

Conflicting cues and competition in subject–verb agreement

Todd R. Haskell^{a,*} and Maryellen C. MacDonald^b

^a *University of Southern California, USA*

^b *University of Wisconsin-Madison, USA*

Received 29 March 2002; revision received 18 December 2002

Abstract

Traditional theories of agreement production assume that verb agreement is an essentially syntactic process. However, recent work shows that agreement is subject to a variety of influences both syntactic and non-syntactic, which raises the question of how these different sources of information are integrated during agreement production. We propose an account of agreement production in which several information sources contribute activation to singular and plural verb forms. Conflict between cues leads to competition which can in turn magnify the influence of subtle cues. Three fragment completion experiments tested key predictions of this constraint satisfaction approach. Experiment 1 demonstrated competition effects on verb choice and sentence initiation latencies. Experiments 2 and 3 demonstrated that conflicts between semantic and grammatical cues allow morphological regularity to exert a small but detectable effect on agreement. These results suggest that the constraint-satisfaction framework may provide a productive approach for understanding agreement production.

© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Agreement; Number; Language production

In English and many other languages, verbs must agree in number with their subjects, as in the contrast between *the dog barks* and *the dogs bark*. In recent years a considerable number of studies have explored the mechanisms by which speakers produce such agreement. This interest in agreement has arisen for several reasons. First, agreement is pervasive: agreement of some sort occurs in about three-quarters of the world's languages (Mallinson & Blake, 1981), and within those languages, it typically occurs in most if not all sentences. For example, English speakers must produce subject–verb agreement for virtually every sentence they speak, though agreement is not always overtly marked. Thus, understanding the mechanisms by which agreement is

produced is an important component of understanding speech production more generally. Second, there is a well-established paradigm for studying agreement production, based on the elicitation of agreement errors under controlled circumstances (Bock & Miller, 1991). Third, and most importantly, how agreement is produced is relevant to several contentious issues within the study of speech production, for example, the degree to which putatively grammatical processes such as agreement are insulated from conceptual or other non-syntactic influences (Bock & Eberhard, 1993; Bock & Miller, 1991; Bock, Nicol, & Cutting, 1999). In the current paper, we propose a constraint satisfaction approach to the production of subject–verb agreement. This approach borrows from work on language comprehension to explain how both syntactic and non-syntactic factors interact during the computation of agreement.

Traditionally, agreement has been treated as an essentially syntactic phenomenon (Corbett, 1994). Such an

* Corresponding author. Present address: Beckman Institute, 405 N. Mathews, University of Illinois, Urbana, IL 61801, USA.

E-mail address: thaskell@uiuc.edu (T.R. Haskell).

approach is consistent with models of language production in which different components function largely autonomously. Perhaps the most widely known model of this sort is that presented in Levelt (1989) and Bock and Levelt (1994). This model involves three major stages or levels. The first stage involves message formulation. The second stage involves the syntactic part of speech planning (including the computation of agreement), and is commonly called grammatical encoding. This stage is followed by phonological encoding, when the actual sounds to be produced are spelled out.

In this model, information flow between the levels is unidirectional, from the message toward phonological encoding (Bock & Levelt, 1994). Furthermore, it is assumed that the system is to a considerable degree modular, in the sense that there are specialized processing mechanisms to carry out particular processes, and these processes are unaffected by information elsewhere in the system (Bock, 1995; Bock & Levelt, 1994). To the extent that agreement is primarily a grammatical process, this model predicts that, aside from initially determining the number of the subject (see, e.g., Bock et al., 1999; Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001), agreement should not be affected by semantic factors. Furthermore, there should be little or no influence of phonological factors on agreement, as phonological encoding is postulated to occur after grammatical encoding, and thus in a strictly feed-forward model there should be no influence of phonology on agreement.

Several studies of agreement production in English have provided support for this model. In one early study, (Bock & Miller, 1991) explored the influence on agreement of a non-head noun in the subject noun phrase. Their procedure involved presenting participants with the initial portion of a sentence, such as *the key to the cabinets*. Participants were asked to repeat this preamble and then complete the sentence in any fashion they desired. A key manipulation involved the number of the noun in the post-modifying prepositional phrase, which either matched or mismatched the number of the head noun, exemplified by *the key to the cabinet* and *the key to the cabinets* respectively. On those occasions when the participant produced a number marked verb such as *was* or *were*, the rate of agreement errors was markedly higher in the mismatch condition, indicating that the number of the second noun (conventionally referred to as the *local* noun) can influence the agreement process. It should also be noted that Bock and Miller found this mismatch effect to be much larger with a singular head noun and plural local noun than vice versa; for that reason many later studies utilized only singular head nouns, and the discussion will be limited to this case here.

In a series of experiments using this procedure, Bock and Miller (1991) and Bock and Eberhard (1993)

examined the role of semantic and phonological factors on the production of subject–verb agreement in English. These experiments generated no evidence that either factor affects agreement processes, supporting a modular, feed-forward account.

More recent work, however, has challenged this view in several domains. First, there is a growing body of evidence that semantic factors may play a more central role in agreement computation. For example, a series of studies by Vigliocco and colleagues (Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996) examined *distributive* noun phrases such as *the label on the bottles*, where the semantics of the phrase implies the existence of multiple labels. Experiments conducted in Italian, Spanish, Dutch and French demonstrated that more plural verbs were produced following distributive phrases than following non-distributive phrases. Eberhard (1999) demonstrated that effects of distributivity can also be found in English under certain circumstances, and she attributed the difference between her results and the failure to find effects of distributivity in Bock and Miller (1991) to differences in the imageability of the preamble phrases. Similarly, Bock et al. (1999) found effects of another semantic factor, collectivity, on subject–verb agreement. In their study, collective head nouns such as *committee*, followed by a plural noun in the post-modifying phrase, elicited 60% plural verbs. For ordinary singular head nouns (again with a plural noun in the post-modifying phrase), this rate was 10%. Finally, Vigliocco and Franck (1999, 2001), working on gender agreement in Italian and French, found differing agreement behavior depending on whether conceptual information was consistent with linguistic gender, conflicted with linguistic gender, or was neutral.

Second, there is evidence for morphophonological influences on agreement processing. Vigliocco et al. (1995), working in Italian, compared rates of agreement errors for invariant plurals and ordinary plurals. Invariant plurals occur when the singular and plural forms of a noun are the same, like English *fish* and *sheep*. For these nouns, morphophonology is not a useful cue to number. The nouns were presented along with a number-marked determiner, so that the number of the noun was never ambiguous (e.g., *la città* ‘the town’ versus *le città* ‘the towns’). Nevertheless, more errors occurred with the invariant nouns than with ordinary nouns. Similarly, Vigliocco and Zilli (1999) found that gender agreement errors in Italian were more common when a noun carried ambiguous gender marking than when it carried clear gender marking.

Finally, there is also evidence for plausibility effects on agreement production. Using the error-elicitation paradigm, Thornton and MacDonald (in press) manipulated the extent to which the head and local nouns were

plausibly related to the agreeing verb. Participants were presented with a noun phrase (e.g., *the album by the classical composers*) and a past participle verb (e.g., *played or praised*), and were asked to create a complete passive sentence (e.g., *The album by the classical composers was played.*). Whereas the head noun was always a plausible object of the verb (e.g., albums can be played or praised), the local noun varied in its relationship to the verb (e.g., composers can plausibly be praised but not played). Participants produced more agreement errors when the past participle could plausibly apply to the local noun (e.g., composers can be praised) than when it could not (composers cannot be played) (but see Barker, Nicol, & Garrett, 2000).

Thus, it is becoming clear that the production of agreement is sensitive to several factors, both syntactic and non-syntactic. What is unclear is how best to account for these effects. One possibility (Bock et al., 1999; Bock et al., 2001) is to continue to view agreement production as a primarily syntactic process. In this approach, other factors may exert an influence on agreement, but the role of these factors is sharply constrained by the architecture of the production system. An alternative, articulated by Vigliocco and colleagues (Vigliocco & Franck, 1999; Vigliocco & Franck, 2001; Vigliocco & Hartsuiker, 2002; Vigliocco & Zilli, 1999), is that the production system utilizes all available information, whether syntactic or non-syntactic, to aid in the production of agreement. Unfortunately, this latter approach is at present still somewhat vague: it is not clear what factors should influence agreement, by what mechanisms they exert their effects, and how they interact.

We believe that useful insights into these issues can be obtained by considering constraint-satisfaction theories developed in the context of work on language comprehension (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994; see also the Competition Model of Bates & MacWhinney, 1982). In the constraint satisfaction framework, the result of processing is determined by the interaction of multiple graded, probabilistic constraints. A strength of this approach is that it provides a natural framework for explaining how highly disparate factors interact during processing (Spivey & Tanenhaus, 1998). This is just the sort of situation that occurs during the computation of subject–verb agreement, which appears to be subject to a variety of different influences. This property of the constraint satisfaction approach gives it a strong intuitive appeal.

However, it has been unclear to what extent this approach can be successfully applied to grammatical encoding, although accounts incorporating certain constraint satisfaction principles exist in studies of phonological encoding (Dell, 1988). Some approaches within linguistics assume that different types of information

must interact during the computation of agreement (e.g., Pollard & Sag, 1988), but the processing mechanisms that would underlie this interaction have remained unspecified.

