Journal of Memory and Language **41,** 427–456 (1999)
Article ID jmla.1999.2653, available online at http://www.idealibrary.com on IDEAL®

Agreement Processes in Sentence Comprehension

Neal J. Pearlmutter

Northeastern University

and

Susan M. Garnsey and Kathryn Bock

University of Illinois

Three experiments examined the processing of subject–verb agreement in sentence comprehension. Experiment 1 used a word-by-word self-paced moving window reading methodology, and participants read sentences such as *The key to the {cabinet/cabinets} {was/were} rusty from many years of disuse*. When the head noun (*key*), local noun (*cabinet*), and verb were all singular, reading times after the verb were faster than when either a plural local noun or plural verb was present. Experiment 2 used an eyetracking paradigm and revealed a pattern like that in Experiment 1, with a finer grain of resolution. Agreement computations influenced processing within one word after encountering the verb, and processing disruptions occurred in response to both agreement violations and locally distracting number-marked nouns, despite the fact that neither is a priori relevant for comprehension in English. Experiment 3 revealed an asymmetry in the pattern of disruptions that parallels error distributions in language production (e.g., Bock & Miller, 1991). The results suggest that agreement is an early, integral component of comprehension, mediated by processes similar to those of production.

Key Words: sentence comprehension; subject–verb agreement; feature processing; plural markedness.

One intuitive way to demonstrate that humans have some abstract knowledge of the languages they speak is to present a contrast like that in (1). Naive English speakers are quite willing to report that (1b) contains some kind of problem not present in (1a). Most notably, speakers can make such judgments despite having no idea what *glorks* means—it is in fact not a word of English—and thus they must have some knowledge of the grammar of their language independent of its particular words.

(1) a. The glorks were in the bucket on the counter.
b. The glorks was in the bucket on the counter.

Example (1) relies on comprehenders' sensitivity to subject–verb agreement: In English, as in a wide variety of languages, the number (singular vs plural) marked on the subject of a clause must agree with the number marked on the verb of that clause. The subject of a clause

The experiments reported in this article were conducted while the first author was a postdoctoral fellow in the Psychology Department and at the Beckman Institute at the University of Illinois. This research was supported by NIH Fellowship MH10592, a University of Illinois Research Board grant, NIH Grant R01-HD21011, and NSF Grant SBR 94-11627. We greatly appreciate the comments and advice of Gary Dell, Ted Gibson, Maryellen MacDonald, Gary Marcus, Elizabeth Myers, Christine Sevald, Michael Tanenhaus, and three anonymous reviewers. (None of the remaining errors are Ted's fault.) We thank Susan Cho, Sandra Fleming, Jennifer Hamblin, Ralda Mnayer, Daynia Sanchez, and Robin Harris for help in conducting the experiments; J. Cooper Cutting, Mera Kachgal, and Corey McCarthy for handling the preamble norming study in Experiment 1; and Steve Fadden and Gary Wolverton for assistance with eyetracking data collection. We are particularly indebted to George McConkie for providing eyetracking facilities (supported by the Army Research Laboratory (DAAL-01096-2-0003) and NSF (CDA 96-24396)). Portions of this work were presented at the 1995 CUNY Sentence Processing Conference (Tucson, AZ) and at the 1995 Conference of the Psychonomic Society (Los Angeles, CA).

Address correspondence and reprint requests to Neal J. Pearlmutter, Psychology Dept., 125 NI, Northeastern University, Boston, MA 02115. Fax: (617) 373-8714. E-mail: pearlmutter@neu.edu.

0749-596X/99 $30.00

is typically a noun phrase (NP), which has as its head a noun marked for number. Thus in (1), comprehenders assume that *glorks* is the (plural) noun head of the subject, so *were* is the appropriate plural verb form in (1a), and *was*, which is singular, is inappropriate in (1b).

Of course, sensitivity to subject–verb agreement is not restricted to cases involving nonsense words. In examples like (2), comprehenders judge versions containing correct agreement (2a) to be better than those with incorrect agreement (2b). Indeed, such judgments for examples like (2) are quite sharp, and thus not surprisingly, linguists have taken agreement phenomena to be an important source of data to be explained by linguistic theories.

(2) a. The sponge was in the bucket on the counter.
b. The sponge were in the bucket on the counter.

From the perspective of language-processing theories, however, the importance of agreement phenomena is not so clear. While language producers clearly must compute agreement in order to generate grammatically correct forms, computation of, for example, subject–verb agreement is not obviously required for comprehension. English in particular has a relatively fixed word order, with the result that the subject NP and the main verb of a clause are nearly always identifiable on the basis of positional and syntactic category information alone (see also MacWhinney, Bates, & Kliegl, 1984). The fact that a subject and a verb must agree in number is then largely redundant from the perspective of identifying clausal constituents and syntactic relations (but cf. some kinds of structural ambiguity, as in Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996, where number agreement alone disambiguates attachment).

The potential availability and therefore usefulness of agreement constraints is further limited by the fact that English has a rather minimal overt agreement system, both for subject–verb agreement and for agreement within NPs. Nearly all nouns are overtly marked for singular versus plural number, but gender marking, common in Germanic and Romance languages, for example, is identifiable in English only in the pronominal system (e.g., *she, itself, him*). Similarly, adjectives and prepositions are marked for neither number nor gender (unlike Romance languages), and even though many determiners (e.g., *many*) are marked for number, *the* is not, and it is the most common determiner in English. It is roughly three times more common than any other determiner and more frequent than all other determiners combined in the Brown corpus (Francis & Kučera, 1982). Most critically, verbs are marked only for number, and only in third-person present-tense forms. The only exception is the copula *be*, which is marked in both first- and third-person and in present- and past-tense forms. In the Brown corpus only 22.4% of all verbs are overtly marked for number (41,727 out of 186,070). Quite often, then, in English comprehension, the information necessary to use agreement constraints will be unavailable to the comprehender, and, even when it is available, the information provided will often be entirely redundant. Thus the comprehension system might be more efficient if it largely ignored agreement information, backtracking to handle it only when other constraints were insufficient (see also Nicol, Forster, & Veres, 1997).

On the other hand, a variety of languages seem to rely more heavily on agreement to provide grammatical constraints. Many languages have a much richer set of overt agreement markings as well, indicating that the parser might rely more heavily on agreement in such languages. The idea that the basic structure of the human language comprehension system is universal (i.e., uniform across speakers of different first languages; e.g., Frazier, 1987a, b; cf. Cuetos & Mitchell, 1988; Gibson et al., 1996; MacWhinney et al., 1984) would then predict that even in English, computation of agreement would be an important part of comprehension. MacWhinney et al.'s results showed a contrasting pattern: In whole-sentence judgments, English-speaking comprehenders tended to rely on word-order constraints, whereas Italian- and German-speaking comprehenders tended to rely more on agreement constraints. However, their stimuli pitted constraints against each other with the result that many of the stimuli were not clearly grammat-

ical, making it difficult to determine whether their results are applicable to normal comprehension. But even if universality turns out not to apply, the parser still might compute agreement in English, both because it will occasionally be necessary to use it and because it provides an additional (even though redundant) source of information.

Questions about the use of agreement constraints are also directly relevant for many linguistic theories (e.g., Bresnan, 1982; Chomsky, 1965, 1995; Pollard & Sag, 1994) because the mechanisms underlying agreement in such theories, typically some form of feature processing, are also used more generally in the computation of many other constraints—to determine the permissibility and placement of arguments, for example. The same is true for many computational and psycholinguistic theories: The primary processing mechanism in unification parsing models (e.g., Kay, 1985/1986; Shieber, 1986; see also Jurafsky, 1996) is feature unification, and Stevenson's (1994) model uses feature-based computation to control all of its structure building. Similarly, recent constraint-based sentence-processing theories (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994), though certainly not explicit about the processing of grammatical constraints, rely heavily on featural information (including nonsyntactic features) to specify constraints, and one of the strongest claims of these models is that different features are all handled by the same underlying processing mechanisms. In other parsing models, including those built around more traditional symbolic tree-construction mechanisms (e.g., Earley, 1970/1986; Frazier, 1979, 1987a; Gibson, 1991; Jurafsky, 1996), agreement computations are typically built into the tree-construction rules (though they need not be) and so operate using the same mechanisms as the rest of the parser. Agreement thus plays a role in all of these theories, and in many of them, feature-processing mechanisms are responsible for substantial proportions of the models' operations. Examinations of agreement phenomena can therefore provide information relevant to the core mechanisms of such theories.

However, despite the potential relevance of information about how people handle agreement, such phenomena have only recently begun to be investigated explicitly in psycholinguistic research. Bock and colleagues (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Bock, Nicol, & Cutting, 1999) examined the computation of agreement during language production in English, and this work has been extended to Italian (Vigliocco, Butterworth, & Semenza, 1995), Spanish (Vigliocco, Butterworth, & Garrett, 1996a), and Dutch and French (Vigliocco, Hartsuiker, Jarema, & Kolk, 1996b). Participants in all of these studies orally completed sentence beginnings [consisting of subject NPs, as in (3)] to form full sentences. Across a variety of manipulations, these experiments have revealed that producing a completion for a beginning like (3b), where the first, or head, noun (*key*) is singular and the second, or local, noun (*cabinets*) is plural, is more likely to result in a subject–verb agreement error (e.g., . . . *were locked in the desk*) than completing a beginning like (3a), where both the head and local noun are singular.

(3) a. The key to the cabinet . . .
b. The key to the cabinets . . .

This pattern is further complicated by the finding that beginnings with plural head nouns do not create the same effect: Completions of (4a) and (4b) have roughly equal probabilities of containing an agreement error.

(4) a. The keys to the cabinets . . .
b. The keys to the cabinet . . .

Bock and Eberhard (1993) ruled out a variety of morphophonological explanations for this asymmetry and proposed that the plural form is explicitly marked with a morphosyntactic feature during on-line processing, whereas the singular is the default unmarked form. On this proposal, the marked plural local noun in (3b) can sometimes inappropriately override the unspecified head noun form, thereby creating errors. In (4b), however, the head noun is much less likely to be inadvertently overwritten, both because it is plural and thus explicitly marked and because the local noun is unmarked and

thus has no feature to override that of the head noun. This *head-overwriting* proposal appears to be generally compatible with the models and theories described above. Eberhard (1997) tested and supported these hypotheses in experimental studies of agreement production, and Vigliocco and Nicol (1997), examining them further, proposed an explicit implementation in terms of a feature-unification mechanism.

Agreement phenomena have received less attention in comprehension, but the effects of both head/local NP number mismatch [as in (3) and (4)] and outright ungrammaticality have recently begun to be examined. Most of these studies have demonstrated readers' sensitivity to real and/or seeming violations (Blackwell, Bates, & Fisher, 1996; Jakubowicz & Faussart, 1995; Kail & Bassano, 1997; Nicol et al., 1997; Sevald & Garnsey, 1995; see also Deutsch, 1998, for related evidence from Hebrew), but in each of these cases an additional concurrent task was involved beyond reading for comprehension—either grammaticality judgment (Blackwell et al., 1996; Kail & Bassano, 1997; Nicol et al., 1997), lexical decision (Jakubowicz & Faussart, 1995), or naming (Sevald & Garnsey, 1995). This raises two relevant concerns: First, with respect to demonstrating basic sensitivity to agreement in comprehension, a subsidiary task (particularly grammaticality judgment) might artificially increase sensitivity. In fact, using eyetracking without any task beyond reading for comprehension, Branigan, Liversedge, and Pickering (1995) found no effects of head/local NP number mismatch with constructions like (3) and (4) followed by verb phrases. The second concern is that the time course of performance for the subsidiary tasks limits how informative they can be about relatively rapid effects of agreement. Typical reaction times following the appearance of the number-marked verb in Nicol et al.'s (1997) grammaticality judgment task (the "maze" task, in which participants choose between two possible words to continue a sentence; see also Freedman & Forster, 1985) were on the order of 650–750 ms, and Sevald and Garnsey's mean naming times were typically 500–550 ms. These time ranges are fairly long relative to normal reading rates (200–250 ms per word), particularly if agreement is computed during initial parsing.

