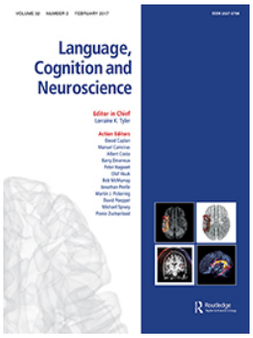

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: http://www.tandfonline.com/loi/plcp21

Referential and lexical forces in number agreement

Laurel Brehm & Kathryn Bock

To cite this article: Laurel Brehm & Kathryn Bock (2017) Referential and lexical forces in number agreement, Language, Cognition and Neuroscience, 32:2, 129-146, DOI: 10.1080/23273798.2016.1234060

To link to this article: http://dx.doi.org/10.1080/23273798.2016.1234060

LANGUAGE, COGNITION AND NEUROSCIENCE, 2017
VOL. 32, NO. 2, 129–146
http://dx.doi.org/10.1080/23273798.2016.1234060

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Referential and lexical forces in number agreement

Laurel Brehm[a] and Kathryn Bock[b]

[a]Department of Linguistics, Northwestern University, Evanston, IL, USA; [b]Department of Psychology, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

ABSTRACT
In work on grammatical agreement in sentence production, there are accounts of verb number formulation that emphasise the role of whole-structure properties and accounts that emphasise the role of word-driven properties. To evaluate these alternatives, we carried out two experiments that examined a referential (wholistic) contributor to agreement along with two lexical-semantic (local) factors. Both experiments gauged the accuracy and latency of inflected-verb production in order to assess how variations in grammatical number interacted with the other factors. The accuracy of verb production was modulated both by the referential effect of notional number and by the lexical-semantic effects of relatedness and category membership. As an index of agreement difficulty, latencies were little affected by either factor. The findings suggest that agreement is sensitive to referential as well as lexical forces and highlight the importance of lexical-structural integration in the process of sentence production.

ARTICLE HISTORY
Received 23 November 2015
Accepted 22 August 2016

KEYWORDS
Language production; number agreement; referential integration; attraction; lexical interference

Highlights

- Theories of sentence production differ in their emphasis on structural and lexical information.
- For subject–verb number agreement, these theories are captured in the Marking and Morphing model and in mechanisms of lexical competition.
- Two experiments varied referential (structurally mediated) and lexical factors to assess effects on the speed and accuracy of number-inflected verb production.
- Results show that both processes are independently critical for agreement.

Introduction

Grammatical agreement is a linguistically productive process that demands the integration and reconciliation of potentially contradictory conceptual, grammatical, lexical, morphological, and phonological information. The usage of agreement can and does deviate from what a speaker intends due to all of these factors, leading to departures from conventionally correct agreement and differences in the fluency of the utterances produced (e.g. pauses, lengthening, and other differences in the speed of speech). These variations offer clues to the cognitive and linguistic mechanisms required for language production. The question in the present work is how speakers compose conceptual, lexical, and structural information to produce agreement. We examine this by measuring variations in the speed and accuracy of agreement (compared to conventionally correct agreement) due to all of these factors. These measured variations show how information is prioritised during sentence production.

Broadly speaking, the process of grammatical agreement involves linking an agreement *controller* (in the current sentence the controller is the subject, *the process of grammatical agreement*) and an agreement *target* (in the current sentence, the verb *involves*). The product of the linkage is that a controller and its target index the same value of an agreement *feature*, such as gender, person, or number. Subject–verb number agreement is a familiar form of grammatical agreement in English, where typical controllers are nouns and typical targets include verbs, pronouns, and determiners.

Deviations from standard agreement patterns have fuelled considerable psycholinguistic research. One major deviation is termed *attraction*, in which the source of a verb's number is not the number of the subject per se but the number of another noun proximal to the verb (e.g. Bock & Miller, 1991). An illustration of attraction is a line from the talk-show host Conan O'Brien, who exclaimed "*The back of my pants* ***are*** *falling off*" (*Conan*, episode 201, 23 January 2012). Here,

CONTACT Laurel Brehm laurel.brehm@gmail.com
Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/23273798.2016.1234060

the subject is the entire phrase *the back of my pants*. Even though the subject is grammatically singular and the subject phrase's referent is a single object (in technical terms, a *notional* singleton), singular number fails to materialise on the target verb. Instead, the verb is plural, seemingly reflecting the plural grammatical number of a subcomponent of the subject (*my pants*). The position of *my pants* in the example typifies the erroneous number source in attraction errors, a noun phrase in the neighbourhood of the verb but not necessarily immediately preceding it. The attractor is often called the *local noun* for the sake of simplicity.

Many variations in number agreement are errors: deviations from the normative, canonical pattern that are truly unintended. Similar to classic types of speech errors, these variations in agreement can be conceptualised as failures of production processes to accurately encode the speaker's meaning due to a linguistic misfire in selecting the words, structures, or morphology to be used in the utterance (Dell, 1986; Garrett, 1975, 1980). Evidence for this is that speakers can recognise unintended deviations in number agreement and sometimes correct them overtly, as in

> "Any of these alternative classifications are … is … "

> "At least as far as this data … these data are concerned … "

> "The breaking of relations in themselves … in itself … "

(examples from Clark & Fox Tree, 2002; Clark & Wasow, 1998; and personal observation).

Viewed as an *error* process, attraction is the product of interference between grammatically plural nouns and grammatically singular sentence subjects (e.g. Eberhard, Cutting, & Bock, 2005; Franck, Lassi, Frauenfelder, & Rizzi, 2006). There need be little involvement of number meaning: speakers ordinarily construe the referent of *pants* as a notional singleton (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001), yet its grammatical plurality causes attraction. In contrast, though *clothing* refers to a notional plural, it fails to attract (Bock & Eberhard, 1993; Bock, Eberhard, & Cutting, 2004) or attracts only weakly (Haskell & MacDonald, 2003).

While number meaning is not a necessary driver of attraction, differences in number meaning can nonetheless create variations in number agreement. These variations are distinguishable from those that are caused by interference during the planning process in that they *do* capture the speaker's intended meaning. This type of variation can be termed *notional agreement*; we walk through several examples of notional agreement below in order to set meaning-driven variations in number agreement apart from those driven by interference.

One common example of notional agreement occurs with collective nouns. Collective nouns (e.g. *gang*, *team*) are canonically treated as grammatically singular in American English, but plural agreement can occur when the referent is construed as a notional plural. A collective sentence subject like *The gang* ***on*** *the motorcycles* is easily construed in terms of distinct individuals, whereas *The gang* ***near*** *the motorcycles* is more likely to be construed as a single group. Consistent with this notional difference, the former is more likely than the latter to elicit plural verb agreement (Humphreys & Bock, 2005). Similar effects occur with respect to distributivity. The phrase *The test for the students* can represent two construals, either a single test type (e.g. the abstract content of the exam) or several tokens of one type (e.g. a stack of exam printouts). The latter construal is distributive and elicits increased plural notional agreement (Eberhard, 1999; Vigliocco, Butterworth, & Semenza, 1995). A final example of notionally sensitive agreement comes from conjunctions of singular nouns. Canonically, conjoined noun phrases take plural agreement, but singular agreement may occur when the referents have a natural singleton construal (e.g. weather: *sleet and freezing rain*; street intersections: *State and Lake*; linked actions denoted by deverbal nouns: *drinking and driving*; see Lorimor, 2007).

In addition to the agreement variations that appear as overt mismatches between subject and verb number, there are other agreement variations are connected with fluency. These reflect changes in *when* words are produced rather than *what* is produced. The finding is that some instances of conventionally correct agreement are accomplished faster than others (e.g. Brehm & Bock, 2013; Haskell & MacDonald, 2003; Staub, 2009). As with the overt variations reviewed above, these changes in fluency reflect how words and structures are accessed and combined in preparation to produce an utterance.

There are several accounts of agreement production that aim to explain the accuracy and timing variations reviewed above in terms of general language production mechanisms. These accounts can be broadly classified as *lexical* approaches or *structural* approaches (Bock & Ferreira, 2014). From a lexical perspective, agreement variability stems from variations in how words are produced, capturing effects due to word-driven properties, while from a structural perspective, agreement variability stems from variations in how structured units larger than words are assembled, capturing effects due to phrase- and message-level properties. A schematic of differences between these is provided in Figure 1, with word-centred processes laid out on the left, and structure-centred processed laid out on the right. Importantly, while these accounts are not mutually exclusive, they do

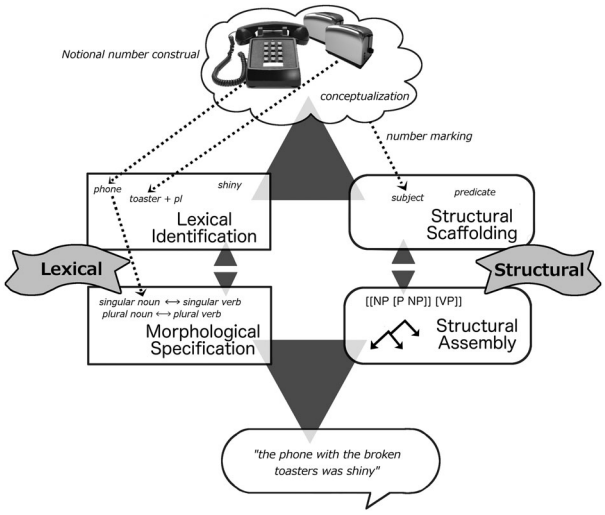

Figure 1. Hypothesised sources of lexical and structural difficulty in agreement. In this sketch, number information (dashed lines) passes from a conceptual level to sentence formulation processes. Plural notional number affects the abstract structural subject's number in parallel with lexical identification and morphological specification, which encode lexical concepts in words with appropriate grammatical number. Other processes are omitted for simplicity.

serve to carve out major lines of psycholinguistic debate. The next section sketches how the two play out in current accounts of agreement.

