Published in final edited form as:
Cognition. 2011 March ; 118(3): 377–397. doi:10.1016/j.cognition.2010.10.008.

Hierarchy and Scope of Planning in Subject-Verb Agreement Production

Maureen Gillespie and **Neal J. Pearlmutter**
Northeastern University

Maureen Gillespie: gillespie.m@neu.edu; Neal J. Pearlmutter: pearlmutter@neu.edu

Abstract

Two subject-verb agreement error elicitation studies tested the hierarchical feature-passing account of agreement computation in production and three timing-based alternatives: linear distance to the head noun, semantic integration, and a combined effect of both (a scope of planning account). In Experiment 1, participants completed subject noun phrase (NP) stimuli consisting of a head NP followed by two prepositional phrase (PP) modifiers, where the first PP modified the first NP, and the second PP modified one of the two preceding NPs. Semantic integration between the head noun and the local noun within each PP was held constant across structures. The mismatch error pattern showed an effect of linear distance to the head noun and no influence of hierarchical distance. In Experiment 2, participants completed NP PP PP stimuli in which both PPs modified the head noun, and both the order of the two PPs and the local nouns' degree of semantic integration with the head noun were varied. The pattern of mismatch errors reflected a combination of semantic integration and linear distance to the head noun. These studies indicate that agreement processes are strongly constrained by grammatical-level scope of planning, with local nouns planned closer to the head having a greater chance of interfering with agreement computation.

Keywords

sentence production; number agreement; syntactic planning; semantic integration; hierarchical feature-passing; scope of planning; speech errors; subject-verb agreement

The study of language production is concerned with how speakers translate non-verbal thoughts into meaningful grammatical utterances. While this is a fairly effortless task that requires little conscious consideration on behalf of the speaker, the nature of the processes that underlie this task are complex. Most language production models (e.g., Bock & Levelt, 1994) separate the production planning process into three main levels: the message level, which represents the speaker's intended meaning; the grammatical encoding level, which translates the meaning into a sequence of words; and the phonological encoding level, which translates the sequence of words into the articulatory plan required to produce the utterance. The current work focuses on the grammatical encoding process and specifically on syntactic planning, which is responsible for creating a syntactic structure encoding word order, hierarchical syntactic relations, and inflections.

Psychology Dept., 125 NI Northeastern University Boston, MA 02115, Phone: (617) 373-3798 (Gillespie), (617) 373-3040 (Pearlmutter), Fax: (617) 373-8714.

Inflectional processes in particular have been investigated in a variety of studies, typically by examining the conditions under which subject-verb agreement errors can be elicited, as a way of gaining insight into syntactic planning (e.g., Bock & Cutting, 1992; Bock & Miller, 1991; Franck, Lassi, Frauenfelder, & Rizzi, 2006; Franck, Vigliocco, & Nicol, 2002; Hartsuiker, Antón-Méndez, & van Zee, 2001; Solomon & Pearlmutter, 2004b; Vigliocco & Nicol, 1994, 1998). Bock and Miller (1991) conducted the first study that elicited subject-verb agreement errors in a laboratory setting. They used sentence preambles that were composed of a head noun followed by a phrase containing a local noun (e.g., as in (1)). Subject-verb agreement errors are commonly produced in sentences containing subject noun phrases with this structure when the head and local noun mismatch in number. Experimental items in Bock and Miller's study manipulated the number marking of the head and local nouns to form four number conditions. Conditions in which the head noun (*key*) and local noun (*cabinet*) had different number markings ((1b), containing the singular-plural (SP) sequence, and (1c), containing the plural-singular (PS) sequence) were considered the mismatch conditions, while conditions in which the head noun and local noun had the same number marking (1a, 1d) were considered the match conditions. Preambles were presented auditorily, and participants were required to repeat them and then complete them as full sentences.

(1)

a. The key to the cabinet (SS)

b. The key to the cabinets (SP)

c. The keys to the cabinet (PS)

d. The keys to the cabinets (PP)

Nearly all agreement errors occurred in the mismatch conditions (1b, 1c). Within these conditions, agreement errors were more common when the head noun was singular and the local noun was plural (1b) than when the head noun was plural and the local noun was singular (1c). This error pattern is referred to as the mismatch effect and has been replicated in essentially all studies examining subject-verb agreement (e.g., Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Bock & Miller, 1991; Bock, Nicol, & Cutting, 1999; Eberhard, 1999; Franck et al., 2006; Hartsuiker et al., 2001; Negro, Chanquoy, Fayol, & Louis-Sidney, 2005). The interference of plural local nouns, and relative lack of interference of singular local nouns, on subject-verb agreement provides support for the hypothesis that plural noun forms are marked with a plural feature, while singular nouns are unmarked (Berent, Pinker, Tzelgov, Bibi, & Goldfarb, 2005; Bock & Eberhard, 1993; Bock & Miller, 1991; Eberhard, 1997; Eberhard, Cutting, & Bock, 2005; Vigliocco & Nicol, 1994, 1998). While this latter pattern, the plural markedness effect, does not provide evidence for a specific mechanism for agreement effects, it does show that mismatch effects are not simply a result of agreement with the nearest noun and that a more complex mechanism is involved.

Most production research assumes that agreement is implemented through hierarchical feature-passing (Eberhard et al., 2005; Franck et al., 2002; Hartsuiker et al., 2001; Vigliocco & Hartsuiker, 2002; Vigliocco & Nicol, 1998). According to this view, agreement is computed once the syntactic tree structure of a sentence is formed, with number features being passed up through the subject NP to the verb phrase. Mismatch effects occur when a plural feature is inadvertently passed too far up the tree, overwriting the number from the head noun with the number from a local noun. Franck et al. (2002) provide the most direct test of the hierarchical feature-passing hypothesis in an error elicitation experiment using subject NP preambles containing two PP modifiers, as in (2). Their stimuli had a descending hierarchical structure in which each PP modified the immediately preceding noun, and the

local nouns (*flight* and *canyon* in (2)) varied in number. Figure 1 shows the syntactic structure as well as the path along which an errant feature from N2 or N3 would have to pass.

(2)

a. The helicopter for the flight over the canyon (SSS)

b. The helicopter for the flights over the canyon (SPS)

c. The helicopter for the flight over the canyons (SSP)

d. The helicopter for the flights over the canyons (SPP)

The hierarchical feature-passing hypothesis predicts a larger mismatch effect for preambles like (2b) than for preambles like (2c). Because N2 (*flight(s)*) is hierarchically closer to the verb than N3 (*canyon(s)*) is, fewer feature-passing errors would have to occur for N2's plural to interfere with agreement in (2b) than for N3's plural to interfere in (2c). Franck et al. (2002) found that the N2 mismatch effect was larger than the N3 mismatch effect in both English and French, and they thus argued for a hierarchical feature-passing account of subject-verb agreement over a linear account in which interference increases with (linear) proximity to the verb.

Current models of agreement computation also assume mechanisms that are consistent with a hierarchical feature-passing account (Eberhard et al., 2005; Vigliocco & Hartsuiker, 2002). Eberhard et al.'s Marking and Morphing model was implemented to account for the findings of a number of agreement studies. According to this model, the marking process assigns number to the subject NP as a whole based on message-level properties. Separately, each noun within the subject NP is also assigned a number specification from its lexical entry, and morphing then combines the subject NP number value set by marking with the number values from all the nouns within the subject NP, to yield an overall specification of number for the subject. This specification in turn determines the probability of singular versus plural agreement on the verb. The morphing process encodes the hierarchical distance assumption: Nouns situated further from the subject NP node in the syntactic tree are stipulated to have a weaker influence on the subject NP's number assignment than nouns closer to the subject NP. The implemented model handles only structures with a head noun and a single local noun at a constant distance from the subject NP (within a PP), but the assumption is that a local noun in a more syntactically distant PP (e.g., N3 in Franck et al., 2002) would have even less of an influence on the subject NP's number assignment.

The hierarchical feature-passing hypothesis can explain various effects seen in previous studies (see Franck et al., 2002, for discussion of Bock & Cutting's, 1992, clause-packaging hypothesis; Vigliocco & Nicol, 1998, for additional support from question production; and Hartsuiker et al., 2001, for cross-linguistic evidence). However, it does not take into account the fact that language production is at least partially sequential or incremental, and that planning of utterances may be as well. In particular, it assumes that a full hierarchical structure for the entire subject NP is available through which features can pass. Although such structure must be computed during the course of producing the subject NP, it may or may not all be present simultaneously, and the part(s) of the structure relevant to creating mismatch errors (the PPs containing the local nouns) might not be present at the point in time when the number of the subject NP is being computed (for a related suggestion, see Haskell & MacDonald, 2005). This possibility suggests a memory-encoding-based alternative to the hierarchical account: Interference with agreement computation might be a function of the extent to which the potentially-interfering element is active in memory at the time when the number-marking of the subject NP is being computed. Thus while some local nouns might be planned close in time to the head, others may be planned relatively later and

thus have a smaller likelihood of creating interference (see also Pearlmutter & Solomon, 2007; Solomon & Pearlmutter, 2004b, for more details on such a timing of activation account). The current work considers two factors, linear distance back to the head noun and semantic integration (Solomon & Pearlmutter, 2004b), which might be expected to affect the relative time of planning of different elements and which would make predictions compatible with Franck et al.'s results.

Thus, for example, while Franck et al. (2002) showed that a local noun's linear proximity to the verb cannot account for mismatch effects, no study to date has examined the influence of a local noun's linear distance back to the head noun (though see Nicol, 1995, for a related proposal). Assuming that the number of the subject NP must be computed and retained in memory, elements linearly closer to the head should be more likely to interfere with this encoding process, predicting correctly that N2 mismatch effects should be larger than N3 mismatch effects for Franck et al.'s stimuli.

Also possible is that, instead of order, the relative timing of planning of words due to semantic relationships among them may affect agreement computation. Solomon and Pearlmutter (2004b) hypothesized that semantic integration (i.e., the degree to which elements within a phrase are linked at the message level) affects the timing of planning of elements within a phrase, such that elements of more semantically integrated phrases are more likely to be planned simultaneously. More integrated cases thus produce more potential interference and a greater possibility for speech errors, including mismatch errors in agreement error elicitation (see also Pearlmutter & Solomon, 2007, for evidence from exchange errors). Solomon and Pearlmutter manipulated local noun plurality in NP PP stimuli and compared integrated cases (e.g., *The pizza with the yummy topping(s)*) to corresponding unintegrated ones (e.g., *The pizza with the tasty beverage(s)*). Across a series of experiments, they found larger mismatch effects for integrated than for unintegrated conditions, supporting the hypothesis that increased semantic integration leads to increases in subject-verb agreement mismatch effects by way of increased interference. This hypothesis can also account for Franck et al.'s (2002) result, as Solomon and Pearlmutter (2004a) obtained ratings of semantic integration for Franck et al.'s Experiment 2 (English) stimuli, which showed semantic integration to be confounded with syntactic distance: N1 and N2 were significantly more integrated than N1 and N3, predicting correctly that the N2 mismatch effect should be larger than the N3 mismatch effect, because N1 and N2 would be more likely to be planned simultaneously than N1 and N3.

An additional possibility is that influences of linear order and semantic integration combine to determine the timing of planning of elements, in which case nouns linearly closer to the head would be more likely to interfere, but the extent of interference would be increased by greater integration with the head and decreased with reduced integration. This account suggests that the scope of planning during grammatical encoding may have an influence on agreement computation. Research examining exchange errors provides clear evidence that multiple elements of an utterance are active simultaneously during production (Garrett, 1975, 1980), suggesting that speakers plan parts of their utterances in advance of articulation. However, there is little agreement about the size of the planning units (cf. Allum & Wheeldon, 2007; Griffin, 2001; Smith & Wheeldon, 1999, 2001; Wheeldon & Lahiri, 2002). Under a scope of planning account, only local nouns that are within the scope of planning of the head noun when the number marking of the subject NP is determined would create mismatch effects, with both decreased head-local linear distance and increased head-local semantic integration contributing to the likelihood of the local noun being within the scope of planning of the head, and thus contributing to the likelihood of the local noun being active simultaneously with the head noun and creating interference in encoding number. This scope account provides an alternative to the hierarchical feature-passing

hypothesis because in Franck et al. (2002), N3 was more likely to have been outside the scope of planning of the head noun when agreement was computed (it was linearly farthest from the head and weakly integrated with the head). This would predict that N3 would be less likely to produce a mismatch effect (or would produce a weaker one than N2), as they found.

The two experiments below examine hierarchical feature-passing and the above three alternatives as influences on agreement computation. Experiment 1 was a direct test of the hierarchical distance hypothesis against a linear distance alternative, manipulating hierarchical distance of N3 within preambles (subject NP stimuli) while controlling semantic integration. Experiment 2 manipulated semantic integration and linear order within preambles while controlling hierarchical distance.

Experiment 1

To examine hierarchical distance directly and address the semantic integration confound in Franck et al.'s (2002) stimuli, Experiment 1 compared two sets of NP PP PP preambles, which controlled semantic integration while manipulating hierarchical distance and local noun number, as in (3) and (4). (The code following each preamble indicates noun number for N1, N2, and N3, with S meaning singular and P plural.) Preambles like (3) had a descending hierarchical structure, such that the PP containing N3 (*on the leather strap(s)*) modified N2 (*buckle(s)*), as in Figure 1 and in Franck et al. (except that the current stimuli also had an adjective or noun modifier of N2 and N3). Preambles like (4) had a flat structure, such that both PPs (*to the western suburb(s), with the steel guardrail(s)*) modified N1 (*highway*), as illustrated in Figure 2.[1] Mean semantic integration of the N1-N2 pair was matched across structures, as was mean semantic integration of the N1-N3 pair.

