Reading the Words Her, His, Him: Implications for Parsing Principles Based on Frequency and on Structure

CHARLES CLIFTON, JR., SHEILA M. KENNISON, AND JASON E. ALBRECHT
University of Massachusetts/Amherst

Two questionnaire studies and two experiments investigated the processing of the pronouns her and him/his. The questionnaires and the first experiment searched for effects of lexical preferences in resolving the temporary syntactic category ambiguity of the word her. Contingent frequency based on a verb's preference for a human vs an inanimate direct object demonstrably affected the final interpretation of an ambiguous her following the verb in an off-line questionnaire study. However, it did not influence on-line processing times for SPEC versus NP uses of her. This result is surprising from the perspective of current lexically based models of ambiguity resolution. Experiment 2 addressed how potential antecedents for the unambiguous pronouns him and his are identified. Both Experiment 1 and Experiment 2 showed that NP uses of personal pronouns like her and him could be processed more easily than SPEC uses of her and his. Experiment 2 showed that this difference largely reflected differences in difficulty of choosing an antecedent for the pronoun. To account for the processing of her, we proposed that the parser initially builds an underspecified syntactic representation, fixing further details and disambiguating the category of her as subsequent information becomes available. Comprehension proceeds by attempting to find an antecedent for the personal pronoun, with the preferred candidate a syntactically available within-sentence antecedent if one exists.

The English personal pronoun her is ambiguous in its syntactic category. It can be a full noun phrase (NP), as in I saw her yesterday or I watched her leave. It can alternatively be a specifier (SPEC), as in I saw her cat. The pronoun her contrasts with its masculine counterparts him and his, which lack the syntactic category ambiguity.

The present paper initially addresses the question of how the syntactic category ambiguity of her is resolved. It tests the possibility that the ambiguity of her is initially resolved by appeal to the lexical context in which it occurs. This possibility is suggested by current parsing theories that emphasize how syntactic ambiguities can be resolved by information stored with individual lexical items, especially argument-assigning heads of phrases (Ford, Bresnan, & Kaplan, 1982; MacDonald, Pearlmuter, & Seidenberg, 1994a, 1994b; Spivey-Knowlton & Sedivy, 1995; Tanenhaus, Garnsey, & Boland, 1990; Tanenhaus, Boland, Mauner, & Carlson, 1993; Tanenhaus & Carlson, 1989; Trueswell & Tanenhaus, 1989). This research was supported in part by Grant HD-18707 and Training Grant MH-07327 to the University of Massachusetts. Sheila M. Kennison is now at the University of Oklahoma, and Jason E. Albrecht is at Texas Tech University. We thank Janina Rado for suggestions she made concerning the experiments and their presentation, and Janina Rado, Sungryung Koh, Jill Lohmeier, Nayan-tara Santhi, Rich Gobeil, Erwin Okwa Foxtree, Ruth Wynta, and Andrea Arricale for assistance in collecting and analyzing the data. Portions of the data presented here were reported at the 1994 meetings of the Psychonomic Society. The full materials are available by contacting any of the authors by mail or sending e-mail to cec@psych.umass.edu. Address correspondence and reprint requests to Charles Clifton, Department of Psychology, University of Massachusetts, Amherst, MA 01003. E-mail: cec@psych.umass.edu.

1 See Boland (in press) for some initial experiments on how full ambiguities like They saw her duck are resolved, emphasizing the effects of lexical biases on the word that follows her; and see Pritchett, 1992, for theoretical discussion of how some structural factors might affect resolution of the ambiguity. Since both these lines of work address questions quite different from those eventually addressed in the present paper, we will not discuss them in detail.
haus, 1994; Trueswell, Tanenhaus, & Kello, 1993). Specifically, the present paper begins by identifying verbs that bias a following her toward NP vs SPEC usage in final interpretation, and determines whether such verbal contexts also bias initial decisions about syntactic category and syntactic structure. The research to be reported demonstrates that the type of verb bias that we examined does not affect the initial resolution of the syntactic category of her as assessed in an on-line reading task, even though it successfully affects readers’ resolution of her in an off-line questionnaire. However, it identified a general advantage of NP over SPEC uses of pronouns.

To account for this NP advantage, the paper turns to an examination of how antecedents of the personal pronouns her and him/his are identified. We propose that the apparent NP preference does not reflect a syntactic category resolution preference, but instead reflects processes of evaluating potential pronoun antecedents for their appropriateness. We present experimental evidence in support of this proposal. In doing so, we provide evidence that syntactic structure constrains the availability of potential antecedents. In particular, our experiments demonstrate that the parser honors the Binding Theory (Chomsky, 1981) in identifying potential antecedents of a pronoun.