Lexical and structural sources of agreement trouble

Lexical sources

On a lexical account, agreement attraction is explained in terms of difficulties in selecting and retrieving words. The architecture of the relevant processes is schematised on the left side of Figure 1. One lexical account posits that in the course of retrieval, probabilistic morphological relationships between nouns and verbs support subject–verb number agreement (e.g. Haskell & MacDonald, 2003). Specifically, if agreement is driven by a speaker's experience in producing plural nouns with plural verbs, the retrieval of a plural noun can increase the probability of producing a plural verb regardless of the structural relationship between the noun and verb. The result is that a verb's number may deviate from the value that would be expected in light of what the speaker intended to convey.

This interpretation of agreement production difficulty is related to a classic view of attraction in which the speaker loses track of the intended agreement controller (Fowler, 1937) and to the currently dominant model of agreement comprehension (see Dillon, Mishler, Sloggett, & Phillips, 2013). The core of this account is that agreement errors occur when the target takes its value from the wrong element of the planned utterance, such as a noun or a phrase that is more accessible in memory. In production, the suggestion is that high-probability lexical associations between successive nouns (e.g. *pants*) and verb forms (e.g. *are*) mask the relationship between the intended subject and its predicate.

The findings that support a lexically controlled model of agreement revolve around a *content-addressable* memory system, where retrieval cues allow direct access of relevant or activated information from memory (e.g. McElree, 1996; McElree, Foraker, & Dyer, 2003; Van Dyke & Lewis, 2003). Such retrieval-based models have persuasively captured a variety of agreement phenomena, in comprehension (Dillon et al., 2013; Wagers, Lau, & Phillips, 2009) and production (Badecker & Kuminiak, 2007; see Haskell & MacDonald, 2003; Solomon & Pearlmutter, 2004; Thornton & MacDonald, 2003; Vigliocco et al., 1995; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco & Franck, 1999 for converging evidence).

A particularly illustrative series of experiments on the role of lexical factors in agreement production comes from Solomon and Pearlmutter (2004). In their account, a core mechanism of attraction is competition between singular and plural nouns during lexical retrieval in the course of sentence formulation. The likelihood of competition increases under circumstances that promote the parallel encoding of lexical information, allowing a noun other than the one that is conventionally associated with verb number to control agreement. A potential consequence of this is that the number of the local noun is reflected in the verb, rather than the number of the head noun of the subject phrase.

In these studies, Solomon and Pearlmutter manipulated a notional property that they expected to affect lexical retrieval. To emphasise its notional-number implications, we term this property *referential integration* (their term was *semantic integration*), where integration refers to the degree of linkage between the parts of a message in a discourse representation or mental model. Under the hypothesis that strong integration yields more overlap in the timing of retrieval for individual nouns, integration can create a lexically driven conflict that disrupts the usual relationship between noun and verb number. This was what they demonstrated: phrases constructed from well-integrated mental models, such as *The phone with the missing buttons*, elicited more attraction than their weakly integrated counterparts, such as *The phone with the broken toasters*, demonstrating the impact of lexically driven conflict on the outcomes on agreement.

Structural sources

In contrast to lexical (word-driven) accounts of agreement difficulty, what we term *structural* accounts of agreement difficulty hinge upon larger pieces of utterances (as schematised on the right of Figure 1). Structurally driven variations might derive directly from the utterance's syntactic structure or from variations in how the controller's number is calculated from the underlying representation of the speaker's message and its interface with syntax.

The supposition of a structural account of agreement is that errors arise in situations where the subject phrase remains a structured whole, with the intended structural relationship between subjects and predicates preserved. In this structured representation, the number feature of the subject is aberrant, and this aberrant number feature is transmitted to the target verb in the ordinary way. As such, the underlying problem is not getting the subject (the controller) wrong, but getting the subject's *number* wrong. Under a structure-centred account, this can occur because subject number is the product of several types of number information reconciled among the components of the subject as a structured whole, including notional, structural, and lexical number properties together (e.g. as illustrated in the *Marking and Morphing* account; Eberhard et al., 2005).

Supporting evidence for the role of structural factors in agreement production comes from the findings that the grammatically "wrong" verb number can result from properties associated with the subject phrase in its entirety. Among these whole-structure properties are notional number (the numerosity of a referent as construed by the speaker; Bock et al., 2004; Humphreys & Bock, 2005; Lorimor, 2007) and lexical-grammatical number modulated by structural (rather than linear) distances among constituents of the subject phrase (Bock & Cutting, 1993; Franck et al., 2006; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Vigliocco & Nicol, 1998). As these properties are based upon the interplay between larger segments of utterances, these phenomena are not well accounted for by the word-driven mechanisms of cued retrieval. This provides evidence that these structural contributions to agreement may be distinct from lexical ones (schematised on the right of Figure 1).

The differing predictions from lexical and structural accounts are nicely illustrated in how referential integration plays out in agreement. As reviewed in the previous section, referential integration creates lexical conflict. However, it also affects notional number. For example, one manipulation of integration capitalises on two different meanings of the preposition *with* (Solomon & Pearlmutter, 2004; Experiment 4). A well-integrated sentence subject uses the attributive *with* (*The phone with the missing buttons*) and describes a referent that is clearly a single individual, while a weakly integrated sentence subject uses the accompaniment sense of *with* (*The phone with the toasters*) and describes a complex set of referents with a semantic reading similar to that of the conjunction *and*. The prediction is then that integration would drive variations in agreement production such that weakly integrated subjects – those with notionally plural referents – are more likely to control plural verbs. This prediction runs counter to the lexical prediction tested by Solomon and Pearlmutter (2004), and it finds support in experiments on both English (Brehm & Bock, 2013) and Dutch (Veenstra, Acheson, Bock, & Meyer, 2014). These notional effects cannot be easily explained by properties of the words in the subject phrase alone, demonstrating the importance of structural factors in agreement.

Disentangling lexical and structural mechanisms

The evidence reviewed in the previous sections suggests that lexical and structural information are both necessary for the normal implementation of number agreement. What is less obvious is how these different sources of information have their impact: Are agreement variations driven directly by lexical difficulty, directly by structural difficulty, or do the two types of difficulty interact?

In order to separate and trace the sources of agreement difficulty, our experiments were designed to assess covariations in the outcomes of agreement (singular or plural verb use) with measures of latency to produce number-agreeing verbs. We did this by manipulating factors attributable to whole-structure properties of sentence subjects (notional number; the number of referents of the sentence subject) and factors attributable to lexical properties of the same subjects (lexical-semantic relatedness between nouns and predicate adjectives in Experiment 1 and between subject nouns in Experiment 2). The grammatical number of local nouns was varied in order to create the conditions for attraction, conditions that are assumed in both lexical and structural accounts. This provides a tracer for the operation of factors that mediate the computation of agreement. To the degree that local-noun number interacts with referential factors (characteristic of whole subjects), there is support for structural control of agreement; to the degree that it interacts with lexical factors (characteristic of single words), there is support for lexical control of agreement.

We used a standard paradigm for looking at agreement in language production. On each trial during the experiment, participants were presented with a subject noun phrase (a *preamble*) that they had to complete as a full sentence. The preambles were designed to have notional and lexical properties that disrupted agreement. We adapted the paradigm to control the final word of the completion produced, a predicate adjective, and to allow measurement of the latency to produce verbs. For example, at the beginning of a trial a participant might see the adjective *ringing* and then hear the preamble *The phone with the missing button*. The task was to complete the preamble as a full sentence containing the adjective (e.g. *The phone with the missing button was ringing*), beginning as rapidly as possible after the preamble's offset. This allowed the verb and its number to vary. Typically, participants used a completion containing a copula verb (*is*, *are*, *was*, *were*) and the designated adjective (e.g. *ringing*).