(3)

a. The backpack with the plastic buckle on the leather strap (SSS)

b. The backpack with the plastic buckles on the leather strap (SPS)

c. The backpack with the plastic buckle on the leather straps (SSP)

(4)

a. The highway to the western suburb with the steel guardrail (SSS)

b. The highway to the western suburbs with the steel guardrail (SPS)

c. The highway to the western suburb with the steel guardrails (SSP)

If hierarchical distance has an independent effect on agreement computation, the difference between the N2 and N3 mismatch effects should be smaller for the flat preambles than the descending preambles, because the distance N2's plural feature would have to travel to affect agreement computation is matched across structures, while the distance N3's plural feature would have to travel is shorter in flat structures than in descending structures. On the other hand, if the effects Franck et al. (2002) attributed to hierarchical distance were instead due to linear distance back to the head noun, then both flat and descending structures should show the pattern Franck et al. found for their descending structures — a higher mismatch

[1]The relatively simple, traditional (Chomsky, 1965) syntactic structures in Figures 1 and 2 are sufficient to illustrate the contrast in attachment height (and thus feature-passing distance) of the second PP. Whether these structures are correct is unknown, and current syntactic theories differ on the details of the structures for both cases. However, we are not aware of any such differences which would alter the general prediction that the difference in feature-passing distance between N2 and N3 is larger for the descending than the flat cases. See the General Discussion for details on the specific case in which a theory enforces binary branching (e.g., Chomsky, 1995) and Solomon and Pearlmutter (2004b) for some discussion of variations in feature-passing predictions depending on structural details.

error rate for N2 mismatches than for N3 mismatches — with no interaction between structure and mismatch position. Finally, if the effects Franck et al. found were due to their stimuli's semantic integration confound (alone), then no interaction should be observed in this experiment; and in addition, the N2 and N3 mismatch effects should differ only to the extent that integration between each of those nouns and N1 varies.

Method

Participants—Fifty-four Northeastern University students and community members participated in the on-line experiment. In this and Experiment 2, all participants were native English speakers and received either course credit or payment ($10) for their participation; no participant provided data for more than one part of any experiment.

Materials and design—Twelve descending stimulus items like (3), chosen from a candidate set of 29, and twelve flat stimulus items like (4), chosen from a candidate set of 34, were used as critical items for the experiment. All preambles consisted of a head NP (always *The* and a singular head noun, N1) followed by two PPs, each of which consisted of a preposition, the determiner *the*, an adjective or modifier noun, and a local noun (N2 or N3). In the descending stimuli, the first PP modified the head NP, and the second PP modified the second NP (containing N2); in the flat cases, both PPs modified the head NP. All nouns in the stimuli were inanimate and had regular plural forms, and each noun's conceptual number matched its grammatical number. The full set of stimuli is shown in Appendix A.

Because simultaneously controlling semantic integration and manipulating hierarchical structure limited the total number of available items, only the three local noun number conditions critical for examining mismatch effects (SSS, SPS, SSP; as shown in (3) and (4)) were included in the critical stimuli, to maximize power. Thus N1 in the critical stimuli was always singular, and the three different versions of each item were created by varying either N2 or N3 number.

Eighty-eight filler preambles were combined with the critical items. Eight consisted of an NP PP PP sequence with a singular head noun and plural N2 and N3, in order to balance the SSS, SPS, and SSP critical items; four of these fillers had descending structures, and four were flat. Of the other 80 fillers, 32 consisted of an NP PP PP sequence (with varying local noun number) but had a plural head noun. The rest had a variety of structures varying in head noun number and were similar in length and complexity to the critical items. The critical items and fillers were combined to form three counterbalanced lists, each containing all fillers and exactly one version of each of the critical items. Each list was seen by 18 participants.

Stimulus norming—The 24 critical stimuli were chosen from the initial 63 candidate stimuli based on two norming studies conducted in advance, one for semantic integration and one for attachment of the second PP. While the critical stimuli ended up instantiating only three local noun number conditions, all four possible local noun number combinations were normed (SSS, SPS, SSP, SPP). Both norming surveys also included an additional 24 filler stimuli with the same NP PP PP format; local noun number was varied between-items in these stimuli (6 items per local noun number condition), as was attachment of the second PP (12 were designed to be N1-attached and 12 N2-attached), and they were intended to have a range of levels of semantic integration between N1, N2, and N3.

The first norming survey, completed by 117 participants, was used to ensure that the preambles controlled semantic integration as desired. The 12 different versions of each of the 63 candidate stimulus items (4 number conditions × 3 possible rating pairs (N1-N2, N1-

N3, N2-N3)), along with the 24 fillers, were rated for integration following the procedure described in Solomon and Pearlmutter (2004b). Participants rated integration of the two underlined nouns in each preamble, using a 1 (loosely linked) to 7 (tightly linked) scale. The instructions included example phrases (*the ketchup or the mustard* and *the bracelet made of silver*) and indicated that although *ketchup* and *mustard* are similar in meaning, they are not closely related in the particular example phrase, in contrast to *bracelet* and *silver*, which are closely related in the example phrase. The 12 versions of each candidate item for rating were counterbalanced across 12 rating lists such that exactly one version of each stimulus item appeared in each list. The 87 preambles in each list were presented over 5 printed pages, and the pages of each list were randomized separately for each participant. Each participant rated the stimuli in one list, and 9–10 ratings were thus obtained for all but one version of one stimulus item (which had only 8).

Table 1 shows mean ratings by condition and rating pair for the 24 critical stimuli used in the on-line experiment. A 2 (structure) × 3 (number) × 3 (rating pair) ANOVA on these data revealed main effects of structure[2] ($F(1, 22) = 7.71$, $MS_e = 2.33$) and rating pair ($F(2, 44) = 31.39$, $MS_e = 1.59$), and an interaction of the two ($F(2, 44) = 8.98$, $MS_e = 1.59$), with no effect nor interactions involving number (all Fs < 1.6, ps > .20). The main effect of structure and the structure interaction with rating pair resulted only from different N2-N3 ratings for descending and flat structures ($F(1, 22) = 31.32$, $MS_e = 0.49$), as neither the N1-N2 nor the N1-N3 ratings differed across structures (Fs < 1); and an additional structure × number × rating pair ANOVA, leaving out the N2-N3 rating pair, showed only a main effect of pair (N1-N3 more integrated than N1-N2; $F(1, 22) = 16.17$, $MS_e = 2.28$). The critical rating pair by structure interaction in this ANOVA did not approach significance ($F < 1$), indicating that N1-N2 versus N1-N3 integration was matched across structures, as desired. There were no main effects of structure or number (Fs < 1) and no interactions involving these factors (Fs < 1.3, ps > .30). These results indicate that semantic integration was controlled for the critical comparison between the N2 and N3 mismatch effects within each structure. The large difference in N2-N3 integration across structures was also expected given the difference in attachment of the PP containing N3.

The second norming survey, completed by 59 participants, ensured that hierarchical distance was manipulated as desired by measuring attachment of the second PP to N1 versus N2. Each preamble was presented followed by a question, which was always *What is* plus a PP from the preamble. The question always asked about the final PP for the candidate stimuli (e.g., *What is on the leather straps*? for (3c), *What is with the steel guardrail*? for (4a)); for the fillers, the question always asked about the first PP (containing N2), to ensure that participants paid attention to the full text of each item and not just the last PP. Participants were instructed to write down the word from the preamble that best answered the question. The 4 different local noun number versions of each candidate preamble were counterbalanced across 4 lists such that exactly one version of each item appeared in each list. The 63 candidate preambles and 24 fillers in each list were presented over 8 printed pages, and the pages were randomized separately for each participant. Each participant completed a single list, resulting in 14–15 usable responses for each version of each candidate stimulus item.

The responses were coded for whether they referred to N1 or to N2 (unclear or uninterpretable responses, 1% of the total, were excluded), and Table 1 shows mean preference for N1 attachment by condition for the 24 critical stimulus items. A 2 (structure) × 3 (number) ANOVA on the %N1 attachment data revealed a stronger N1 attachment preference for flat stimuli than for descending stimuli ($F(1, 22) = 2201$, $MS_e = 64.09$), as

[2]Throughout all experiments, patterns reported as reliable were significant at or beyond the .05 level unless otherwise noted.

desired. No main effect of local noun number and no interaction were present (Fs < 1.2, ps > .30). In addition to differing in direction as desired, the flat and descending attachment preferences were equally strong in their respective directions: In the flat stimuli, the second PP attached to N1 over 90% of the time, while the second PP in the descending stimuli attached to N1 less than 10% of the time; the strength of these preferences relative to 50% did not differ (F < 1.1, p > .30).

Apparatus and procedure—Each participant was run individually in the on-line experiment using the visual-fragment completion paradigm (e.g., Bock & Eberhard, 1993; Solomon & Pearl-mutter, 2004b; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco & Nicol, 1998; cf. Haskell & MacDonald, 2003). Participants were instructed to begin reading each visually-presented preamble aloud as soon as it appeared and add an ending that formed a complete sentence. Participants were not instructed as to how they should formulate a completion, only that they should form a complete sentence for each preamble.

On each trial, a fixation cross appeared at the left edge of the computer screen for 1000 ms, followed by the preamble. As soon as the preamble appeared, the participant began speaking it aloud, continuing it as a complete sentence. Each preamble was presented for the longer of 1000 ms or 50 ms/character. After the preamble disappeared, the screen was blank for 2000 ms, followed by a prompt to begin the next trial. A PC running the MicroExperimental Laboratory software package (Schneider, 1988) presented the preambles, and participants' responses were recorded uncompressed onto CD-R for analysis, using a Shure SM58 microphone connected to a Mackie 1202-VLZ Pro mixer/preamp and an Alesis Masterlink ML-9600 (OS v2.20) CD recorder. Five practice items preceded the 112 trials. If at any point the participant's speech rate slowed, the experimenter encouraged the participant to speak more quickly.

Scoring—All completions were transcribed and assigned to one of four coding categories: (1) correct, if the participant repeated the preamble correctly, only once, produced an inflected verb immediately after the preamble, and used a verb form that was correctly marked for number; (2) error, if all the criteria for a correct response were met, but the verb form failed to agree in number with the subject; (3) uninflected, if all the criteria for a correct response were met, but the verb was uninflected; and (4) miscellaneous, if the participant made an error repeating the preamble, if a verb did not immediately follow the preamble, or if the response did not fall into any of the other categories. Trials in which a participant made no response were excluded from all analyses. If the participant produced a dysfluency (e.g., pauses, coughs) during or immediately after producing the preamble and went on to produce a correct, error, or uninflected response, the scoring category and the dysfluency were recorded. On miscellaneous trials, dysfluencies were not scored.

Results

Across all critical trials, there were 788 correctly-inflected responses, 43 agreement errors, 199 uninflected responses, 265 miscellaneous cases, and 1 trial with no response. Table 1 shows the counts for each analyzed response type by condition. Separate analyses were performed for error rates (the percentage of error responses out of error plus correct responses), the number of uninflected responses, and the number of miscellaneous responses. All analyses except those involving miscellaneous counts included dysfluency cases, and unless otherwise noted, the statistical patterns were identical if dysfluency cases were excluded. We also computed supplemental analyses on error counts and on arcsine-transformed proportions of errors (Cohen & Cohen, 1983); these are detailed only when they produced significance patterns different from those for the main error rate analyses. For each of these measures, we computed mismatch effects by subtracting from each condition with a

plural local noun (SPS, SSP) the corresponding (structure-matched) singular baseline (SSS). The relevant mismatch effects for each of the measures above were then analyzed in corresponding 2 (structure) × 2 (plural position) ANOVAs, one with participants (F_1) and one with items (F_2; Clark, 1973) as the random factor.

Agreement errors—Figure 3 shows the mismatch effects as a function of structure and plural position. N2 plurals produced larger mismatch effects than N3 plurals ($F_1(1, 53) = 16.04$, $MS_e = 313.04$; $F_2(1, 22) = 13.92$, $MS_e = 93.22$), but the main effect of structure was not significant (Fs < 1.7, ps > .15), and, critically, neither was the interaction (Fs < 1).

Uninflected and miscellaneous responses—There were no reliable main effects or interactions in either the uninflected response count analyses (all Fs < 1) or the miscellaneous response count analyses (all Fs < 2.2, ps > .15).

Discussion

These error patterns, and especially the lack of an interaction, provide strong evidence in favor of a linear distance to the head account of agreement errors and against a hierarchical distance account: The linear distance account correctly predicted a larger mismatch effect for N2 than for N3 mismatches for both descending and flat structures, and it correctly predicted that the size of this difference would be equal for the two structures. The hierarchical account also predicted a larger mismatch effect for N2 than for N3 mismatches, at least in the descending structure cases (following Franck et al., 2002), but it also required an interaction, with a larger difference in mismatch effects for the descending than the flat cases.

In addition, these results indicate that the semantic integration confound in Franck et al.'s (2002) stimuli is not solely responsible for the pattern they found, because semantic integration alone cannot account for the larger N2 than N3 mismatch effect. Integration norming showed that N1-N2 integration was matched across structures, as was N1-N3 integration, and N1-N3 integration was higher than N1-N2 integration in both structures. An account of the results based solely on semantic integration thus predicts a larger N3 than N2 mismatch effect in both structures (as well as no interaction). However, Experiment 1's results do not rule out the Solomon and Pearlmutter (2004b) proposal that semantic integration influences agreement computation in combination with other factors; because semantic integration was not varied across structures, this proposal is consistent with the lack of an interaction in Experiment 1, and the difference between N1-N2 and N1-N3 integration may have contributed (equally) to the difference between the N2 and N3 mismatch effects in the two structures. Experiment 2 was designed to examine this possibility and how semantic integration might interact with linear distance to the head.