Several event-related potential recording studies have also considered agreement violations (Coulson, King, & Kutas, 1998; Hagoort, Brown, & Groothusen, 1993; Kutas & Hillyard, 1983; Osterhout & Mobley, 1995) and suggest that the above concerns warrant some attention. With no task involved beyond reading for comprehension, Coulson et al. (1998), Hagoort et al. (1993), and Osterhout and Mobley (1995) found that subject–verb agreement failure resulted in increased late positivity (peaking at least 500 ms after the appearance of the incorrectly marked verb), a pattern also associated with some other kinds of syntactic processing difficulty (e.g., Osterhout & Holcomb, 1992; cf. Kutas & Hillyard, 1983). However, when Osterhout and Mobley required participants to judge the acceptability of each sentence, they also found differences in two earlier components, P2 and a left anterior negativity peaking around 400 ms.[1] This change in timing might be taken to suggest that, if anything, sensitivity to agreement violations is actually delayed in normal comprehension.

In the studies reported below, we investigated comprehenders' sensitivity to agreement violations using self-paced reading and eyetracking methodologies. Participants were not required to perform any additional tasks beyond reading for comprehension and then answering a comprehension question and did not receive any instructions or examples focusing on agreement phenomena or violations. Thus at a minimum, these methodologies can provide additional converging evidence about the processing of agreement. In addition, these methods permit relatively fine-grained tracking of processing

[1] Osterhout and Mobley were careful to instruct their participants to make "acceptability" judgments rather than grammaticality judgments; but of the 210 stimuli read by their participants, 95 (45%) contained number- or gender-feature violations involving subject–verb or pronoun–antecedent pairs, and 10 (5%) contained phrase-structure violations, while the remainder were both grammatical and sensible, containing neither semantic nor typographical anomaly. Thus the relevant discrimination does appear to have been on the basis of grammaticality.

difficulty over time (e.g., Just, Carpenter, & Woolley, 1982; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989) with at least individual-word resolution, and thus they have the potential to provide strong constraints on the timing of readers' sensitivity to agreement in normal comprehension.

A second question is how violations are handled if and when they are detected. In addition to expected increases in processing time, Experiment 2, using eyetracking, permits an examination of other aspects of eye-movement patterns, such as how often movements back to earlier portions of text (regressions) occur, as well as their starting and ending positions. We can specifically examine how often regressions beginning at the agreement violation end on the subject head noun as opposed to a locally distracting noun. A hypothesis according to which eye movements are strongly controlled by higher level cognitive processes (e.g., Carpenter & Just, 1977; Frazier & Rayner, 1982; Just & Carpenter, 1980; see Rayner & Pollatsek, 1989, for discussion) would predict that the head noun should be refixated often, as it specifies the information relevant for checking agreement.

The following experiments also investigated an additional set of questions about the relationship between language production and comprehension, both by considering agreement violations and by considering head/local NP mismatch constructions like those given above in (3), where local noun number is varied [*cabinet(s)*]. If the increase in production errors created by the mismatch in number between *key* and *cabinets* in (3b) (e.g., Bock & Miller, 1991) is the result of an interference effect in handling syntactic representations, then to the extent that the comprehension and production systems rely on the same processes (or make use of identical processes in separate implementations), a head/local NP number mismatch [as in (3b)] should result in increased difficulty for comprehenders relative to the corresponding head/local match case (3a). The alternative to this is that the two systems are quite distinct: The comprehension system might not display sensitivity to agreement phenomena at all in normal comprehension or sensitivity might be noticeably delayed. Even if the comprehension system is sensitive to agreement, the relevant processor might be organized differently from that in production, so that head/local mismatches have little effect (as in Branigan et al., 1995). For example, the comprehension system might be better able to isolate the head noun's agreement features or it might simply ignore them, perhaps only checking agreement if some later property of the incoming sentence (e.g., ambiguity) forced it to backtrack and do so. A third alternative is that both systems display sensitivity to agreement, but as a result of different underlying mechanisms. We investigate this possibility in Experiment 3, pitting the head-overwriting explanation of head/local mismatch effects from production, described earlier, against an explanation based on comprehension sensitivity to word-to-word transition probabilities instead of feature passing.

One final question examined in the experiments was about the relationship between grammatical number and so-called notional number (numerosity in the conceptual representation of a speaker or hearer). This relationship is far from perfect, which complicates questions about number agreement. For example, the subject NP *the picture on the postcards* is singular in grammatical number but ambiguous in its notional number properties. It is notionally plural if the conceptual referent of the expression consists of tokens of the same picture distributed across multiple postcards. This is the distributive sense. However, if the conceptual referent is the picture itself, the expression is notionally singular. This is the nondistributive sense. Notional number clearly influences agreement in some circumstances (see Pollard & Sag, 1994, for discussion and Gernsbacher, 1991, for an empirical demonstration), but there are conflicting findings in the production literature about the predominance of grammatical and notional agreement in unstudied speech (compare Bock & Miller, 1991; Eberhard, 1996; and Vigliocco et al., 1996a). Nothing is known about the immediate effects of notional number on comprehension. For this reason, we included a notional-number contrast in the stimulus materials to explore whether and how differences in notional number influence the processing of agreement during reading.

EXPERIMENT 1

This experiment was intended to examine, first, whether comprehenders are sensitive to agreement violations during the course of normal reading. The relatively fixed word order of English and its minimal agreement system conspire to limit the potential usefulness to a comprehender of knowledge about the expected number-marking on an upcoming verb. Second, if readers were sensitive to agreement, the relative timing of this sensitivity could be considered. The third issue of interest was whether the pattern of interference created by number mismatches in language production also appeared in comprehension. Finally, to examine the effects of notional number, we included among the materials subject NPs whose dominant interpretation was nondistributive, or notionally singular (e.g., *the key to the cabinets*), and others whose dominant interpretation was distributive, or notionally plural (e.g., *the picture on the postcards*). All of these subject NPs were grammatically singular. If notional plurality has an immediate and sizable impact on the ongoing computation of agreement, we should see greater disruptions from agreement violations after notionally (and grammatically) singular subject NPs than after notionally plural (but grammatically singular) subjects.

Method

Participants. Eighty-two University of Illinois undergraduates participated for class credit or $5. All (in this experiment and the following two) were native English speakers. Two participants were excluded from all analyses due to poor comprehension question performance (less than 90% correct across all items in the experiment).

Materials. Sixteen stimulus sets like that shown in (5) were constructed. Each consisted of a head NP [e.g., *the key* in (5)] followed by a preposition (*to*) and a local NP [*the cabinet(s)*], which was the object of the preposition. Following the subject of the sentence was a past-tense copula (*was* or *were*) and then a four- to nine-word completion. The word immediately following the copula was either a past-participle verb, a nonparticipial adjective, a preposition, or the determiner *a* (four of each type). The head NP was always singular, and the four different versions of an item were created by varying the local NP number (*cabinet* vs *cabinets*) and the verb number (*was* vs *were*), as shown in (5). The subject NPs were the short, prepositional phrase (PP) versions of Bock and Miller's (1991, Experiment 1) stimuli; half of these were distributive and half were nondistributive, according to Bock and Miller's classification.

The full-sentence versions of the stimuli are shown in Appendix A. All items were written to fit on a single 80-character display line, and each item had an associated yes/no comprehension question. Because we did not wish to draw participants' attention to agreement issues, the comprehension questions asked about other parts of the sentences.

(5) a. The key to the cabinet was rusty from many years of disuse.
b. The key to the cabinets was rusty from many years of disuse.
c. The key to the cabinet were rusty from many years of disuse.
d. The key to the cabinets were rusty from many years of disuse.

Plausibility norming. To insure that any effects of local noun number were due to differences in number marking rather than more general plausibility differences, a separate group of 68 University of Illinois undergraduates rated the subject NPs for plausibility as sentence-initial fragments like *The key to the cabinet . . . .* The 16 items were combined with 32 fillers with similar syntactic structures as well as the other 16 Bock and Miller (1991, Experiment 1) subject NPs, which contained a relative clause rather than a PP modifying the head NP. Ratings were conducted for four versions of each item, created by crossing head noun number (singular vs plural) with local noun number (singular vs plural). (The plural-head versions were not relevant for Experiments 1 and 2, but were included in Experiment 3.) Participants rated exactly one version of each item using a scale of 1 (plausible) to 5 (implausible). Participants were explicitly instructed to "judge plau-

TABLE 1

Subject NP Plausibility Ratings

Head noun number	Local noun number	Mean (*SD*)
Singular	Singular	1.11 (.23)
	Plural	1.16 (.29)
Plural	Singular	1.13 (.22)
	Plural	1.10 (.19)

Note. Rating scale was 1 (plausible) to 5 (implausible).

sibility" and were given a clearly plausible and clearly implausible example.

The mean plausibility ratings for all four versions are shown in Table 1, and the ratings for each version of each item are shown in Appendix A. An error in construction of the rating forms resulted in one item [*The name(s) on the billboard(s) . . .*] being rated in a slightly different form than that in which it was presented in the reading time experiment: *sign(s)* was used in place of *billboard(s)* for rating. None of the results reported below differed when the ratings of this item were excluded.

Plausibility ratings did not differ significantly between either the two singular-head versions [$t_1(67) = -1.16$, $p > .20$; $t_2(15) = -1.25$, $p > .20$] or the two plural head versions [$t_1(67) = .97$, $p > .30$; $t_2(15) = .75$, $p > .45$]. However, as an additional check on possible plausibility effects, correlations between reading times and plausibility ratings are considered below.

Design. We refer to conditions as combinations of the factors grammaticality and head/local NP number match (NP-match), so the four conditions in the experiment were NP-match grammatical [as in (5a)], NP-mismatch grammatical (5b), NP-match ungrammatical (5c), and NP-mismatch ungrammatical (5d). The 16 experimental stimuli were placed into four lists such that each list contained exactly one version of each item, and each list contained four items in a given condition. Each list also contained 94 other (filler) items, 74 of which were part of unrelated experiments reported elsewhere, involving ambiguous and unambiguous versions of one of two different temporary ambiguities (direct object vs sentential complement, or main verb vs reduced relative clause). The remainder of the items incorporated a variety of structures and were always grammatical sentences. Thus each subject saw 8 ungrammatical sentences out of 110 items total (7.3%).

Apparatus. An IBM-compatible computer running the MicroExperimental Laboratory (MEL) software package controlled stimulus presentation and response collection.

Procedure. Participants read 10 initial practice items followed by one of the 110-item lists. Neither the instructions for the experiment nor the practice items contained any reference to or example of agreement violations or other ungrammaticality. The stimuli were presented in a random order using a noncumulative word-by-word self-paced moving window paradigm (Just et al., 1982). At the beginning of a trial, an item was displayed on the screen with all nonspace characters replaced by dashes. When the participant pressed the space bar, the first word of the item was displayed, replacing the corresponding dashes. When the participant pressed the space bar a second time, the first word reverted to dashes, and the second word was displayed in place of the appropriate dashes. Each subsequent press of the space bar revealed the next word and removed the previous word. Pressing the space bar on the last word of the item caused the item to be replaced by its Yes/No comprehension question, which the participant answered by pressing one of two keys above the space bar on the keyboard. The computer recorded the time between each button-press as well as the comprehension question response and presented feedback about the participant's answer to the question. Most participants completed the experiment in approximately 35 min.

Results

Comprehension performance was 96% correct for the NP-match ungrammatical condition and 95% correct for the other three conditions; these values did not differ (all *p*s $> .20$). Trials on which the participant answered the comprehension question incorrectly were excluded from reading time analyses.