In the current paper, we focus on two predictions that arise out of the constraint satisfaction framework: the existence of competition between alternative verb forms, and interactions between different information sources. With respect to competition, the constraint satisfaction account claims that a number of sources of information contribute toward promoting or inhibiting particular alternatives. For example, the presence of a singular subject noun ‘horse’ would promote a singular verb form such as ‘runs,’ and inhibit a plural verb form such as ‘run.’ In most situations, the different sources of information will be correlated. ‘Horse’ is not only grammatically singular, but also singular in meaning and phonological form (in that it does not carry the plural suffix *-s*). However, in some situations, different sources of information may promote different verb forms, resulting in competition (Berg, 1998). A key claim of the constraint satisfaction approach is that competition between two different verb forms should increase processing time and/or variability of the number marking on the verb. This prediction is tested in Experiment 1.

The constraint satisfaction approach also predicts that different information sources should interact. That is, the contribution of a given source of information may depend on other sources of information (see, e.g., McClelland, 1987). In particular, in cases where one or more factors strongly promote a certain alternative (say, the singular form), then a weaker factor which promotes the plural form will have little effect. If, however, other factors are evenly divided in which form they promote, then even a weak factor can play a decisive role. This prediction is explored in Experiments 2 and 3.

Experiment 1

To test the prediction of competition effects, collective nouns such as *fleet* were contrasted with ordinary singular nouns such as *ship*. Agreement behavior of collectives is variable, in that they are sometimes used with singular and sometimes with plural verbs. However, a variety of other syntactic tests suggest that collectives should be treated as grammatically singular. For example, they can be used with the article *a*, the quantifiers *each* and *every*, the demonstrative *this*, and they have distinct plural forms (e.g., *team* vs. *teams*, *army* vs. *armies*). Furthermore, collective nouns are distinctly unacceptable in contexts where a plural is required (**many family*, **these/those family*, **three family*), regardless of whether agreement is singular or plural. Thus, both ordinary singular nouns and collective

singular nouns are probably best analyzed as grammatically singular, and in the constraint satisfaction framework, grammatical number should promote use of a singular verb in both cases.

However, collective nouns share some conceptual characteristics with plural nouns, in that they refer to a collection of several individuals (e.g., a team consists of several players). Thus, in the constraint satisfaction framework, conceptual factors will potentially contribute some activation to the plural verb form. Since both verb forms are at least partially activated, competition will arise between them. This conflict should slow the production of a number-marked verb which agrees with a collective subject, and the conflict should sometimes be resolved in favor of a plural verb form, as is commonly observed in British English, and as Bock et al. (1999) also found for American English. In contrast, ordinary singular nouns are conceptually singular. Thus, conceptual factors will converge with grammatical factors in promoting a singular verb form, and there will be no competition.

There is already some support for the first half of the competition prediction, that is, that competition should lead to variability in choice of verb form. Bock et al. (1999) reported that when speakers were presented with sentence preambles with collective subject nouns, a situation that should engender competition, they continued those preambles with a plural verb about half the time (averaging across singular and plural local nouns). In other words, both singular and plural verbs were produced with considerable frequency. With ordinary singular nouns, and thus no competition, speakers continued with plural verbs only about 6% of the time; in this condition, singular verbs were clearly dominant. Experiment 1a aimed to replicate this finding, while also testing the prediction of longer processing times by means of a modified procedure that allows the collection of speech initiation latencies.

Experiment 1a also further investigated the mismatch effect reported by Bock & Miller (1991). Bock & Miller (1991) found that phrases like *the key to the cabinets*, where there are two nouns that mismatch in number, elicit a measurable number of (erroneous) plural verbs, while phrases like *the key to the cabinet*, where the number on the two nouns is the same, elicit very few errors. In the constraint satisfaction framework, the presence of a plural local noun can lead to partial activation of the plural verb form. This would result in competition between the singular and plural verb forms. In this view, the elevated error rates for the mismatch phrases reflect the higher variability in verb form selection that results from this competition. If this analysis is correct, then a mismatch between the number of the head and local nouns should also lead to longer processing times. This prediction was also tested in Experiment 1a.

To augment the data provided by the production task, off-line normative ratings were also collected. One set of ratings (Experiment 1b) assessed the conceptual plurality of noun phrases used in the production task. This allowed us to further explore the relationship between conceptual plurality and the effects of collectivity. A second set of ratings (Experiment 1c) involved an off-line graded grammaticality judgment task. The relationship between this measure, conceptual plurality, and performance on the production task provides further insight into the nature of the agreement process.

Several previous studies have used initiation times to quantify the difficulty of the production task (F. Ferreira, 1991; F. Ferreira, 1994; V. Ferreira, 1996; Stallings, MacDonald, & O'Seaghdha, 1998). In particular, Stallings et al. (1998) linked increased initiation times to competition between two possible syntactic forms. To extend these methods to agreement production, several modifications were made to the standard preamble completion task originally developed by Bock & Miller (1991).

First, participants were asked to form questions rather than declarative sentences. The advantage of using questions is that verbs are frequently fronted in English questions, and when this occurs, the sentence initiation latencies should reflect the time required to select a verb form. Vigliocco & Nicol (1998) directly compared agreement in questions and declarative sentences, and found that the pattern of agreement errors was the same. Thus, findings based on production of questions should generalize to production in general.

Second, participants were asked to produce the utterance immediately upon receiving the last piece of material necessary to formulate it. This was done to maximize time pressure, and thereby to increase sensitivity to agreement processing difficulty. Crucially, this last modification required that participants not receive sufficient information to choose a verb until the last moment. Participants were first presented with a subject noun phrase for the question they would utter, e.g., *the plates for the party guests*. This phrase logically provided enough information to determine the number on the verb. However, participants were then presented with an additional word to include in the sentence. This word was varied to promote the use of either a form of 'to do' or a form of 'to be' in the sentence. For example, if the word was *break*, participants would be biased to say *Did the plates for the party guests break?*; if the word was *broken*, the expected response would be *Were the plates for the party guests broken?* Thus, participants could not know what the structure of their response would be until the final word was presented, so that initiation latencies should be sensitive to the difficulty of sentence-initial planning, including verb agreement.

Method—Experiment 1a

Participants

Fifty-six young adults completed the experiment. All participants were native speakers of English. Participants either received class credit or monetary compensation for their participation.

Materials

The stimulus lists contained 106 items. Each item consisted of a noun phrase (e.g., *the towel in the guest bathroom*) and a word which could be plausibly related to the noun phrase (e.g., *wet*). Of the stimulus items, 20 were experimental items, 78 were fillers, and eight items were included as part of a separate manipulation (see Experiment 2).

For the experimental items, all the noun phrases had singular head nouns followed by a prepositional phrase containing an adjective and a noun. The head noun was always preceded by *the*, and the prepositional phrase contained either *the* or, occasionally, a demonstrative or possessive pronoun, if this resulted in a more natural phrase. All the experimental noun phrases contained the same number of words.

Each experimental item had four variants, depending on whether the head noun was a collective (e.g., *gang*) or a semantically related non-collective (e.g., *criminal*), and whether the local noun was singular or plural. For all the experimental noun phrases, an adjective or past participle was chosen as the associated word, e.g., *dirty* or *convicted*. For five of the items, the singular local noun variant was conspicuously less plausible or meaningful than the plural local noun variant; to address this, the prepositional phrase for the singular local noun variant was modified slightly to better balance plausibility in these cases. Table 1 shows one of the items in all four versions; all the experimental items are listed in Appendix A.

The collective and non-collective head nouns did not significantly differ in log frequency [3.28 for collectives, 3.19 for non-collectives; $t(19) < 1$] or number of syllables [1.90 for collectives, 2.05 for non-collectives; $t(19) < 1$].

Of the fillers, 24 contained only a short noun phrase consisting of *the*, an adjective, and a noun. Other fillers included sixteen examples each of singular–singular, plural–singular, and plural–plural combinations of head

and local nouns. This was done to ensure that the particular structure of the experimental items was not overly conspicuous. The remaining fillers had a singular–plural structure. Across the entire list, participants saw an equal number of singular and plural heads.

In contrast to the experimental items, the associated word for the fillers could be an uninflected verb, as well as an adjective or past participle. We varied the type of word for two reasons. First, it ensured that participants would produce a variety of sentence structures in their responses, making the task more natural. Second, it forced participants to process the entire phrase before formulating a response, so that information about the local noun, and not just the head noun, would influence processing.

Four stimulus lists were generated. Each stimulus list contained exactly one variant of each experimental item. Local noun number on the experimental items was held constant within each list, but each list contained an equal number of both collective and non-collective variants. Thus, collectivity was a within-participants manipulation, while local noun number was a between-participants manipulation. The experimental and filler items were presented in a different random order in each list, preceded by eight practice items of varying types. Equal numbers of participants were presented with each list.

Procedure

Testing took place individually. Each participant wore a lapel microphone and sat in front of a computer screen. The entire session was recorded on audio tape for later scoring. Trials were controlled by a Macintosh computer. All text presented on the computer screen appeared in the 12 pt Chicago font.