Experiment 2

Experiment 1 suggested that linear distance back to the head noun, rather than hierarchical distance, is a critical factor modulating mismatch effects, at least in NP PP PP preambles; but it also left open the possibility that semantic integration has an influence. Solomon and Pearlmutter (2004b) showed that the size of the mismatch effect for NP PP preambles was partially determined by how semantically integrated the head noun and local noun were within the subject NP, irrespective of a local noun's distance back to the head noun and of hierarchical distance. Experiment 2 was thus designed to test whether linear distance to the head noun and semantic integration combine to influence agreement error production, using flat structures like (4) that controlled hierarchical distance. Under a combined linear distance and semantic integration account, the likelihood of interference would be a function of whether the interfering element was within the scope of planning of the head noun; only

local nouns planned close enough in time to the head would be likely to create mismatch effects, with both decreased head-local linear distance and increased head-local semantic integration increasing the chance of overlap in planning. To our knowledge, the effect of scope properties on agreement computation has not previously been investigated.

In Experiment 2, each preamble had an NP PP PP structure, and the number of N2 and N3 was varied, as in (5). One PP (e.g., *with the torn page(s)*) was designed to be highly integrated with the head noun, while the other (e.g., *by the red pen(s)*) was designed to be weakly integrated, and examining both possible PP orderings allowed linear order and semantic integration to be manipulated orthogonally. The stimuli equated hierarchical distance between the head noun (and thus also the verb) and each of the two local nouns by ensuring that both PPs modified the head.

(5)

a. The book with the torn page by the red pen (SSS)

b. The book with the torn pages by the red pen (SPS)

c. The book with the torn page by the red pens (SSP)

d. The book with the torn pages by the red pens (SPP)

e. The book by the red pen with the torn page (SSS)

f. The book by the red pens with the torn page (SPS)

g. The book by the red pen with the torn pages (SSP)

h. The book by the red pens with the torn pages (SPP)

If only linear distance to the head noun is responsible for agreement error rates in these stimuli, the N2 mismatch effect should be larger than the N3 mismatch effect in both the early- and late-integrated versions (5a–d and 5e–h, respectively), as in Experiment 1, because N2 would always be planned closer to N1 than N3 would be.

Although Experiment 1's results cannot be explained by semantic integration, agreement error effects in the Experiment 2 stimuli still might only show effects of integration, in which case, collapsing over integration version, the N2 and N3 mismatch effects should be equal, because the mismatch effect of a given noun should be the same regardless of where the noun appears linearly within the stimulus. However, the N2 and N3 mismatch effects within each integration condition should differ: The N2 mismatch effect should be larger than the N3 mismatch effect for early-integrated cases (5a–d), because the more integrated noun (N2) should be planned with N1, while N3 should be planned later. For late-integrated cases (5e–h), the pattern should reverse, because N3 is the noun more tightly integrated with N1.

The scope of planning alternative, a combined effect of linear distance to the head and semantic integration, predicts that linear distance to the head noun will partially determine the timing of planning of nouns within the phrase, but semantic integration should shift the relative planning time of more and less integrated nouns as well. Figure 4A shows the timing of planning of nouns according to the order in which they are to be produced (this corresponds to the predictions of linear distance to the head alone), and Figure 4B includes the shifting of the timing of planning due to semantic integration. In early-integrated cases (5a–d), because N2 is more integrated with N1, it would be planned at roughly the same time as N1; and because N3 is less integrated with N1, N3 would be planned later. This predicts a large N2 mismatch effect and a very small N3 mismatch effect. In late-integrated cases (5e–h), because N2 is not very integrated with N1, it would be planned later; and

because N3 is more integrated with N1, it would be planned sooner. Thus N2 and N3 should be planned at roughly the same time and relatively late after N1. This predicts that the N2 and N3 mismatch effects should be about equal and fairly small.

Method

Participants—One hundred five Northeastern University students and community members participated in the on-line experiment, but the data from one participant were lost because of a CD recording failure.

Materials and design—Forty stimulus items like (5), chosen from a candidate set of 60, were used as critical items. All preambles consisted of a head NP (always *The* and a singular head noun, N1; e.g., *book* in (5)) followed by two PPs, each of which consisted of a preposition, the determiner *the*, an adjective or modifier noun, and a local noun (N2 or N3), as in the Experiment 1 stimuli. One PP always described an attribute of the head noun using the preposition *with* (e.g., *with the torn page*), while the other PP specified a location for the head noun and used a locative preposition (e.g., *by the red pen*). The eight different versions of an item were created by varying N2 and N3 number and PP order, as shown in (5). The versions with the attribute PP first were the early-integrated versions (5a–d), and those with the attribute PP last were the late-integrated versions (5e–h). All nouns in the stimuli were inanimate and had regular plural forms, and each noun's conceptual number matched its grammatical number. The full set of stimuli is shown in Appendix B.

The 40 critical stimuli were combined with 80 of the fillers from Experiment 1 (all but the 8 SPP fillers used in Experiment 1 to balance the critical items' SSS, SPS, and SSP cases) to form 8 counterbalanced presentation lists, each containing all the fillers and exactly one version of each of the critical items. One list was seen by 14 participants while the other seven were seen by 13 participants each.

Stimulus norming—Semantic integration and attachment were normed similarly to Experiment 1. In addition to the 60 candidate stimuli, 33 NP PP PP fillers were included. The fillers had a descending hierarchical structure and did not have adjectives within the PPs; but they varied local noun number as in the candidate stimuli, and they included a range of levels of semantic integration between N1, N2, and N3.

Semantic integration ratings were obtained from 186 participants, using the same procedures and instructions as in Experiment 1. The 24 different versions of the 60 candidate stimuli (2 PP orders × 4 number conditions × 3 possible rating pairs) were combined with the 33 fillers in 24 counterbalanced lists, such that each list included exactly one version of each item. The 93 preambles in each list were presented over 6 printed pages, which were randomized separately for each participant. Each participant rated the stimuli in one list, yielding 7–8 ratings for nearly all versions of all critical stimuli (3 different items had only 6 ratings for one of their versions).

Table 2 shows mean ratings by condition and rating pair for the 40 critical stimuli used in the on-line experiment. These were analyzed in a 2 (PP order) × 2 (N2 number) × 2 (N3 number) × 3 (rating pair) ANOVA, which showed the desired interaction between PP order and rating pair ($F(2, 78) = 1025$, $MS_e = .97$): N1-N2 integration was greater than N1-N3 integration in the early-integrated stimuli ($F(1, 39) = 1137$, $MS_e = .22$), while this pattern reversed in the late-integrated stimuli ($F(1, 39) = 1140$, $MS_e = .21$), and N2-N3 integration was equal for early- and late-integrated cases ($F < 1$). Paired tests also confirmed that, as intended, N1-N2 integration was greater for early- than for late-integrated versions ($F(1, 39) = 1157$, $MS_e = .22$), N1-N3 integration was greater for late- than for early-integrated versions ($F(1, 39) = 1016$, $MS_e = .24$), and integration did not differ for the same local noun

in different positions (i.e., N1-N2 for early-integrated vs. N1-N3 for late-integrated, and N1-N2 for late-integrated vs. N1-N3 for early-integrated; both Fs < 1).

The overall ANOVA on integration ratings also revealed a main effect of rating pair ($F(2, 78) = 908$, $MS_e = .41$), which arose because the N2-N3 pair was overall less semantically integrated than the N1-N2 or N1-N3 pair; given the modification structure of the stimuli and the nature of the PP order by rating pair interaction described above, this was expected. However, in addition to these very strong effects, the overall ANOVA also yielded two weaker results involving local noun number: First was a reliable interaction between PP order, N2 number, and rating pair ($F(2, 78) = 3.51$, $MS_e = .21$, $p < .05$), which arose because the PP order by rating pair interaction described above was stronger when N2 was plural than when it was singular: The N1-N2 versus N1-N3 difference for early-integrated was 3.58 for plural N2 and 3.48 for singular N2; the corresponding differences for late-integrated were −3.65 and −3.38. The second result was a marginally significant N2 number by N3 number interaction ($F(1, 39) = 3.13$, $MS_e = .20$), which arose because the difference in integration ratings between the SPS and SPP conditions was slightly larger than the difference between the SSP and SSS conditions. We will return to both of these unexpected results below, with the discussion of the results of the on-line experiment. No other main effects nor interactions appeared in the overall ANOVA (all Fs < 2.6, ps > .10).

To ensure that all critical stimuli had a flat structure, attachment of the second PP was normed by 150 participants. The procedure was slightly different from Experiment 1's attachment norming, in that the final PP of each preamble was underlined, and participants were instructed to write the word that the underlined portion of the preamble "gives more information about". The 8 versions of each candidate stimulus were counterbalanced across 8 lists in combination with the 33 fillers, and the preambles of each list were presented over 6 printed pages, which were randomized separately for each participant. This yielded 18–19 responses for each version of each critical item.

As in Experiment 1, the responses were coded for whether they referred to N1 or to N2 (excluding unclear or uninterpretable responses, < 1% of the total), and Table 2 shows mean preference for attachment to N1 by condition for the 40 critical stimulus items. Attachment of the second PP was strongly to N1 in all conditions, as desired, and a 2 (PP order) × 2 (N2 number) × 2 (N3 number) ANOVA on the %N1 attachment data revealed only that plural N2 cases (98.8%) were slightly more strongly N1-attached than singular N2 cases (97.3%; $F(1, 39) = 4.91$, $MS_e = 32.24$). No other main effects and no interactions were significant (all Fs < 1.8, ps > .15).

Apparatus, procedure, and scoring—The apparatus, procedure, and response scoring for Experiment 2 were identical to Experiment 1, except that there were 120 trials.

Results

Across all critical trials, there were 2,354 correctly-inflected responses, 123 agreement errors, 796 uninflected responses, 923 miscellaneous cases, and 1 trial with no response. The remaining 3 trials were lost due to a recording failure for one subject. Table 2 shows the counts of each analyzed response type by condition. All analyses were conducted as in Experiment 1, except that the (within-items) integration factor replaced Experiment 1's (between-items) structure factor in the ANOVAs.

Agreement errors—Figure 5 shows N2 and N3 mismatch effects for each integration condition. The mismatch effect ANOVA showed no main effect of integration (Fs < 1), but N2 plurals created larger mismatch effects than N3 plurals ($F_1(1, 104) = 13.74$, $MS_e = 237.78$; $F_2(1, 39) = 17.57$, $MS_e = 65.84$). Furthermore, plural position and integration

interacted ($F_1(1, 104) = 9.31$, $MS_e = 203.33$; $F_2(1, 39) = 4.13$, $MS_e = 78.65$), because the N2 mismatch effect was larger than the N3 mismatch effect for early-integrated stimuli ($F_1(1, 104) = 22.24$, $MS_e = 227.88$; $F_2(1, 39) = 15.76$, $MS_e = 85.94$), while the two mismatch effects did not differ for late-integrated stimuli ($F_1 < 1$; $F_2(1, 39) = 2.18$, $MS_e = 58.55$, $p > .10$).

The other four pairings of the four mismatch conditions were also tested with planned comparisons: The early-integrated N2 plural condition generated larger mismatch effects than the late-integrated, though this was reliable only in the analysis by participants ($F_1(1, 104) = 4.42$, $MS_e = 292.87$; $F_2(1, 39) = 1.76$, $MS_e = 117.97$, $p > .15$; the analysis of error counts by participants was also only marginal, and the analysis by participants excluding dysfluencies did not reach significance). The pattern reversed for N3 plurals, where the late-integrated condition showed larger mismatch effects than the early-integrated ($F_1(1, 104) = 4.43$, $MS_e = 147.07$; $F_2(1, 39) = 4.68$, $MS_e = 26.25$; the difference was only marginal in the analysis of error counts by participants). The tests comparing the same phrases in different linear positions also showed differences: The early-integrated N2 plural condition created larger mismatch effects than the late-integrated N3 plural condition ($F_1(1, 104) = 5.96$, $MS_e = 349.55$; $F_2(1, 39) = 8.34$, $MS_e = 79.32$; this was nonsignificant when dysfluencies were excluded), and the late-integrated N2 plural condition created larger mismatch effects than the early-integrated N3 plural condition ($F_1(1, 104) = 9.92$, $MS_e = 124.84$; $F_2(1, 39) = 9.62$, $MS_e = 52.09$).

Uninflected responses—There were no reliable main effects or interactions for uninflected responses (all Fs < 1).

Miscellaneous responses—N2 mismatches generated more miscellaneous responses than N3 mismatches ($F_1(1, 104) = 6.67$, $MS_e = .86$; $F_2(1, 39) = 7.88$, $MS_e = 1.90$), but there was no main effect of integration (Fs < 1) and no interaction (Fs < 1.4, ps > .20).

Discussion

Experiment 2 examined how effects of linear distance to the head and of semantic integration might be involved in determining the size of mismatch effects. Because N2 was always linearly closer to N1 than N3 was, linear distance alone predicted that the N2 mismatch effect should be uniformly larger than the N3 mismatch effect. While this difference appeared for the early-integrated stimuli, the N2 and N3 mismatch effects did not differ for the late-integrated stimuli. The semantic integration account predicted that the mismatch effects for the more integrated cases (N2 in the early-integrated stimuli, N3 in late-integrated) should not differ and should both be larger than for the less integrated cases (N3 in early-integrated, N2 in late-integrated; these also should not differ from each other). Instead, the early-integrated N2 mismatch effect was larger than the late-integrated N3 mismatch effect (at least when dysfluencies were included), and the late-integrated N2 mismatch effect was larger than the early-integrated N3 mismatch effect, showing that the more integrated cases did differ, as did the less integrated cases; and the two late-integrated cases did not differ in mismatch effects despite differing in semantic integration. Semantic integration alone also predicted no main effect of plural position (the N2 and N3 mismatch effects should have been equal). Thus, neither linear distance to the head noun nor semantic integration on its own is sufficient to explain the pattern of effects.