To adjust for differences in word length

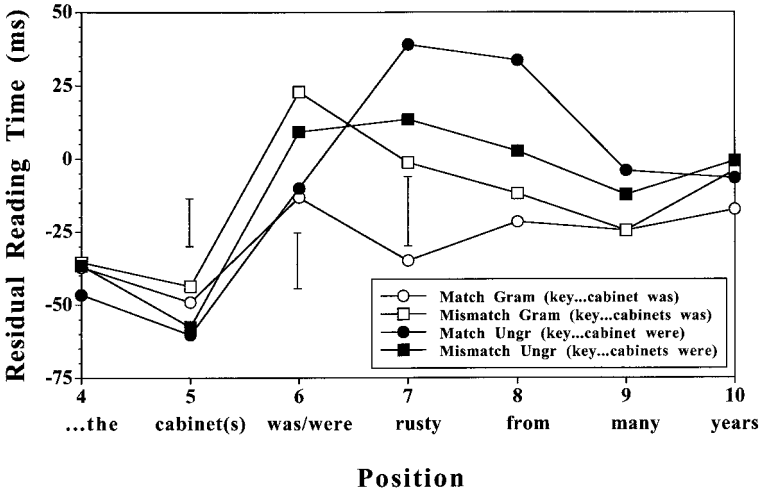

FIG. 1. Experiment 1 grand mean residual reading time by position. Error bars show 95% confidence intervals for differences between cell means, computed from the analyses by participants.

across conditions as well as overall differences in participants' reading rates, a regression equation predicting reading time from word length was constructed for each participant, using all filler and experimental items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). At each word position, the reading time predicted by the participant's regression equation was subtracted from the actual measured reading time to obtain a residual reading time. Thus each participant's mean reading time per word across the entire experiment was transformed to 0 ms residual reading time, and negative residual times indicate faster-than-average times. Residual reading times beyond 5 *SD* from the corresponding condition × position cell mean were excluded, affecting less than 0.6% of the data, and all analyses were conducted on the resulting data set. Appendix B reports the raw reading times trimmed at 5 *SD*.

Figure 1 shows grand mean residual reading times by condition at each word from the determiner preceding the local noun (position 4) through the 10th word of the sentence. We report results for the verb (*was/were*, position 6), the following word (e.g., *rusty*, position 7), and the local noun [e.g., *cabinet(s)*, position 5]. The data were analyzed in 2 (grammaticality) × 2 (head/local NP number match) ANOVAs conducted separately for participants (F_1) and items (F_2; Clark, 1973) at each word position. Where interactions are present, we describe the pattern of reliable differences among cell means based on the pooled MS_e from the corresponding ANOVA (Winer, 1971). Figure 1 shows the relevant 95% confidence interval for the size of the difference between cell means, computed over participants (Loftus & Masson, 1994) at each word. Appendix C lists the cell means and confidence intervals computed over participants and over items for each reported analysis. In the text below and in subsequent experiments, effects reported as reliable were significant at or beyond the .05 level unless otherwise noted.

Verb (position 6). At the verb (always *was* or *were*), NP-match conditions were faster than NP-mismatch conditions [$F_1(1,79) = 15.85$; $F_2(1,15) = 30.09$], but neither the main effect of grammaticality nor the interaction approached significance (Fs < 2, ps $> .20$).

Word following the verb (position 7). An effect of grammaticality did not appear until the word following the verb (e.g., *rusty*), where grammatical conditions were faster than ungrammatical ones (Fs > 24). The effect of NP-match was no longer reliable on this word (Fs < 1), but an interaction was present [$F_1(1,79) = 15.71$; $F_2(1,15) = 24.77$] because the NP-match effect in the ungrammatical con-

ditions was the reverse of that in the grammatical conditions: As Fig. 1 shows, the NP-match grammatical condition was easier than the other three, but the NP-match ungrammatical condition was the most difficult. Thus across the verb and the following word, comprehenders did show sensitivity to both agreement violations and distracting local nouns, but the sensitivity to agreement violations appeared later than that to head/local mismatches.

Local noun (position 5). At the local noun, grammatical conditions were read more slowly than ungrammatical ones [$F_1(1,79) = 5.32$; $F_2(1,15) = 4.86$], but neither an NP-match effect nor an interaction was present at this position (Fs < 1). Because the stimuli were still identical at this point with respect to grammaticality and because this effect did not replicate in Experiment 2, it is likely to be spurious, and we do not discuss it further.

Correlations with plausibility. Although the two different versions of the subject NPs (singular head–singular local and singular head–plural local) did not differ significantly in plausibility, the singular head–singular local noun versions were rated as slightly more plausible, which might have made them easier to read. However, correlations computed across items, at the verb and at the following word, in each of the four conditions, revealed no reliable relationships (all ps $> .35$).

Distributivity effects. To determine whether distributive and nondistributive items differed in the degree to which they displayed difficulty for NP-match versus NP-mismatch conditions, analyses of residual reading times at the verb and the next word were conducted including distributivity as a factor. Because of the small number of items (2) in each cell of this design, five participants were missing all data in at least one cell and were excluded. At both the verb and the following word, distributive items tended to be read more quickly than nondistributive items [verb: $F_1(1,74) = 7.28$, $MS_e = 6445$, $p < .01$; $F_2(1,14) = 2.45$, $MS_e = 2246$, $p < .15$; following word: $F_1(1,74) = 3.60$, $MS_e = 7818$, $p < .10$; $F_2(1,14) = 3.71$, $MS_e = 1194$, $p < .10$], but distributivity did not interact with any other factors (all Fs < 1).[2]

Discussion

The clearest result from Experiment 1 was that readers were sensitive to both ungrammaticality and a locally distracting number-marked noun. Sensitivity to head/local NP number match appeared immediately on the verb, and sensitivity to grammaticality, as controlled by the number marking on the verb, appeared one word later. Thus readers did attend to agreement information during comprehension, and processing was disrupted when agreement constraints were violated in cases of ungrammaticality or appeared to be violated as in the NP-mismatch grammatical condition. Furthermore, these effects occurred despite the fact that agreement constraints were logically redundant for the purposes of computing syntax and interpreting the stimuli and despite the fact that no grammaticality judgment or other metalinguistic task was required. These results suggest that computation of agreement is a necessary part of language comprehension, even in a language like English, where it is often not particularly useful. This provides some support for the linguistic, computational, and psycholinguistic theories described earlier, which make extensive use of agreement and other feature-based computation.

Beyond the general finding of sensitivity to both agreement violations and head/local NP mismatches, Experiment 1 also revealed a specific interaction between the two factors, visible at the word following the verb (position 7). In the two grammatical conditions, the pattern duplicated that at the verb itself (position 6): A mismatch in number between the head noun and the local noun created processing difficulty, which parallels the increases in verb number-marking errors found in production studies. For the two ungrammatical conditions, however, an NP mismatch eased processing, so that the NP-

[2] Note that the lack of reliable interactions is unlikely to be the result of (just) a lack of power, at least for the participants' analysis: This design's power to detect a main effect (as at the verb) is the same as its power to detect two-factor interactions, assuming comparable effect sizes.

mismatch ungrammatical condition was easier than the NP-match ungrammatical condition. This pattern across the four conditions is consistent with the idea that the effect of a head/local NP-mismatch is to increase the probability of an error in computing the number of the subject NP, resulting in more mismatch-induced seeming errors in the grammatical conditions but fewer mismatch-induced seeming errors in the ungrammatical conditions. We return to this point in Experiment 3.

Experiment 1 also tested the possibility that effects of agreement might be attributable to semantic factors. However, differences in plausibility between the NP-match and NP-mismatch conditions were minimal, and they did not predict differences in reading time. Distributivity did have some effect at the verb and the following word, with those items allowing a distributive interpretation (e.g., *The slogan on the posters . . .*) tending to be read more quickly. However, distributivity never interacted with either head/local NP number match or with grammaticality. Thus there was no evidence that these particular semantic factors could account for the observed agreement effects. Similar analyses in Experiment 2 also revealed the same patterns, so we return to this issue only in the General Discussion.

EXPERIMENT 2

This experiment was intended to replicate Experiment 1 using a different methodology, eyetracking. From the standpoint of the tight timing made possible by eyetracking, one particularly interesting finding from Experiment 1 is the one-word delay in effects of grammaticality. This pattern might be ascribed to properties of the self-paced reading methodology or might instead reflect the underlying timing of such effects. Reading times in eyetracking are generally shorter than in self-paced moving-window paradigms and presumably more closely reflect reading in nonlaboratory settings. This should allow us to measure the timing of effects more tightly and also to examine the responses to real or seeming violations in more detail by considering regressive eye movements.

Method

Participants. Seventy-eight University of Illinois students participated for $5. All had normal, uncorrected vision. Fourteen participants were excluded from all analyses because of a combination of excessive trackloss and a high error rate on comprehension questions, leaving 64 participants, all of whom had at least one usable trial in each condition.

Materials and design. The materials and design were identical to those in Experiment 1.

Apparatus and procedure. Eye movements were monitored with a Fifth Generation SRI Dual Purkinje Eyetracker interfaced with an IBM-compatible PC. Stimuli were displayed on a Conrac 1000 color monitor such that four characters subtended 1° of visual angle. Viewing was binocular but only the right eye was monitored, and vertical and horizontal eye position was sampled every millisecond. A bite bar was prepared for each participant to minimize head movements, and the room was slightly dimmed to provide a comfortable viewing environment. The eyetracker was initially calibrated by having participants fixate three positions distributed across the middle line of the screen until stable position values were obtained. The calibration procedure took approximately 5 min.

Each trial began with a trial number at the left side of the screen, and participants were instructed to fixate the number before pressing a button to indicate that they were ready to read a sentence. The sentence was then presented in its entirety across the middle line of the screen, along with a fixation box at the right edge of the screen. Participants read at their own pace and were instructed to fixate the box before pressing a button to indicate when they had finished reading the sentence. This was intended to minimize recording of eye movements made to position the eyes at the left margin following completion of the sentence. As in Experiment 1, each item was followed by its Yes/No comprehension question with feedback; participants used two handheld buttons to answer.

Participants read 10 practice items followed by one of the 110-item lists; a single pseudo-

random order was used for all participants. Participants took a scheduled break following the practice items and every 30 items thereafter, and they were free to take other breaks as needed. The eyetracker was recalibrated after each break. Most participants completed the experiment in approximately 40 min.

Results

As in Experiment 1, trials on which the participant answered the comprehension question incorrectly were excluded. Comprehension performance was 90% correct in the NP-match grammatical condition, 91% in the NP-match ungrammatical condition, and 93% in the two NP-mismatch conditions. These values did not differ ($ps > .10$). Trials involving trackloss in a region of interest were also excluded, resulting in the loss of 11% of NP-match grammatical trials, 15% of NP-mismatch grammatical trials, 16% of NP-match ungrammatical trials, and 13% of NP-mismatch ungrammatical trials.[3] These values did not differ significantly, although the interaction of grammaticality and NP-match was marginal by participants only [$F_1(1,63) = 2.98$, $MS_e = 275$, $p < .10$; $F_2(1,15) = 2.56$, $MS_e = 79$, $p > .10$; other effects: all $ps > .25$].

The data were analyzed in 2 (grammaticality) $\times$ 2 (head/local NP number match) ANOVAs conducted separately for participants (F_1) and items (F_2). The ANOVAs and cell-mean-difference confidence intervals are reported as in Experiment 1 (see Appendix C). We focus on analyses of the verb (position 6) and the following word (position 7). However, individual words (especially short high-frequency ones such as *was* and *were*) are not always fixated during normal reading, and thus for many analyses, we excluded additional participants who had no data for a particular condition. For several measures, we therefore also conducted analyses of the *verb region*, which consisted of the verb and the following word combined. The number of participants excluded from analyses is noted where relevant.

[3] Tracklosses were identified using a combination of the eyetracker's trackloss signal and a trial-filtering algorithm (available from the authors).