At the beginning of the session, the participant was informed about the format of the trials, and instructed to respond as quickly and accurately as possible. Once the participant was ready, the experimenter initiated the trials. The format of each trial was as follows. First, a warning signal (****) was presented in the middle of the computer screen for 600 ms. This was followed by the noun phrase, in all lowercase, for 2000 ms. The screen was then cleared for 600 ms, after which the associated word appeared, again in lowercase. The word remained on the screen until the voice key was triggered or 2500 ms passed without the participant responding. The

Table 1
Example stimulus item for Experiment 1

Head type	Local noun type	Phrase	Word
Non-collective	Singular	The actor in the weekend performance	Famous
Collective	Singular	The cast in the weekend performance	Famous
Non-collective	Plural	The actor in the weekend performances	Famous
Collective	Plural	The cast in the weekend performances	Famous

computer recorded the latency to respond. The experimenter then indicated an initial coding of verb number by pressing a key on a button box, and the next trial began. Each session lasted about 20 min.

Scoring

Responses to experimental trials were partitioned into three groups, based on the on-line scoring and later verification via the audiotapes: singular verb, plural verb and miscellaneous responses.

Singular responses consisted of all cases where the response began with a singular verb, and included the noun phrase at least as far as the head noun, e.g., *was the cast*—in other words, enough of the noun phrase to unambiguously determine the agreement relation. In 29% of these cases, the participant later made some sort of error, usually changing or forgetting a preamble word. A total of 69% of responses were scored as singular.

Plural responses (4%) consisted of all cases where the response began with a plural verb, and included the noun phrase at least as far as the head noun, e.g., *were the cast*—again, enough of the noun phrase to unambiguously determine the agreement relation. In 37% of these cases, the participant later made some sort of error.

All remaining trials (28%) were classified as miscellaneous responses; these included failures to use a number-marked verb, disfluencies at the verb or head noun, and a few trials on which the voice-key did not trigger at the appropriate time.

Results—Experiment 1a

Verb number analyses

Since the rate of miscellaneous responses was almost identical in the collective (0.27) and non-collective (0.28) conditions, analyses were conducted on the raw proportion of plural responses.¹ These plural verb proportions, by condition, are illustrated in the left panel of Fig. 1. 2×2 ANOVAs were conducted by participants and by items, with collectivity as a within-groups variable, and local noun number as a between groups (analysis by participants) or within groups (analysis by items) variable.

These analyses revealed a main effect of collectivity, such that collective head nouns elicited more plural-agreeing verbs than non-collective heads [$F_1(1, 54) = 17.3, p < .001$; $F_2(1, 19) = 14.3, p < .01$]. The main effect

of local noun number was also significant, with more plural verbs produced when the local noun was plural (and thus mismatched in number with the subject noun) [$F_1(1, 54) = 5.2, p = .03$; $F_2(1, 19) = 12.7, p < .01$]. However, these main effects were largely due to the elevated plural verb rates in the collective, plural local noun condition, as reflected in a significant interaction [$F_1(1, 54) = 4.9, p = .03$; $F_2(1, 19) = 7.9, p = .01$]. The results of this analysis replicate the pattern of data in Bock et al. (1999), though the overall proportion of plural verbs is somewhat lower in the present study.

Latency analyses

Latency analyses were conducted on singular responses only. Latencies over 2500 ms (when the stimulus was removed from the screen) were excluded from the analyses; 2.3% of singular responses were excluded in this manner. The results from the remaining trials are shown in the right panel of Fig. 1. Conducting latency analyses by participants proved difficult, as several participants had no singular responses in at least one condition. Thus, we report only the results of the items analysis here, and item means are shown in the figure; for consistency, item means are also shown in the figures for Experiments 2 and 3.

A 2×2 ANOVA showed a main effect of collectivity [$F_2(1, 19) = 8.9, p < .01$]: participants took longer to initiate a sentence with a singular verb when they were presented with a collective head noun. There was also a main effect of local noun number [$F_2(1, 19) = 24.3, p < .001$]. Participants took longer to respond when the local noun mismatched in number with the head noun. In contrast to the analyses of number marking above, there was no evidence of an interaction [$F_2(1, 19) < 1$]. There was, however, a reliable positive correlation over items between plural rates and latencies [$r = .37, t(38) = 2.44, p = .02$], such that items with high plural rates also tended to be ones that yielded long latencies to produce singular verbs.

Discussion—Experiment 1a

Experiment 1a tested two predictions of the constraint satisfaction approach. First, English collective nouns, while grammatically singular, have elements of plural semantics. According to the constraint satisfaction approach, these characteristics should cause partial activation of both singular and plural verb forms. This in turn will result in competition between the two verb forms, leading to increased variability in verb form, and longer processing times. Bock et al. (1999) reported data supporting the first part of this prediction (increased variability); the current experiment replicated this result, while demonstrating that collective head nouns also lead to longer processing times. Thus, this prediction of the constraint satisfaction approach was fully supported.

¹ In this and the following experiments, the analysis was repeated using the proportion of non-miscellaneous responses which began with plural verbs. In all cases the pattern of significance was identical to the first analysis, so only the first analysis is reported in each experiment.

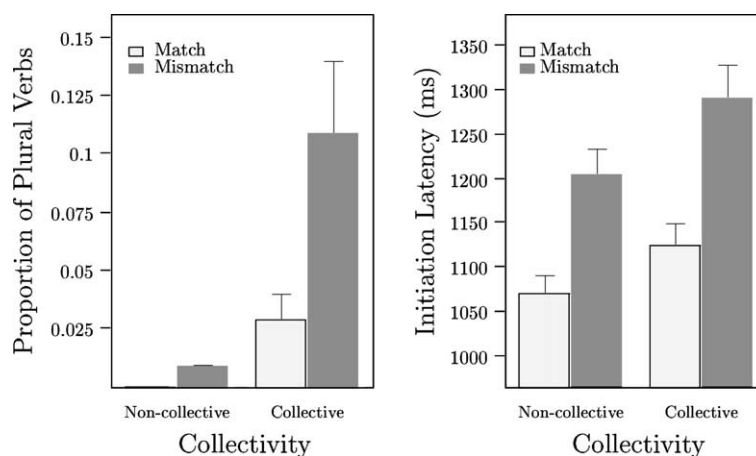


Fig. 1. The proportion of plural marked verbs (left panel) and initiation latencies (right panel) for collective and non-collective subjects with local nouns matching or mismatching in number for Experiment 1. Latencies are for utterances beginning with singular verbs only. Means are across items.

The second prediction of the constraint satisfaction approach was that the mismatch effects observed by Bock & Miller (1991) and others can also be interpreted as competition effects. In the case of a mismatching local noun, the competition arises because the number information on the local noun can partially activate the incorrect verb form, resulting in partial activation of both verb forms. Again, the signs of competition—increased variability in verb form and longer latencies—were clearly evident. Thus, the second prediction of the constraint satisfaction approach was also supported.

Although both the plural verb proportions and the latency data revealed largely the same pattern of data, there was one important inconsistency, involving the collectivity-by-local-noun-number interaction, which was present in the plural verb proportion data but not the latency data. A plausible interpretation of this result is that the interaction in the plural verb proportion data emerged because these data suffer from floor effects. That is, there is an effect of collectivity in both the match and the mismatch condition, but it is obscured in the match condition because plural verb rates are extremely low. This is a vexing problem in research using the fragment completion paradigm, because rates for the target response type (here, plural verbs) are typically near zero in at least one cell of the design. Alternative measures (such as latencies) are thus an important adjunct. Another advantage of the latency measure in particular is that it allows comparisons between normal (error-free) processing across conditions. Thus, measuring latencies should prove useful as alternative accounts of agreement and grammatical encoding are tested.

Although Experiment 1a demonstrated that collectivity of head nouns leads to both increased plural verb

rates and longer processing times, the mechanisms underlying these effects are not entirely clear. The constraint satisfaction approach claims that the behavior of collective head nouns arises because they have conceptual characteristics of both singulars and plurals—i.e., that they lie somewhere in between singulars and plurals in terms of conceptual number. In Experiment 1b, we collected off-line normative ratings on the conceptual plurality of the stimuli used in Experiment 1a. This measure was then compared to the plural verb rate and latency data.

Method-Experiment 1b

Participants

Forty-four participants completed the conceptual plurality task. All participants were native speakers of English, and none participated in Experiment 1a. Participants were volunteers, or received token monetary compensation for their participation.

Materials and procedure

The ratings task consisted of an untimed, paper-and-pencil survey. Stimuli consisted of the noun phrases from Experiment 1a, as well as sixteen items included as part of a separate manipulation (“fillers”). There were four lists. Each contained all of the items from Experiment 1a in either the collective or non-collective version, such that each participant saw 10 of each type. Two of the lists contained the plural local noun version of the items (e.g., *the cast in the weekend performances*), and two contained the singular local noun version (e.g., *the cast in the weekend performance*). Two of the lists (one with the singular local noun versions and one with the plural local noun versions)

contained one set of fillers. The other two lists contained a second set of fillers. The fillers consisted of distributive and non-distributive complex noun phrases. The order of items in each list was randomized separately. Three practice items were placed at the beginning of each list. Equal numbers of participants saw each list.