Instead, error rates showed a pattern reflecting a combination of linear distance to the head and semantic integration, which fits with the scope of planning account. The scope account in general suggests that mismatch effects will be more likely to occur for local nouns planned in closer proximity to the head noun, and factors that affect planning proximity, in

the combination case including both linear distance from the head and semantic integration, can thus affect mismatch error rates. More specifically, this account predicts a relatively large difference in mismatch error rates between the two early-integrated cases, because N2 is both more integrated and linearly closer to the head than N3; and it predicts a relatively small difference between the two late-integrated cases, because N2 is linearly closer to the head than N3 is, but N3 is more integrated with the head than N2 is. Thus semantic integration reinforces the linear distance difference in the early-integrated cases, but it counteracts the linear distance difference in the late-integrated cases, and this is the pattern seen in mismatch effects: For early-integrated stimuli, the N2 mismatch effect was larger than the N3 mismatch effect, but for late-integrated stimuli, they did not differ.

Although the results point to a combination of linear distance to the head and semantic integration as responsible for the mismatch effect pattern, and neither of these alone is sufficient to explain the full pattern of results, the planned comparisons also provided evidence for the influence of each factor when the other was controlled: In particular, the comparisons of the early- versus late-integrated cases within each plural position (N2, N3) are a direct test of the effect of semantic integration when linear distance was controlled. These comparisons revealed that the more integrated noun at each plural position generated larger mismatch effects than the corresponding less integrated noun (early-integrated N2 vs. late-integrated N2; late-integrated N3 vs. early-integrated N3). Direct tests of the effect of linear distance when semantic integration was controlled come from the comparisons in which the same PP occurred in the two different linear positions: Within the two attribute PPs (early-integrated N2 vs. late-integrated N3), the N2 cases generated a larger mismatch effect; and the same was true for the two locative cases (late-integrated N2 vs. early-integrated N3). Because the scope of planning account incorporates effects of both linear distance to the head and semantic integration, it predicts these results as well.

The evidence that both linear distance to the head and semantic integration are relevant for a general scope of planning account also raises the issue of exactly how they combine. While they can both be seen as influencing the timing of planning, whether they are independent influences or not is an open question. For example, if the difference in timing of planning for N2 versus N3 created by linear distance is large enough that N3 is essentially always planned too long after N1 to interfere with agreement, then semantic integration might only have an effect for mismatches at N2, or it might have a much weaker effect at N3 compared to N2. The relevant statistical test of the general interaction — whether the difference between more- and less-integrated cases was different at the N2 position versus at the N3 position — is equivalent to the main effect of integration, and this effect did not approach significance. However, the numerical pattern of mismatch effects did include a larger effect of integration at N2 compared to N3; future work designed specifically to investigate this issue should therefore be enlightening.

Before concluding that these patterns are entirely the result of scope of planning effects, however, we need to consider whether the two unexpected interactions involving noun number found in the integration rating analyses — N2 number × N3 number and PP order × N2 number × rating pair — might be relevant. Both of these interactions involved numerically very small differences relative to the intended integration manipulations, and the former was in fact only marginal rather than significant. The N2 number × N3 number interaction also does not account for any of the critical patterns in the mismatch effects, as it does not involve PP order, and it is linked to the SPP condition ratings, which were irrelevant for mismatch effects. But the other of the two interactions, involving PP order, could in principle be involved in the pattern of paired comparisons for the same noun in different linear positions (early-integrated N2 vs. late-integrated N3; early-integrated N3 vs. late-integrated N2). These comparisons provided evidence for linear distance effects when

semantic integration was controlled, but the interaction in the norming shows that integration was not perfectly controlled, and this might provide an alternative explanation for the differences in mismatch effects. In fact, the N1-N2 rating for the early-integrated SPS condition was slightly higher than the N1-N3 rating for the late-integrated SSP condition, which correctly predicts the direction of the corresponding mismatch effect comparison (a larger mismatch effect for the early-integrated N2 case than for the late-integrated N3 case). However, for the other pair, the mismatch effect continued to follow the linear distance prediction, with the late-integrated N2 mismatch effect larger than the early-integrated N3 mismatch effect, whereas the interaction in the ratings resulted from the opposite pattern: The N1-N2 rating for the late-integrated SPS condition was slightly lower than the N1-N3 rating for the early-integrated SSP condition. As a result, the rating interaction pattern diverges from the mismatch effect pattern. Thus neither of the unexpected interactions from the rating analyses seems likely to be responsible for the mismatch effects.

There was also a main effect of plural position on miscellaneous responses, with a greater increase in such responses over the SSS baseline for N2 plurals compared to N3 plurals. This suggests that N2 plural cases were generally more complex or more difficult than N3 plural cases, at least in Experiment 2, and it obviously matches the main effect of plural position in the agreement error analyses, but not the critical interaction. Nevertheless, if N2 plural cases were generally more difficult to produce than N3 plural cases, this would provide a potential alternative explanation for the effects attributed to linear distance to the head. Experiment 2's results do not rule out this possibility, but Experiment 1 also showed clear support for effects of linear distance to the head, without any comparable effects or interactions in miscellaneous errors. Thus the linear distance to the head account appears to be the more robust explanation.

General Discussion

Together, the results of Experiments 1 and 2 provide no support for hierarchical feature-passing. In Experiment 1, when factors known to influence agreement computation were controlled and hierarchical distance was manipulated, there was no effect of structure on the size of mismatch effects; instead, only a local noun's linear proximity to the head noun affected error rates. In addition, when structure was controlled in Experiment 2, a combination of linear distance and semantic integration controlled error rates. The results of these experiments point to an account of agreement computation that relies on memory encoding and a limited scope of planning. Mismatch effects are the result of the extent to which the head noun and interfering local noun(s) are simultaneously active in memory when the number of the subject NP is being computed. The timing of planning of elements within a phrase is determined by the order in which elements are to be produced, and semantic integration shifts the relative timing of planning. Agreement errors are likely to occur when a number-mismatching local noun is planned within the scope of (i.e., close in time to) the head noun: Because of the overlap in planning, the nouns and their corresponding number elements are likely to be active simultaneously and are likely to interfere at the time when the number marking of the subject NP is set.

We discuss the implications of these results further below, but there are three potential concerns with the current evidence to be considered, related to the syntactic structure of the stimuli. The first concern is that while the offline norming established that the stimuli were interpreted as having the desired syntactic structures (flat or descending), initial interpretations during the online experiments could have been different, particularly in the flat cases. English comprehenders prefer to attach new material to more recent over less recent (otherwise-equivalent) sites (e.g., Frazier & Clifton, 1996; Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996), and in the Experiment 1 and 2 stimuli this would have

resulted in the second PP modifying the more recent N2 (within the first PP) instead of the less recent N1. In Experiment 1, this was the correct interpretation for the descending stimuli but would have been an incorrect interpretation for the flat stimuli; it would have been the incorrect interpretation for all the critical stimuli in Experiment 2, which all had a flat structure. Although there was no specific test of attachment of the second PP during the online experiments, the miscellaneous errors provide a measure of general difficulty of the preambles; and if flat structures were initially interpreted incorrectly and then reanalyzed, or if they were simply left incorrectly attached without reanalysis, yielding semantic anomalies, then flat structures should have generated more miscellaneous errors than descending structures in Experiment 1. But this was not the case: There was no effect of structure nor an interaction in the analysis of mismatch effects, and as Table 1 shows, the overall counts of miscellaneous errors were nearly identical (135 for descending, 130 for flat; Fs < 1). Coupled with the relative weakness of the recency effect in English comprehension for attachments to noun sites in particular (e.g., Cuetos & Mitchell, 1988; Gibson et al., 1996; Pearlmutter & Gibson, 2001), the lack of any difference in mismatch errors for flat versus descending structures suggests that the second PP was attached as intended (and as normed) for both structures and in both experiments.

The second possible concern related to the structure of the stimuli is that the argument/adjunct status of the PPs containing the local nouns might have varied across conditions, creating confounds. Solomon and Pearlmutter (2004b) showed that argument/adjunct status and related structural properties in their NP PP stimuli could not account for their semantic integration results, but it still might be involved in the effects here. Determining the argument/adjunct status of PP modifiers of NPs is often difficult, but Schütze and Gibson (1999) offer six diagnostics of PP argumenthood, although they note that individual speakers may disagree in applying a particular diagnostic to a particular case, and there may be disagreement across diagnostics even when speakers agree. When we applied the diagnostics to the Experiment 2 stimuli, there was in fact strong agreement (across the authors and the diagnostics) that both the first and second PPs were adjuncts, eliminating any basis for a confound. In the descending stimuli of Experiment 1, both PPs were similarly clearly adjuncts; and this was also the case for the second PP of the flat stimuli in Experiment 1. The only case where there appeared to be any support for an argument classification was for two of the first PPs in the flat stimuli (these both involved *of*, as in *The postcard of the roller coaster with the foreign stamp*), which two of the six diagnostics (Schütze & Gibson's "ordering" and "iterativity") seemed to classify as arguments. If these PPs were indeed arguments rather than adjuncts, then according to syntactic theories which encode the argument/adjunct distinction in structure (e.g., Carnie, 2005; Chomsky, 1995; Pollard & Sag, 1994), they would be attached farther from the head NP node than corresponding adjuncts. For versions of the hierarchical feature-passing hypothesis that compute distance based on all intervening nodes (versus only major categories; see Solomon & Pearlmutter, 2004b), this would in turn predict that N2 would be less likely to create errors in these cases. But the result of this potential confound is that the difference between the sizes of the N2 and N3 mismatch effects should be even smaller for flat cases than for descending cases, because in the flat cases, N2 will be more distant from the head NP than N3 (the difference in the size of the mismatch effects will be negative), while in the descending cases, N2 will be closer to the head NP than N3. Thus, if it has any effect, it should be to strengthen the interaction prediction for hierarchical feature-passing. There was no hint of such an interaction in Experiment 1, and thus an argument/adjunct confound cannot explain the results.

The final concern related to the structure of the stimuli is that in some syntactic theories (e.g., Chomsky, 1995), flat structures are ruled out; typically this is because such theories enforce binary branching (Kayne, 1984), stipulating that no syntactic node may have more

than two daughters. The evidence for a binary-branching requirement is largely theory-internal, depending primarily on what other assumptions a theory holds about the relationship between syntactic structure, meaning, and discourse modification structure; and many syntactic theories allow flat structures (e.g., Culicover & Jackendoff, 2006; Goldberg, 2006; Pollard & Sag, 1994). But if the structure of the stimuli in the current experiments were required to be binary-branching, both the descending (Figure 1) and flat (Figure 2) stimulus structures would be altered. In the case of the descending stimuli, this would have no influence on the predictions of hierarchical feature-passing: N3 would still be more deeply-embedded and more distant from the subject NP node than N2. Changing the flat structures to enforce binary branching, on the other hand, would alter the predictions of hierarchical feature-passing. The exact structural alterations would depend on the specific theory, but because the first PP modifies N1 and the second PP is independent of the first PP, the second PP will attach above the first PP, by adjunction to the subject NP node (e.g., Carnie, 2005). The result of this will be that the hierarchical path for a feature from N3 to reach the top of the full subject NP will be shorter than the path for a feature from N2, and thus a hierarchical feature-passing approach built on binary-branching syntactic structures would predict an even clearer interaction than the one described for Experiment 1: The descending structures should show a larger mismatch effect for N2 than for N3, while the "flat" structures should show a larger mismatch effect for N3 than for N2. Of course, Experiment 1 showed no interaction at all: The difference between the N2 and N3 mismatch effects was equivalent for the two different structures, with N2 yielding a larger mismatch effect than N3. Thus while we cannot evaluate hierarchical feature-passing against every conceivable set of syntactic structures, its key prediction of an interaction in Experiment 1 fails using either the flat or the most likely binary-branched structures.

Assuming, then, that the structures in the Experiment 1 and 2 stimuli were constructed as intended, the experiments provide strong evidence that hierarchical distance and hierarchical feature-passing are not relevant in determining agreement error patterns. This claim permits reconsideration of some results in the literature, where the scope of planning account might provide an alternative explanation. As discussed above, Franck et al.'s (2002) results are one such case, but three other agreement error patterns have been explained at least in part with reference to hierarchical mechanisms: First, sentential objects can produce mismatch effects when they intervene between the subject NP and the verb (Chanquoy & Negro, 1996; Fayol, Largy, & Lemaire, 1994; Franck et al., 2006; Hartsuiker et al., 2001; see also Hemforth & Konieczny, 2003, for related data from German); second, the relative order of the subject and verb can affect error rates (Franck et al., 2006; Vigliocco & Nicol, 1998); and, third, local nouns in PP modifiers produce larger mismatch effects than local nouns in clausal modifiers (Bock & Cutting, 1992; Negro et al., 2005; Solomon & Pearlmutter, 2004b). How the scope of planning account addresses these patterns is discussed below.

Hartsuiker et al. (2001) showed that local nouns in direct objects can interfere with subject-verb agreement when they appear between the subject noun and the verb, though they produce smaller mismatch effects than local nouns in PP modifiers (cf. Franck et al., 2006; Hartsuiker et al.; and references therein; for more complex cross-linguistic patterns involving object pronouns and clitics). Examining Dutch, Hartsuiker et al. presented participants with preambles like (6), consisting of a matrix clause followed by the start of a subordinate clause and then an uninflected verb stem (*WIN*) to be used to complete the subordinate clause (Dutch subordinate clauses are verb-final). Both pairs produced mismatch effects, but larger effects occurred for subordinate clauses containing a subject-modifying PP (*met de krans(en)* in (6a)) than for those containing a direct object (*de krans(en)* in (6b)). Hartsuiker et al. interpreted the presence of mismatch effects for both cases as support for hierarchical feature-passing, and they interpreted the difference in mismatch effect size as a consequence of the direct object noun's greater hierarchical

distance from the subject NP node: An interfering feature from the direct object would have to pass up out of the direct object NP, to the subordinate clause VP and/or the top of the subordinate clause itself, and then to the subject NP; whereas in the PP-modifier case, an interfering feature would only have to pass up out of the prepositional-object NP and the PP itself (just as in corresponding English cases).