Analyses were conducted on several related reading time and regressive saccade measures (see Rayner et al., 1989, for discussion). We report results for first-pass reading time, which is the sum of the time spent fixating a region from the first occasion when it is fixated until some other region is fixated, and for total reading time, which is the total time spent fixating a region during a trial. First-pass time is a *forward-sweep* measure: It does not include fixations which occur after some later (farther right) position has been fixated, whereas total times include all fixations in a region during a trial.

For each of these measures separately, we computed trimmed residual reading times as in Experiment 1. Less than 0.4% of the data were excluded by trimming, and all reading-time analyses were conducted on the trimmed residual times. Appendix B reports raw first-pass and total reading times by position trimmed at 5 *SD*.

In addition to processing-time measures, we also analyzed the probability of a regressive saccade following fixations in the verb region. Regressive saccade measures are generally coarser than reading-time measures, but they have been shown to index processing difficulty, particularly for fairly severe disruptions (e.g., Altmann, Garnham, & Dennis, 1992; Frazier & Rayner, 1982). They thus may provide some indication of participants' difficulty with real or seeming agreement violations. We also examined the endpoints of regressions originating in the verb region, which are possible indicators of where participants expected to find resolutions for problems which triggered regressions (Carpenter & Just, 1977; Frazier & Rayner, 1982; Just & Carpenter, 1980).

First-Pass Reading Times

Figure 2 shows first-pass residual reading times for the words around the verb. To examine how rapidly sensitivity to ungrammaticality and head/local NP number mismatches appeared, we examined first-pass times at the verb position, the word following the verb, and these two words combined (the verb region). For the verb region, we also performed a regression-contingent first-pass time analysis (Altmann et al., 1992; see also Altmann, 1994; Rayner &

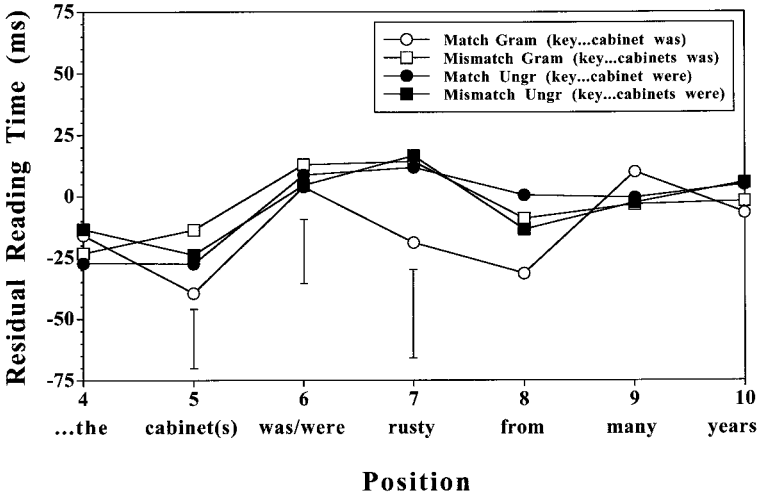

FIG. 2. Experiment 2 grand mean first-pass residual reading time by position. Error bars show 95% confidence intervals for differences between cell means, computed from the analyses by participants.

Sereno, 1994) because of the possibility that some readers were responding to apparent violations by regressing back to an earlier region while others were spending additional time in the verb region. Condition means for the verb region and regression-contingent analyses are reported in the text; confidence intervals are detailed in Appendix C, as for Experiment 1.

Verb (position 6). At the verb, as can be seen in Fig. 2, first-pass times in all conditions were quite similar. Although there was a very small numerical advantage for the NP-match grammatical condition (1–9 ms, depending on the comparison), there was no hint of a reliable main effect or interaction (all Fs < 2, ps $> .25$; 17 participants excluded).[4]

Word following the verb (position 7). At the word following the verb, neither main effect was reliable (13 participants excluded; all Fs < 3, ps $> .10$), and their interaction was marginal by items only [$F_1(1,50) = 2.53$, $p = .11$; $F_2(1,15) = 3.95$, $p < .10$]. The NP-match grammatical condition was reliably faster than the other three conditions for the analysis by items, but the cell means computed by participants did not differ.

Verb region. Because neither the reading time patterns at the verb nor at the following word displayed particularly clear sensitivity to agreement, we considered the pattern across the two words combined. The first-pass time pattern in this region was very similar to that at the word following the verb: The NP-match grammatical condition ($M = -6$ ms) appeared relatively fast compared to the other three (12, 36, and 21 ms for the NP-mismatch grammatical, NP-match ungrammatical, and NP-mismatch ungrammatical conditions, respectively). However, the NP-match effect and interaction were not reliable (all Fs < 3, ps $> .10$), and the main effect of grammaticality was only marginal, with ungrammatical conditions slower than grammatical ones [$F_1(1,63) = 3.46$, $p < .10$; $F_2(1,15) = 3.97$, $p < .10$].

The lack of reliability in these analyses, despite the numerical similarity of the pattern to that in Experiment 1, suggested that there might be some important uncaptured variation in participants' responses to real or seeming violations. One possibility, identified in ambiguity resolution studies by Rayner and Sereno (1994) and Altmann (1994), is that readers may respond to difficulty in a region in either of two

[4] An additional analysis in which the last four characters of the local noun were combined with the verb to form a region (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994) also yielded no interaction (two participants excluded; Fs < 1) or reliable effects (Fs < 3, ps $> .15$).

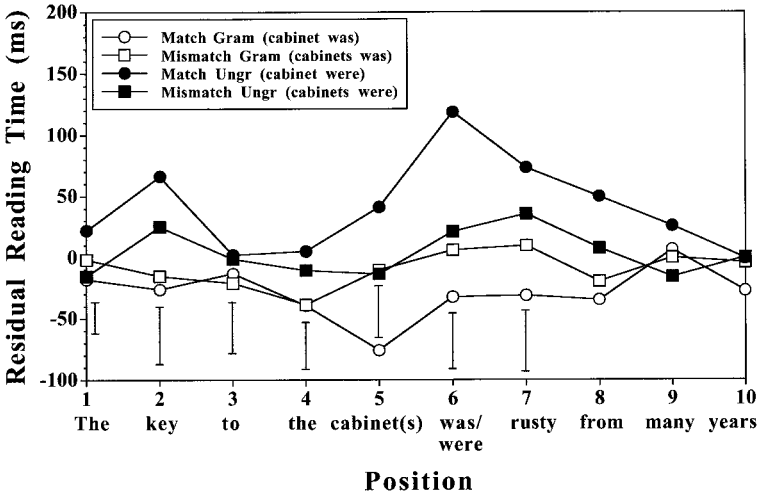

FIG. 3. Experiment 2 grand mean total reading time by position. Error bars show 95% confidence intervals for differences between cell means, computed from the analyses by participants.

ways: by spending more time in it, or by leaving it with a regressive saccade. In order to separate effects of regressions from first-pass times, we examined first-pass times separately depending on whether the first pass in the verb region was ended by a regression or not (Altmann et al., 1992). Collapsing over grammaticality and NP-match, first passes in the verb region were substantially shorter when they were ended by a regression ($M = -50$ ms) than when they were ended by a forward saccade [$M = 32$ ms; $F_1(1,41) = 14.41$; $F_2(1,15) = 11.98$; only the 42 participants with both regression-ended and nonregression-ended first-pass data were included]. This matched the findings of Altmann (1994, p. 288) and Rayner and Sereno and indicated that some participants were regressing rather than spending time in the verb region.

We did not analyze the regression-ended passes separately by condition because they comprised only about 14% of the overall first-pass data. However, the nonregression-ended first-pass data were analyzed further. The resulting pattern was numerically similar to, but statistically stronger than, that for the full first-pass data set: A reliable interaction was present [eight participants excluded; $F_1(1,55) = 5.31$; $F_2(1,15) = 8.86$] as was a main effect of grammaticality (Fs > 4). The NP-match grammatical condition ($M = -10$ ms) was faster than the others (32, 51, and 27 ms for the NP-mismatch grammatical, NP-match ungrammatical, and NP-mismatch ungrammatical conditions, respectively).

Overall then, across first-pass times considered for the verb, the following word, and the two words combined, approximately the same pattern appeared. However, effects at the verb were extremely small and did not approach significance. The pattern was stronger at the word following the verb, but for that word alone as well, effects were statistically unreliable. For the two words combined, the pattern was clearest and most stable, and once regression-ended passes were removed, the advantage for the NP-match grammatical condition over the other three was reliable. There was thus good evidence for effects of agreement processes within the span of a single word after the verb, as in Experiment 1.

Total Reading Times

The first-pass time analyses showed how rapidly readers displayed sensitivity to real or seeming violations. We next examined total reading times in order to provide additional comparisons with the results of Experiment 1 and in order to track readers' responses to violations in more detail. Figure 3 presents total reading time by word position. We first discuss

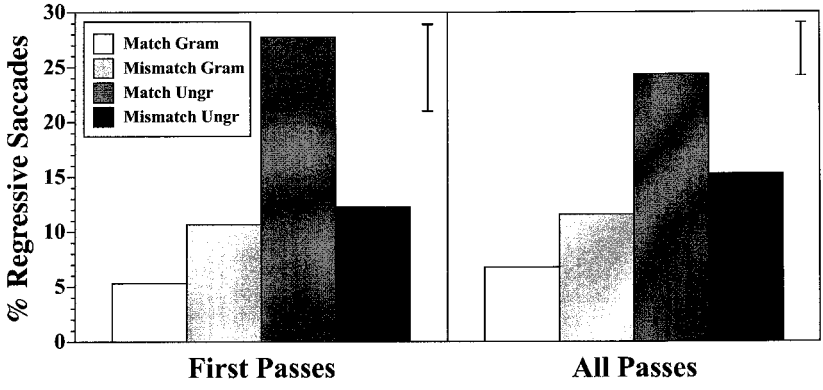

FIG. 4. Experiment 2 grand mean regressive saccade probability from the verb region following first passes (left panel) and all passes (right panel). Error bars show 95% confidence intervals for differences between cell means, computed from the analyses by participants.

reading time results at the verb and the following word and then briefly summarize the effects across the words preceding the verb.

Verb (position 6). As indicated in Fig. 3, a reliable interaction was present at the verb [eight participants excluded; $F_1(1,55) = 21.43$; $F_2(1,15) = 24.62$], and total reading times were longest in the NP-match ungrammatical condition. The NP-match grammatical condition was also reliably faster than the NP-mismatch ungrammatical condition. Main effects of grammaticality (Fs > 19) and NP-match (Fs > 4) were also present.

Word following the verb (position 7). The pattern at the word following the verb was very similar to that at the verb, with the NP-match ungrammatical condition more difficult than the other three. In addition, the NP-match grammatical condition was reliably less difficult than the other conditions. An interaction was again present [10 participants excluded; $F_1(1,53) = 13.44$, $F_2(1,15) = 7.05$] as was a main effect of grammaticality (Fs > 8).

Words preceding the verb. In addition to the verb and the following word, we also examined the positions leading up to the verb, because one possible response to a perceived agreement violation would be to reread earlier parts of the sentence. In first-pass times, there were no reliable main effects or interactions on words preceding the verb, so effects in total times can be attributed to rereading.

As Fig. 3 indicates, the pattern of total reading times was similar across the five words preceding the verb, with grammatical conditions reliably faster than ungrammatical ones at both the head noun (position 2) and local noun (position 5), as well as at the local noun's determiner (position 4). There were no effects of NP-match, but reliable interactions appeared at both the head noun's determiner (position 1) and the local noun (position 5). At the head noun's determiner, this was because the NP-match ungrammatical condition was relatively slow. At the local noun, the pattern of differences matched that at the word following the verb.

Regressive Saccade Probability

Because of the possibility that processing difficulty might be reflected in an increase in regressive saccades (e.g., Altmann, 1994; Altmann et al., 1992; Frazier & Rayner, 1982; Rayner & Sereno, 1994), we analyzed the probability of a regressive saccade following first passes and following all passes on the verb region (Fig. 4).[5] In addition, we examined the endpoints of regressions from the verb region

[5] We also examined regression probabilities for the verb alone and on the region formed by combining the last four characters of the local noun with the verb. While the patterns of regressions using these regions were generally very similar to those reported in the text, differences between conditions usually did not approach significance.