**Syntactic Category Ambiguity Resolution**

The syntactic category ambiguity of her has structural consequences. The NP realization of the word results in a simpler syntactic structure than the SPEC of NP realization does. Assuming that a noun, possibly preceded by a specifier or determiner, is an NP (as opposed to a DP; Abney, 1987) the structure of the NP usage of her is the relatively simple \([s \cdot \cdot \cdot [np \ her] \cdot \cdot \cdot]\), while the structure of the SPEC usage is the more complex \([s \cdot \cdot \cdot [np [spec \ her] [s \cdot \cdot \cdot ]]]\). Since her is attached with fewer nodes as NP than as SPEC, theories that posit a preference to analyze a word with the smallest possible amount of additional structure (e.g., Frazier’s Minimal Attachment principle; Frazier, 1979, 1987) would predict a structural preference for the NP analysis. Note, though, that some specific versions of such theories would not make this prediction. Frazier and Rayner (1987), for instance, claim that the resolution of syntactic category ambiguity is slightly delayed, permitting structure-building to be delayed until the proper syntactic category is unambiguously determined.

Alternative accounts appeal to projection of the initial resolution of the syntactic category of her as assessed in an on-line reading task, detailed information stored with individual lexical items. Such accounts, in their most explicit form (as in MacDonald et al., 1994a), claim that reading a word such as her activates a variety of forms and structures to a degree determined by their frequency of occurrence as well as by what else is currently activated (among other factors). More highly activated forms are more likely to be represented in the final interpretation of a phrase or sentence. The SPEC use of her is approximately twice as frequent as the NP use (Francis & Kučera, 1982). Frequency-based accounts might therefore predict a preference for the SPEC usage of her.

However, “fine-grained” frequency-based accounts can appeal to the frequency with which various forms and structures are experienced in specific contexts, making the relevant frequency specific to particular lexical items or to particular contexts (MacDonald, 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1993; but cf. Gibson & Pearlmutter, 1994, for a discussion of the need for careful analysis in determining what to count in calculating relevant frequencies, and cf. Mitchell, Cuetos, Corley & Brysbaert, 1995, for critical discussion of such accounts). We raise the possibility that the frequency of use of her as SPEC vs NP may differ substantially following different verbs. In particular, we conjecture (and later provide supporting evidence) that her is more likely to be used as NP than as SPEC when it follows a verb that is commonly used with a human direct object, and that it is more likely to be used as SPEC than as NP when it follows a verb that is most commonly used with an inanimate direct object (see Table 1 for examples). Theories such as those
TABLE 1
EXAMPLES OF VERBS CHOSEN IN QUESTIONNAIRE 1, OF SENTENCE FRAGMENTS USED IN QUESTIONNAIRE 2, AND OF SENTENCE COMPLETIONS FOUND IN QUESTIONNAIRE 2

<table>
<thead>
<tr>
<th>Verbs and sentence fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Verbs that preferred human direct object</td>
</tr>
<tr>
<td>II. Verbs that had no strong preference</td>
</tr>
<tr>
<td>III. Verbs that preferred inanimate direct object</td>
</tr>
</tbody>
</table>

Sample completions (expected form marked with a √ and completion type indicated)

| I. Verbs that preferred human direct object | √(NP) They helped her move all her furniture on Friday. |
| II. Verbs that had no strong preference | √(NP) They disturbed her early in the morning. |
| III. Verbs that preferred inanimate direct object | √(NP) They painted her face for the birthday party. |

The subjects were instructed to write the first sensible completion that came to mind, without spending too much time on any one item.

Results

Each sentence completion was scored according to the type of direct object that was used: human direct object, inanimate direct object, and nonhuman animate direct object. The verbs chosen for further study were never, or nearly never, used with nonhuman animate direct objects, so will be classified in terms of the human versus inanimate contrast. Three verb categories were identified: verbs that preferred a human direct object (percentage of human object usage ranged from 79 to 100%, with a mean of 93%); verbs that preferred an inanimate direct object (human direct object usage from 0 to 13%, mean of 2.4%); and verbs that were approximately equally balanced between human and inanimate direct object usage (human direct object usage from...
13 to 79%, mean of 41%). Twenty-four verbs were chosen in each category. Examples appear in Table 1 (and all 72 verbs appear in Appendix A).