(6)

a. Karin zegt dat het meisje met de krans(en) WIN

Karen said that the girl with the garland(s) WIN

b. Karin zegt dat het meisje de krans(en) WIN

Karen said that the girl the garland(s) WIN

Although a hierarchical feature-passing account can explain these results, they do not provide direct evidence for such an account, and the scope of planning explanation can also explain them. In particular, if constituents are generally planned in their utterance order (Bock, Loebell, & Morey, 1992, provide some support for this view), the direct object NP in (6b), like the corresponding NP in (6a), will be planned between the head noun of the subject NP and the verb, allowing plurals in either case to create some interference in encoding the number of the subject. The difference between the two cases is likely created by semantic integration: Most of the subject-modifying PPs in (the English translations of) Hartsuiker et al.'s (2001) Experiment 1a and 1b stimuli appear to be strongly integrated, which would yield relatively large mismatch error effects. In the direct object cases, we assume that verbs can potentially create high levels of integration between their arguments (e.g., Solomon & Pearlmutter, 2004b, Exp. 5), but even assuming Hartsuiker et al.'s uninflected target verbs did so, the link created by the target verb in the direct object conditions would have been unavailable until relatively late in the processing of the two NPs, compared to the preposition in the subject-modifying PP case. The result would be that the subject and object would behave as if they were relatively unintegrated, limiting the size of mismatch effects. Thus the scope of planning account can account for both the presence of mismatch effects in each case, as well as the relatively larger effects in the subject-modifier case (6a).

An additional finding that the scope of planning account can address without requiring a hierarchical component concerns word order variation. In Franck et al. (2006; Exp. 1), participants were presented with an NP PP subject phrase (in Italian) followed by an infinitival verb form (e.g., 7a), and participants were required to construct a sentence using those materials, inflecting the verb, that had either a subject-verb (7b) or verb-subject (7c) order. More agreement errors were observed in the subject-verb cases, where the local noun (*ragazzi*) appeared between the head noun (*vicino*) and the verb (*viene*), than in the verb-subject cases. Franck et al. make several specific assumptions about the syntactic representations and processes involved in producing the subject-verb and verb-subject structures, in order to fit them into their framework; but the critical property they use to differentiate the two structures is linear precedence, and specifically whether the local noun intervenes between the head noun and verb. Under a scope of planning account, with hierarchical factors irrelevant, essentially this same linear order factor would differentiate the two cases: Verb planning in verb-subject cases would begin and could potentially be completed before the local noun is planned, yielding few opportunities for interference; while in subject-verb cases, the local noun will essentially always be planned before the verb and will thus have ample opportunity to interfere with head number tracking. (An order of planning account can similarly explain Franck et al.'s Experiment 3 result with mismatching local nouns in direct objects, in which French speakers produced more errors in inflecting verbs within object-verb-subject sequences than within object-subject-verb sequences. See

also Hupet, Fayol, & Schelstraete, 1998, for related evidence from the French PP-inversion construction, which also creates an object-verb-subject sequence.)

(7) a. il vicino dei ragazzi venire
the neighbor of the boys to come
b. Il vicino dei ragazzi viene.
c. Viene il vicino dei ragazzi.

Franck et al. (2006) point out that their Italian subject-verb versus verb-subject construction results contrast with Vigliocco and Nicol's (1998) finding that local nouns in English declaratives (subject-verb word order; e.g., *The helicopter for the flights is safe*.) and interrogatives (verb-subject; e.g., *Is the helicopter for the flights safe?*) produce equal agreement error mismatch effects. Franck et al. account for the difference in patterns with the idea that the English interrogative (unlike the Italian verb-subject case) is created from the declarative after agreement has been computed, so the difference in produced word order has no effect. A scope of planning account would not predict this lack of an interaction with word order; but the Vigliocco and Nicol experiments also used an altered version of the typical agreement error elicitation task, which may have contributed to the lack of an interaction: In both of Vigliocco and Nicol's experiments, participants were presented with an adjective followed by a subject NP. In all cases, including fillers, the response was the subject NP and then the adjective, with either *is* or *are* inserted in the appropriate position (after or before the subject noun phrase). Because declaratives and interrogatives were produced in separate experiments, a given participant always produced the verb in the same position, and the task may have operated essentially as a forced choice procedure between the two possible verb forms, with an additional memory component. Thus their observed interference effects may not have been the result of natural agreement processing. Further experiments will be needed to determine whether these constructions present a problem for a scope of planning account.

One effect that the scope of planning account cannot easily explain is the difference in mismatch effect size for local nouns embedded in phrases versus clauses. Bock and Cutting (1992, Exp. 1) compared PP modifier preambles (e.g., *The editor of the history book(s)*) with corresponding length-matched clausal modifier preambles (e.g., *The editor who rejected the book(s)*) and showed that mismatch effects were larger for the phrasal cases. They suggested an explanation in terms of clause boundedness: the idea that clauses are planned independently, and elements within separate clauses cannot interfere with each other. Solomon and Pearlmutter (2004b, Exp. 5) replicated this result and showed that it could not be explained by semantic integration; integration between the head noun (e.g., *editor*) and the local noun (*book(s)*) did not differ across structures in either the Bock and Cutting or the Solomon and Pearlmutter stimuli. Differences in linear distance to the head noun also cannot explain the effect, as the phrasal and clausal modifier stimuli in both experiments were matched on number of intervening syllables.

One possible account of the phrasal versus clausal modifier results is that clause boundedness influences agreement processes independently of scope of planning factors, essentially as Bock and Cutting (1992) described. However, Franck et al. (2002) showed that the phrasal versus clausal modifier effect could instead be explained by a hierarchical distance account, because the verb phrase and potentially other structure needed to instantiate the clausal modifier embeds the local noun more deeply than in the phrasal modifier case. Solomon and Pearlmutter (2004b) noted that their own results were compatible with an independent effect of either clause boundedness or hierarchical feature-passing, and recent work by Bock and colleagues (Eberhard et al., 2005) has assumed a hierarchical feature-passing approach in general, although they did not discuss clausal

modifiers in particular. The current results suggest hierarchical feature-passing is unlikely to be the source of errors in these constructions (see below), but neither Experiment 1 nor 2 examined clausal cases, so this leaves open the possibility that the mechanism underlying agreement computation is hierarchical feature-passing, but hierarchical distance depends only on syntactic nodes associated with clausal structure (e.g., CP, IP or S, VP). This would be a substantially different version of hierarchical feature-passing than what has previously been proposed, but we know of no data that rule it out.

In addition to alternative explanations for some results in the literature, Experiments 1 and 2 also suggest a reconsideration of some aspects of current agreement models. In particular, because many of the results in the literature have been connected to hierarchical feature-passing, current models of agreement computation often incorporate a hierarchical component (Eberhard et al., 2005; Vigliocco & Hartsuiker, 2002; cf. Stevenson, 1994, in comprehension). In Eberhard et al.'s (2005) marking and morphing model, for example, as described above, the morphing process weights the effect of each individual noun's number information on the overall subject NP's number, based on that noun's hierarchical distance from the subject NP node. This weighting is critical for the model, because it enables it to capture the controlling influence of the head noun over the local noun, and thus the production of primarily grammatical agreement, when head noun number and conceptual number diverge, as in "distributive" subject NPs like *The label on the bottles* (e.g., Bock et al., 2001; Eberhard, 1999; Vigliocco et al., 1996). The current results would thus have two main implications for this model: First, Experiment 1 suggests that hierarchical distance might not be the appropriate determinant of weights, as the model cannot account for the lack of an interaction with structure in Experiment 1 if local noun weights are based on hierarchical distance. Second, the scope account suggests the alternative of setting weights on the basis of the relative timing of planning of number-bearing elements, including the combination of linear distance to the head and semantic integration seen in Experiment 2. This change would allow the model to handle both the Experiment 1 and 2 results, as well as Franck et al.'s (2002) pattern and the effects attributed to hierarchical distance in Hartsuiker et al. (2001), while still accounting for the cases specifically modelled by Eberhard et al. (always NP PP preambles).[3]

In fact, simply basing weights on linear distance to the head would be sufficient to account for most of the results considered by Eberhard et al. (2005). Incorporating semantic integration into the weighting process is probably necessary to account for the current Experiment 2 results as well as the difference Hartsuiker et al. (2001) found between local nouns in direct objects and in subject-modifying PPs; but the handling of semantic integration in the marking and morphing model is also complicated by two factors: First, there may be unidentified differences in integration in the stimuli modeled by Eberhard et al. (see Solomon & Pearlmutter's, 2004b, meta-analyses, for some discussion), which could obviously affect the model's performance. This can only be handled by gathering additional semantic integration data and evaluating its effect on the model's fit to the human error data.

The second complication is that Eberhard et al. (2005) did account for Solomon and Pearlmutter's (2004b) overall semantic integration effect, but they did so by way of the marking process: More integrated phrases were assumed to mark the subject NP as a whole as more strongly plural, and, specifically, they were treated as cases with "ambiguous notional number, such as subject phrases denoting masses… collections, or distributions" (Eberhard et al., p. 543). This approach to semantic integration is in principle a possibility,

[3]Eberhard et al. note (p. 551) that their model "takes no account of time, incrementality, or variations in syntactic complexity apart from structural distance"; so our proposal might be seen as remedying this while simultaneously removing any need for sensitivity to structural distance.

but it appears to run into some difficulty in its application to various experiments: For example, in Solomon and Pearlmutter's Experiment 1, the conceptual representation of the overall preamble for the integrated versions (e.g., *The drawing of the flowers*) could be argued to be more like a mass because the local noun's referent is incorporated into the head noun's; but it is still a single entity, and the unintegrated versions (e.g., *The drawing with the flowers*) appear to refer more clearly to multiple entities, which would suggest the opposite of what Solomon and Pearlmutter found. Solomon and Pearlmutter's Experiment 4 stimuli have similar properties, and it is difficult to see how their Experiment 2 and 3 stimuli involve incorporation or masses at all; intuitions suggest that both the integrated (e.g., *The chauffeur for/of the actors*) and unintegrated stimuli (e.g., *The chauffeur with the actors*) refer about equally to multiple individuals.

The use of marking to handle semantic integration also runs into problems with the current Experiment 2: Marking by definition applies to the subject NP as a whole, and while we did not separately define an overall integration measure, the early- and late-integrated cases are matched on average integration of the three noun-noun relationships, so if marked number depends on this, the model will fail to account for either of the within-position comparisons between the early-and late-integrated conditions (the mismatch effect from N2 was larger for early- than for late-integrated cases, and the pattern reversed for N3 mismatch effects). One could instead stipulate that marking's sensitivity to semantic integration is based more heavily on the integration of N1 and N2, or instead on the N1-N3 relationship, but either of these will make the wrong prediction for the mismatch effect difference at the other position. Using the maximum semantic integration (or the minimum) of the pairs in the subject NP fails as well, because the early- and late-integrated stimuli are matched on these.

Given these issues, reliance on the marking process to handle the full range of semantic integration effects thus seems problematic. While conceptual differences associated with semantic integration manipulations can have correlated consequences for conceptual number (e.g., *The engine for the cars* is probably both conceptually more plural and more semantically integrated than is *The engine beside the cars*), the two are separable. Marking is certainly needed within the marking and morphing model to handle cases where conceptual number of the subject NP diverges from the number of any of its constituent lexical items (see, e.g., Eberhard et al.'s (2005, p. 536) discussion of the metonymic example *The hash browns at table nine is getting angry*); but semantic integration effects appear to be independent of marking, and the current work suggests that they can instead best be captured in Eberhard et al.'s model by altering the weighting mechanism needed for morphing.

The idea that both linear distance to the head and semantic integration affect Eberhard et al.'s (2005) weighting parameter is a specific implementation of the general claim that both of these properties specifically affect timing of planning. If language production proceeds at least somewhat incrementally (e.g., Bock & Levelt, 1994; Brown-Schimdt & Konopka, 2008; Brown-Schmidt & Tanenhaus, 2006; Griffin, 2001), the linear distance to the head manipulation straightforwardly varies timing of planning of the local nouns in the current studies; however, the link between timing and semantic integration is not as clear. Solomon and Pearlmutter (2004a, b) proposed that semantic integration effects on agreement error rates resulted from more highly integrated elements being planned with more overlap in timing, but while their results were compatible with a timing account, they did not have direct evidence for one. Pearlmutter and Solomon (2007) argued for a timing account of semantic integration based on exchange error patterns — more integrated phrases yielded more exchanges, suggesting that their elements were more likely to be available simultaneously — and the current findings provide additional support for interpreting semantic integration effects as reflecting timing, in that doing so allows for a unified

explanation of the results. Placing both linear order and semantic integration on the same temporal scale also suggests the possibility that they might interact, with stronger or weaker effects of semantic integration at N2 versus at N3. This turned out not to be the case, although the numeric pattern of effects suggested a larger integration effect at N2 than at N3. The lack of a significant interaction may only indicate that the two factors are independent influences on timing, but further research will be required to establish that semantic integration affects timing of planning, and if it does, how it interacts with other planning phenomena.