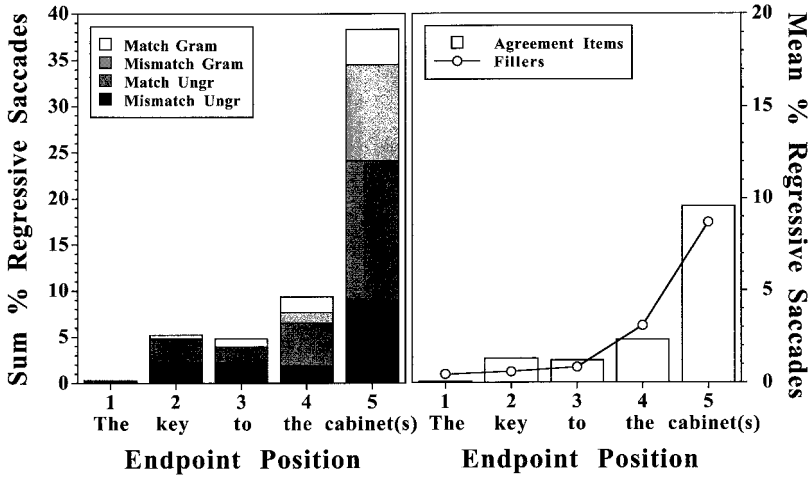

FIG. 5. Experiment 2 endpoint distribution of regressive saccades from the verb region. The left panel shows regression percentage by condition and endpoint position. Each bar segment shows the number of regressions to that position in that condition out of the total number of saccades leaving the verb region for that condition. The right panel shows the regression probability averaged across the four conditions at each position (bars) as well as the probability of a regression backward from 1–5 words (positions 5–1, respectively) in the filler stimuli (line).

(Fig. 5) as an additional indicator of readers' responses to real or seeming violations.

First-pass regressions. The left panel of Fig. 4 presents the percentage of first passes in the verb region that ended with saccades to a preceding region, out of all first passes in the verb region. Regressive saccades were reliably more likely in the NP-match ungrammatical condition than in the other three, resulting in an interaction [$F_1(1,63) = 8.93$; $F_2(1,15) = 17.02$] and reliable main effects (grammaticality: Fs $>$ 12; NP-match: Fs > 5).

Total regressions. The right panel of Fig. 4 shows the percentage of regressive saccades out of all saccades leaving the verb region, computed separately for each condition. The pattern was similar to that for first-pass regressions and for total reading times in the region: An interaction was present [$F_1(1,63) = 11.85$; $F_2(1,15) = 8.72$], along with a main effect of grammaticality (Fs > 21). The effect of NP-match was reliable only by participants. This pattern was primarily the result of the relatively high probability of a regression in the NP-match ungrammatical condition, but the NP-match grammatical condition also yielded a lower regression probability than the NP-mismatch ungrammatical condition.

Regression endpoints. The left panel of Fig. 5 shows the percentage of regressions back to each of the first five words out of all passes through the verb region. It is a breakdown by ending position of the pattern across conditions summarized in the right panel of Fig. 4: The percentage associated with a particular combination of position and condition is the number of regressions in the condition ending at that position out of the total number of saccades leaving the verb region in that condition. Because of the limited data available at each position for each condition separately, we focused on the pattern with all four conditions pooled together (the total height of each bar in the left panel of Fig. 5). This pattern makes quite clear that readers' targeting of regressions was not strongly controlled by syntactic processes: A larger percentage of the movements from the verb region ended on the local noun than on the four earlier words combined, and in particular, far more regressions ended on the local noun than on the head noun. If the needs of syntactic reanalysis controlled regression targeting, then

the head noun, which contained the relevant number information for agreement, should have been targeted more often. Instead, the probability of ending a regression on a given position appeared to be primarily a rapidly decreasing function of distance, although the head noun regression probability was slightly higher than might be expected on such an account, suggesting that the needs of reanalysis may have played a small role in directing regressions.

To provide a comparison pattern, we computed the probability of a regression back one to five words out of all movements leaving any of positions 6 through 11 in the 20 filler items that were not part of any other experiment. These values are plotted as the line in the right panel of Fig. 5, which also shows (as bars) the regression percentage averaged across the four experimental conditions. The pattern from the experimental stimuli appears to be quite similar to that from the fillers, indicating that the processes controlling regression targeting were largely independent of specifically syntactic (re-)analysis processes. Furthermore, even using the filler probabilities as a baseline and subtracting each from the corresponding probability for the experimental items in the right panel of Fig. 5 (i.e., performing an approximate correction for distance), the local noun still received a slightly higher percentage of regressions than the head noun. Thus, even though the total reading times and total regression percentages in Figs. 3 and 4 indicated that the decision about when to trigger a regression was influenced by syntactic processes, the decision about where to target the regression appeared not to be.

On the other hand, readers did eventually reexamine the head noun, as suggested by a comparison of its total time to that on the local noun (see Fig. 3). This can also be seen clearly in Fig. 6, which shows the probability of a second pass on each of positions 1 through 5 (i.e., the probability of fixating a position after moving beyond it during the forward sweep through the sentence). In contrast to the right panel of Fig. 5, the two nouns each received relatively many passes compared to the preposition and determiners. Nevertheless, the local noun still received slightly more passes than the head noun. Together with the endpoint data above, these results indicate that regression targeting was not tightly controlled by the needs of syntactic processing, at least soon after the detection of a violation, but that readers did eventually target the appropriate locations for reanalysis.

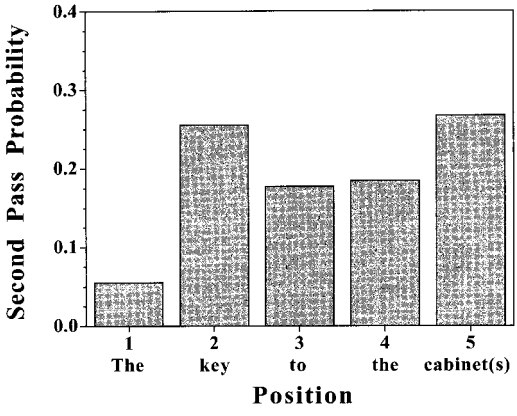

FIG. 6. Experiment 2 probability of a second pass on words preceding the verb (positions 1 through 5).

Discussion

Experiment 2 revealed a variety of notable results. First, it replicated the pattern from Experiment 1 in that both ungrammaticality and head/local NP number mismatches had an effect on reading times, with ungrammaticality creating additional difficulty, and a head/local mismatch creating difficulty in grammatical conditions but easing difficulty in ungrammatical conditions. This again supported the idea that the presence of a head/local NP mismatch increased the probability of an error in computing the subject NP's number from the head noun's number-marking, as has been claimed in production.

In addition, this experiment revealed more about the time course of sensitivity to real and seeming violations. The first-pass time measure excluding regression-ended cases showed that early in processing, the NP-match grammatical condition was relatively fast, and the other three conditions all created about the same amount of disruption. Over time, however, as indicated in

the total reading times, the difference between the two grammatical conditions tended to decrease, while the relative difficulty of the ungrammatical conditions, and particularly the NP-match ungrammatical condition, increased. Thus overall, NP-mismatch tended to be less disruptive than ungrammaticality, and an ungrammaticality with unmistakable number-marking created the most severe problems.

This experiment also confirmed the Experiment 1 finding that effects primarily appeared at the word following the verb, but not at the verb itself, indicating that properties of the self-paced reading task from Experiment 1 were not artificially creating a delay. Thus agreement computations do not seem to have reliably influenced eye movements during the first fixation, which included information about the verb. However, the reliable effects in first-pass time excluding regression-ended passes for the verb region indicated that agreement computations had influenced eye movements before the eyes moved beyond the verb region. Given that other syntactic effects (e.g., garden-path effects, as in Frazier & Rayner, 1982, and Garnsey, Pearlmutter, Myers, & Lotocky, 1997; and syntactic complexity effects, as in Holmes & O'Regan, 1981) appear reliably within one word following their earliest possible locations, these results are clearly compatible with the requirements of theories which rely on feature-based computation for all of their syntactic processing (e.g., feature-unification parsers) as well as with theories which rely less on feature-based processing.

Beyond evidence about time course, Experiment 2 also showed that the response to seeming ungrammaticality could be either an increased probability of a regression or an increase in time spent reading the seemingly ungrammatical region (replicating Altmann et al.'s, 1992, results in regions other than the disambiguation of their referential context condition; see Altmann, 1994, p. 288, and Rayner & Sereno, 1994, for discussion). The fact that the first-pass time effects were stronger after the exclusion of regression-ended trials suggested that these alternative responses tended to be in complementary distribution—readers either slowed down or regressed, but not both. Furthermore, the overall regression measures indicated that while real or seeming agreement violations could certainly influence the probability of a regression, the targets of those regressions were not controlled by knowledge of where potentially useful information might be found (cf. Carpenter & Just, 1977): Regressions from the verb region were mostly directed back to the local noun just preceding the region and not to the head noun of the subject NP, which carried the relevant number marking. The choice of regression target appeared to be independent of the needs of syntactic processing, instead following a rapidly and monotonically decreasing function of distance much like that seen across a sample of fillers which did not contain any violations. Readers did eventually make their way back to the head noun, however, as indicated by the similar mean number of passes and total time on the head and local nouns.

EXPERIMENT 3

The combination of Experiments 1 and 2 provided strong evidence about readers' responses to agreement violations and to seeming violations created by a head/local NP number mismatch. This latter result parallels the most basic result from the language production literature (e.g., Bock & Miller, 1991), but Experiments 1 and 2 only examined half of the pattern: the comparison between NP-match and NP-mismatch conditions when the head NP was singular. When the head NP was instead plural, the proportion of production errors did not increase in response to a head/local NP-mismatch. Experiment 3 examined the effect of NP mismatches for both singular and plural head NPs, replicating the grammatical conditions of the first two experiments and allowing a comparison of the complete pattern with production results. Because the different methodologies of Experiments 1 and 2 yielded essentially the same pattern in the primary region of interest, Experiment 3 used the simpler one, self-paced reading. In addition, Experiment 3 did not contain any ungrammatical items, so a replication of the NP-mismatch effect found in the first two experiments would rule out the idea that it depended on participants being sensitive to agreement violations.

The use of plural head conditions in Experiment 3 also allows us to examine two possible explanations for NP-mismatch effects. One possi-

TABLE 2

NP-Mismatch Effect Predictions of Head-Overwriting and Word-to-Word Transition Probability

Exp.	Text	Grammaticality	Match	Head number	Local number	Verb number	Head-overwriting & markedness	Transition probability
1–3	*key to the cabinet was*	Gram.	Match	S	S	S	Easy	Easy
1–3	*key to the cabinets was*	Gram.	Mismatch	S	P	S	Hard	Hard
1, 2	*key to the cabinet were*	Ungr.	Match	S	S	P	Hard	Hard
1, 2	*key to the cabinets were*	Ungr.	Mismatch	S	P	P	Easy	Easy
3	*keys to the cabinets were*	Gram.	Match	P	P	P	Easy	Easy
3	*keys to the cabinet were*	Gram.	Mismatch	P	S	P	Easy	Hard

Note. Easy and Hard refer to difficulty within each Match/Mismatch pair. Abbreviations: Exp = experiment; S = singular; P = plural.

bility, mentioned earlier, is that the effect of a head/local NP mismatch is to increase the probability of an error in computing the number of the subject NP. On such a head-overwriting approach, the number specification of the local noun can occasionally incorrectly replace the head noun's number-marking as the number-marking for the whole subject NP. In the NP-match conditions of Experiments 1 and 2, where the head and local noun were both singular, this would have no noticeable effect because the number-marking on the whole subject NP would still be singular. However, in the NP-mismatch conditions, if the local noun's specification replaces that of the head noun on the subject NP, the subject NP will be marked plural. In the grammatical conditions, when the singular verb is then encountered, it will appear to violate agreement more often in the NP-mismatch condition than in the NP-match condition, resulting in additional difficulty in the NP-mismatch condition. In the ungrammatical conditions, on the other hand, the verb is plural, and thus it will appear to agree properly more often in the NP-mismatch condition than in the NP-match condition, making the former less difficult. This head-overwriting approach can thus explain the pattern of NP-mismatch effects in Experiments 1 and 2 with essentially the same mechanism proposed for production (Bock & Eberhard, 1993; Vigliocco & Nicol, 1997). Table 2 summarizes the predicted pattern.