Adjacent to each item was a 7-point rating scale. Participants were instructed to circle a low number if the phrase clearly referred to only a single entity, a high number if the phrase clearly referred to multiple entities, and an intermediate number if “it depends on how you think about it.” To clarify the purpose of the intermediate numbers, participants were presented with two example phrases, one involving a collective noun (*stamp collection*), the other involving a distributive noun phrase (*the stripe on the signs*). The experimenter explained to the participants that these phrases had two possible interpretations, and described the distributive and non-distributive senses of the phrases. Participants were told that if a phrase had multiple interpretations, they should select a number based on the relative dominance or naturalness of the two interpretations. The questionnaire took approximately 5 min to complete.

Results and discussion—Experiment 1b

Phrases with collective head nouns had a significantly higher mean plurality rating ($M = 3.34$, $SE = 0.24$) than phrases with non-collective head nouns ($M = 1.69$, $SE = 0.13$), confirming that our manipulation of collectivity did succeed in affecting the underlying conceptual

number of the phrases [$F_1(1, 42) = 38.1$, $p < .001$; $F_2(1, 19) = 84.3$, $p < .001$]. Importantly, the effect of local noun number on conceptual plurality ratings did not approach significance [$F_1(1, 42) < 1$, n.s.; $F_2(1, 19) < 1$, n.s.]. This result is not surprising because we deliberately chose phrases such that the number semantics of the phrase was independent of the number of the local noun. Thus, for the remainder of the analyses the data from the match and mismatch conditions were pooled.

Comparison with singular and plural verb rates

More important than the conceptual plurality ratings themselves is a direct comparison of the conceptual plurality ratings with the plural verb rates and latency data from Experiment 1a. The plural verb rates are considered in this section, and the latency data in the next. As an initial step, the mean plural verb rates for each item (averaged across subjects) from the production task were plotted against the mean conceptual plurality ratings for the same items (also averaged across subjects). The correlation between the two measures proved significant [$r = .35$, $t(78) = 3.34$, $p < .01$], such that higher conceptual plurality ratings were associated with higher plural verb rates. This correlation is shown in the left panel of Fig. 2.

One concern about this analysis is that for 83% of the items in the match condition and 63% of the items in the mismatch condition the mean plural verb rate was zero. Thus, the correlation is being carried by a small subset of the items. One approach to this issue is to use a measure that also reflects agreement processes, but does not suffer from such severe floor effects. The natural

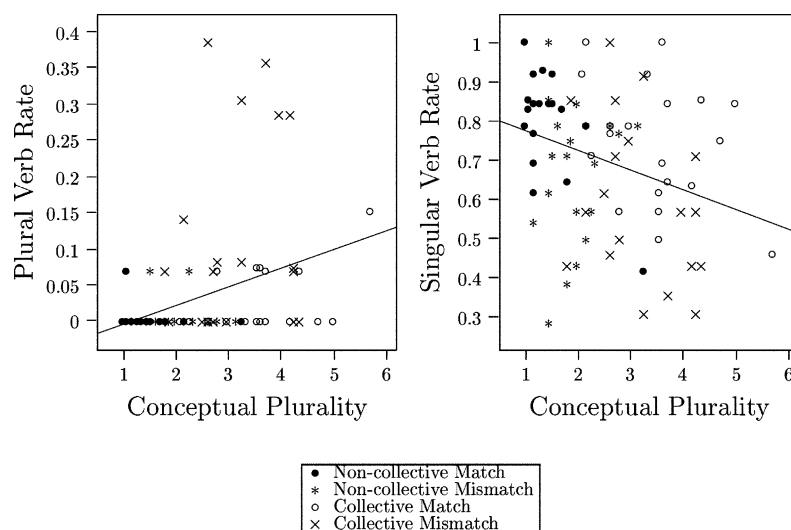


Fig. 2. Correlation of plural verb rate (left panel) and singular verb rate (right panel) from Experiment 1a with conceptual plurality ratings from Experiment 1b.

candidate for such a measure is singular verb rate, i.e., the proportion of trials on which speakers began their question with a singular verb. Not surprisingly, this measure has a strong inverse relationship with plural verb rate [$r = -.47$, $t(78) = 4.67$, $p < .01$] (the correlation is not perfect because, in addition to plural and singular responses, participants also produced responses in the miscellaneous category). Thus, the analysis was reconducted using singular verb rate rather than plural verb rate. The correlation between singular verb rate and the conceptual plurality ratings was in fact significant [$r = -.31$, $t(78) = 2.91$, $p < .01$], with higher conceptual plurality ratings associated with lower singular verb rates. These data are presented in the right panel of Fig. 2.

Comparison with latencies

A second set of analyses was conducted which compared conceptual plurality ratings for the items in Experiment 1b with the latencies for the same items from Experiment 1a. These two measures exhibited a significant correlation [$r = .25$, $t(78) = 2.28$, $p = .03$], such that higher conceptual plurality ratings were associated with longer latencies. This relationship is shown in Fig. 3.

To summarize, according to the constraint satisfaction approach, the more conceptually plural a phrase is, the more it should tend to activate a plural verb form. This increased activation should lead to increased competition between the singular and plural verb forms, and the variability in verb selection and longer processing times that this entails. Experiment 1b explored this issue by using conceptual plurality ratings to predict

the plural verb rates and latencies observed in Experiment 1a. The significant correlations between the ratings and the on-line measures lend support to the idea that conceptual plurality underlies performance in the production task.

Appealing to the notion of conceptual plurality also helps provide a unified account of collectivity effects and distributivity effects. Distributive noun phrases such as *the label on the bottles* have two senses: a concrete sense in which *label* refers to the physical labels attached to each bottle, and a sense in which it refers to the label as an abstract entity divorced from any particular instantiation. Thus, like collectives, distributive noun phrases have elements of both singular and plural meaning. In English, the abstract (singular) sense appears to be strongly dominant, making conceptual effects with distributive noun phrases much more elusive than with collectives (Bock & Miller, 1991; Vigliocco et al., 1996; Eberhard, 1999). However, Eberhard (1999) was able to find such effects with more imageable noun phrases. In the constraint-satisfaction approach, more imageable phrases would emphasize the concrete over the abstract sense, thus making the phrases more conceptually plural.

An additional aspect of the constraint-satisfaction approach is that it does not make a clear distinction between “correct” and “erroneous” agreement. The likelihood of producing a singular verb versus a plural verb can vary along a continuous dimension based on the degree of support for one alternative or the other. The frequency of plural verbs increases as one moves from ordinary singular nouns, to distributive noun phrases, to collectives, to plural nouns; however, in each case the choice of verb form is governed by the same mechanisms. Probabilistic selection of verb form is a fundamental characteristic of the constraint satisfaction approach.

If this view is correct, then the acceptability of the plural form relative to the singular should depend on the degree of support for each alternative, in the same way as choice of verb form does. To test this prediction, Experiment 1c asked English speakers to make untimed, deliberate judgments about the acceptability of plural verb forms with the stimuli from Experiment 1a. If plural verbs occurring with collective subjects are errors or artifacts of a timed production task, then we should find that they are considered unacceptable in the judgment task.

Method—Experiment 1c

Participants

Twenty native speakers of English completed the ratings. None of the participants had participated in Experiment 1a or b. Participants either received class credit or monetary compensation for their participation.

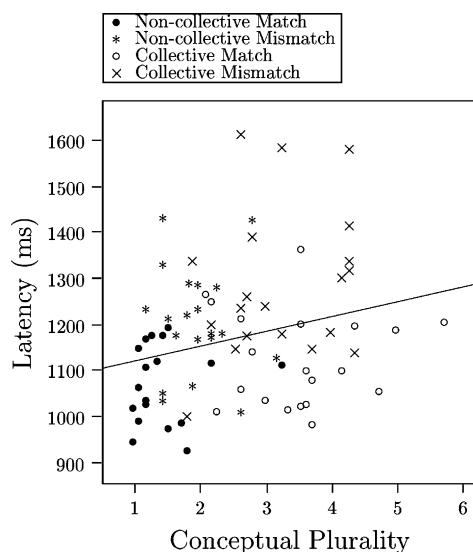


Fig. 3. Correlation of latencies from Experiment 1a with conceptual plurality ratings from Experiment 1b.

Materials and procedure

The ratings task consisted of an untimed, paper-and-pencil survey. Participants were presented with the plural local noun variants of the experimental items from Experiment 1a, in the following form:

The cast in the weekend performances was/were famous

Adjacent to each item was a 7-point rating scale. Participants were instructed to consider the relative acceptability of *was* and *were* in the sentence, after which they should circle a low number if only *was* was acceptable, a high number if only *were* was acceptable, or an intermediate number if either verb form could be used. Participants were told that there were no right or wrong answers, and that they should rely on their intuitions.

The survey was divided into two separate forms; each form contained all the experimental items, with half appearing in the collective version and half in the non-collective version, such that all items appeared in each version exactly once across both forms. Only the plural-local-noun (mismatch) version of each item was used. The order of items in each form was randomized separately. Three additional items were included at the beginning of each form as practice items. Equal numbers of participants completed each form. Completing the survey took approximately 5 min.

Results and discussion—Experiment 1c

While the singular verb form was clearly preferred for both collective and non-collective head nouns, plural verbs were nonetheless significantly more acceptable in the context of singular collective heads ($M = 2.28$, $SE = 0.19$) than singular non-collective heads ($M = 1.20$, $SE = 0.07$), [$F_1(1, 19) = 10.8$, $p < .01$; $F_2(1, 19) = 32.0$, $p < .001$].