**Questionnaire 2**

**Method**

Thirty-three undergraduates at the University of Massachusetts were presented with a written list containing eight verbs of each of the three categories identified in Questionnaire 2, each in a fragment of the form “They verbed her ———.” The subject was always the plural *They*, to make it impossible for *her* to be taken as coreferential with the subject. Table 1 presents examples, and all the verbs used appear in Appendix A. Each list contained 70 filler fragment sentences in addition to the 24 verbs of present interest. There were three such lists, each counterbalanced to contain a different 24 verbs, so that each of the 72 verbs identified in Questionnaire 1 was seen by 11 of the 33 Questionnaire 2 subjects. The filler fragments contained 70 otherwise-unused verbs with a variety of different subject types (e.g., *Somebody, Nobody, The student*, etc.) and a variety of different postverbal continuations (e.g., verb particles, prepositions, determiners, quantifiers, etc.).

The subjects’ task was as it was in Questionnaire 1: To write down the first sensible completion of each fragment that came to mind.

**Results**

The percentages of NP and SPEC completions were computed for each verb category. Example completions appear in Table 1. The mean percentage of direct object (NP) uses of *her* was 81% for verbs that preferred human direct objects, 62% for verbs that had no clear preference, and 27% for verbs that preferred inanimate direct objects. These values differed sharply from one another ($F(2,69) = 26.81$, $p < .001$), and each pairwise comparison was significant at the .01 level or greater. The mean preference for NP uses was 56%. While this appeared to be greater than a presumed “chance” value of 50%, the difference was not significant ($t(71) = 1.58, p > .10$).

Questionnaire 2 provides clear support for our conjecture that verbs that prefer human direct objects will bias toward the NP usage of a following *her*, compared with verbs that do not have such a preference, at least in an off-line sentence completion task. The apparent (but nonsignificant) bias for an overall preference for the NP usage does appear to be inconsistent (from a frequency-based parsing perspective) with the overall greater frequency of SPEC usage (Kucera & Francis, 1982). However, this inconsistency may be only apparent: Given the strong effect of verb biases, the observed mean preference may depend on the particular set of verbs we included.

**Experiment 1**

Experiment 1 tested the suggestion that the different normative biases for different categories of verbs will affect on-line parsing preferences. Such an influence would be expected on the basis of theories (discussed earlier) that claim that parsing difficulty reflects relative degrees of lexically based activation.

A phrase-by-phrase self-paced reading task was used to study the difficulty of comprehending sentences with either the temporarily ambiguous pronoun *her* or its unambiguous counterpart *his/him*. The pronoun began a postverbal NP and either served as the entire direct object NP (as in the (a) forms of Table 2, which continued with an adverbial after the pronoun) or as the SPEC of the direct object NP (as in the (b) forms of Table 2, in which the pronoun was followed by a head noun). The main verb was either one of the 20 most strongly biased in favor of preferring human objects as identified in Questionnaire 1, or one of the 20 most strongly biased prefer-inanimate object verbs, or one of the 20 least biased.

---

2 If *her* were taken as coreferential with the subject, it would have to be used as SPEC, not as NP, to satisfy the binding conditions described by Chomsky (1981). This would presumably bias readers into favoring the SPEC usage.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence form</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP/ambiguous</td>
<td>The producers /VERBED/ her /frequently/ during rehearsals /of the play/ about World War II/</td>
</tr>
<tr>
<td>NP/unambiguous</td>
<td>The producers /VERBED/ him /frequently/ during rehearsals /of the play/ about World War II/</td>
</tr>
<tr>
<td>SPEC/ambiguous</td>
<td>The producers /VERBED/ her /secretary/ during rehearsals /of the play/ about World War II/</td>
</tr>
<tr>
<td>SPEC/unambiguous</td>
<td>The producers /VERBED/ his /secretary/ during rehearsals /of the play/ about World War II/</td>
</tr>
</tbody>
</table>

Note. Division into presentation regions indicated by /; disambiguating region in boldface. Sample “prefer human object” verb = harass; sample “no preference” verb = hurry; sample “prefer inanimate object” verb = prepare.

forms (types I, III, and II in Table 1, respectively).

Consider the predictions that a lexically based theory would make for sentences containing the ambiguous her. Questionnaire 2 demonstrated the existence of a preference to analyze this word as a human direct object NP when it follows a verb that prefers a human direct object, but as a SPEC of NP when it follows a verb that prefers an inanimate direct object. If these preferences, which presumably reflect frequency of experience, guide the resolution of ambiguities encountered when parsing a sentence, then her should initially be analyzed as NP when it follows a human-preference verb but as SPEC when it follows an inanimate-preference verb. Later material in the sentence (the disambiguating region, boldfaced in Table 2) can be either consistent or inconsistent with this initial analysis. When it is inconsistent, it will force a reanalysis, disrupting reading.