However, an alternative explanation for NP-mismatch effects is that they are the result of a mismatch in number between the local noun and the verb, as opposed to the local noun and the head noun. Because of English word order constraints, head nouns often immediately precede their verbs, and thus the parser will quite often encounter singular noun–singular verb and plural noun–plural verb sequences (e.g., *cabinet was*, *cabinets were*), whereas singular noun–plural verb and plural noun–singular verb sequences (e.g., *cabinets was*, *cabinet were*) will be rarer. In the first two experiments, the two more common sequences occurred in the NP-match grammatical (*cabinet was*) and the NP-mismatch ungrammatical (*cabinets were*) conditions, whereas the two rarer sequences occurred in the NP-match ungrammatical (*cabinet were*) and NP-mismatch grammatical (*cabinets was*) conditions. Thus this *word-to-word transition probability* approach might explain NP-mismatch effects as a matter of fairly simple serial association, with the parser having less difficulty processing more common sequences (see Table 2): The NP-match grammatical condition involved a more common sequence than the NP-mismatch grammatical condition, whereas the NP-match ungrammatical condition involved a less common sequence than the NP-mismatch ungrammatical condition. This explanation still requires the postulation of an independent sensitivity to grammaticality to account for the overall greater difficulty with un-

grammatical conditions, but it provides an alternative to the head-overwriting explanation of NP-mismatch effects. On a transition probability approach, the comprehension system would track head number and hence agreement essentially perfectly, but it would also be sensitive to more local inconsistencies (i.e., low-frequency word-to-word transition probabilities).

The stimuli in Experiments 1 and 2 were not sufficient to distinguish between these two possibilities, but the plural head conditions in Experiment 3 can do so. As Table 2 indicates, the transition probability explanation predicts an NP-mismatch effect for plural head nouns just as for singular head nouns because the local noun–verb transition for the plural head NP-match condition will be relatively more common (plural noun–plural verb, e.g., *cabinets were*) than the corresponding transition for the plural head NP-mismatch condition (singular noun–plural verb, e.g., *cabinet were*). If instead comprehension parallels production in failing to show an NP-mismatch effect for plural head nouns, this would be evidence against a transition probability explanation. Of course, the head-overwriting explanation alone also predicts an NP-mismatch effect for plural head nouns, but the combination of the head-overwriting explanation and the proposal that the plural is explicitly marked (Eberhard, 1997) predicts an NP-mismatch effect only for singular heads, and no effect for plural heads. This is because in the plural head case, the head is explicitly marked, making it less likely to be overwritten, and the mismatching local noun is not marked, so that it has no feature available to overwrite that of the head.[6]

Method

Participants. Fifty University of Illinois undergraduates participated for class credit or $5. To balance the number of participants across lists, two participants were excluded. One of these had relatively poor comprehension question performance across all items in the experiment (less than 87% correct), and the other had poor comprehension performance for the experimental stimuli in particular (88% correct). Thus, data from 48 participants were analyzed.

Materials and design. The experimental stimuli were the same as in Experiments 1 and 2, except that head noun number was manipulated in place of grammaticality. Thus, the head noun was either singular or plural, and the verb always agreed with the head noun in number, so all sentences were grammatical. As in Experiments 1 and 2, the local noun could be either singular or plural. An example stimulus set is shown in (6), where the four conditions, formed by crossing the factors head number and head/local NP number match, are singular head NP-match (6a), singular head NP-mismatch (6b), plural head NP-match (6c), and plural head NP-mismatch (6d). The two singular head conditions are the same as the two corresponding grammatical conditions in Experiments 1 and 2. The plausibility of the four different subject NPs, obtained in the plausibility rating study in Experiment 1, is shown for each item in Appendix A.

(6) a. The key to the cabinet was rusty from many years of disuse.

b. The key to the cabinets was rusty from many years of disuse.

c. The keys to the cabinets were rusty from many years of disuse.

d. The keys to the cabinet were rusty from many years of disuse.

The 16 experimental stimuli were placed into four lists as in Experiments 1 and 2. Each list also contained 94 other (filler) items, 60 of which were part of an unrelated experiment involving direct object versus sentential complement ambiguities. The remainder of the items incorporated a variety of structures and were always grammatical sentences.

Apparatus and procedure. The apparatus and procedure were identical to those in Experiment 1. Participants typically completed the experiment in approximately 30 min.

[6] The predictions of the transition probability explanation for plural head cases do not change if plurals are marked because head noun number and local noun number do not directly interact. To explain the Experiment 1 and 2 ungrammatical singular head cases, a plural local noun must ease processing when the verb is also plural, and this should be the case regardless of head noun number.

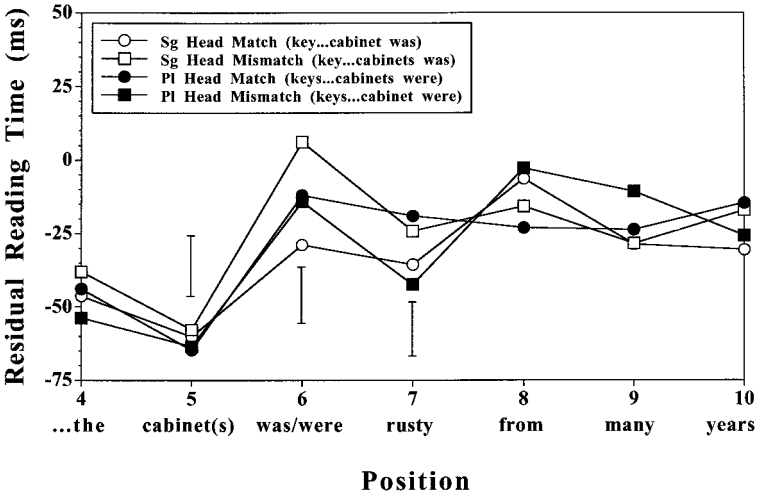

FIG. 7. Experiment 3 grand mean residual reading time by position. The singular head conditions (open symbols) correspond to the grammatical conditions in Experiments 1 and 2. Error bars show 95% confidence intervals for differences between cell means, computed from the analyses by participants.

Results

As in Experiments 1 and 2, trials on which the participant answered the comprehension question incorrectly were excluded. Comprehension performance was 95% correct for the singular head NP-mismatch condition and 94% correct for the other three conditions; these values did not differ (all Fs < 1). Residual reading times were computed and trimmed at 5 *SD* as in Experiment 1, affecting less than 0.6% of the data. Appendix B reports the raw reading times trimmed at 5 *SD*.

Figure 7 shows residual reading times by condition at each word position, and analyses are presented as in Experiment 1 for the verb (*was* or *were*, position 6) and the following word (e.g., *rusty*, position 7). The data were analyzed in 2 (head number) $\times$ 2 (head/local NP number match) ANOVAs conducted separately for participants (F_1) and items (F_2) at each word position, and cell-mean-difference confidence intervals are reported as in Experiment 1 (see Appendix C).

Verb (position 6). At the verb, there was an interaction between head number and head/local number match [$F_1(1,47) = 11.91$; $F_2(1,15) = 9.92$] as well as a head/local number match main effect (Fs > 6), but no effect of head number (Fs < 1). As in Experiment 1 at this position, the singular head NP-match condition was faster than the singular head NP-mismatch condition, but there was no NP-match effect at all for the plural head conditions.

Word following the verb (position 7). At the word following the verb, the singular head NP-match effect was still present numerically but was not reliable. However, a reverse NP-match effect for the plural head conditions appeared: The plural head NP-match condition was slower than the plural head NP-mismatch condition. This resulted in a significant head number $\times$ NP-match interaction [$F_1(1,47) = 9.11$; $F_2(1,15) = 6.62$] and no main effect of either head number or NP-match (all Fs < 2, ps $> .30$).[7]

Discussion

This experiment revealed three primary results: First, it replicated the Experiment 1 and 2 effects of head/local NP number match when

[7] We also analyzed positions 8 and 9, where the plural head NP-mismatch condition appeared to be slower than the plural head NP-match condition, suggesting a late (nonreversed) NP-match effect. However, there were no reliable effects at these positions.

the head noun and verb were singular and did so in the absence of any ungrammatical items or instructions or references to agreement phenomena. Second, in the plural head conditions, NP-match also had an effect, but the effect was the reverse of that for the singular head conditions: The plural head NP-match condition was more difficult than the plural head NP-mismatch condition. Third, the timing of the effects in the singular and plural head conditions was different: In the singular head conditions, the NP-match effect appeared immediately at the verb, whereas in the plural head conditions, the (reverse) effect did not appear until a word later, and the two plural head conditions did not differ at the verb itself. This combination of effects is compatible with that found by Bock and colleagues in production (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Eberhard, 1997), where singular head conditions showed an NP-mismatch effect, but plural head conditions did not differ in probability of a subject–verb agreement error. Of course, this leaves open the question of why the plural head conditions showed a reverse NP-mismatch effect on the word following the verb, and we return to this point below.

The combination of this experiment and the previous two also provided evidence about the mechanism by which NP-mismatch effects operate in comprehension, suggesting that such effects are the result of interference between the head noun's number-marking and the local noun's number-marking rather than the result of a (mis-)match between the local noun and the immediately following verb (i.e., word-to-word transition probability). The latter explanation predicted that the plural head conditions in this experiment should be impaired by an NP mismatch, and there was no hint of such an effect.

GENERAL DISCUSSION

The three experiments in this article provide a range of data about readers' sensitivity to real and seeming agreement violations. Despite the lack of any task beyond reading to answer a comprehension question and the fact that computation of agreement was not a priori necessary to comprehend the stimuli, readers displayed substantial disruption in response to both actual agreement violations and seeming violations created by head/local NP number mismatches, at least when the head noun was singular. This sensitivity clearly affected processing by the time the word following the verb was being read. The general pattern of the disruptions was that real violations (the ungrammatical conditions; e.g., *the key . . . were*) overall created more difficulty than seeming violations (NP-mismatch grammatical; *the key to the cabinets was*), with cases of unmistakable violation being the most severe (NP-match ungrammatical; *the key to the cabinet were*). These disruptions were reflected in reading times in all three experiments, and mismatch-induced disruptions appeared in Experiment 3 even when no ungrammatical stimuli were presented. These disruptions also appeared in regression probabilities in Experiment 2 (see also Altmann et al., 1992, and Frazier & Rayner, 1982), but the choice of regression target did not appear to be influenced at least initially by syntactic constraints.

This pattern of sensitivity, and particularly the timing revealed in Experiment 2, is consistent with the linguistic, computational, and psycholinguistic theories described earlier. Of course, these theories suggest a variety of different (and often theory-internal) mechanisms for agreement computation, and the current results do not provide strong evidence implicating any one in particular. However, we can distinguish between two general processing categories: In a *compute-on-the-fly* system, agreement features are processed by the comprehension system as they are encountered, so that the subject NP's number-marking has already been computed when the verb is processed. (The head-overwriting and word-to-word transition probability explanations for mismatch effects discussed above would both apply within this category.) Alternatively, if a *backtracking* mechanism is used to handle agreement, number-marking would only be checked after initial parsing and only when possible (e.g., when a verb overtly marked for number is encountered).[8] In English, given the paucity of check-

[8] Note that Nicol et al.'s (1997) proposed backtracking mechanism for comprehension is a compute-on-the-fly sys-

able cases noted earlier and the redundancy of agreement information with other syntactic constraints, a backtracking system might turn out to be more efficient, depending on how onerous the agreement-checking process is.