In Experiment 1, we examined the interaction of local-noun grammatical number (which triggers attraction) with structurally mediated, referential number (from integration, a source of notional number) and lexical-semantically conveyed number (mediated by semantic relatedness) in determining agreement attraction. The structural approach to agreement predicts that whole-structure properties will be a strong force in agreement, exhibited in interactions between notional number and grammatical number. In contrast, the lexical approach to attraction predicts that variations in lexical processes will be the chief determinant of attraction, exhibited in interactions between grammatical number and lexical-semantic relatedness.

In Experiment 2, we tested the same predictions with the same paradigm but with another type of utterance in order to gain converging evidence of the processes behind number agreement. We again used manipulations of local-noun grammatical number, referentially conveyed notional number, and lexical-semantic relatedness, analogous to those in Experiment 1. Notional number was manipulated in terms of abstract vs. concrete conceptual combinations, and lexical-semantic relatedness was manipulated in terms of semantic category membership. Critically, the syntactic structure of the subject differed from Experiment 1, consisting of conjoined noun phrases. With conjunctions as subjects, lexical relationships can come into play more directly than in the structurally complex subjects used in Experiment 1. Furthermore, with conjoined subjects the conventionally correct verb number is plural, with the consequence that local-noun plurality should promote correct responses rather than errors. The question again was whether this tendency would be modulated more by referential or lexical sources of number information.

Experiment 1

The first experiment was designed to examine variations in number agreement due to lexical and referential factors. The variables examined were the sentence subjects' local-noun number, referential integration, and the lexical-semantic compatibility between the required predicate and the head and local nouns. Referential integration was designed to change the notional valuation of the subject phrase's referent, targeting structurally derived agreement variations. Lexical-semantic compatibility was designed to alter the degree of interference from the attractor local noun, targeting lexically derived agreement variations.

For example, a referentially integrated subject might be a phrase like *The phone with the missing button*[*s*], and might require the head-compatible predicate *ringing* (compatible with *phone* but not *button*) or the local-compatible predicate *plastic* (more compatible with *button* than *phone*) to elicit *The phone with the missing button*[*s*] *was/were ringing* or *The phone with the missing button*[*s*] *was/were plastic*. Alternatively, a referentially unintegrated subject might be a phrase like *The phone with the broken toaster*[*s*] requiring the predicates *ringing* (head-compatible) or *shiny* (local-compatible) to elicit *The phone with the broken toaster*[*s*] *was/were ringing* or *The phone with the broken toaster*[*s*] *was/were shiny*.

In this experiment, use of the singular verb *was* is conventionally correct, whereas *were* indicates an attraction error or a variation due to referential number. The measure of attraction is the difference between singular and plural local nouns in the frequency of using *were*; critically, the two accounts of agreement make differing predictions with regard to how strongly agreement is perturbed by variations in referential integration (referential) and predicate compatibility (lexical).

Method

Participants

In exchange for course credit or $7.00 compensation, 165 undergraduates from the University of Illinois participated in the experiment. Participants were excluded if they had fewer than 66% usable experimental trials ($N = 21$) or were non-native English speakers ($N = 3$). An additional 13 participants were excluded due to technical difficulties ($N = 4$) and counterbalancing errors ($N = 9$). This left 128 participants.

Equipment

Stimuli were presented using PsyScope X B53 (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Macintosh Mini computer with a 17-inch LCD flat-screen monitor. Audio was presented to participants over Koss headphones, and their speech was digitally recorded to a computer using a Sennheiser directional microphone run through a USB button box and Tube MP preamplifier. PsyScope recorded the latency of vocal response onsets through the button box.

Materials

There were 24 experimental items, all based on those with the highest integration differences in Solomon and Pearlmutter's ratings (2004, Experiment 4). These items were designed to serve as sentence subjects (*preambles*) and were made up of complex noun phrases. Preambles varied in integration, manipulated by using an attributive *with* (integrated) or an accompaniment *with* (unintegrated). All had singular heads and local-noun phrases that varied in grammatical number (singular, plural). Preambles were paired with predicate adjectives that differed in their likelihood of modifying the head (head compatible) or local-noun phrase (local compatible). This yielded eight versions of all 24 items varying in integration, compatibility, and local-noun number. See Table 1 for an example and Appendix 1 for a full list of stimuli.

To construct the lexical-semantic compatibility manipulation for each item, candidate adjectives were rated on a five-point Likert scale for their fit with the head and local nouns. For the head-compatible condition, selected adjectives were judged to be better modifiers of the head noun (e.g. *phone*) than either local noun (e.g. *button*, *toaster*). For the local-compatible condition, selected adjectives were judged to be better modifiers of the local noun (e.g. *button* or *toaster*) than the head noun (e.g. *phone*). The selections were made on the basis of paper-and-pencil ratings in which judges ($N =$ between 5 and 21 per item) assessed the likelihood of a given adjective modifying the accompanying noun (e.g. *How likely is it that a phone is ringing*?). Noun–adjective pairings were presented in lists in which no noun or predicate was repeated. Integration was fully balanced within these lists, but only the singular form of the local nouns was presented. Ratings were iterated until adjectives with the appropriate biases (head > local or local > head) were identified. For the final set of adjectives, the overall compatibility advantage for the head-compatible adjectives was 1.08 (the average difference in adjective compatibility ratings for the head vs. local noun); the overall compatibility advantage for the local-compatible adjectives was 1.14 (the average difference in adjective compatibility ratings for the local vs. head noun). Mean ratings of compatibility in each condition are shown in Table 2. Any differences between means larger than 0.80 fall outside of the margin of error and can be considered reliable.[1]

In addition to the experimental items, there were 61 filler stimuli. These were designed to increase the variability of sentence types shown in the experiment and to balance the positions of singular and plural nouns across items. Filler preambles included prepositional phrases, conjoined noun phrases, and simple noun phrases. Fillers were paired with predicate adjectives different from those in the critical trials, determined using acceptability judgments from the first author and a research assistant. Of the fillers, 14% took singular agreement, so that 42% of all stimuli required singular agreement.

Eight lists were created from the eight versions of each critical preamble and the fillers. Each list contained one version of each experimental item and all the filler items. List order was determined quasi-randomly, with fillers in fixed positions across lists and critical items assigned randomly to slots between fillers. Ordering was constrained so that no more than two experimental items appeared consecutively and no semantically similar items were adjacent. Experimental items were

Table 2. Mean lexical-semantic compatibility ratings for Experiment 1.

Predicate adjective	Preamble (e.g. "The phone with the missing button/broken toaster")		
	Head noun (e.g. "phone")	Integrated local (e.g. "button")	Unintegrated local (e.g. "toaster")
Head-compatible "ringing"	3.81	2.91	2.55
Integrated local-compatible "plastic"	2.55	3.52	–
Unintegrated local-compatible "shiny"	2.47	–	3.77

Note: Margin of error for differences between means = 0.80.

Table 1. Example stimuli from Experiment 1.

	Preamble	Head-compatible	Local-compatible
Integrated	The phone with the missing button(s) (head) (local)	ringing	plastic
Unintegrated	The phone with the broken toaster(s) (head) (local)	ringing	shiny

counterbalanced so that every item was represented once and only once on each list, with an equal number of item versions in each condition. Each list was also divided into two halves, with the order of the halves counterbalanced over participants, for a total of 16 experimental lists. Every list began with 30 of the filler items in order to form a covert practice block.

Integration and sensibility norming. Norms were also collected for the integration and sensibility properties of the final set of experimental items in paper-and-pencil tasks. To establish that the integration difference remained for all adjective-head combinations, 16 participants who did not participate in the main part of the experiment were asked to rate the integration of the preambles combined with their predicates on a seven-point Likert scale. For this task, participants saw the preambles for each of the 24 items in one of four versions, with integration and lexical-semantic compatibility fully crossed, and with only singular local nouns. Items were presented to participants in one of four lists, so that each participant received only one version of each item and an equal number of version types in each condition. Task instructions and examples were adapted from Solomon and Pearlmutter (2004) (see Appendix 2).

To check variations in sensibility for each preamble-adjective combination, sensibility norms were collected from 32 additional participants. In these, participants were asked to rate the likelihood of statements on a seven-point Likert scale. The same lists were used as in the integration norming. Full instructions are shown in Appendix 2.

Ratings from both tasks were analysed using a by-items repeated measures ANOVA; margins of error were calculated as described above. Table 3 shows the results. Integration ratings differed significantly for complete sentences in the integrated and unintegrated item versions (integrated $M = 5.98$, range 4.5–6.5; unintegrated $M = 2.42$, range 1.13–4.00), but not for complete sentences in the head-compatible-predicate and local-compatible-predicate item versions (head-compatible $M = 4.18$, range 3.00–5.13; local-compatible $M = 4.21$, range 3.38–5.63). The integration ratings of items collapsed across lexical-semantic compatibility were highly correlated with the ratings of Solomon and Pearlmutter ($r(46) = .89$), and the ratings for the head- and local-compatible item versions were highly correlated with each other ($r(46) = .90$). This suggests that for this set of materials, compatibility differences did not change the relative levels of integration of the completed sentences.