This pattern did not arise from some participants uniformly using plural verbs with collectives, while others used singulars. Only three participants actually preferred plural verbs to singular verbs (on average) for the collective items.

The acceptability ratings also correlated significantly with the conceptual number ratings for the mismatch items from Experiment 1b [$r = .65$, $t(38) = 5.21$, $p < .01$], as well as the plural verb proportions [$r = .4$, $t(38) = 2.68$, $p = .01$], singular verb proportions [$r = .34$, $t(38) = 2.22$, $p = .03$] and latency data [$r = .43$, $t(38) = 2.97$, $p < .01$] from Experiment 1a.

The fact that speakers rated plural verbs as being somewhat acceptable with collective subjects, even in a contemplative, off-line task, is consistent with the constraint satisfaction account of collectivity effects. In this view, speakers did not strongly prefer either the singular or the plural verb forms in the ratings task because both verb forms were partially activated.

To summarize, the findings of Experiment 1 support an important prediction of the constraint satisfaction approach: that competition between different sources of information leads to variability in verb selection and longer processing time. Off-line normative measures provided additional insight into the relationship between conceptual plurality and verb number, and between the on-line computation of agreement and speakers' intuitions about English grammar.

We now turn to a second important prediction of the constraint satisfaction approach, the interaction of different information sources. In particular, it is claimed that subtle factors which do not ordinarily visibly influence agreement processing may nevertheless have a significant effect when other, stronger factors conflict. In the present case, we consider whether morphological regularity influences agreement processing in English.

Although the constraint satisfaction approach does not require that morphological regularity (or any other factor) affect agreement processing, to the extent that regularity is correlated with other relevant factors (such as grammatical number), and thus provides a useful source of information, it might affect agreement. Initially, we supposed that morphological regularity might affect agreement processing because of the phonological correlates of regularity. In English, the presence of an *-s* suffix on the subject noun (as in the regular plural *rats*) is highly correlated with use of a plural verb. In contrast, the presence of idiosyncratic (irregular) plural marking (as in the *-ice* ending of *mice*) is not strongly predictive of a plural verb (consider *rice*, *advice*, *price*, etc.). Thus, if morphophonological information is being utilized during agreement processing, we would expect that regular plural nouns should promote a plural verb more strongly than irregular plural nouns. As will be seen, however, there may be other avenues by which regularity could affect agreement as well.

Behavioral evidence on effects of morphological regularity has been mixed. Bock & Eberhard (1993) failed to find evidence for regularity effects on agreement in English. However, the nature of their data precludes forming strong conclusions. In their Experiments 1 and 2, pseudo-plural forms such as *cruise* and *hose* were compared to singulars and true plurals. Plural verbs were produced only with true plural local nouns. However, one cannot conclude from this that plural verb proportions in the singular and pseudo-plural conditions were equivalent, as they were both at floor. In their Experiment 3, irregular plurals (*mice*) were compared with semantically matched regular plurals (*rats*). Although the plural verb proportions were non-zero in all the relevant cells, the numbers were still very small. Furthermore, Vigliocco et al. (1995) and Vigliocco & Zilli (1999), discussed earlier, did find effects of morphophonological factors on agreement in Italian.

Given that a distributivity effect was eventually found in English (Eberhard, 1999), despite initial null results (Bock & Miller, 1991; Vigliocco et al., 1996), it is possible that regularity effects may also appear in English number agreement under some circumstances. The constraint satisfaction framework makes a specific prediction about what these circumstances should be, i.e., cases where the factors promoting singular and plural verbs are fairly evenly balanced. If effects of morphological regularity emerged under precisely these conditions, the constraint satisfaction account would be supported.

As a first step in testing this prediction, we replicated Bock & Eberhard's (1993) Experiment 3 using the revised procedure introduced in Experiment 1 and a new set of stimuli. The collection of latencies (as was done in Experiment 1) provides a potentially more sensitive measure than plural verb proportions alone. Experiment 3 repeated the morphological manipulation under conditions that, according to the constraint satisfaction account, should magnify the influence of subtle factors.

Experiment 2

Method

Participants and Procedure

This experiment was run concurrently with Experiment 1; the items for Experiment 2 were randomly intermingled with those for Experiment 1.

Materials

The eight experimental items followed the same format as the experimental items in Experiment 1, with the exception that the head noun was singular (non-collective) and the same for all versions of an item, and the local noun was either a regular or irregular plural. Thus, all of the experimental items contained a mismatch between the number of the head and local nouns. The result is that there were only two variants of each item, rather than four, e.g., *the cage for the spotted mice* versus *the cage for the spotted rats*. The experimental items are listed in Appendix B.

The small number of items was necessitated by the constraints on irregular plurals in English. Although English has a number of irregular plurals, many of these were undesirable as stimuli for various reasons. For example, though native speakers generally know that *dice* and *lice* are plurals, they do not necessarily associate them with their singular versions *die* and *louse*. Similarly, there is a class of items exemplified by *wives* and *shelves* which involves an unpredictable stem change in the plural; these items are technically irregular, in that they must be learned by rote, but they do carry regular plural marking, and were therefore excluded from the current experiment. Other words such as *data* and

alumni and used as both singular and plural nouns across speakers (and sometimes by the same speaker). Elimination of these and other undesirable irregular items reduced the list to eight usable irregular plurals. The regular and irregular local nouns did not significantly differ in log frequency [3.63 for regulars, 4.43 for irregulars; $t(14) < 1$] or number of syllables [1.25 for regulars, 1.38 for irregulars; $t(14) < 1$].

Stimulus lists. The four stimulus lists were the same as in Experiment 1. Since there were only two variants for each item in Experiment 2 (i.e., the local noun was always plural), the same variant appeared on two of the lists. Each list contained each item exactly once, and an equal number of regular and irregular plural variants appeared in each list.

Scoring

Scoring was the same as for Experiment 1.

Results

Verb number analyses

The raw proportion of plural responses was used in the analyses, as in Experiment 1.

The proportion of plural verbs for each condition is shown in the left panel of Fig. 4. There were numerically more plural verbs with regular as opposed to irregular plural local nouns, but this effect did not approach significance. Overall, the plural verb rate was very low (0.01), comparable with the plural verb rate in the mismatch non-collective condition of Experiment 1 (also 0.01).

Miscellaneous responses were slightly more common in the regular (0.29) than the irregular (0.23) condition, but the difference was non-significant [$F_1(1, 55) = 2.8$, $p = .10$; $F_2(1, 7) = 1.2$, n.s.].

Latency analyses

Latencies were computed over singular responses only and are shown in the right panel of Fig. 4. Responses with latencies over 2500 ms (3.4%) were excluded from the analyses. Several subjects did not have data in at least one cell; therefore, only an items analysis was conducted. As can be seen, there was essentially no difference in latencies as a function of local noun type [$F_2(1, 7) < 1$].

Discussion

In the current experiment, although the plural verb proportion for regulars was roughly double that for irregulars, both were very low, and the difference was not significant. In the latency data, there was essentially no difference between sentences with regular and irregular plurals as local nouns. Thus, the use of the latency

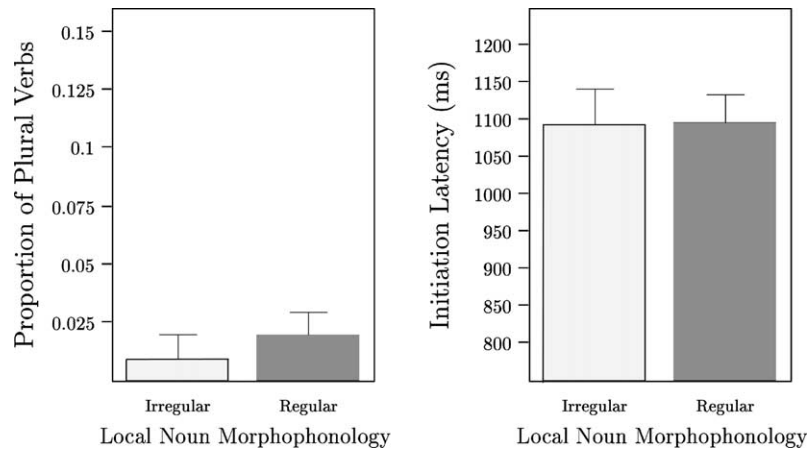


Fig. 4. Proportion of plural verbs (left panel) and initiation latencies (right panel) for regular and irregular plural local nouns in Experiment 2. Item means are shown.

measure in addition to the plural verb rate measure was not sufficient to reveal effects of morphophonology on agreement in English. In replicating Bock & Eberhard's (1993) Experiment 3, which also failed to find a reliable effect of morphophonology, our data are compatible with the claim that, under the conditions of these experiments, morphological regularity does not noticeably influence agreement production in English.