The current data are most straightforwardly handled by a compute-on-the-fly process in which both reading times and the incidence of regressions simply reflect the probability of detecting an agreement violation. The backtracking hypothesis specifically has difficulty handling the Experiment 2 regression data. Because it assumes that the comprehension system ignores agreement information until it encounters an overtly number-marked verb, the NP-match and NP-mismatch conditions should be identical with respect to agreement properties, but in Experiment 2, regressions were more frequent in NP-match than NP-mismatch conditions, and NP-match also interacted with grammaticality. The backtracking hypothesis can explain NP-match effects in first-pass *reading time* around the verb by assuming that subject NP number is computed without making any regressive saccades. However, trying to use this same assumption to explain why the probability of regressions after first passes varied creates two interacting problems: First, first passes through the verb region were much shorter (by 82 ms) when ended by a regression than when ended by a forward saccade, suggesting that readers were not spending time computing the subject NP's number before regressing. Second, and more critically, if readers did manage to compute the subject NP's number in the regression-ended trials, then first-pass times on those trials should have been affected. But the first passes ended by regressions were the cases which did not contribute to the first-pass time effects at the verb region (reliable effects appeared only when those trials were removed, in the regression-contingent analyses). Thus the number of the subject NP was probably computed prior to encountering the verb.

tem in our terms because the subject NP's number is computed as the NP is initially processed. For Nicol et al., the backtracking arises in the actual check of the verb's number against the subject NP's, while the verb is being processed.

Comprehension and Production

In addition to the evidence they provide about comprehension and related theories, of course, these results can also be compared to those from production studies, where a head/local NP mismatch creates a higher proportion of verb-number-marking errors only when the head NP is singular, not when it is plural. All three of the current experiments also showed difficulty engendered by a head/local mismatch when the head NP was singular and the verb was grammatically correct. In Experiments 1 and 3, an NP-mismatch effect for singular head NPs appeared immediately at the verb; and in Experiment 2, despite the absence of significant effects at the verb itself, a reliable NP-mismatch effect emerged at the next word.

The three experiments together also suggested that the same kind of head-overwriting explanation of NP-mismatch effects as in production (e.g., Vigliocco & Nicol, 1997) might apply as well in comprehension. Such an explanation, along with the proposed markedness of the plural relative to the singular, accounted for both the NP-match interaction with grammaticality in the first two experiments and for the lack of an NP-match effect (at the verb) for plural heads in the third experiment. An alternative, that NP-mismatch effects result from differences in the frequency of local noun–verb sequences, predicted an NP-match effect for both singular and plural heads and was not supported. Together, these patterns suggest that the comprehension and production systems call on similar mechanisms for agreement and that the markedness of the plural (Bock & Eberhard, 1993; Eberhard, 1997) applies to both systems.

Experiment 3, however, showed that the NP-mismatch effect for plural-head NPs was more complicated: At the verb, a head/local mismatch had no effect, but a surprising reverse mismatch effect (NP-match harder than NP-mismatch) appeared at the following word. The fact that the NP-mismatch effects for singular and plural heads were in opposite directions and had different onset times with respect to the appearance of the verb suggests that they may have different origins, and one possibility is that the

complexity of the discourse model required for the different subject NPs of Experiment 3 played a role. For example, in the plural head NP-match condition, two multiple-member sets (e.g., *keys* and *cabinets*) must be postulated, whereas only one multiple-member set is needed in the plural head NP-mismatch condition (*keys, cabinet*). Complexity differences are reversed for the singular head cases: The singular head NP-match condition (*key, cabinet*) requires no multiple-member sets, but the singular head NP-mismatch condition (*key, cabinets*) does require one. Although the difficulty of *constructing* these discourse representations should be spread across the words of the subject NP (and is correlated with subject NP length), the effects in the verb region might reflect the difficulty of combining the discourse representation of the subject with the content of the predicate. Critically for our stimuli, this content only appeared following the verb, because the verb itself (always the copula) provided essentially none. Thus discourse effects of this sort could conceivably explain the reversal of the plural head mismatch effect at the word following the verb. We are examining this possibility in ongoing work.

An additional point to consider with respect to similarities between the comprehension and production systems concerns the idea that the underlying computation mechanism in both cases involves discrete features and "slots"; that is, that an NP is identified as either definitely singular or plural and that production errors and comprehension difficulty result from an inadvertent overwriting process in which one value is replaced by another (e.g., the head NP's number specification is replaced by the local NP's specification). This approach straightforwardly predicts production results (in part because the responses are discrete): When incorrect overwriting occurs, producers generate the incorrect verb form. If the same mechanism is used in comprehension, then the difficulty of a condition will be a function of its *detected-error probability:* the probability of noticing that an agreement error is present in the condition, regardless of whether an error is actually present. In Experiments 1 and 2, in the NP-match grammatical condition, the subject NP will almost always be correctly identified as singular (even inappropriate overwriting will not result in an error), as will the verb, so readers will almost never think an error is present. Thus the detected-error probability will be very low. In the NP-match ungrammatical condition, however, while the subject NP will again almost always be correctly identified as singular, the verb will essentially never be misidentified as singular, so the detected-error probability will be very high (1.00 minus the detected-error probability in the NP-match grammatical condition). In the two NP-mismatch conditions, the detected-error probability will depend on the probability of the local NP's number overwriting the head NP's number: The higher this probability, the higher the NP-mismatch grammatical condition's detected-error probability, and the *lower* the NP-mismatch ungrammatical condition's detected-error probability. Thus as the probability of incorrect overwriting increases toward .50, the detected-error probabilities in the two NP-mismatch conditions will approach each other.

The concern here is that in both Experiments 1 and 2, while the NP-match grammatical condition was easier and the NP-match ungrammatical condition was harder than the other conditions, the two NP-mismatch conditions were mostly indistinguishable. A discrete slot-and-feature approach could account for this by setting the probability of inadvertent overwriting of a head NP's number by a local NP close to .50. This would straightforwardly predict that the two conditions should be similar in detected-error probability and thus difficulty. Of course, this suggests a fairly inept agreement computation system—it fails close to half the time—but more directly relevant, the data from production suggest that mismatch-induced errors are much less frequent: Bock & Miller (1991), for example, found a maximum number-marking error rate of about 25% for their singular head NP-mismatch condition, and other studies have found comparable or lower rates.

An alternative to the discrete slot-and-feature approach is to allow number features to take on continuous activation (or probability) values

during processing, as in constraint-based models of ambiguity resolution (e.g., MacDonald et al., 1994; see Trueswell, 1996, for similar treatment of verb tense markers; and Dell, 1986, and Roelofs, 1992, for conceptually similar production models). On this approach, the possibility of perceptual error and the existence of noise within the activation system make feature processing, even for objectively unambiguous words (e.g., *keys*), a matter of ambiguity resolution. For example, in most cases, when *keys* is processed in the lexicon, the plural feature will become very active, but because of occasional errors and general noise within the system, the singular possibility may also receive some small amount of support. This situation is the same as that for a strongly biased semantically ambiguous word (e.g., *boxer*): One possible meaning is very strongly preferred, but it is not the only possibility. Other words (e.g., *deer, you*, most verbs) can be more clearly ambiguous, and in addition to the uncertainty associated with activating the appropriate feature when a word is identified, potential uncertainty will also arise because later words (e.g., a local noun) must be processed while the marking of earlier words is maintained.

Because of these sources of interference and ambiguity, identifying the subject NP's number when it must be checked against the verb's will be more difficult and more error prone to the extent that its activation level is no longer clearly differentiated from that of any alternatives. Noun phrase-mismatch effects will thus arise as interference increases, and grammaticality effects will arise when the verb's number specification fails to correspond to that retrieved for the subject NP. The combination of these effects will match those described above in terms of detected-error-probability, but the competition between alternatives on individual trials provides a source of nonlinearity which can account for the similar difficulty of the two NP-mismatch conditions, despite their difference in grammaticality.

This approach can handle the Experiment 3 result that plural head nouns are insulated from NP-mismatch effects, essentially by implementing Eberhard's (1997; Bock & Eberhard, 1993) plural-markedness proposal. Instead of having both a plural feature and a singular feature which can be activated and which can compete with each other, only a plural feature would be available. When the plural feature is activated above a set threshold, the element in question is identified as plural; when the plural is not sufficiently activated, the system assumes that the element in question is singular. In such a model, all sources of interference (noise, decay, etc.) act directly on the plural feature, either driving its activation up or down.

While this description is just a sketch of an alternative to a discrete slot-and-feature approach, it does provide a potential explanation for the lack of difference between the two NP-mismatch conditions in Experiments 1 and 2, which is problematic for the discrete approach. In addition, this framework has the potential to explain ambiguity resolution and a variety of feature-processing effects, including interference and grammaticality effects, with the same mechanisms that have been used in modeling general syntactic comprehension (e.g., Stevenson, 1994) as well as word- and sentence-level effects in production (e.g., Dell, 1986; Roelofs, 1992).

Empirical and Methodological Issues

In addition to theoretical implications, the current results have several empirical and methodological consequences. First, the results of Experiments 1 and 2 bear on questions about the role of notional number in the computation of agreement in English. We found no interaction between notional number of the subject NP (distributivity) and either grammaticality or NP-match at the local noun position, at the number-marked verb, or at the following word, indicating that notional number had no measurable impact on the computation of agreement. This accords with production results for similar English materials (Bock & Miller, 1991; Vigliocco et al., 1996a), results which in turn contrast interestingly with findings in several other languages (Vigliocco et al., 1995, 1996a,b). However, Eberhard (1996) offers evidence that the appearance of distributivity effects in English production may depend on the relative con-

creteness of the speaker's number representation. This was not manipulated in our materials. Coupled with the conflicting cross-linguistic patterns in the production data and the absence of converging evidence from other work on comprehension, our null interactions obviously warrant further scrutiny. But from this first glance at how notional number affects agreement computation during reading, the conclusion must be that notional number has little power to deflect the robust feature-matching processes set in train by grammatical number.

Second, the use of both self-paced and eyetracking methodologies with the same materials in the current studies allows a direct comparison of the two. The clearest result of such a comparison is that the differences are minimal. The patterns of difficulty in the two measures were mostly identical, except that eyetracking allowed for finer-grained timing and a chance to examine a wider range of measures. Both reading times and regression probabilities in Experiment 2 revealed the same pattern as the self-paced studies.

Overall, then, the three reported experiments provide substantial data about the handling of agreement in comprehension, about the relationship between agreement in comprehension and in production, and about the timing of agreement processing. The pattern of results for singular versus plural head NPs paralleled effects in language production using identical subject NPs (e.g., Bock & Miller, 1991) and suggested that the two systems rely on closely related mechanisms for processing agreement information. In both systems there is evidence that NP-mismatch effects are mediated by processes responsible for subject-number computation, rather than by mere serial association. The time course of agreement computation during comprehension appears to be compatible with many linguistic, computational, and psycholinguistic theories, but it challenges theories that assume agreement to be checked after the fact. In particular, comprehenders in the present experiment displayed early sensitivity to both real and seeming violations of agreement. Because the agreement system of English is comparatively meager and rarely constrains the interpretation of sentences, this early sensitivity argues for a system that continuously integrates grammatical features during comprehension, irrespective of their eventual relevance to understanding.

APPENDIX A

Stimuli

The stimuli used in the experiments are listed below in their fully singular versions (singular head noun, singular local noun, singular verb). For Experiments 1 and 2, the other versions were created by varying the number of the local noun and/or the verb. For Experiment 3, the other versions were created by varying the number of the head noun and/or local noun (the verb's number-marking matched that on the head noun). Items 1–8 have a distributive interpretation in their singular head–plural local noun versions; items 9–16 are nondistributive. The plausibility ratings described in Experiment 1 for the four different versions of the subject NPs are shown in parentheses after each item (1 = plausible, 5 = implausible) in the order: singular head–singular local noun, singular head–plural local noun, plural head–singular local noun, plural head–plural local noun. The ratings shown for item 6 (*The name on the billboard . . .*) were actually collected using the fragment *The name on the sign . . .* as described in the Method section of Experiment 1.