Table 3. Mean integration and sensibility norming ratings for Experiment 1.

Integration	Predicate	Integration	Sentence sensibility
Integrated "The phone with the missing button"	Head-compatible "ringing"	5.9	4.18
	Local-compatible "plastic"	6.05	5.06
Unintegrated "The phone with the broken toaster"	Head-compatible "ringing"	2.45	3.5
	Local-compatible "shiny"	2.38	3.83
Margin of error		1.1	1.39

Note: Bottom row contains the margin of error of differences between means for the rating.

For sensibility, the completed sentences had an average rating of 4.14, slightly above the midpoint of the scale. Within levels of integration, differences in whole-sentence sensibility between the head- and local-compatible predicates were roughly comparable, 0.88 for integrated and 0.33 for unintegrated, against a margin of error of 1.39. This indicates that the relative sensibility of the completed sentences was similar at both levels of integration, and should not compromise the integration manipulation. Concerns about overall sensibility are also attenuated by the fact that these items or similar ones have been used previously and normed for plausibility in various ways (e.g. Brehm & Bock, 2013; Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004).

Sensibility did vary jointly by compatibility and integration, but in an unsurprising way. Overall, the completed sentences for local-compatible integrated item versions were rated as more sensible than those for head-compatible unintegrated item versions, as a logical and unavoidable consequence of a predicate's relationship to parts and wholes. For an integrated referent, a property of one of its parts (e.g. of the local-noun phrase) must also be a property of the whole: *plastic button* is an attribute of *phone*. For an unintegrated referent, however, a property of a part is not necessarily a property of the whole: *shiny toaster* is not an attribute of *phone*.

Preamble recording. Audio stimuli were recorded in a quiet room on a Sennheiser directional microphone run through a Tube MP preamplifier. The talker was a woman from northern Illinois. She produced the phrases in the carrier phrase "*The next sentence is X disappeared yesterday*" (e.g. *The next sentence is the phone with the missing button disappeared yesterday*). These carrier phrases were removed from the audio files and stimuli were edited using Audacity to shorten the length of continuants and reduce pauses between words, in order to increase speech rate while keeping natural-sounding stimuli.

Procedure

The experimental procedure was a version of a preamble completion paradigm with two types of trials, standard

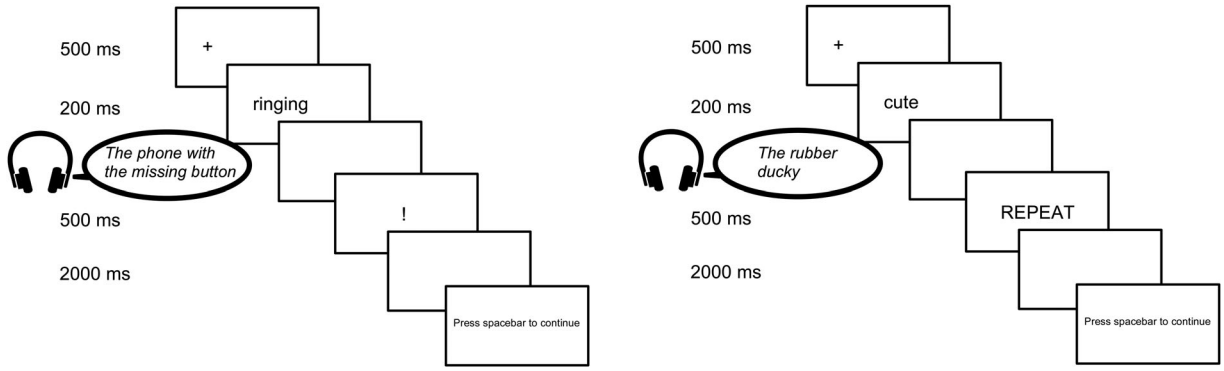

Figure 2. Trial diagram, for standard trials (left) and catch trials (right).

and catch trials (see Figure 2). The standard sequence (Figure 2(a)) was used for all experimental trials and a subset of the fillers. Trials began with a fixation cross presented for 500 ms. The cross was placed 10% of the screen width relative to the screen's left margin, midway between top and bottom. The predicate adjective then appeared in the same location for 200 ms in 36-point lowercase black Arial font. Next, the preamble was presented over headphones. Immediately after the offset of the preamble, the cue "!" appeared at the centre of the screen for 500 ms, prompting the participant to speak. A blank screen then appeared for 2 seconds, giving participants a total of 2.5 seconds to respond with a complete predicate (e.g. "*was ringing*").

Catch trials were included to encourage participants to attend carefully to preambles on every trial. The catch-trial sequence (Figure 2(b)) occurred on 31% of the filler trials. Events were the same as on standard trials up to the point at which the response was cued. At that point, participants were prompted to repeat the preamble before completing it with a predicate. This was signalled with the word *Repeat* in place of the exclamation point. After 500 ms a blank screen appeared and remained for 3.5 seconds, giving a total of 4 seconds for a response (e.g. "*The rubber ducky was cute*").

Participants were instructed to complete all preambles as quickly and accurately as possible. They received two explicit practice trials, one of each type, and were given the opportunity to adjust the volume of the audio to a comfortable level. Before starting the experiment, they were queried about their understanding of the procedure. The experimenter stayed in the room for the entire session.

Scoring

Responses on the critical trials were scored as valid, miscellaneous, or missing. Valid responses consisted of an inflected form of the copula (*was*, *is*, *were*, *are*) and the correct predicate adjective for the trial, where the utterance contained no additional modifiers (e.g. *very* or *really*) and had no disfluencies (uh, um) or non-speech noises (lip-smacks, coughs) before the verb. Valid responses were scored as singular or plural according to their verb number; only valid responses were submitted to analyses.

Design

Each participant received 1 of the 16 experimental lists. Each list contained 1 version of all 24 items (3 in every condition) and item order within lists was counterbalanced by list halves. Every list was presented to 8 participants, so that every version of each item was tested on 32 participants. The fixed effects in the statistical analyses were integration (integrated–unintegrated), lexical-semantic compatibility (head–local), and local-noun number (singular–plural), all fully crossed.

Analysis

Response types (accuracy), correct response latencies, and error response latencies were analysed statistically using multi-level models in R with the package lme4 (version 1.1-7; Bates, Maechler, Bolker, & Walker, 2014; R Core Team, 2014), with random intercepts for subjects and items (the maximal random effect structure justified by the data). Effects-coding (contrasts of 0.5, −0.5) was used for all fixed effects.

Results

Figure 3 summarises response types (accuracy) and correct latencies by condition; we follow the custom of describing all plural responses as "errors" to mark their deviation from the conventionally correct number inflection. In both the response-type and response-latency measures, robust effects of local-noun number were observed. Local plural nouns elicited effects indicative of attraction with slower, less accurate responses as compared to local singular nouns (local plural: 79% correct,

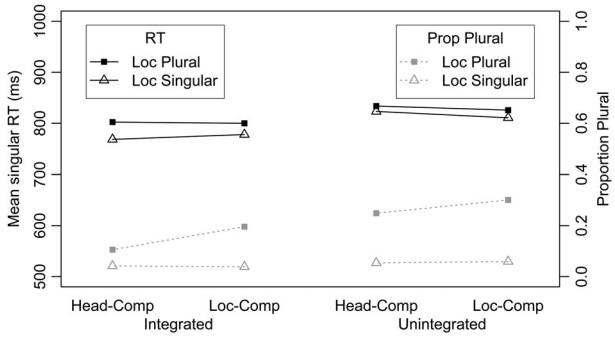

Figure 3. Experiment 1 singular verb production latencies (in ms; dark lines) and plural verb use (proportions; light lines) for preambles varying by integration, local-noun number, and predicate compatibility.

mean correct RT = 815 ms; local singular: 95% correct, mean correct RT = 795 ms). These were reflected statistically in mixed-effect analyses (see Table 4).

Predicted effects of notional number were also observed. Overall, integrated preambles elicited faster, more accurate responses than unintegrated preambles (integrated: 91% correct, mean correct RT = 786 ms; unintegrated: 84% correct, mean correct RT = 823 ms). Consistent with structural effects on agreement, this was further reflected in an interaction between integration and local-noun number in the response-type measure such that the most plural responses were elicited in the unintegrated-local plural condition. See Table 4 for mixed-effect analyses and Table 5 for mean latencies.

Error response latencies also showed an interaction of integration and local-noun number, but the effect fell short of statistical significance (see Table 4 for mixed-effect analyses and Table 5 for mean latencies). The pattern was that integrated (notionally singular) preambles elicited slower error responses with local plural than local singular nouns (integrated-local plural: M = 920 ms, SEM = 30 ms; integrated-local singular: M = 837 ms, SEM = 47 ms).[2] In contrast, unintegrated (notionally plural) preambles disclosed a smaller effect in the reverse direction (unintegrated-local plural: M = 849 ms, SEM = 20 ms; unintegrated-local singular: M = 878 ms, SEM = 46 ms). These effects are indicative of conflict between notional and grammatical number, the resolution of which is critical to the structural hypothesis.

