1191

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review, 96*, 523.

Sloman, S., & Malt, B. (2003). Artifacts are not ascribed essences, nor are they treated as belonging to kinds. *Language and cognitive processes,* 18, 563-582.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive psychology, 49*, 1-46. doi:10.1016/j.cogpsych. 2003.10.001

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of memory and language, 48*, 740-759. doi: http://dx.doi.org/10.1016/ S0749-596x(03)00003-2

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of memory and language, 50*, 289-335. doi: 10.1016/j.jml.2003.10.003

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of memory and language, 34*, 186-215.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and cognitive processes*, *11*, 407-442. doi:10.1080/016909696387169

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. doi:10.1037//0278-7393.28.1.46