However, a crucial prediction of the constraint satisfaction approach is that effects of subtle factors may only emerge when other, stronger factors are in conflict. The indicators of such conflict (variability in verb selection, longer latencies) were distinctly absent in Experiment 2. Experiment 3 explored whether effects of morphological regularity would emerge under conditions which, according to the constraint satisfaction approach, should maximize its effect. This was done by replicating Experiment 2 using materials designed to make the factors promoting singular and plural verb forms more evenly balanced. Specifically, the collective mismatch condition of Experiment 1 resulted in a plural verb rate of 0.11 (about ten times the plural verb rate in the current experiment), suggesting that the combination of a collective head noun and plural local noun at least partially balances the grammatical number of the head noun in its ability to influence agreement. Thus, a new set of stimuli was developed using collective head nouns, while again manipulating the regularity of the plural local noun. The expectation was that by increasing the overall amount of competition, morphological regularity might play a larger role in the outcome.

Experiment 3

In Experiment 3, the stimuli were modified by using collective head nouns rather than ordinary singulars, as

was done in Experiment 2. This modification was intended to increase the competition between different factors in determining number agreement, perhaps allowing morphophonology to play a more important role.

Method—Experiment 3a

Participants

Thirty-four college students were paid for their participation. All were native speakers of English.

Materials

The general design of the materials was as in Experiment 2, but with collective head nouns for the experimental items, e.g., *the family of mice* versus *the family of rats*. In addition, the items did not contain adjectives between the preposition and the local noun. This change was made to reduce the memory burden on the participants, in an attempt to reduce the high rate of miscellaneous responses seen in the previous two experiments. The eight experimental items are listed in Appendix C.

The fillers were adapted from those used in Experiment 2. However, a much smaller number of fillers was used (42), and generally these were chosen from the easier fillers used in the previous experiments. Again, this was done in an attempt to reduce the high rate of miscellaneous responses seen in the previous two experiments. The previous experiments used fillers that were generally designed to encourage choice of verbs such as *do*, which are not marked for number in the past tense; these fillers may have influenced the overall pattern of responding, resulting in a lower proportion of usable experimental responses. A higher proportion of usable responses might provide a greater opportunity for subtle effects of morphophonology to be evident.

Procedure

The procedure was the same as for Experiment 2.

Scoring

Scoring was the same as in Experiment 2, with a few exceptions. Several subjects produced plural verbs when the verb was not sentence-initial. It was decided to include these as plural verb responses in the plural verb proportion analysis (responses of this sort had not occurred in Experiments 1 and 2). In addition, trials where the voice key was not triggered at the appropriate time (4.4%) were coded separately, and excluded from all analyses.

Results—Experiment 3a

Verb number analyses

The proportion of plural verbs for each condition is shown in Fig. 5. Analyses were conducted in the same manner as for Experiments 1 and 2. One-way ANOVA's revealed a main effect of regularity [$F_1(1, 33) = 9.9$, $p < .01$; $F_2(1, 7) = 8.3$, $p = .02$]; plural verbs were produced far more commonly when the local noun was a regular plural (0.63) than when it was an irregular plural (0.45). Miscellaneous responses were equally frequent in the regular (0.14) and irregular (0.15) conditions.

Latency analyses

Due to the very high plural verb proportions, only sixteen participants and six items had at least one singular response in both conditions. It would be inappropriate to draw firm conclusions based on the data for only those participants. However, for those sixteen participants, latencies for responses beginning with a singular verb were longer in the regular local noun

condition (1292, $SE = 82$) than in the irregular local noun condition (1183, $SE = 96$), though this difference was only marginally significant [$t_1(1, 15) = 1.49$, $p = .08$, one-tailed] (One tailed analyses are justified in this study given the clear direction of predictions for latencies and error rates.) Similarly, for the six items with data in both conditions, latencies in the regular local noun condition (1360, $SE = 89$) were longer than latencies in the irregular local noun condition (1223, $SE = 90$), though of course, with only six items, the difference was not significant [$t_1(1, 5) < 1$]. It is unclear whether these results would attain significance with more power; however, they are at least consistent with our hypotheses.

Discussion—Experiment 3a

The use of collective head nouns had the desired effect—plural verb proportions were much higher in the current experiment than in Experiment 2. With higher overall plural verb proportions, the data demonstrated a clear effect of morphological regularity. Importantly, this effect emerged under precisely the conditions where the constraint satisfaction approach claims it should be most evident.

One interpretation of this finding is that the presence of typical plural marking on a noun enhances its ability to interfere with agreement processing, relative to an irregular plural carrying idiosyncratic marking. However, there is an alternative explanation as well. Because the fragment completion paradigm involves both a comprehension and a production component, it is possible that the regularity effects we observed actually originated during the comprehension portion of the task, or in the conceptual representation that mediates between the two processes. For example, it might be the case that the phrases with regular plural local nouns are represented as being somewhat more plural in meaning than the phrases with irregular plural local nouns. This could occur because of phonological effects during comprehension, or because irregular plurals are inherently less plural in meaning than regular plurals. To address this possibility, in Experiment 3b we collected conceptual plurality ratings for the items used in Experiment 3a, analogous to those collected in Experiment 1b. If morphological regularity affects the comprehension process, or if regular and irregular plurals have different conceptual representations, this should be revealed in the ratings.

Method—Experiment 3b

Participants

Thirty-six students from the University of Wisconsin-Madison completed the conceptual plurality task in exchange for credit in an undergraduate psychology class. All participants were native speakers of English.

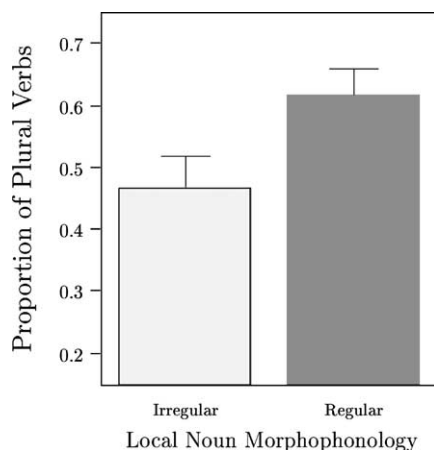


Fig. 5. Proportion of plural verbs in Experiment 3. Item means are shown.

Materials

The stimuli for the ratings task were the two lists of phrases used in Experiment 3a.

Procedure

The procedure was the same as in Experiment 1b.

Results and discussion—Experiment 3b

Phrases with regular local nouns had a significantly higher mean plurality rating ($M = 3.97$, $SE = 0.24$) than phrases with irregular local nouns ($M = 3.16$, $SE = 0.29$). This finding was somewhat surprising, in that it suggests that morphological regularity might actually influence the conceptual representations of the stimulus phrases. Given the abundant evidence for effects of conceptual plurality on agreement, this raises the possibility that the effects of morphological regularity that we observed were actually due in whole or in part to semantic differences across the regular and irregular items. It is therefore important to ask whether morphophonology affected agreement directly, above and beyond any effect mediated by conceptual representations. To address this question, we conducted an analysis examining the effect of regularity after first partialling out the effect of conceptual plurality in a regression. This analysis indicated that a weak effect of regularity remained after differences in conceptual plurality had been taken into account [for plural verb rates, $t(13) = 1.72$, $p = .055$, one-tailed; for singular verb rates, $t(13) = 1.73$, $p = .053$, one-tailed]. Thus, there is some suggestion that regularity affects agreement directly (perhaps via morphophonological cues, as proposed earlier), but a definitive answer to this question awaits further research.

The fact that a morphological manipulation affected the conceptual plurality ratings suggests that apparently non-semantic factors may nevertheless have strong semantic correlates. This suggests that other apparently non-semantic manipulations may also be confounded with meaning differences. Given the abundant evidence for semantic effects on agreement, it would seem that the collection of conceptual plurality ratings is as important for experiments where a semantic manipulation is not intended as it would be for experiments where semantics is intentionally manipulated.

Regardless of whether the regularity effect ultimately turns out to be a morphophonological effect, a semantic effect, or some combination of the two, the current results provide clear support for a key prediction of the constraint satisfaction approach. When multiple strong factors (grammatical, semantic) promoted a singular verb form, as in Experiment 2, no effect of morphological regularity was observed. When semantic and grammatical factors were put into conflict, however, effects of regularity were observed. This finding suggests that a failure to find effects of a certain factor on

agreement should be interpreted with caution, and that it is important to consider how the interaction of different factors will influence the ability to observe an effect.

It is interesting to note that the plural verb proportions in Experiment 3a were considerably higher than those in the collective mismatch condition of Experiment 1a. We suspect this is because the items in Experiment 3a emphasized the distributive sense (Bock & Eberhard, 1993) of the collectives by specifying what was being collected, e.g., *the family of rats*, whereas the postmodifiers in Experiment 1a (e.g., *the fleet near the islands*) did not significantly affect the meaning of the head. However, this is probably just one of several factors affecting plural verb rate. Bock et al. (1999) observed plural verb rates similar to those in the current experiment in their collective head, plural local noun condition, even though most of their items were similar to those in Experiment 1a, with only a few strongly distributive items like those in our Experiment 3. The claim of the constraint satisfaction approach is that the more variability there is in verb form, no matter how it is achieved, the easier it should be to observe the effects of subtle factors like morphological regularity.

General discussion

The primary goal of this paper was to test two key predictions of a constraint satisfaction approach to agreement production. Experiment 1 examined the prediction that conflicting grammatical and semantic factors will lead to competition between verb forms. Experiment 1a's verb choice and latency data supported this claim of competition by showing that conflicts between semantic and grammatical information lead to increased variability in verb form selection and longer processing times. Experiments 2 and 3 tested a second claim of the constraint satisfaction approach, that multiple factors interact to determine agreement. The influence of morphological regularity was shown to depend on whether other factors provided consistent cues to verb number, or conflicted.