Predicted effects of lexical-semantic compatibility were also observed for response types (89% correct for head-compatible vs. 85% correct for local-compatible predicates). Consistent with lexical effects on agreement, these patterns were particularly clear in combination with local-noun number (head-compatible, local singular 95% correct, head-compatible, local plural 82% correct; local-compatible, local singular: 95% correct; local-compatible, local plural, 75% correct). Mixed-effect analyses of these data are in Table 4.

However, despite the differences in accuracy, lexical-semantic compatibility elicited similar correct response latencies (local-compatible: 802 ms; head-compatible: 805 ms) and similar error response latencies (local-compatible: 869 ms; head-compatible: 874 ms). Mixed-effect

Table 4. Experiment 1 parameter estimates for response types (logistic regression) and response latencies (linear regression).

Response types	Estimate	S.E.	z-Value	$p(\chi^2)$
Intercept	2.74	0.17	15.75	<0.001
Local number	1.90	0.16	12.03	<0.001
Integration	0.69	0.15	4.53	<0.001
Predicate compatibility	0.31	0.15	2.01	<0.05
Local number × integration	−0.61	0.31	−2.01	<0.05
Local number × predicate compatibility	−0.62	0.31	−2.05	<0.05
Integration × predicate compatibility	0.14	0.31	0.46	0.65
Local number × integration × predicate compatibility	−0.80	0.61	−1.31	0.19
Correct response latencies	**Estimate**	**S.E.**	**t-Value**	**$p(\chi^2)$**
Intercept	812.27	15.89	51.11	<0.001
Local number	−23.16	9.42	−2.46	<0.01
Integration	−40.86	9.41	−4.34	<0.001
Predicate compatibility	3.07	9.40	0.33	0.74
Local number × integration	−14.03	18.79	−0.75	0.46
Local number × predicate compatibility	7.58	18.78	0.40	0.69
Integration × predicate compatibility	−13.73	18.82	−0.73	0.47
Local number × integration × predicate compatibility	−21.73	37.59	−0.58	0.56
Error response latencies	**Estimate**	**S.E.**	**t-Value**	**$p(\chi^2)$**
Intercept	903.14	25.99	34.75	<0.001
Local number	−1.94	32.00	−0.06	0.95
Integration	10.42	31.60	0.33	0.74
Predicate compatibility	40.40	31.61	1.28	0.21
Local number × integration	−103.48	62.86	−1.65	0.11
Local number × predicate compatibility	78.98	63.86	1.24	0.22
Integration × predicate compatibility	−57.93	63.45	−0.91	0.36
Local number × integration × predicate compatibility	−109.49	126.49	−0.87	0.39

Note: Random effects are the maximum justified by the data (random slopes and intercepts for all fixed factor main effects). p-Values are calculated from model comparison, leaving one factor out at a time.

Table 5. Mean response latencies (in milliseconds) from Experiment 1.

		Singular (correct) response		Plural (error) response	
Integration	Predicate compatibility	Local singular	Local plural	Local singular	Local plural
Integrated	Head-compatible	768	802	893	916
	Local-compatible	778	800	776	921
Unintegrated	Head-compatible	823	834	927	841
	Local-compatible	811	826	833	855

analyses of these data are in Table 4 and mean latencies are in Table 5.

Discussion

In the critical interaction with grammatical number, referential integration (a structural factor) and predicate compatibility (a lexical factor) both had strong effects on the rate of agreement variations. This was revealed in variable increases in plural verb use after plural local nouns (relative to singulars), representing modulation of agreement by referential and lexical factors. All three factors also created separate substantial changes in response types (accuracy), with more plural verb use when the subject phrase had a plural local noun, when the phrase had a weakly integrated referent, and when the local noun was more compatible with the predicate than the head noun. For response latencies, only the separate slowing effects due to plural local number and weak referential integration were evident. Notably absent were interactions between the referential and lexical factors on either latencies or response type.

In line with previous work, the interactions between grammatical number and referential integration and between grammatical number and predicate compatibility point to roles for referential and lexical factors in driving agreement production. However, the absence of interactions between lexical and referential factors implies that these have separable influences on agreement, where referential integration and lexical compatibility affect different components of the formulation process. This in turn suggests against the strongest possible lexical or structural accounts. With respect to a strong lexical account, it is unlikely that controller confusion is the full explanation of all agreement variation (e.g. Fowler, 1937). Conversely, with respect to a strong structural account, it is also unlikely that semantic misanalysis drove both effects in their entirety (e.g. misinterpreting the *with* in integrated and local-compatible sentences as *and* due to semantic similarity; see Bergen, Levy, & Gibson, 2012 for work supporting a similar hypothesis).

The main effects of response slowing due to integration and local-noun number suggest that both of these properties also impede the production of inflected verbs in different ways. One straightforward hypothesis is that integration complicates the construal of the subject's referent itself, while local-noun number alternately facilitates or disrupts production of the output of the agreement process. For instance, the grammatical number compatibility or incompatibility between the local noun and the verb could affect the operation of a monitoring mechanism.

In the second experiment, we sought to generalise the findings from Experiment 1 by using different manipulations of notional and lexical properties in a different type of structure with a different conventionally correct inflection. We did this by manipulating two well-studied variables, concreteness and semantic relatedness. In parallel to Experiment 1, these variables allowed us to contrast the effects of notional number and lexical processing on agreement.

Experiment 2

Experiment 2 was designed to further examine the contributions of structurally mediated (referential) and word-mediated (lexical) sources of number information to agreement, to provide converging evidence for their impact on number agreement production. This experiment examined agreement variations with conjoined noun phrases. As number agreement controllers in English, conjunctions have the property of being conventionally plural, allowing us to explore notional and lexical effects with a different grammatically correct response than in Experiment 1. To manipulate the grammatical number of local nouns, the first noun in the experimental preambles was singular and the second noun was varied between plural and singular such that the plural local noun facilitated correct responses (e.g. Haskell & MacDonald, 2005). We relied upon the same method as Experiment 1, measuring the speed and accuracy of producing number-specified verbs after preambles with varying notional and lexical properties.

The structurally mediated variable in Experiment 2 was concreteness and the lexical variable was semantic relatedness. These are factors that may play roles in agreement that are roughly parallel to integration and lexical-semantic predicate compatibility in Experiment 1. High concreteness, like low integration, promotes notional plurality due to the individuation of the two referents, while low concreteness (abstractness) more readily admits the coalescence of separate referents into one (e.g. Lorimor, 2007). The prediction is that concreteness will increase the uncertainty of notional-number evaluation and encourage singular verb inflections.

Lexical-semantic relatedness is a strong promoter of interference during retrieval (e.g. Wheeldon & Monsell, 1994), eliciting competition due to shared semantic category or associative features of words (Abdel Rahman & Melinger, 2007). Competition between concurrently processed nouns differing in grammatical number has been hypothesised to be critical in eliciting attraction (e.g. Solomon & Pearlmutter, 2004). The prediction then is that the well-established impact of semantic relatedness

on word production will offer a potent source of disruption to agreement such that more erroneous singular verb inflections are produced. This may be further enhanced by the more equivalent structural status of nouns in conjoined phrases as compared to the complex noun phrases used in the previous experiment.

Method

Participants

In exchange for course credit or $7.00 compensation, 113 undergraduates from the University of Illinois participated in the study. Of these participants, 13 were excluded from the study for having less than 84% usable experimental trials and an additional four participants were excluded due to technical difficulties. This left 96 participants.

Equipment

Equipment was identical to Experiment 1, except that headphones were not needed.

Materials

There were 32 experimental items. These were conjunctions of nouns varying in the concreteness of their referents, category relatedness (related or unrelated nouns), and local-noun number (singular or plural). See Table 6 for sample items. Stimuli were created using a free-association word database containing ratings for both nouns, including forward association (the number of times the second noun was generated from the first noun prompt) and concreteness (Nelson, McEvoy, & Schreiber, 1998).

Abstract preambles contained nouns rated between one and three on a seven-point concreteness rating scale, while concrete preambles contained nouns rated between five and seven on the same scale. Related and unrelated versions of these preambles were developed by changing the second noun, with related preambles having a forward association rating between 30% and 50% for the pair and unrelated preambles having a forward association rating between 0.01% and 10% for the pair. All but one of the nouns in the preambles had a regular plural form and a regular singular–plural alternation, regardless of their status as the first or second (local) noun. The exception was *tooth*, which occurred as a local noun. A full list of preambles is in Appendix 1.

Table 6. Example stimuli from Experiment 2.