1. The slogan on the poster was designed to get attention. (1.18, 1.06, 1, 1.06)
2. The picture on the postcard was of a village church in the south of France. (1, 1, 1, 1)
3. The mistake in the program was disastrous for the small software company. (1, 1.12, 1, 1.12)
4. The label on the bottle was a warning about the toxic effects of the drug. (1, 1.06, 1.24, 1)
5. The problem in the school was solved by firing the superintendent. (1.18, 1.12, 1.12, 1)
6. The name on the billboard was of a prominent local politician. (1.18, 1.24, 1, 1)
7. The crime in the city was a reflection of the violence in today's society. (1.06, 1.06, 1.06, 1.12)
8. The defect in the car was unknown to consumers and government regulators. (1.12, 1.23, 1.18, 1.06)
9. The door to the office was left unlocked by the cleaning service. (1, 1, 1, 1)
10. The memo from the accountant was about the delinquent tax return. (1.06, 1.18, 1.29, 1.06)

11. The check from the stockbroker was a dividend on a long-term bond. (1.06, 1.47, 1.53, 1.29)
12. The key to the cabinet was rusty from many years of disuse. (1.06, 1.18, 1.12, 1.24)
13. The letter from the lawyer was received in San Francisco in late March. (1.12, 1, 1, 1.06)
14. The entrance to the laboratory was hard to locate on the diagram. (1, 1, 1.06, 1)
15. The warning from the expert was a shock to the residents of the city. (1.25, 1, 1.06, 1.29)
16. The bridge to the island was about ten miles off the main highway. (1.59, 2, 1.47, 1.41)

APPENDIX B

Trimmed Raw Reading Times

TABLE B1

Experiment 1 Trimmed Raw Reading Time (in Milliseconds)

	Position						
Condition	4	5	6	7	8	9	10
Grammatical							
Match	297	317	324	313	324	328	329
Mismatch	299	336	361	344	329	327	342
Ungrammatical							
Match	291	311	333	389	382	345	338
Mismatch	300	326	356	361	344	337	346

TABLE B2

Experiment 2 Trimmed Raw First-Pass Reading Time (in Milliseconds)

	Position						
Condition	4	5	6	7	8	9	10
Grammatical							
Match	219	246	241	247	237	282	257
Mismatch	213	289	251	281	258	269	259
Ungrammatical							
Match	207	253	257	283	259	267	268
Mismatch	219	274	255	284	248	272	271

TABLE B3

Experiment 2 Trimmed Raw Total Reading Time (in Milliseconds)

	Position									
Condition	1	2	3	4	5	6	7	8	9	10
Grammatical										
Match	240	306	257	233	280	248	286	280	331	281
Mismatch	256	314	249	241	374	284	337	298	322	316
Ungrammatical										
Match	273	383	268	277	385	409	391	351	333	310
Mismatch	237	353	268	263	361	318	356	315	319	313

TABLE B4

Experiment 3 Trimmed Raw Reading Time (in Milliseconds)

	Position						
Condition	4	5	6	7	8	9	10
Singular head							
Match	275	288	290	293	319	300	298
Mismatch	281	300	325	310	308	305	313
Plural head							
Match	276	301	320	316	308	311	321
Mismatch	265	286	312	291	321	314	306

APPENDIX C

Condition Means and Confidence Intervals by Participants and Items

TABLE C1

Experiment 1 Residual Reading Time Means and Confidence Intervals by Participants and Items (in Milliseconds)

	Position					
Condition	5		6		7	
Grammatical						
Match	−50	−48	−12	−13	−36	−34
Mismatch	−44	−43	22	23	−1	1
Ungrammatical						
Match	−60	−60	−11	−10	40	38
Mismatch	−57	−58	9	8	13	13
CI	16	19	19	16	24	21

Note. At each position, the left and right columns show analyses by participants and items, respectively. Abbreviation: CI = 95% confidence interval for individual mean comparisons.

TABLE C2

Experiment 2 First-Pass Residual Reading Time Means and Confidence Intervals by Participants and Items (in Milliseconds)

	Position						Verb region			
Condition	5		6		7		All		NoRegr	
Grammatical										
Match	−35	−41	5	3	−18	−20	−4	−2	−9	−7
Mismatch	−17	−15	13	18	11	22	17	14	36	33
Ungrammatical										
Match	−25	−27	10	5	14	16	35	38	71	52
Mismatch	−27	−24	−1	7	4	18	19	21	29	27
CI	24	24	26	27	36	36	37	34	49	33

Note. At each position, the left and right columns show analyses by participants and items, respectively. Abbreviations: All = including all first passes; NoRegr = including only first passes not ended by regressive saccades; CI = 95% confidence interval for individual mean comparisons.

TABLE C3

Experiment 2 Total Reading Time Means and Confidence Intervals by Participants and Items (in Milliseconds)

	Position													
Condition	1		2		3		4		5		6		7	
Grammatical														
Match	−11	−10	−26	−26	−14	−18	−45	−45	−70	−79	−26	−28	−43	−36
Mismatch	2	0	−23	−16	−26	−17	−41	−36	−9	−12	4	4	7	12
Ungrammatical														
Match	30	22	66	74	−25	−15	9	30	42	37	125	131	66	70
Mismatch	−10	−19	29	31	−4	1	−1	−17	−19	−19	22	22	21	27
CI	26	29	47	43	42	42	38	44	42	39	45	47	50	39

Note. At each position, the left and right columns show analyses by participants and items, respectively. Abbreviation: CI = 95% confidence interval for individual mean comparisons.

TABLE C4

Experiment 2 Regressive Saccade Condition Means and Confidence Intervals by Participants and Items (in %)

	Regression measure			
Condition	First pass		Total	
Grammatical				
Match	6.9	4.9	5.7	6.1
Mismatch	10.5	9.7	8.7	10.4
Ungrammatical				
Match	27.3	27.1	20.3	23.9
Mismatch	12.1	11.8	11.0	14.3
CI	8.0	7.0	4.7	6.1

Note. For each measure, the left and right columns show analyses by participants and items, respectively. Abbreviation: CI = 95% confidence interval for individual mean comparisons.

TABLE C5

Experiment 3 Residual Reading Time Means and Confidence Intervals by Participants and Items (in Milliseconds)

	Position					
Condition	5		6		7	
Singular head						
Match	−62	−59	−30	−29	−37	−34
Mismatch	−57	−57	7	6	−24	−24
Plural head						
Match	−66	−63	−12	−12	−20	−17
Mismatch	−65	−61	−17	−14	−43	−42
CI	21	19	19	18	18	20

Note. At each position, the left and right columns show analyses by participants and items, respectively. Abbreviation: CI = 95% confidence interval for individual mean comparisons.

REFERENCES

Altmann, G. T. M. (1994). Regression-contingent analyses of eye movements during sentence processing: Reply to Rayner and Sereno. *Memory & Cognition,* **22,** 286–290.

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language,* **31,** 685–712.

Blackwell, A., Bates, E., & Fisher, D. (1996). The time course of grammaticality judgement. *Language and Cognitive Processes,* **11,** 337–406.

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., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology,* **23,** 45–93.

Bock, K., Nicol, J. L., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language,* **40,** 330–346.

Branigan, H. P., Liversedge, S. P., & Pickering, M. J. (1995). Verb agreement in written comprehension and production. Poster presented at the Eighth Annual CUNY Conference on Human Sentence Processing, Tucson, AZ.

Bresnan, J. (Ed.). (1982). *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.

Carpenter, P. A., & Just, M. A. (1977). Reading comprehension as eyes see it. In M. A. Just, & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 109–139). Hillsdale, NJ: Erlbaum.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* **12,** 335–359.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes,* **13,** 21–58.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition,* **30,** 73–105.

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

Deutsch, A. (1998). Subject-predicate agreement in Hebrew: Interrelations with semantic processes. *Language and Cognitive Processes,* **13,** 575–597.

Earley, J. (1970/1986). An efficient context-free parsing algorithm. In B. Grosz, K. S. Jones, & B. L. Webber (Eds.), *Readings in natural language processing* (pp. 25–33). Los Altos, CA: Morgan Kaufmann. (Reprinted from *Communications of the Association for Computing Machinery,* **13,** 94–102.)

Eberhard, K. M. (1996). Conceptual accessibility and its control of syntactic processing in sentence production. Poster presented at the 37th Annual Conference of the Psychonomic Society, Chicago, IL.

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

Ehrlich, S. F., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior,* **22,** 75–87.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language,* **25,** 348–368.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton–Mifflin.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* University of Connecticut doctoral dissertation. Bloomington, IN: Indiana Univ. Linguistics Club.

Frazier, L. (1987a). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.

Frazier, L. (1987b). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory,* **5,** 519–559.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology,* **14,** 178–210.

Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition,* **19,** 101–131.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language,* **37,** 58–93.

Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language,* **33,** 39–68.

Gernsbacher, M. A. (1991). Comprehending conceptual anaphors. *Language and Cognitive Processes,* **6,** 81–105.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown.* Unpublished doctoral dissertation, Carnegie Mellon Univ., Pittsburgh, PA.

Gibson, E., Pearlmutter, N. J., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition,* **59,** 23–59.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of

syntactic processing. *Language and Cognitive Processes,* **8,** 439–483.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior,* **20,** 417–430.

Jakubowicz, C., & Faussart, C. (1995). Agreement phenomena in the processing of spoken French. Paper presented at the Eighth Annual CUNY Conference on Human Sentence Processing, Tucson, AZ.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science,* **20,** 137–194.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review,* **87,** 329–354.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General,* **111,** 228–238.

Kail, M., & Bassano, D. (1997). Verb agreement processing in French: A study of on-line grammaticality judgments. *Language and Speech,* **40,** 25–46.

Kay, M. (1985/1986). Parsing in functional unification grammar. In B. Grosz, K. S. Jones, & B. L. Webber (Eds.), *Readings in natural language processing* (pp. 125–138). Los Altos, CA: Morgan Kaufmann. (Reprinted from D. R. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 251–278), Cambridge, UK: Cambridge Univ. Press.)

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition,* **11,** 539–550.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review,* **1,** 476–490.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review,* **101,** 676–703.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior,* **23,** 127–150.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language,* **36,** 569–587.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language,* **31,** 785–806.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language,* **34,** 739–773.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar.* Chicago, IL: Univ. of Chicago Press.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Hillsdale, NJ: Erlbaum.

Rayner, K., & Sereno, S. C. (1994). Regressive eye movements and sentence parsing: On the use of regression-contingent analyses. *Memory & Cognition,* **22,** 281–285.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes,* **4,** SI21–49.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition,* **42,** 107–142.

Sevald, C. A., & Garnsey, S. M. (1995). Safe syntax: Encapsulation of number-marking information in sentence comprehension. Paper presented at the Eighth Annual CUNY Conference on Human Sentence Processing, Tucson, AZ.

Shieber, S. M. (1986). *An introduction to unification-based approaches to grammar* (CSLI Lecture Notes No. 4). Stanford, CA: Stanford Univ., Center for the Study of Language and Information.

Stevenson, S. (1994). *A competitive attachment model for resolving syntactic ambiguities in natural language parsing* (Tech. Rep. RuCCS TR-18). New Brunswick, NJ: Rutgers Univ., Center for Cognitive Science.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language,* **35,** 566–585.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language,* **33,** 285–318.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996a). 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., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996b). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes,* **11,** 407–442.

Vigliocco, G., & Nicol, J. L. (1997). The role of syntactic tree structure in the construction of subject–verb agreement. Unpublished manuscript.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw–Hill.

(Received March 3, 1997)
(Revision received February 3, 1999)