Such a theory would predict relatively fast reading of the disambiguating region (and perhaps following regions if there is spillover) in the NP/Prefer Human Object and the SPEC/Prefer Inanimate Object conditions and relatively slow reading in the SPEC/Prefer Human Object and the NP/Prefer Inanimate Object conditions. Reading time would presumably be intermediate in the No-Preference conditions. To the extent that reading time differences reflect errors of initial analysis of the word her and the difficulty of revising these analyses, they should disappear in the unambiguous cases with the pronoun his or him.³

Method

Materials. A total of 60 sets of 12 sentences each were constructed. One set is illustrated in Table 2. Each sentence had a plural subject and contained an NP beginning with a pronoun her/his/him) immediately after the main verb. The pronoun was immediately followed by disambiguating material, either an adverbial (for the NP usage of the pronoun) or a head noun (for the SPEC usage). The disambiguating material was closely matched in length between conditions, averaging 7.4 characters when it was an adverb that disambiguated the pronoun to NP and 7.7 characters when it was a noun that disambiguated the pronoun to SPEC.

Each sentence set had 12 sentences: 4 using

³ Although the frequency-based theories cited in the text have been developed as theories of ambiguity resolution (theories of how conflict among different alternative analyses is resolved), one could develop a theory that claims that unpreferred interpretations are read slowly, regardless of the existence of competing alternatives. As noted by a reviewer of this paper, such a theory would predict slow reading time for SPEC-N as well as for NP uses of her following a prefer-inanimate verb, since the SPEC-N phrase is animate (and the theory would presumably predict fast reading for all animate phrases following a prefer-animate verb). This was not observed; see Table 3.
one of the 20 most biased prefer-human verbs, 4 using one of the 20 most biased prefer-inanimate verbs, and 4 using 20 of the most equi-biased verbs. The mean human-object biases from Questionnaire 1 are reported in Appendix A for the verbs that were used in Experiment 1. The mean percentages of NP usage of her from Questionnaire 2 were 85, 62, and 25% for the actually used verbs with prefer-human bias, no bias, and prefer-inanimate bias, respectively ($F(2,69) = 26.8, p < .001$). The 4 sentences used for each verb were those described by the factorial combination of her versus his/him (ambiguous vs unambiguous) and NP vs SPEC. Each triple of verbs was used in three sets of sentences. A list was constructed with these 60 experimental sentence sets together with 44 filler items, 24 of which examined a reduced relative clause ambiguity and 20 of which were of a variety of forms. Twenty of the experimental sentences and 36 of the fillers were followed by a two-choice wh-question to ensure accurate comprehension. A practice list of six items (four of which were questioned) was also constructed.

Subjects and procedure. Sixty undergraduates at the University of Massachusetts were tested individually in sessions of just over half an hour. After receiving instructions and being trained with the practice list, each subject read an individually randomized list of 104 sentences (60 experimental and 44 fillers). Each received just 1 sentence from each of the 60 experimental sets. Full counterbalancing procedures were used in assigning sentences to subjects, so that each subject read each of the 60 verbs just one time and read 5 different sentences in each of the 12 conditions. Further, across the 60 subjects, each set of sentences was tested equally often in each of the 12 conditions defined by the factorial combination of verb bias, ambiguity (her vs him/his), and final interpretation (NP vs SPEC).

A phrase-by-phrase self-paced moving window procedure was used (Kennedy & Murray, 1984). At the beginning of a trial, underscore marks appeared where each word was to appear. Pulling a lever with the right hand displayed the first presentation region of the sentence, and successive pulls of the lever caused this display to revert to underscore marks and the next region to appear. The time taken to examine each successive region was recorded. The question, if there was one, appeared on the computer monitor immediately after the last region of the sentence was read and was answered by pulling the lever under the correct answer on the monitor. Accuracy feedback was provided by displaying the word ERROR for 500 ms when a mistake was made.