	Related	Unrelated
Abstract	The hypothesis and the theory(ies)	The hypothesis and the thought(s)
Concrete	The dish and the plate(s)	The dish and the cat(s)

As in Experiment 1, filler stimuli were added to the critical preambles. The 192 filler preambles included a mixture of simple noun phrases and noun phrases modified by prepositional phrases. As in Experiment 1, some of the filler trials (18%) were catch trials. Among all fillers, 63% took singular agreement. Together with the experimental preambles, this meant that correct responses were plural on 46% of all trials.

The four different versions of each experimental item were divided among four lists. As in Experiment 1, lists were counterbalanced such that they contained only one version of each item and an equal number of items of each type. The sequence of the preambles in the lists was determined in the same way as in Experiment 1, with the same counterbalancing of presentation order, for a total of eight lists. All lists began with a fixed set of 12 fillers as a covert practice block.

Norming. Norming was carried out to establish the fit of the preamble nouns with a set of suitable adjectives to use as predicates. The adjectives were the ones used in Brehm and Bock (2013): *good*, *bad*, *ready*, and *true*. These adjectives were judged in a fashion similar to the sensibility norming in Experiment 1, with the singular form of each noun presented with each possible adjective in the phrase *How likely is it that X is Y* (e.g. *How likely is it that a cat is bad*?). Items were divided into four lists, each with ten instances of all four adjectives counterbalanced across lists, and with each list containing one token of every noun. Items were presented in a fixed random order with no fillers.

Ratings were collected from 20 participants using paper-and-pencil surveys or a computer-presented Excel workbook. We calculated the average rating for the head and local nouns in a phrase across the set of four adjectives and for the highest-rated adjective only. These ratings are displayed in Table 7. As in Experiment 1, margins of error were determined from the MSE of the highest-level interaction in an ANOVA by items with Scheffé corrections. The ratings imply that every noun had at least one suitable adjective among the four alternative completions. Numerically, the abstract pairs were rated as better fits with the adjectives than the concrete pairs on both metrics, but differences were small and within the margin of error for the highest-rated adjective for each pair.

To establish the relationship between these measures and those used in Experiment 1, norming was also carried out to assess how integration mapped to concreteness and relatedness. The instructions and procedure were the same as the integration rating that was conducted for Experiment 1, except that preambles were presented in noun phrases rather than complete

Table 7. Mean predicate sensibility and integration ratings for Experiment 2.

		Predicate sensibility		Integration	
Concreteness	Relatedness	All adjectives	Best adjective	Local *S*	Local *P*
Abstract "hypothesis"	Related "theory"	3.04	3.85	5.06	4.60
	Unrelated "thought"	3.03	3.92	4.05	3.95
Concrete "dish"	Related "plate"	2.82	3.68	3.99	4.44
	Unrelated "cat"	2.88	3.65	3.21	3.40
Margin of error		0.14	0.23	1.18	

Note: Bottom row contains the margin of error of differences between means for the rating.

sentences. The within-item variables (relatedness, local plurality) were fully crossed, divided across four lists with items presented once per list. Ratings were collected from 20 participants and results are displayed in Table 7. Low concreteness was associated with higher integration ratings, as expected, and the difference in integration ratings between levels of lexical relatedness was similar for abstract and concrete items (abstract = 0.8; concrete = 0.9). This suggests that relatedness and concreteness contributed independently to integration judgments, serving as dissociable, independent factors with respect to referential number. Ratings were equivalent when local nouns were singular or plural.

Procedure

Preambles were presented visually for an interval equal to the longer of 1000 or 40 ms per character. Participants were instructed to read these silently and complete them aloud with the best of a memorised set of four adjectives, *good*, *bad*, *ready*, and *true*. Previous work has shown a small memorised set of adjectives to be effective for eliciting a relatively homogenous group of utterances suitable for use with a reaction time measure (e.g. Bock, Carreiras, & Meseguer, 2012; Brehm & Bock, 2013). As in Experiment 1, there were standard (completion only) and catch (repeat and complete preamble) trials, with the same cuing, timing, and sequence of events (see Figure 2).

Scoring

Scoring was the same as in Experiment 1, except that plural verbs were scored as correct and singulars as incorrect.

Design

The variables of concreteness (abstract–concrete), relatedness (related–unrelated), and local-noun number (singular–plural) were fully crossed in a mixed design, with concreteness as a between-item factor. Every participant received one version of each of the 32 experimental preambles, divided equally across the eight combinations of relatedness, concreteness and local-noun number. Every item in each of its versions was presented to 24 participants.

Analysis

Analysis was identical to Experiment 1.

Results

Figure 4 shows response-type (accuracy) and response-latency results across conditions; we again follow the custom of describing all singular responses as "errors" to mark their deviation from the conventionally correct number inflection. Of all the factors, local-noun number had the strongest effect on response types, with local singular nouns eliciting more singular responses vs. local plural nouns (local singular, 79% plural; local plural, 95% plural). Local number interacted with the referential factor, consistent with structurally mediated effects on agreement: abstract phrases increased singular responding, especially when paired with a local singular noun (concrete-local singular, 84% plural; abstract-local singular, 74% plural). The interaction between relatedness and local number was not significant, but it trended in the same direction as in Experiment 1, consistent with lexically mediated effects on agreement: there were more singular verbs elicited for the attraction-inducing lexical factor than the attraction-reducing one (related-local singular, 78% plural; unrelated-local singular, 80% plural). The main effects of concreteness and relatedness on error rates were not reliable (see Table 8 for analyses). No reliable effects of any factor were observed for correct response latencies (see Table 8 for analyses and Table 9 for means).

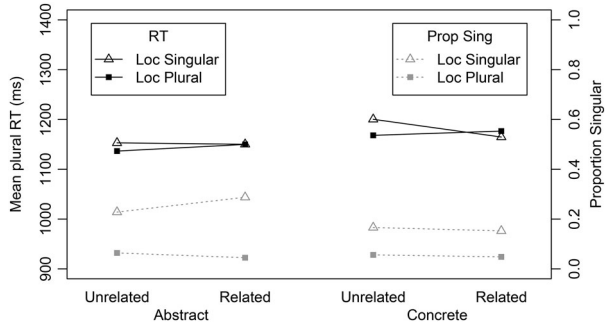

Figure 4. Experiment 2 plural verb production latencies (in ms, dark lines) and singular verb use (proportions, light lines) for preambles varying by concreteness, local-noun number, and relatedness.

In the analysis of error response latencies, grammatical number interacted both other factors, significantly with concreteness and marginally with relatedness. Consistent with structurally mediated referential effects on agreement, there was a significant interaction between local-noun number and concreteness such that mismatching local number and notional number made errors particularly slow. In concrete (notionally plural) phrases, local singular nouns elicited slower error responses than local plural nouns (concrete-local singular, $M = 1277$, SEM = 38; concrete-local plural, $M = 1184$, SEM = 76). The difference was smaller and in the reversed direction for abstract (notionally singular) phrases (abstract-local singular, $M = 1216$, SEM = 26; abstract-local plural, $M = 1256$, SEM = 73). This is the same pattern we observed in Experiment 1.

There was also a marginal interaction on error latencies between local-noun number and relatedness, consistent with lexical effects on agreement. The pattern was that related-local singular condition elicited the slowest errors (related-local singular, $M = 1261$, SEM = 29; unrelated-local singular, $M = 1216$, SEM = 32; related-local plural, $M = 1219$, SEM = 80; unrelated-local plural $M = 1222$, SEM = 71). See Table 8 for analyses and Table 9 for means.

Table 8. Experiment 2 parameter estimates for response types (logistic regression) and response latencies (linear regression).

Response types	Estimate	SE	*z*-Value	$p(\chi^2)$
Intercept	2.60	0.18	14.74	<0.001
Local number	1.76	0.14	12.47	<0.001
Concreteness	0.42	0.26	1.58	0.12
Relatedness	0.08	0.14	0.59	0.56
Local number × concreteness	−0.66	0.28	−2.41	<0.05
Local number × relatedness	0.41	0.27	1.47	0.15
Concreteness × relatedness	0.15	0.27	0.55	0.59
Local number × concreteness × relatedness	−0.77	0.55	−1.39	0.18
Correct response latencies	**Estimate**	**SE**	***t*-Value**	$p(\chi^2)$
Intercept	1164.64	27.54	42.28	<0.001
Local number	−12.09	12.65	−0.96	0.34
Concreteness	21.38	29.17	0.73	0.47
Relatedness	−4.85	12.59	−0.39	0.70
Local number × concreteness	13.45	25.20	0.53	0.59
Local number × relatedness	29.18	25.16	1.16	0.25
Concreteness × relatedness	−10.71	25.18	−0.43	0.67
Local number × concreteness × relatedness	23.45	50.32	0.47	0.64
Error response latencies	**Estimate**	**SE**	***t*-Value**	$p(\chi^2)$
Intercept	1278.16	41.16	31.05	<0.001
Local number	1.25	40.69	0.03	0.98
Concreteness	14.11	54.88	0.26	0.80
Relatedness	−15.10	41.46	−0.36	0.72
Local number × concreteness	−217.76	81.65	−2.67	0.01
Local number × relatedness	−133.16	81.54	−1.63	0.10
Concreteness × relatedness	41.44	81.02	0.51	0.61
Local number × concreteness × relatedness	89.59	160.48	0.56	0.58

Note: Random effects are the maximum justified by the data (random slopes and intercepts for all fixed factor main effects). *p*-Values are calculated from model comparison.