Beyond the specific consequences of our results for the existing literature is the more general question about the source of agreement errors and the mechanism of agreement computation. Experiments 1 and 2 do not directly rule out hierarchical feature-passing as the mechanism implementing agreement; but they show that the evidence purportedly supporting it in the literature is confounded and thus inconclusive, and we argued above that with a scope of planning account, feature passing is not needed to explain known results. In addition, hierarchical feature-passing cannot be the source of the error patterns in Experiments 1 and 2; the experiments implicate factor(s) related to temporal planning distance from the head of the subject NP, rather than syntactic distance from the subject NP. Furthermore, if feature passing is the mechanism implementing agreement, the current results severely limit its application and especially its potential as a source of errors, because the explanation for the lack of an influence of structure in Experiment 1 is that the structure associated with the second PP — through which errant features from N3 would have to pass — is not yet present when agreement properties of the subject are computed. At the same time, the structure corresponding to the first PP must be present to account for the preponderance of N2 errors. For a feature-passing model, this means that feature passing (and consequent feature-passing errors) can occur only within a very limited scope; namely, within a phrase or so. While feature passing might thus be responsible for errors in the PP construction from a PP versus RC contrast (Bock & Cutting, 1992; Solomon & Pearlmutter, 2004b), it would not be responsible for errors from the RC construction (or, to the extent that they occur, from full embedded clauses): We do not yet have data on the precise scope that might be relevant here, but if structure has been planned for only a single PP in the current experiments, it seems unlikely that structure for an entire RC will have been planned in corresponding cases from Bock and Cutting's and Solomon and Pearlmutter's experiments. And if the point of a feature-passing mechanism is to deliver subject number to the verb (e.g., Bock & Levelt, 1994; Eberhard et al., 2005; Franck et al., 2002; Vigliocco et al., 1996), some additional explanation for that part of the process will be needed, given that the structure connected with the predicate will not have been created when the subject's number is computed. These are not insurmountable problems, but they indicate that feature-passing accounts cannot fully explain apparent effects of structure in the literature, and that they are not complete even as accounts of the core phenomena or operations for which they were developed.

As we noted above, the Experiment 1 and 2 results also point to an alternative source of errors, which is the encoding of subject number into memory: If a plural feature happens to be active in the planning system around the time that a singular subject is being encoded, the plural can interfere, eventually increasing the probability of producing a plural instead of a singular verb (see also Nicol, 1995). Along with the current experiments, much of the existing data in the agreement production literature (e.g., effects attributed to hierarchical feature-passing, semantic integration effects, and most or all of the effects on verb number captured by marking and morphing and their interplay in the Eberhard et al. (2005) model) is compatible with the idea that errors occur when interference arises in setting or tracking the number of the subject prior to whatever agreement target (e.g., a verb) is eventually

planned. Indeed, the Eberhard et al. model itself can be seen as primarily a model of the memory encoding process for subject number.

However, if memory processes are responsible for agreement errors, then not only encoding interference but also retrieval interference must be considered, either as an alternative or in addition: At the time of planning an agreement target (e.g., a verb), the encoded source for that target must be retrieved from memory, and this process may be susceptible to interference from other elements (e.g., intervening local nouns) that might be incorrectly retrieved instead of the correct source (Badecker & Kuminiak, 2007; Badecker & Lewis, 2007; Lewis & Badecker, 2010). Detailed models of memory retrieval processes in subject-verb agreement production are only beginning to be discussed, but they have so far primarily focused on two factors which have been proposed to play important roles in retrieval models of long-distance dependency processing in comprehension: recency (or decay) of activation and similarity-based interference (Badecker & Lewis; Gordon, Hendrick, Johnson, & Lee, 2006; Lewis & Badecker; Lewis, Vasishth, & Van Dyke, 2006).[4] However, the effect of linear order in Experiment 2 (and in the flat conditions of Experiment 1) seems to be a problem for both of these factors: Recency of activation predicts that N3 should be more likely to be incorrectly retrieved than N2, because N3 has been more recently produced and thus more recently activated; but this is the reverse of the pattern in Experiment 2. An alternative to recency of activation is frequency of activation (an element retrieved multiple times during processing would be more likely to interfere than one retrieved less often; cf. Lewis & Badecker); but in producing the flat conditions, N2 and N3 should be retrieved equally often, predicting no difference in interference. This also incorrectly predicts the same interaction in Experiment 1 as hierarchical feature-passing, because in the descending cases, unlike the flat cases, N2 should be retrieved more frequently than N3 (N2 must be retrieved as a modifier of N1 and then again when it is modified by the PP containing N3); yet the descending and flat cases yielded the same difference between N2 and N3 mismatch effects. Similarity-based interference, similarly, does not differentiate N2 and N3, as neither noun (or perhaps local NP) is tagged as a grammatical subject (e.g., Bock & Levelt, 1994), is in a subject-head position (e.g., Lewis & Badecker), or is marked with nominative case (Badecker & Kuminiak), and both nouns are inanimate and at the same structural depth (but see Badecker & Kuminiak (p. 82) for a suggestion about how a retrieval model might use "planning chunks" as cues).

While memory retrieval thus appears to be insufficient on its own to explain the full range of error patterns, it obviously could nevertheless play a role in combination with memory encoding: Badecker and Kuminiak (2007) argue for a retrieval-interference model of agreement based on an interaction between case-marking and gender agreement in Slovak; and Haskell and MacDonald (2005) show that for disjunctive subjects (e.g., *the horse or the clocks*), in which the two nouns might be matched in subject properties, the verb overwhelmingly agrees with the noun that is proximal to the verb. Furthermore, effects of phrase length (Bock & Cutting, 1992) are more naturally explained by retrieval interference

[4]A separate question for retrieval-based models is whether the retrieval process attempts to retrieve number information independently or instead attempts to retrieve the head of the subject phrase, which in turn yields number information; presumably the system might attempt to retrieve both, as this might provide an alternative approach to combining number derived (in Eberhard et al.'s (2005) terms) by marking and by morphing. These variations make different predictions about the extent to which subject-verb agreement errors are correlated with head mis-selection errors (cases in which the predicate is about a noun other than the head of the subject; e.g., *The baby on the blanket had tangled plaid fringe*.). We attempted to evaluate this correlation by looking for head mis-selection errors, using the continuations of all of the critical items from 12 participants in Experiment 1 (288 sentences), but the majority of predicates did not unambiguously specify which noun was selected. Of the 42 cases that were unambiguous, only 4 were head mis-selection errors, and none of them (nor any of the correct head-selection cases) was also an agreement error case. These data are obviously very limited, but they certainly do not suggest that head mis-selection during retrieval was the source of the error patterns in the current experiments.

than by encoding interference, as local nouns in longer PPs cause more interference than local nouns in shorter PPs.

The same contrast between retrieval and encoding effects arises in agreement comprehension as well, where initial work focused on what were essentially encoding-based models (e.g., Nicol, Forster, & Veres, 1997; Pearlmutter, 2000, Pearlmutter, Garnsey, & Bock, 1999), while several more recent investigations of agreement have suggested that retrieval-based errors may also occur (e.g., Häussler, 2006; Wagers, Lau, & Phillips, 2009). The general comprehension model presented by Lewis, Vasishth, and Van Dyke (2006) incorporates both memory encoding and retrieval processes, and it may be that agreement production models will need to incorporate both encoding and retrieval interference as potential sources of agreement errors.

But whether or not memory retrieval is needed in addition to encoding to explain agreement errors, this still leaves open the question of the mechanism of agreement computation, if there is no feature passing over structure. In fact, however, a memory-based system for encoding and retrieving subject number may provide most of the needed machinery on its own, with the rest provided by properties and processes independently needed: In addition to encoding relevant agreement information in working memory when the subject is planned and retrieving it when the predicate is planned, an agreement system must determine the relevant information to be encoded, which means identifying the correct subject phrase, identifying the head of that phrase, and combining the relevant message-level and lexical number information associated with these elements. The combination process is the focus of Eberhard et al.'s (2005) model, and we discussed above how it could operate without any need for hierarchical feature-passing; the mechanism for this part of the process might just be a weighted combination (e.g., implemented in terms of activation). The other two aspects depend on the syntax and semantics to link message-level elements to the syntactic subject phrase (usually an NP) and to the lexical head for that phrase (and to link the subject phrase to its head), but they are not specific to agreement processes and will be needed in any production system that generates grammatically well-formed and discourse- and semantically-appropriate sentences (e.g., Bock & Levelt, 1994; Bock et al., 1992), regardless of agreement. Using them specifically for agreement only requires stipulating that part of the information linked to the subject phrase from the message, and part of the information activated in the lexicon for the head, is information needed for agreement (at least conceptual and lexical number, respectively, for English). Similarly at the time of retrieval, the retrieved number information must be associated with the appropriate predicate phrase and then applied appropriately to the relevant head; but identification of the predicate and its head are also processes needed independently. This is obviously barely a sketch of a mechanism for agreement computation, and future research will have to examine it in more detail; but it suggests how agreement might operate without any need for hierarchical feature-passing, and it takes advantage of the cascading, activation-based properties of the Bock and Levelt and Eberhard et al. models.

One final potential limitation of these studies is that the fragment-completion task used to elicit errors necessarily involves a comprehension component, which is presumably not usually a part of the production process. However, to the extent there is an issue here, it holds for nearly all agreement error elicitation studies to date: Almost every study has used a version of one of two basic methods, either presenting preambles first (auditorily or visually) and having speakers remember then recite and complete them (e.g., Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Eberhard, 1999; Fayol et al., 1994; Franck et al., 2002; Haskell & MacDonald, 2003), or else presenting them (visually) and having speakers read them aloud and complete them (e.g., the current experiments, Bock & Eberhard, 1993; Solomon & Pearlmutter, 2004b; Vigliocco et al., 1996; Vigliocco

& Nicol, 1998). In either variant, speakers must comprehend the presented preamble in order to use it during production.

On the other hand, the comprehension aspects of these tasks seem unlikely to have much of an influence on the results: First, both versions of the task have revealed clear influences of message-level representations (e.g., distributivity effects in Eberhard, 1999, and Hartsuiker, Kolk, & Huinck, 1999; and semantic integration effects in both Experiment 2 and Solomon & Pearlmutter, 2004b), indicating that production is being at least partly driven by its normal source (the message). Second, although both tasks require comprehension, its potential influence during production is likely to be greater in the read-aloud version (in which comprehension overlaps with production) than in the comprehend-then-repeat version; yet the result patterns seen with the two different tasks generally show few differences (e.g., Bock & Cutting, 1992, vs. Solomon & Pearlmutter's Exp. 5; Bock & Eberhard, 1993; Gillespie & Pearlmutter, 2010).

Nevertheless, we cannot entirely rule out possible influences of comprehension, and because the scope of planning account in particular relies on timing of planning of elements of the subject noun phrase to explain agreement errors, future work will have to examine whether these tasks alter timing of availability and thus the extent to which different factors are relevant. One possibility would be to conduct more detailed comparisons of the task variants that involve comprehension; an alternative is to develop a paradigm requiring speakers to formulate their utterances from the message level without any comprehension involved (see Haskell & MacDonald, 2005, for one possibility).

Finally, the current results show that agreement studies can inform us about the scope and units of planning in language production. Much of the research on scope of planning has focused on properties that affect phonological encoding by measuring effects of semantically- and phonologically-related distractors and syntactic complexity on speech onset times (Allum & Wheel-don, 2007; Martin, Crowther, Knight, Tamborello II, & Yang, in press; Martin, Miller, & Vu, 2004; Smith & Wheeldon, 1999, 2001; Wheeldon & Lahiri, 2002). While these findings show that phonological encoding is affected by properties specified at other levels, there is so far little direct evidence of scope of planning effects at the grammatical encoding level (cf. Gómez Gallo & Jaeger, 2009, for some recent suggestive evidence; and Watson, Breen, & Gibson, 2006, for evidence of grammatical effects on prosodic structure). The results of the current studies provide evidence that planning proceeds at least somewhat incrementally, such that elements (phrasal heads, at a minimum) are generally planned in the order in which they are to be produced, multiple elements may be overlappingly activated based on conceptual-level factors, and the advanced planning of elements can influence grammatical encoding processes such as agreement computation. A major advantage of this account of agreement error production is that it is in accord with explanations of other types of speech errors. For example, exchange and other ordering errors are thought to occur when the interacting elements are simultaneously active, and the wrong element is selected for production (Garrett, 1975, 1980; Pearlmutter & Solomon, 2007; see, e.g., Dell, 1986; Dell, Burger, & Svec, 1997, for corresponding explanations of phoneme ordering errors). Thus, the proposed scope of planning account, which relies on the degree to which elements are overlappingly planned, links the findings of agreement error production studies to the rich tradition of research examining lexical and phonological errors in spontaneous and experimentally elicited speech, suggesting that different kinds of speech errors can be linked to the same source.

References

Allum PH, Wheeldon LR. Planning scope in spoken sentence production: The role of grammatical units. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33:791–810.

Badecker W, Kuminiak F. Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. Journal of Memory and Language. 2007; 56:65–85.

Badecker, W.; Lewis, R. A new theory and computational model of working memory in sentence production: Agreement errors as failures of cue-based retrieval. Paper presented at the 20th Annual CUNY Conference on Human Sentence Processing; La Jolla, CA. 2007 Mar.

Berent I, Pinker S, Tzelgov J, Bibi U, Goldfarb L. Computation of semantic number from morphological information. Journal of Memory and Language. 2005; 53:342–358.

Bock K, Cutting JC. Regulating mental energy: Performance units in language production. Journal of Memory and Language. 1992; 31:99–127.

Bock K, Eberhard KM. Meaning, sound and syntax in English number agreement. Language and Cognitive Processes. 1993; 8:57–99.

Bock K, Eberhard KM, Cutting JC, Meyer AS, Schriefers H. Some attractions of verb agreement. Cognitive Psychology. 2001; 43:83–128. [PubMed: 11527432]

Bock, K.; Levelt, W. Language production: Grammatical encoding. In: Gernsbacher, M., editor. Handbook of psycholinguistics. San Diego, CA: Academic Press; 1994. p. 945-984.

Bock K, Loebell H, Morey R. From conceptual roles to structural relations: Bridging the syntactic cleft. Psychological Review. 1992; 99:150–171. [PubMed: 1546115]

Bock K, Miller CA. Broken agreement. Cognitive Psychology. 1991; 23:45–93. [PubMed: 2001615]

Bock K, Nicol J, Cutting JC. The ties that bind: Creating number agreement in speech. Journal of Memory and Language. 1999; 40:330–346.