These findings suggest that the constraint satisfaction approach, which has proved useful in studying sentence comprehension, may also be profitably applied to grammatical encoding. This idea is not entirely novel; in fact, Vigliocco & Franck (1999) presented a proposal which "is similar in nature to 'constraint satisfaction accounts' put forward in the literature concerning sentence comprehension." (p. 475). However, significant questions remain unanswered regarding the applicability of the constraint satisfaction approach to agreement and to grammatical encoding more generally. We first address some potential alternative explanations for our findings, and then consider the implications of the

constraint satisfaction approach for agreement production and speech production more generally.

Interaction between semantics and syntax

In the constraint satisfaction approach, the competition effects observed in Experiment 1 arise because the contributions of semantic and syntactic factors combine to partially activate both singular and plural verb forms. However, there is a possible alternative explanation that is much closer in spirit to the production model of Bock & Levelt (1994).

Earlier it was noted that collective nouns have some elements of plural semantics. Reid (1991) argued that deciding whether a word refers to one or multiple entities requires defining what is meant by an entity, a point that highlights the dual nature of collectives. English collective nouns are arguably polysemous: a collective can refer to either the individual entities it contains (what Bock & Eberhard, 1993 call the *distributive* sense), or an undifferentiated whole (Bock and Eberhard's *collective* sense). In the constraint-satisfaction approach, the distributive sense promotes a plural verb while the collective sense promotes a singular verb, with the contribution of each depending on the relative dominance of the distributive or collective sense. The net effect of conceptual number would then reflect the weighted average of the contribution from each sense. Such a view is consistent with findings presented by Humphreys & Bock (1999), who were able to influence the agreement behavior of collectives by introducing contextual manipulations emphasizing one sense or the other.

Given this characterization of collectivity, one could imagine that the competition effects observed in Experiment 1 actually arise at the message level. In this view, it is necessary to determine the number of the subject at the message level, so that this number can then be assigned to the subject noun (Levelt, 1989) or the subject noun phrase (Bock et al., 2001). Competition would occur at the message level to resolve the conflict between the collective (singular) and distributive (plural) senses of the collective. This approach would leave the standard model largely intact, while offering a claim (that is, competition) for how some production processes work within the message level. While this view is not necessarily inconsistent with the general constraint satisfaction approach, the specific claim made in this paper is that the competition is not limited to the message level, but reflects interaction between conceptual and grammatical processes. It is therefore important to assess the adequacy of this more modest "intra-message competition" proposal.

As a general claim, the idea that concepts are identified as singular or plural at the message level is somewhat problematic. This is because the conceptual distinctions that are relevant in a given instance depend

on the language and lexical item involved. In English, the singular–plural contrast is relevant for some nouns (e.g., *noodle*) but not others (e.g., *pasta*). Some languages code collectivity (or distributivity) in their inflectional morphology (Barlow, 1992). In the Native American language Papago, different inflections are applied to the word meaning 'sheep' to indicate a single sheep, multiple sheep belonging to the same flock, and multiple sheep belonging to different flocks. In contrast, there are only two forms for the word meaning 'cow,' for which the distinction between a single cow and multiple cows belonging to the same herd is collapsed (Ojeda, 1998). Thus, the distinctions that are relevant in English (e.g., between singular and plural) are somewhat different than the distinctions that are relevant in Papago, and within a language they depend on the lexical item involved. Furthermore, the distinctions relevant for particular lexical items are not always easy to predict from their semantics.

However, this does not rule out the possibility that for collective nouns in particular there are two concepts at the message level, one for the collective and one for the distributive sense. If this turns out to be the case, there will remain the question of whether the competition effects are truly localized to the message level, or spill over into grammatical encoding. There is at least some evidence in the literature which leads us to favor the latter possibility. The choice between distributive and collective senses of a collective is in many respects similar to a version of the picture-word interference task in which a speaker must name a picture while a semantically related distractor word is superimposed on it. Two semantically related concepts are activated, and competition (or interference) is observed (e.g., Glaser & Dungan, 1984). Levelt, Roelofs, & Meyer (1999) argue that this effect reflects competition during lemma selection, rather than competition at the conceptual level, on the grounds that the interference goes away when the task doesn't require actually producing one of the words. Thus there are several reasons not to ascribe our competition effects to the message level, but clearly more research investigating the precise locus of competition effects in production would aid in clarifying these issues.

Weighing semantics and phonology

A second question raised by the current results is why, if morphophonology can sometimes be a good cue to agreement, morphophonological effects on agreement seem to be so much weaker than semantic ones. In the Bock & Levelt (1994) model, this is explained by architectural constraints on the flow of information through the system. However, another possible reason involves the time at which semantic and phonological information become available during production. During utterance planning, the semantic content of messages is

formulated before the articulatory plan is developed. This means that even in the most interactive production model, semantic information will tend to be available earlier than phonological information, so that at the point when alternative forms of the verb are being activated, semantic information is more likely to be available.

Incorporating this observation into the constraint satisfaction model presented earlier allows us to propose three criteria for predicting if and when a given factor will noticeably affect agreement processing (see, e.g., Bates & MacWhinney, 1982; McClelland, 1987; Spivey & Tanenhaus, 1998, for more general statements of these principles). First, in order to influence agreement processing, a factor must exhibit a reliable correlation with verb marking. Second, the effect of a given factor will be modulated by its interaction with other factors. An individual factor will have little effect when several other factors strongly activate a particular alternative. On the other hand, even a relatively weak factor may have considerable influence if the other factors are not decisive. Finally, the effect of a factor will be modulated by when it is available to the processing mechanisms computing agreement. All other things being equal, factors available earlier on will have a greater influence than factors available later.

A potential criticism of the constraint satisfaction approach is that it allows unreliable, redundant information sources to play a role in agreement processing. For example, phonology does not provide any information which is not available from other, more reliable information sources, such as grammar and semantics. Vigliocco & Franck (1999) suggested that the use of redundant sources of information helps ensure accurate production. If information from one source is missing or lost, the correct determination of marking on the verb can (usually) still proceed based on the contribution of other sources. Vigliocco and Franck also noted that the use of semantic information may facilitate simultaneous encoding of multiple elements, as the marking on an agreement target could conceivably be selected before encoding of the agreement controller is complete.

This approach extends naturally to other effects in the literature, such as finding by Thornton & MacDonald (in press) that local nouns interfered with agreement more when they were plausible subjects of the sentence than when they were implausible. Plausibility of noun-verb relationships is a very good cue to which noun the verb should agree with. This information is redundant with that provided by syntax, but making use of it helps to reduce the likelihood of errors.

Syntax versus semantics

Although the constraint satisfaction approach departs from the traditional account of agreement in

significant ways, it also maintains certain aspects of it. In particular, the constraint satisfaction approach is not inherently inconsistent with the claim that agreement is essentially a syntactic phenomenon. Non-syntactic factors come to play a role because they facilitate the reliable and error-free production of agreement. However, the sizable effect of semantic factors on agreement might suggest a more radical departure from the traditional approach, in which syntactic information does not play a privileged role. One alternative, evident in several proposals in the linguistics literature, is that agreement is primarily based on semantics or discourse factors rather than on syntax (e.g., Barlow, 1992; Pollard & Sag, 1988; Reid, 1991). However, this alternative does not seem to offer a real improvement over syntactic approaches; in reviewing the range of proposals, Corbett (1994) commented that “Just as there are cases where the syntactic view of agreement has to appeal to semantic factors, so it would seem that a semantic view will need to appeal to syntactic factors. ... The evidence suggests that both syntax and semantics are involved in agreement” (p. 59). A strength of more interactive approaches, including constraint satisfaction, is that they do not require deciding whether agreement is primarily semantic or syntactic; rather, the relative contribution of each factor may be different in different cases.

The effect of local nouns

One important issue in the study of agreement production is how local nouns come to influence agreement. In the traditional approach to agreement, attraction errors occur when features originating on the local noun are mistakenly allowed to percolate to the top of the subject noun phrase (Vigliocco & Nicol, 1998). In the constraint satisfaction approach, however, factors should influence agreement to the extent they exhibit a reliable correlation with verb marking. For correct agreement, local noun number is largely independent of verb marking, which would seem to predict that local noun number should play no role. However, there is an interesting subset of cases for which local noun number, rather than head noun number, appears to govern agreement: compare *a bunch of marbles were rolling around the floor* to *a bunch of sand was blowing around*. In such expressions, the head noun (which is often a collective) acts much like a quantifier (Michaux, 1992).

Thus, this particular construction (*a* ⟨singular noun⟩ of ⟨plural noun⟩, hereafter the ‘a number of’ construction) is correlated with the use of a plural verb. In the constraint satisfaction framework, the more similar a given construction is to this one, the more strongly the plural verb form will be activated. For a highly dissimilar phrase such as *my winter jacket*, the plural verb form will not be activated at all, and plural verbs should be

produced very rarely. In contrast, a more similar construction such as *the key to the cabinets* will result in the plural form being slightly activated, and plural verb forms should occasionally be produced—which is in fact the case (Bock & Miller, 1991). Greater similarity to the ‘a number of’ construction should make the production of plural verbs even more likely. Note that this proposal essentially amounts to the claim that distributional information, which has been shown to play a prominent role in language comprehension, also has an impact in production as well. This possibility was considered in more detail by Thornton, Haskell, & MacDonald (2001).