Table 9. Mean response latencies (in milliseconds) from Experiment 2.

		Plural (correct) response		Singular (error) response	
Concreteness	Relatedness	Local plural	Local singular	Local plural	Local singular
Abstract	Unrelated	1137	1153	1270	1189
	Related	1150	1150	1236	1238
Concrete	Unrelated	1168	1200	1167	1254
	Related	1177	1165	1203	1303

Discussion

The effects seen in Experiment 2 were similar to those in Experiment 1. Local-noun number contributed substantially to agreement: When local-noun number was consistent with the correct verb number, verb production occurred with limited variation; when local-noun number was inconsistent with the correct verb number, more variation was observed. Notional number likewise mattered, such that when the subject noun phrase had concrete (notionally plural) referents, plural local nouns elicited more plural (conventionally correct) verb number. Lexical competition had a similar effect as in Experiment 1, such that related nouns led to more agreement variation; however, this effect failed to reach significance. Again, as in Experiment 1, these interactions appeared on error rates but not correct latencies.

Error latencies followed the same pattern as in Experiment 1, being fastest in the conditions where local grammatical number mismatched the verb that was produced and notional number matched it. This shows again that congruent notional and local grammatical number facilitated controller number selection. In addition, there was an impact from lexical relatedness on error latencies, though this was less pronounced than the notional effects.

There was one other intriguing lexical difference that emerged from an exploration of item effects. Previous findings (Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004) have been interpreted in terms of a relationship between lexical competition in agreement and scope of planning. In a small subset (six) of our items, an increased possibility of a narrow planning window may have allowed semantic relatedness to particularly disrupt agreement, consistent with lexical competition modulated by timing.

Specifically, for four concrete and two abstract items, the conjoined subject noun phrase had only one determiner (e.g. *Their destiny and fate*). In the 26 remaining items, there were two determiners (e.g. *The hypothesis and the theory)*. The one-determiner configuration places both nouns in the same phrase, which may

allow a wider scope of planning (see Konopka, 2012). The one-determiner configuration is also notionally more unitary (Lorimor, 2007). Both of these properties could increase the likelihood of contemporaneous preparation of the words for a phrase. Consistent with this possibility, in the abstract-related-local singular condition, one-determiner responses were 220 ms slower than two-determiner responses (1414 vs. 1194 ms) and twice as error-prone (52–26%).

Because contemporaneous preparation is the crux of the scope-of-planning hypothesis about lexical disruptions to agreement, this structural difference is highly suggestive, despite its restrictions. The small one-determiner item sample naturally limits the value of inferential statistics or modelling, and apart from this single condition there were no striking differences between the one- and two-determiner items. Still, the combined impact on these items of abstractness, lexical-semantic relatedness, and structural simplicity points to a promising avenue for further exploration.

General discussion

In these two studies, we examined the ways that grammatical number, notional number, and lexical-semantic relatedness are used in subject–verb number agreement. By examining measures of response types (accuracy) and response latency, we looked at how number-inflected verbs were produced after complex sentence subjects. Our aim was to evaluate the predictions of structural and lexical accounts of number agreement.

In Experiment 1, sentence subjects had singular head nouns with local nouns that varied between singular and plural (e.g. *The phone with the broken toaster*; *The phone with the broken toasters*). This local number variation created the standard attraction effect, with local plural nouns eliciting increased rates of erroneous plural agreement and slowing responses. We also manipulated two other factors in order to trace the origin of agreement difficulty to structure-centred referential factors or word-centred lexical factors. The structural factor was referential integration: preambles were integrated and notionally singular (e.g. *The phone with the missing button*) or unintegrated and notionally plural (e.g. *The phone with the broken toaster*). Consistent with previous effects of referentially driven notional number (e.g. Brehm & Bock, 2013; Eberhard, 1999; Humphreys & Bock, 2005; Vigliocco et al., 1995), integration decreased attraction, slowed correct responses and marginally speeded error responses when local grammatical number mismatched the subject's number. The lexical factor was lexical-semantic compatibility, manipulated by using predicates semantically suited to properties of the head noun (e.g. *ringing* goes well with *phone* but not *button* or *toaster)* or the local noun (e.g. *plastic* goes better with *button* than *toaster*, *shiny* goes better with *toaster* than *button*). Consistent with previous lexically driven competition effects (e.g. Solomon & Pearlmutter, 2004; Thornton & MacDonald, 2003), this manipulation boosted attraction, though it did not affect response latencies.

In Experiment 2, sentence subjects were conjoined noun phrases with singular first nouns combined with second (local) nouns that varied between singular and plural (e.g. *The dish and the plate*; *The dish and the plates*). Again, this local number variation affected accuracy, with local plurals associated with increased rates of correct plural agreement. We likewise manipulated referential and lexical factors to trace the origin of agreement difficulty. The referential factor was concreteness, which affects the ease of individuating (counting) referents. This also influenced agreement. Consistent with effects of referentially driven notional number (e.g. Lorimor, 2007), concrete preambles (e.g. *The dish and the plate*) elicited more correct plural agreement than abstract ones (e.g. *The hypothesis and the theory*), particularly when local grammatical number was plural. The lexical variable was lexical-semantic relatedness of the two conjoined nouns (e.g. related, *The dish and the plate*; unrelated, *The dish and the cat*) and it disclosed no reliable effect on either error rates or correct response latencies, though it did show a marginal effect on error latencies such that the conditions with tighter semantic relationships and mismatching local number elicited the slowest error responses.

In both experiments, the goal was to disentangle how speakers use structural (wholistic) sources of information from how they use lexical sources of information during subject–verb agreement. Agreement involves linking features of controllers and targets, which could be accomplished within the syntactic representations constructed from global message properties or through combinations of individual word meanings. The issue that separates structural and lexical accounts is which of these dominates agreement.

Implications for structural accounts of agreement

The present findings showed that structural control was a critical force in the production of agreement. In a structural account, agreement is one product of a process in which a wholistic mental model of the subject's referent is evaluated for numerosity and unpacked (in essence, *dis*integrated) into linguistically viable pieces. Sentences are composed with an emerging syntactic frame that supports yet-to-be specified morphological and phonological properties of words (e.g. Dell, 1986). As this unpacking and mapping occur, agreement is created

through a combination of referential indices, structural properties, and words. Agreement difficulty arises when the construed number of the subject's referent, its notional number, conflicts with the grammatical number associated with retrieved pieces of words.

Our test of this account relied on two manipulations of notional number, referential integration in Experiment 1 and concreteness in Experiment 2. These notional factors were designed to alter message-level properties in similar ways: low levels of integration and high levels of concreteness tend to individuate referents and make them notionally more plural. This follows from other work on notional factors in agreement (e.g. Bock et al., 2004; Eberhard, 1999; Lorimor, 2007; Vigliocco et al., 1995). Both notional manipulations interacted with the grammatical number of the local noun to affect the outcome of agreement, showing clearly that a property of the sentence subject as a whole swayed the overt linguistic signals of agreement.

In addition to interactions between notional number and grammatical number, we obtained effects of notional number on the overall time course of production – with respect to correct latencies (Experiment 1) and error latencies (Experiments 1 and 2). The general pattern is that verb-matching notional number speeded correct responses, and local-noun matching notional number speeded erroneous responses. The suggestion is that referential information affected the sentence planning process as a whole, in addition to its impact on agreement implementation. We speculate that these timing effects might be due to the ease of mentally representing coherent referents or to processes involved with error monitoring, but leave these as an open question for future research.

Implications for lexical accounts of agreement

The present findings also pointed towards a role for lexical mechanisms in agreement production. In contrast to the structural view, a lexically driven account claims that lexical-semantic associations among words in the sentence are central to the production of agreement. Lexical accounts of sentence production and agreement hinge on assembling components: words must be retrieved and then pieced together into a syntactically acceptable sequence relying on the words' co-occurrence restrictions (such as subcategorisations and conditional probabilities). Finally, the features of words in certain positions must be given agreeing values. On this view, the locus of agreement difficulty is in assembling the pieces of a phrase and identifying a controller.

The lexical-semantic compatibility variable in Experiment 1 was inspired by Thornton and MacDonald's results (2003). Their basic finding was that predicates that could plausibly modify a local noun promoted attraction. When paired with *The album by the classical composers*, the predicate *praised* caused more difficulty than *played*, as shown by increased production errors and slower reading times. We aimed for an operationalisation of these factors that rested on lexical-semantic compatibility and replicated the effect, but for error rates only.