Brown-Schmidt S, Konopka AE. Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. Cognition. 2008; 109:274–280. [PubMed: 18842259]

Brown-Schmidt S, Tanenhaus MK. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. Journal of Memory and Language. 2006; 54:592–609.

Carnie, A. Syntax: A generative introduction. Malden, MA: Blackwell; 2005.

Chanquoy L, Negro I. Subject-verb agreement errors in written productions: A study of French children and adults. Journal of Psycholinguistic Research. 1996; 25:553–570.

Chomsky, N. Aspects of the theory of syntax. Cambridge, MA: MIT Press; 1965.

Chomsky, N. The minimalist program. Cambridge, MA: MIT Press; 1995.

Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior. 1973; 12:335–359.

Cohen, J.; Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. 2. Hillsdale, NJ: Lawrence Erlbaum; 1983.

Cuetos F, Mitchell DC. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. Cognition. 1988; 30:73–105. [PubMed: 3180704]

Culicover PW, Jackendoff R. The simpler syntax hypothesis. Trends in Cognitive Sciences. 2006; 10:413–418. [PubMed: 16899400]

Dell GS. A spreading activation theory of retrieval in language production. Psychological Review. 1986; 93:283–321. [PubMed: 3749399]

Dell GS, Burger LK, Svec WR. Language production and serial order: A functional analysis and a model. Psychological Review. 1997; 104:123–147. [PubMed: 9009882]

Eberhard KM. The marked effect of number on subject-verb agreement. Journal of Memory and Language. 1997; 36:147–164.

Eberhard KM. The accessibility of conceptual number to the processes of subject-verb agreement in English. Journal of Memory and Language. 1999; 41:560–578.

Eberhard KM, Cutting JC, Bock K. Making syntax of sense: Number agreement in sentence production. Psychological Review. 2005; 112:531–559. [PubMed: 16060750]

Fayol M, Largy P, Lemaire P. When cognitive overload enhances subject-verb agreement errors: A study in French written language. Quarterly Journal of Experimental Psychology. 1994; 47A:437–464.

Franck J, Lassi G, Frauenfelder UH, Rizzi L. Agreement and movement: A syntactic analysis of attraction. Cognition. 2006; 101:173–216. [PubMed: 16360139]

Franck J, Vigliocco G, Nicol J. Subject-verb agreement errors in French and English: The role of syntactic hierarchy. Language and Cognitive Processes. 2002; 17:371–404.

Frazier, L.; Clifton, C, Jr. Construal. Cambridge, MA: MIT Press; 1996.

Garrett, MF. The analysis of sentence production. In: Bower, G., editor. Psychology of learning and motivation. Vol. 9. New York: Academic Press; 1975. p. 133-177.

Garrett, MF. Levels of processing in sentence production. In: Butterworth, B., editor. Language production. Vol. 1. London: Academic Press; 1980. p. 177-220.

Gibson E, Pearlmutter NJ, Canseco-Gonzalez E, Hickok G. Recency preference in the human sentence processing mechanism. Cognition. 1996; 59:23–59. [PubMed: 8857470]

Gillespie, M.; Pearlmutter, NJ. Simultaneity of planning increases interference during subject-verb agreement production. Poster presented at the 23rd Annual CUNY Conference on Human Sentence Processing; New York, NY. 2010 Mar.

Goldberg, AE. Constructions at work: The nature of generalization in language. Oxford: Oxford Univ. Press; 2006.

Gómez Gallo, C.; Jaeger, TF. Early verb choice and fluency as evidence for moderately incremental or possibly limited parallel sentence production. Paper presented at the 22nd Annual CUNY Conference on Human Sentence Processing; Davis, CA. 2009 Mar.

Gordon PC, Hendrick R, Johnson M, Lee Y. Similarity-based interference during language comprehension: Evidence from eye tracking during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32:1304–1321.

Griffin ZM. Gaze durations during speech reflect word selection and phonological encoding. Cognition. 2001; 82:B1–B14. [PubMed: 11672707]

Hartsuiker RJ, Antón-Méndez I, van Zee M. Object attraction in subject-verb agreement construction. Journal of Memory and Language. 2001; 45:546–572.

Hartsuiker RJ, Kolk HHJ, Huinck WJ. Agrammatic production of subject-verb agreement: The effect of conceptual number. Brain and Language. 1999; 69:119–160. [PubMed: 10447988]

Haskell TR, MacDonald MC. Conflicting cues and competition in subject-verb agreement. Journal of Memory and Language. 2003; 48:760–778.

Haskell TR, MacDonald MC. Constituent structure and linear order in language production: Evidence from subject-verb agreement. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:891–904.

Häussler, J. Disrupted agreement checking in sentence comprehension. Proceedings of the Eleventh ESSLLI Student Session; 2006. p. 39-50.

Hemforth, B.; Konieczny, L. Proximity in agreement errors. Proceedings of the 25th Annual Conference of the Cognitive Science Society; 2003. p. 557-562.

Hupet M, Fayol M, Schelstraete M. Effects of semantic variables on the subject-verb agreement processes in writing. British Journal of Psychology. 1998; 89:59–75.

Kayne, RS. Connectedness and binary branching. Dordrecht: Foris; 1984.

Lewis, RL.; Badecker, W. Sentence production and the declarative and procedural components of short term memory. Paper presented at the 23rd Annual CUNY Conference on Human Sentence Processing; New York, NY. 2010 Mar.

Lewis RL, Vasishth S, Van Dyke JA. Computational principles of working memory in sentence comprehension. Trends in Cognitive Sciences. 2006; 10:447–454. [PubMed: 16949330]

Martin RC, Crowther JE, Knight M, Tamborello FP II, Yang C. Planning in sentence production: Evidence for the phrase as a default planning scope. Cognition. in press.

Martin RC, Miller M, Vu H. Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. Cognitive Neuropsychology. 2004; 21:625–644. [PubMed: 21038225]

Negro I, Chanquoy L, Fayol M, Louis-Sidney M. Subject-verb agreement in children and adults: Serial or hierarchical processing? Journal of Psycholinguistic Research. 2005; 34:233–258. [PubMed: 16050444]

Nicol JL. Effects of clausal structure on subject-verb agreement errors. Journal of Psycholinguistic Research. 1995; 24:507–516. [PubMed: 8531170]

Nicol JL, Forster KI, Veres C. Subject-verb agreement processes in comprehension. Journal of Memory and Language. 1997; 36:569–587.

Pearlmutter NJ. Linear versus hierarchical agreement feature processing in comprehension. Journal of Psycholinguistic Research. 2000; 29:89–98. [PubMed: 10723713]

Pearlmutter NJ, Garnsey SM, Bock K. Agreement processes in sentence comprehension. Journal of Memory and Language. 1999; 41:427–456.

Pearlmutter NJ, Gibson E. Recency in verb phrase attachment. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2001; 27:574–590.

Pearlmutter, NJ.; Solomon, ES. Semantic integration and competition versus incrementality in planning complex noun phrases. Paper presented at the 20th Annual CUNY Conference on Human Sentence Processing; San Diego, CA. 2007 Mar.

Pollard, C.; Sag, IA. Head-driven phrase structure grammar. Chicago: Univ. of Chicago Press; 1994.

Schneider W. Micro Experimental Laboratory: An integrated system for IBM PC compatibles. Behavior Research Methods, Instruments, & Computers. 1988; 20:206–217.

Schütze CT, Gibson E. Argumenthood and English prepositional phrase attachment. Journal of Memory and Language. 1999; 40:409–431.

Smith M, Wheeldon L. High level processing scope in spoken sentence production. Cognition. 1999; 73:205–246. [PubMed: 10585515]

Smith M, Wheeldon L. Syntactic priming in spoken sentence production: An online study. Cognition. 2001; 78:123–164. [PubMed: 11074248]

Solomon, ES.; Pearlmutter, NJ. Semantic integration and hierarchical feature-passing in sentence production. Poster presented at the 17th Annual CUNY Conference on Human Sentence Processing; College Park, MD. 2004a Mar.

Solomon ES, Pearlmutter NJ. Semantic integration and syntactic planning in language production. Cognitive Psychology. 2004b; 49:1–46. [PubMed: 15193971]

Stevenson, S. Tech. Rep. No. 18. New Brunswick, NJ: Rutgers Univ. Center for Cognitive Science; 1994. A competitive attachment model for resolving syntactic ambiguities in natural language processing.

Vigliocco G, Butterworth B, Garrett MF. Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. Cognition. 1996; 61:261–298. [PubMed: 8990974]

Vigliocco G, Hartsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. Psychological Bulletin. 2002; 128:442–472. [PubMed: 12002697]

Vigliocco, G.; Nicol, J. Unpublished manuscript. University of Arizona; Tucson: 1994. The role of syntactic tree structure in the construction of subject verb agreement.

Vigliocco G, Nicol J. Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? Cognition. 1998; 68:B13–B29. [PubMed: 9775519]

Wagers MW, Lau EF, Phillips C. Agreement attraction in comprehension: Representations and processes. Journal of Memory and Language. 2009; 61:206–237.

Watson D, Breen M, Gibson E. The role of syntactic obligatoriness in the production of intonational boundaries. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32:1045–1056.

Wheeldon LR, Lahiri A. The minimal unit of phonological encoding: Prosodic or lexical word. Cognition. 2002; 85:B31–B41. [PubMed: 12127702]

Appendix A: Experiment 1 Stimuli

The purely singular versions of the Experiment 1 stimuli are shown below. Items 1–12 are the descending stimuli; items 13–24 are the flat stimuli. The other versions were created by making either the the second noun or the third noun plural (but not both).

1. The bookcase with the ornate carving on the wooden shelf
2. The car with the silly sticker on the chrome bumper
3. The uniform with the silver star on the felt badge
4. The coat with the nylon tag on the folded cuff
5. The castle with the flaming torch on the stone moat
6. The magazine with the accurate illustration in the lengthy article
7. The purse with the pink button on the side pocket
8. The suit with the jagged rip in the wide lapel
9. The bracelet with the glass bead on the tiny clasp
10. The backpack with the plastic buckle on the leather strap
11. The watch with the sparkling jewel on the hour hand
12. The apartment with the full closet in the narrow hallway
13. The catalog for the department store with the ripped binding
14. The fax about the bankrupt company with the torn cover sheet
15. The safe for the pricy necklace with the combination lock
16. The ball for the rowdy game with the thick stripe
17. The keyboard for the modern computer with the chipped key
18. The cucumber for the fresh salad with the brownish spot
19. The tomato for the tasty sandwich with the nasty bruise
20. The postcard of the roller coaster with the foreign stamp
21. The diagram of the giant skyscraper with the tricky graph
22. The hose for the gorgeous garden with the replaceable nozzle
23. The tunnel through the steep mountain to the gold mine
24. The highway to the western suburb with the steel guardrail

Appendix B: Experiment 2 Stimuli

The purely singular early-integrated versions of the Experiment 2 stimuli are shown below. The other versions were created by varying the number of the noun in each PP and by varying the order of the two PPs.

1. The book with the torn page by the red pen
2. The shirt with the snug sleeve under the cardboard box
3. The ring with the fake diamond near the blueberry muffin
4. The apple with the brown bruise beside the wicker basket

5. The lamp with the halogen bulb beside the antique vase
6. The drill with the titanium bit under the wool sweater
7. The receipt with the blurry price near the dirty towel
8. The tree with the dead branch by the old building
9. The pizza with the yummy topping beside the broken toaster
10. The blanket with the soft seam behind the filing cabinet
11. The bowl with the noticeable crack under the flannel sheet
12. The bike with the bent spoke behind the rickety shed
13. The chair with the wobbly leg near the pruned shrub
14. The laptop with the loud speaker near the framed mirror
15. The staircase with the iron railing by the narrow hallway
16. The fork with the crooked prong by the fresh peach
17. The rose with the prickly thorn near the gold bracelet
18. The candle with the long wick beside the oak bookcase
19. The church with the tall steeple by the grassy park
20. The plane with the icy wing behind the massive truck
21. The skirt with the tattered hem behind the closet door
22. The printer with the ink cartridge by the ticking clock
23. The drawer with the wooden inlay under the tangled wire
24. The newspaper with the controversial headline beside the plastic bag
25. The purse with the full pocket near the remote control
26. The sign with the wooden post near the deep puddle
27. The glove with the tight finger under the gaudy necklace
28. The coat with the ripped cuff by the orange ball
29. The radio with the cracked knob beside the leather belt
30. The store with the vintage register near the crowded sidewalk
31. The train with the piercing whistle beside the peaceful lake
32. The shoe with the knotted lace behind the mossy log
33. The sink with the leaky faucet under the ceiling fan
34. The cake with the gooey filling near the electric blender
35. The boat with the nylon sail behind the granite monument
36. The oven with the hot burner beside the metal shelf
37. The plant with the yellowing leaf by the spiral notebook
38. The jacket with the faulty zipper near the sturdy desk
39. The razor with the rusty blade beside the purple brush

40. The chain with the tarnished link behind the oil tank

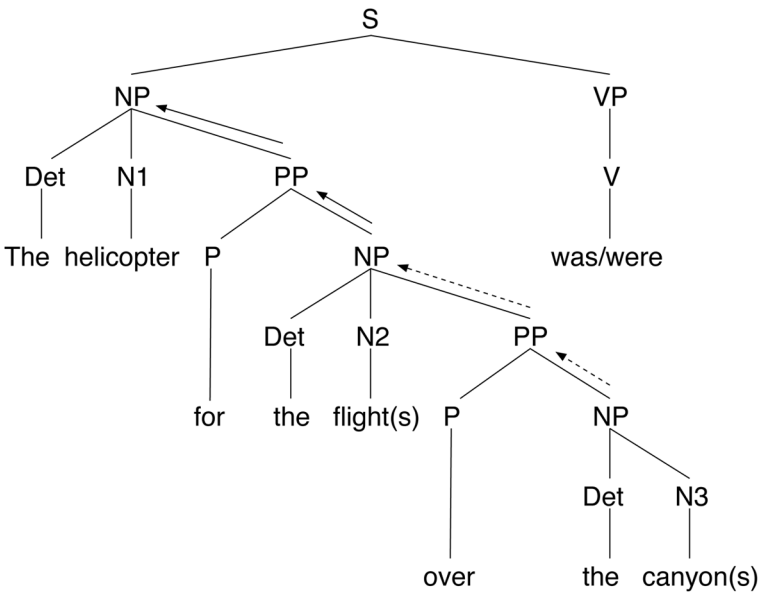

Figure 1.
Syntactic path a plural feature must travel to interfere with agreement in Franck et al.'s (2002) stimuli. The route for a feature from N2 is shown with solid arrows; the route for a feature from N3 includes the route from N2 as well as the dashed arrows, so additional feature-passing errors would have to occur before N3's plural feature could influence verb number, predicting fewer subject-verb agreement errors when N3 is plural compared to when N2 is.