Solomon & Pearlmutter (2001) presented data which are consistent with these claims, though these authors focused on the semantic rather than the distributional properties of noun phrases. Using the fragment completion paradigm, they contrasted noun phrases like *the drawing of the flowers*, containing the preposition ‘of,’ and *the drawing with the flowers*, which contains a different preposition, in this case ‘with.’ Note that the items in the ‘of’ condition were more similar to the ‘a number of’ construction than the items using ‘with.’ As the constraint satisfaction account would predict, plural verbs were significantly more likely following the ‘of’ phrases than following phrases with ‘with.’ Haskell (2002) replicated this finding with a more general contrast between ‘of’ and several other prepositions.

In sum, the constraint satisfaction account suggests that previous experience producing agreement in particular constructions affects agreement production in subsequent utterances. Similar suggestions have been offered for other aspects of grammatical encoding (Stallings et al., 1998) and for effects of structural persistence across utterances (Bock & Griffin, 2000). Thornton et al. (2001) considered some of the additional implications of this account for agreement production, but it is clear that significant work remains to be done on the role of previous experience in speech production.

Limitations of the paradigm

The fragment completion task used in the current set of experiments is a well-established paradigm and has been used in a large number of studies (e.g., Bock & Miller, 1991; Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock et al., 1999; Bock et al., 2001; Eberhard, 1997; Eberhard, 1999; Vigliocco et al., 1995; Vigliocco et al., 1996; Vigliocco et al., 1996; Vigliocco & Nicol, 1998; Vigliocco & Franck, 1999; Vigliocco & Zilli, 1999). However, as with any experimental paradigm, this task has both strengths and weaknesses. One concern with fragment completion is that participants must comprehend the fragment before carrying out the production portion of the task. This is of particular concern when a

particular factor is known or suspected to affect comprehension processes, and the goal is to determine if it affects production in a similar manner. It is also not clear how similar fragment completion is to natural production, in which speakers generate a message on their own. Finally, use of this paradigm requires the assumption that the factors which influence error rates are also relevant in normal production. The use of additional measures (latencies and ratings data) was intended to address some of the weaknesses of the fragment completion paradigm. As the alternative accounts of agreement become more complicated and the crucial data become more subtle, the use of these or other converging measures would clearly benefit researchers from all theoretical perspectives.

Conclusion

In sum, the results on collective head nouns, distributive noun phrases, regularity effects, and plausibility effects all find a natural interpretation within the constraint satisfaction framework, strongly suggesting that agreement can be viewed as a constraint satisfaction process. Whether such an approach can be applied to grammatical encoding in general, as well as phonological encoding, remains to be seen. There is some evidence for competition phenomena in choices of syntactic structure (Stallings et al., 1998). The persistence of structural priming observed by Bock & Griffin (2000) is suggestive of frequency effects in the choice of structure. However, a constraint satisfaction account of grammatical encoding would need to account for the large body of data that has motivated the development of the traditional model (Bock & Levelt, 1994). With respect to phonological encoding, several researchers (e.g., Dell, 1988; Dell, Juliano, & Govindjee, 1993; Vousden, Brown, & Harley, 2000) have developed connectionist-style models which share many properties with the constraint satisfaction approach presented here. An important avenue of research will be to explore the extent to which similar computational accounts can be developed for agreement production and other aspects of grammatical encoding.

Acknowledgments

We would like to thank Kodi Hirst for assistance in conducting the experiments, Robert Thornton for helpful discussion, and Zeni Griffin and two anonymous reviewers for insightful comments on earlier versions of the manuscript. This research was partially supported by NSF grant SBR-9511270 to the second author. Portions of this research were presented at the 14th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA.

Appendix A. Experimental Items for Experiment 1

Noun Phrase	Word
the gang/criminal behind the vicious attack(s)	convicted
the committee/mayor with the controversial idea(s)	successful
the audience/spectator in the crowded stadium(stands)	cheering
the choir/singer at the gospel concert(s)	talented
the cast/actor in the weekend performance(s)	famous
the team/athlete with the enthusiastic fan(s)	victorious
the crew/sailor on the muddy deck(with the muddy shoes)	disciplined
the fleet/ship near the small island(s)	destroyed
the squad/agent on the back porch(steps)	ready
the swarm/bee over the colorful flower(s)	angry
the anthology/book by the feminist author(s)	popular
the platoon/soldier across the corn field(s)	hiding
the orchestra/musician beside the energetic dancer(s)	loud
the herd/cow by the abandoned tractor(s)	well-fed
the flock/bird near the tall bush(es)	startled
the ensemble/violinist with the rave review(s)	arrogant
the panel/speaker from the local school(s)	knowledgeable
the band/rapper with the offensive song(lyrics)	apologetic
the council/leader behind the disastrous proposal(s)	corrupt
the coalition/candidate with the extremist agenda(supporters)	well-funded

Appendix B. Experimental Items for Experiment 2

Noun Phrase	Word
the cage for the spotted mice/rats	dirty
the blanket on her frigid feet/hands	warm
the pen for the baby geese/ducks	finished
the room for the sick children/kids	cheerful
the story about the clumsy men/boys	funny
the tree near the tired oxen/horses	old
the beggar with the crooked teeth/toes	hungry
the sign over the excited people/tourists	bright

Appendix C. Experimental Items for Experiment 3

Noun Phrase	Word
the family of mice/rats	hungry
the pair of feet/hands	protruding
the flock of geese/birds	noisy
the class of children/kids	cheerful
the mob of men/boys	violent
the team of oxen/horses	tired
the row of teeth/toes	visible
the group of people/tourists	excited

References

- Barker, J., Nicol, J., & Garrett, M. (2000). *Semantic and probabilistic factors in the processing of number of agreement*. Paper presented at the Thirteenth Annual CUNY Conference on Human Sentence Processing.
- Barlow, M. (1992). *A situated theory of agreement*. New York: Garland.
- Bates, E., & MacWhinney, B. (1982). Functional approaches to grammar. In E. Wanner, & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.
- Berg, T. (1998). The resolution of number conflicts in English and German agreement patterns. *Linguistics*, 36, 41–70.
- Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129, 177–192.
- Bock, K. (1995). Producing agreement. *Current Directions in Psychological Science*, 4, 56–61.
- Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99–127.
- Bock, K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, 8, 57–99.
- Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, 43, 83–128.
- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93.
- Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, 40, 330–346.
- Corbett, G. G. (1994). Agreement. In R. Asher, & J. Simpson (Eds.), *Encyclopedia of language and linguistics*. New York: Pergamon Press.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124–142.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17, 149–195.
- Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, 36, 147–164.
- Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language*, 41, 560–578.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210–233.
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33, 715–736.
- Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35, 724–755.

- Glaser, W., & Dungenhöff, F. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–654.
- Haskell, T. R. (2002). *Distributional cues and subject identification in the production of subject-verb agreement*. Doctoral Dissertation. University of Southern California, Los Angeles, CA.
- Humphreys, K. R., & Bock, K. (1999). *The role of conceptual number in subject-verb agreement*. Paper presented at the University of Southern California Speech Production Conference.
- Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Mallinson, G., & Blake, B. (1981). *Language typology: Cross-linguistic studies in syntax*. Amsterdam: North-Holland.
- McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading*. The Hove, England: Erlbaum.
- Michaux, C. (1992). The collectives in French: A linguistic investigation. *Linguisticae Investigationes*, 16, 99–124.
- Ojeda, A. E. (1998). The semantics of collectives and distributives in Papago. *Natural Language Semantics*, 6, 245–270.
- Pollard, C., & Sag, I. A. (1988). An information-based theory of agreement. In D. Brentari, G. Larson, & L. MacLeod (Eds.), *Papers from the 24th annual regional meeting of the Chicago Linguistics Society: Parasession on agreement in grammatical theory*. Chicago: Chicago Linguistics Society.
- Reid, W. (1991). *Verb and noun number in English: A functional explanation*. New York: Longman.
- Solomon, E. S., Pearlmutter, N. J. (2001). *Effects of linking prepositions in the production of subject-verb agreement*. Paper presented at the Fourteenth Annual CUNY Conference on Human Sentence Processing.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1521–1543.
- Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, 39, 392–417.
- Thornton, R., Haskell, T., & MacDonald, M. C. (2001). *A distributional account of agreement production*. Paper presented at the 7th Annual Conference on Architectures and Mechanisms for Language Processing.
- Thornton, R., & MacDonald, M. C. (in press). Plausibility and grammatical agreement. *Journal of Memory and Language*.
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, & L. Frazier (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, 61, 261–298.
- Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, 34, 186–215.
- Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, 40, 455–478.
- Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, 45, 368–390.
- Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound and syntax in language production. *Psychological Bulletin*, 128, 442–472.
- Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles. Notional concord in Dutch and French. *Language and Cognitive Processes*, 11, 407–442.
- Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, 68, B13–B29.
- Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: The case of gender disagreement in Italian language-impaired and unimpaired speakers. *Journal of Psycholinguistic Research*, 28, 623–648.
- Vousden, J., Brown, G., & Harley, T. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, 41, 101–175.