In Experiment 2, we manipulated lexical-semantic relatedness between the head and local nouns. In previous work, it has been shown that semantic relationships between successive nouns increase interference in the production of later nouns (e.g. Abdel Rahman & Melinger, 2007; Damian, Vigliocco, & Levelt, 2001; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Schriefers, Meyer, & Levelt, 1990; Wheeldon & Monsell, 1994). In agreement, lexical competition has been hypothesised to raise processing difficulty as the speaker prepares to produce the second noun (Solomon & Pearlmutter, 2004). This in turn disrupts selection of a lexical controller, especially in the presence of a mismatching grammatical number feature (e.g. as outlined by Dillon et al., 2013). We showed a numerical effect consistent with these previous findings, with higher error rates and increased error latencies in interference-inducing conditions.

The lexical effects in Experiments 1 and 2 are suggestive of how content-addressable memory retrieval mechanisms might support agreement (e.g. Dillon et al., 2013; Lewis & Vasishth, 2005), suggesting some possible differences between language comprehension and production. In a retrieval-based account of attraction, cues (such as noun and verb inflections) serve as pointers to other words that can be directly retrieved from a content-addressable memory store. In the process of number agreement, the presence of a number-inflected target verb (in comprehension) or the need to inflect a target verb (in production) cues the retrieval of a previously occurring noun controller in order to recover the noun's number. If a cue leads to the retrieval of a wrong noun, agreement problems like attraction may result.

A retrieval process of this kind has had more success in explaining agreement disruptions in language comprehension than structural or message-based phenomena have (Tanner, Nicol, & Brehm, 2014; Wagers et al., 2009; see Shen, Staub, & Sanders, 2013 for similar results in listening comprehension). Retrieval-based accounts also provide a clear mechanism for differences in subject–verb and reflexive pronoun agreement comprehension (e.g. Dillon et al., 2013). In comprehension, the presence of a number-specified target verb or reflexive pronoun prompts a retrieval operation. In production, retrieval of noun controllers may play a smaller part due

to the normally top-down, forward-looking nature of sentence formulation. The relative weakness of lexical effects in our results is compatible with this possibility.

Structural and lexical interactions

While we found effects consistent with lexical and structural contributions to agreement, we saw no sign of an interaction between the two. This suggests that the two contribute separately to processing, influencing agreement in different ways. This aligns with other work. Veenstra, Meyer, and Acheson (2015) obtained a similar agreement result in the presence of a direct, strong, empirically verified degree of lexical interference, showing that even high levels of difficulty in lexical retrieval do not compromise structurally mediated referential effects. Underscoring this, identical manipulations of semantic integration have shown both lexical interference (Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004), and referential effects on number agreement (Brehm & Bock, 2013). The suggestion is that both effects are representative of cognitive processes required for speech – but that they are separable and influenced by different planning constraints.

Note that the separation between lexical and structural sources of information is generally consistent with many models of sentence planning and production (see Bock & Ferreira, 2014 for discussion). We schematised this in Figure 1 by distinguishing the lexical and structural contributions to agreement. The principle is that if lexical and structural effects on agreement arise before words and structures are combined, then they may be measurably separable, without interactions between them. This should hold even when a single utterance is affected by both properties. The present results support this framework, allocating difficulties in agreement formulation to distinct lexical and structural factors.

Conclusion

Language production requires combining structures and words. For the production of agreement, we found that variations in verb number are strongly influenced by wholistic number properties (here, notional number) that are conveyed in sentence-subject structures. Lexical-semantic relationships also affected verb number, though less consistently and independently of notional factors. The implied separation of structurally and lexically mediated components of sentence formulation underscores the complex cognitive architecture behind the production of simple, everyday utterances. Before speakers start to talk, they typically have messages that contain the basic internal features and relationships characteristic of human cognition. These messages need to be translated into language, where the speaker retrieves words from memory and maps the mental relationships in thoughts onto linguistic structures. Agreement – whether erroneous or correct – reflects the synthesis of all of these processes.

Notes

1. Margin of error is defined as the half-width of the 95% confidence interval for differences between condition means. For these experiments, we calculated this confidence interval from the mean-squared error (MSE) of the highest-level interaction in a repeated measures ANOVA by items, using a Scheffé correction and the type III sum of squares.
2. The numbers of items going into each of these cells are extremely unbalanced. Standard error of the mean (SEM) gives a sense of the reliability of each mean estimate while taking the sample size into account.

Acknowledgements

We would like to thank Ben Carter for collecting, transcribing, and coding the Experiment 2 data, Taylor Eighmy for providing voice recordings, transcription, and coding, and Megan Krull, Alexa Mazza, and Stacey Cohen for data collection and transcription. We are particularly grateful for advice from anonymous reviewers and from Gary Dell, Maureen Gillespie, Maryellen MacDonald, and Darren Tanner.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by Division of Behavioral and Cognitive Sciences NSF grant BCS-0843866 and National Institute of Child Health and Human Development NIH grant T32-HD055272.

References

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 604–614. doi:10.1037/0278-7393.33.3.604

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement, and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*, 65–85. doi:10.1016/j.jml.2006.08.004

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4

Bergen, L., Levy, R., & Gibson, E. (2012). Verb omission errors: Evidence of rational processing of noisy language inputs. In *Proceedings of the 34th annual meeting of the Cognitive Science Society* (pp. 1320–1325). Sapporo, Japan.

Bock, J. K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, *66*, 17–37. doi:10.1016/j.jml.2011.07.008

Bock, K., & Cutting, J. C. (1993). 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, K., & Eberhard, K. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, *8*, 57–99. doi:10.1080/01690969308406949

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, *51*, 251–278. doi:10.1016/j.jml.2004.04.005

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

Bock, J. K., & Ferreira, V. (2014). Syntactically speaking. In M. Goldrick, V. Ferriera, & M. Miozzo (Eds.), *The oxford handbook of language production* (pp. 21–46). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199735471.013.008

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93. doi:10.1016/0010-0285(91)90003-7

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

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111. doi:10.1016/s0010-0277(02)00017-3

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201–242. doi:10.1006/cogp.1998.0693

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271. doi:10.3758/bf03204507

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*, B77–B86. doi:10.1016/s0010-0277(01)00135-4

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

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*, 85–103. doi:10.1016/j.jml.2013.04.003

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: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

Fowler, H. W. (1937). *A dictionary of modern English usage*. Oxford: Oxford University Press.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, *101*, 173–216. doi:10.1016/j.cognition.2005.10.003

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*, 329–374. doi:10.1080/01690960701467993

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 9, pp. 133–177). doi:10.1016/S0079-7421(08)60270-4

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–220). 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

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, *48*, 760–778. doi:10.1016/s0749-596x(03)00010-x

Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: Evidence from subject–verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 891–904. doi:10.1037/0278-7393.31.5.891

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464–482. doi:10.1016/j.cognition.2005.02.006

Humphreys, K. R., & Bock, J. K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, *12*, 689–695. doi:10.3758/bf03196759

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*, 143–162. doi:10.1016/j.jml.2011.08.003

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419. doi:10.1207/s15516709cog0000_25

Lorimor, H. (2007). *Conjunctions and grammatical agreement* (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign, Champaign.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time course analysis. *Memory and Cognition*, *24*, 173–187. doi:10.3758/bf03200879

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91. doi:10.1016/s0749-596x(02)00515-6

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

R Core Team. (2014). *R: A language and environment for statistical computing* [Version 3.1.2]. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

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

Shen, E. Y., Staub, A., & Sanders, L. D. (2013). Event-related brain potential evidence that local nouns affect subject-verb agreement processing. *Language and Cognitive Processes*, *28*, 498–524. doi:10.1080/01690965.2011.650900

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

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*, 308–327. doi:10.1016/j.jml.2008.11.002

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215. doi:10.1016/j.jml.2014.07.003

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*, 740–759. doi:10.1016/s0749-596x(03)00003-2

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285–316. doi:10.1016/s0749-596x(03)00081-0

Veenstra, A., Acheson, D. J., Bock, K., & Meyer, A. S. (2014). Effects of semantic integration on subject–verb agreement: Evidence from Dutch. *Language, Cognition and Neuroscience*, *29*, 355–380. doi:10.1080/01690965.2013.862284

Veenstra, A., Meyer, A. S., & Acheson, D. J. (2015). Effects of parallel planning on agreement production. *Acta Psychologica*, *162*, 29–39. doi:10.1016/j.actpsy.2015.09.011

Vigliocco, G., Butterworth, B., & Garrett, M. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*, 261–298. doi:10.1016/s0010-0277(96)00713-5

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. doi:10.1006/jmla.1995.1009

Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*, 455–478. doi:10.1006/jmla.1998.2624

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, *68*, B13–B29. doi:10.1016/s0010-0277(98)00041-9

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237. doi:10.1016/j.jml.2009.04.002

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language*, *33*, 332–356. doi:10.1006/jmla.1994.1016