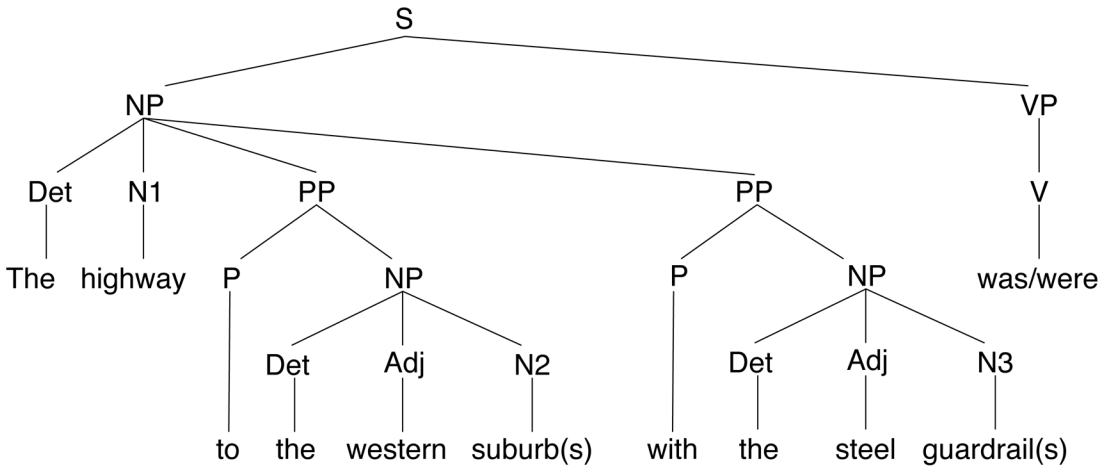

Figure 2.
Syntactic structure for flat stimuli, in which both PPs attach to the first NP.

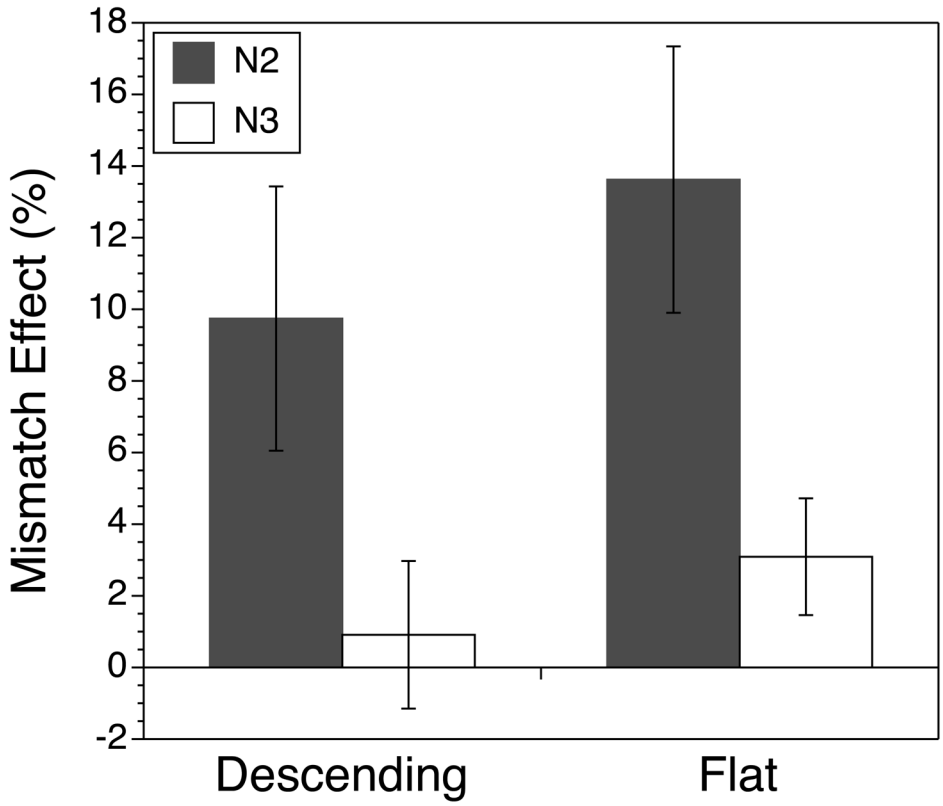

Figure 3.
Experiment 1 grand mean mismatch error rates by structure and mismatch position. Error bars show ±1 *SEM*, computed from the analyses by participants.

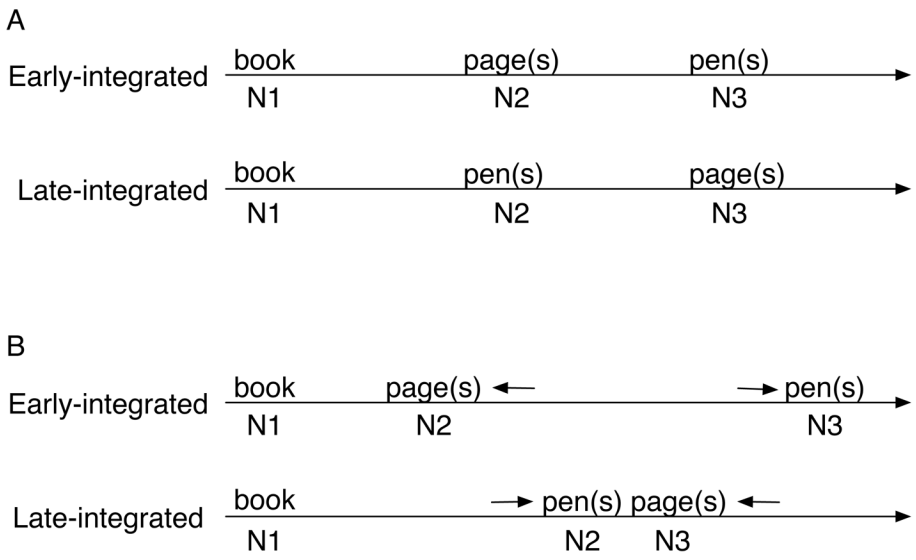

Figure 4.
Timelines depicting the predicted timing of planning of nouns in the Experiment 2 early-integrated and late-integrated stimuli. Panel A shows the nouns planned according to the order in which they are to be produced, corresponding to the predictions of the linear distance to the head account. Panel B shows the timing of planning of the nouns after semantic integration shifts their relative timing, corresponding to the scope of planning account.

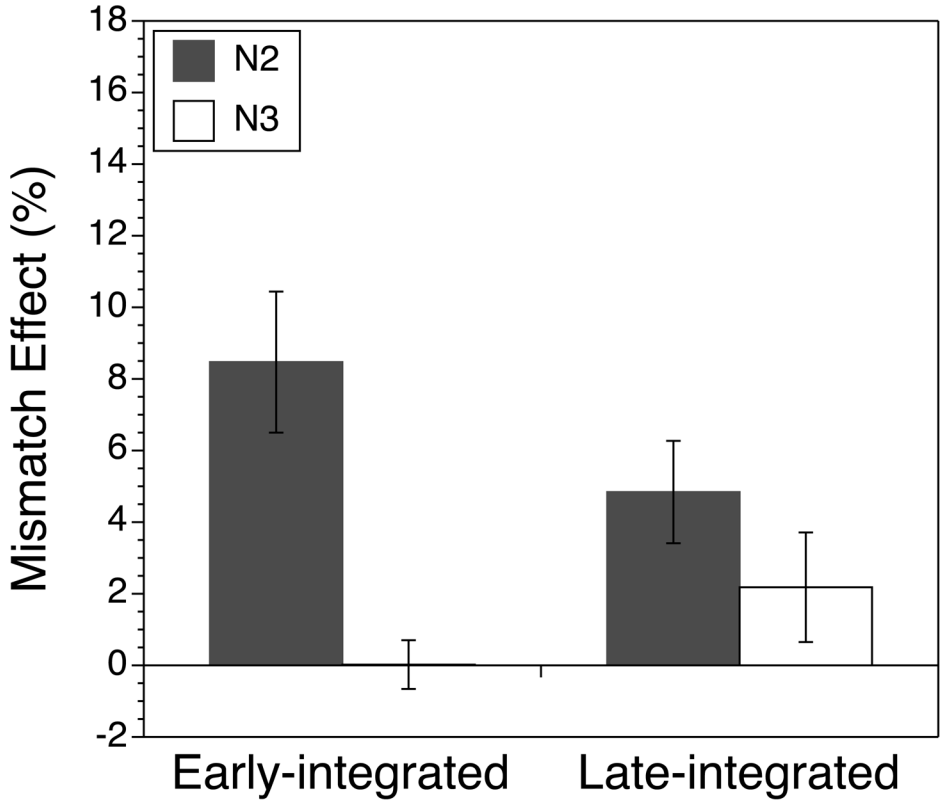

Figure 5.
Experiment 2 grand mean mismatch error rates by integration and mismatch position. Error bars show ±1 *SEM*, computed from the analyses by participants.

Table 1

Experiment 1 Mean Semantic Integration Ratings, Mean Attachment Preferences, and Response Counts by Condition

	Semantic Integration Rating				Response Count			
Condition	**N1-N2**	**N1-N3**	**N2-N3**	**%N1 Attachment**	**Error**	**Correct**	**Uninflected**	**Misc**
Descending								
SSS	4.13 (.94)	5.23 (.49)	4.46 (.83)	4.7 (7.01)	1 (0)	145 (19)	37 (5)	33
SPS	4.34 (.79)	5.08 (.47)	4.22 (.72)	5.6 (5.57)	15 (2)	129 (13)	29 (5)	43
SSP	4.02 (.77)	5.21 (.45)	4.20 (.53)	4.1 (5.52)	2 (0)	124 (12)	31 (5)	59
M	4.16 (.83)	5.17 (.46)	4.29 (.70)	4.8 (5.93)				
Flat								
SSS	4.21 (1.38)	5.24 (.98)	2.81 (.94)	93.6 (5.68)	1 (0)	148 (20)	38 (4)	29
SPS	4.11 (1.36)	4.96 (.99)	2.62 (.91)	91.9 (6.68)	19 (2)	114 (10)	31 (5)	52
SSP	3.97 (1.23)	5.13 (.97)	2.64 (.75)	94.4 (5.57)	5 (2)	128 (10)	33 (2)	49
M	4.10 (1.29)	5.11 (.96)	2.69 (.85)	93.3 (5.92)				
Grand Mean	4.13 (1.08)	5.14 (.75)	3.49 (1.12)	49.0 (45.0)				

Note. Conditions are indicated by noun number (N1, N2, N3), with S = singular and P = plural. Semantic integration rating scale was 1 (loosely linked) to 7 (tightly linked). For semantic integration ratings and attachment, means are computed by-items and standard deviations are in parentheses. For response counts, dysfluency counts are in parentheses. Misc = Miscellaneous.

Table 2

Experiment 2 Mean Semantic Integration Ratings, Mean Attachment Preferences, and Response Counts by Condition

	Semantic Integration Rating				Response Count			
Condition	**N1-N2**	**N1-N3**	**N2-N3**	**%N1 Attachment**	**Error**	**Correct**	**Uninflected**	**Misc**
Early-integrated								
SSS	5.74 (.74)	2.24 (.62)	2.08 (.59)	96.4(8.21)	2(1)	322 (45)	109(15)	91
SPS	5.70 (.56)	2.16 (.62)	2.00 (.45)	98.9 (6.05)	27(8)	270 (33)	95(11)	132
SSP	5.63 (.63)	2.21 (.60)	2.04 (.48)	97.6 (5.49)	2(0)	312 (39)	91 (10)	119
SPP	5.75 (.60)	2.12 (.54)	2.05 (.57)	98.0 (5.06)	36(5)	253 (29)	108 (16)	128
M	5.70 (.63)	2.18 (.59)	2.04 (.52)	97.7 (5.63)				
Late-integrated								
SSS	2.25 (.65)	5.63 (.66)	1.98 (.54)	97.7 (3.79)	1(1)	318 (54)	105 (20)	100
SPS	2.03 (.47)	5.71 (.64)	2.00 (.62)	98.9 (2.74)	15(5)	276 (42)	96(15)	138
SSP	2.18 (.58)	5.57 (.75)	2.18 (.62)	97.7 (3.98)	8(0)	313 (35)	102 (17)	102
SPP	2.19 (.56)	5.81 (.80)	2.17 (.53)	99.2 (3.29)	32(3)	290 (42)	90 (16)	113
M	2.16 (.57)	5.68 (.71)	2.08 (.58)	98.4 (3.52)				
Grand Mean	3.93 (1.87)	3.93 (1.87)	2.06 (.55)	98.0(5.12)				

Note. Conditions are indicated by noun number (N1, N2, N3), with S = singular and P = plural. Semantic integration rating scale was 1 (loosely linked) to 7 (tightly linked). For semantic integration ratings and attachment, means are computed by-items and standard deviations are in parentheses. For response counts, dysfluency counts are in parentheses. Misc = Miscellaneous.