Results

Question-answering accuracy exceeded 90% in each of the 12 conditions and will not be discussed further. Reading times (eliminating times less than 100 ms and greater than 2000 ms, 0.4% of all responses in the regions from the verb on) are presented in Table 3 and (averaged over verb bias) in Fig. 1. Reading times are expressed in terms of milliseconds taken to read each region, unadjusted for length of region. Since the contrasts of interest compare regions that are identical to one another across conditions (or in the case of the disambiguating region, regions that are closely matched in length), it was not necessary to adjust for differences in region length. However, analyses in which a linear regression equation was used to adjust statistically for length differences (Clifton & Ferreira, 1987; Trueswell, Tanenhaus, & Garnsey, 1994) yielded the same conclusions as the analyses reported here.

The reading times were analyzed using analyses of variance (with the factors verb bias, ambiguity (her vs him/his) and NP vs SPEC resolution) at each of several presentation regions of interest, as well as an analysis of variance that added three critical regions as a factor. In addition, correlational analyses that treated verb bias as a continuous factor were conducted.

No differences in reading time approached significance in the region containing the pronoun. In the disambiguating region, mean reading time was less for NP than for SPEC sentences, 528 vs 549 ms ($F(1,59) = 9.44, p < .005$; $F(2,59) = 4.70, p < .05$). No


`TABLE 3

MEAN READING TIMES (MS) BY REGION, EXPERIMENT 1`

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subj</th>
<th>Verb</th>
<th>Pro</th>
<th>Disam</th>
<th>Disam +1</th>
<th>Disam +2</th>
<th>Disam +3</th>
<th>Mean D, D + 1, D + 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ambiguous her conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Her-NP</td>
<td>708</td>
<td>620</td>
<td>515</td>
<td>534</td>
<td>667</td>
<td>607</td>
<td>731</td>
<td>603</td>
</tr>
<tr>
<td>Her-SPEC</td>
<td>722</td>
<td>613</td>
<td>507</td>
<td>556</td>
<td>686</td>
<td>612</td>
<td>725</td>
<td>618</td>
</tr>
<tr>
<td>No preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Her-NP</td>
<td>731</td>
<td>607</td>
<td>494</td>
<td>523</td>
<td>649</td>
<td>606</td>
<td>687</td>
<td>593</td>
</tr>
<tr>
<td>Her-SPEC</td>
<td>698</td>
<td>592</td>
<td>505</td>
<td>563</td>
<td>682</td>
<td>644</td>
<td>728</td>
<td>630</td>
</tr>
<tr>
<td>Prefer inanimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Her-NP</td>
<td>720</td>
<td>601</td>
<td>504</td>
<td>528</td>
<td>689</td>
<td>628</td>
<td>742</td>
<td>615</td>
</tr>
<tr>
<td>Her-SPEC</td>
<td>708</td>
<td>567</td>
<td>497</td>
<td>531</td>
<td>675</td>
<td>631</td>
<td>717</td>
<td>613</td>
</tr>
<tr>
<td><strong>Unambiguous his/him conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Him-NP</td>
<td>688</td>
<td>617</td>
<td>513</td>
<td>526</td>
<td>640</td>
<td>612</td>
<td>715</td>
<td>593</td>
</tr>
<tr>
<td>His-SPEC</td>
<td>705</td>
<td>585</td>
<td>502</td>
<td>544</td>
<td>661</td>
<td>616</td>
<td>702</td>
<td>607</td>
</tr>
<tr>
<td>No preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Him-NP</td>
<td>710</td>
<td>585</td>
<td>508</td>
<td>517</td>
<td>647</td>
<td>593</td>
<td>700</td>
<td>586</td>
</tr>
<tr>
<td>His-SPEC</td>
<td>716</td>
<td>593</td>
<td>507</td>
<td>545</td>
<td>688</td>
<td>614</td>
<td>710</td>
<td>616</td>
</tr>
<tr>
<td>Prefer inanimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Him-NP</td>
<td>723</td>
<td>602</td>
<td>505</td>
<td>536</td>
<td>665</td>
<td>604</td>
<td>728</td>
<td>602</td>
</tr>
<tr>
<td>His-SPEC</td>
<td>693</td>
<td>592</td>
<td>500</td>
<td>552</td>
<td>708</td>
<td>644</td>
<td>730</td>
<td>635</td>
</tr>
</tbody>
</table>

Interactions or effects involving verb bias approached significance (largest $F$ was 1.59, $p > .20$).

The same basic pattern held up in the region following the disambiguating item. Reading was faster for NP than for SPEC sentences, 660 vs 684 ms ($F1(1,59) = 4.12, p < .05$; $F2(1,59) = 7.08, p < .02$). No other effect or interaction approached significance. The pattern persisted even two regions past the disambiguating item, with continued fast reading for sentences with NP ($F1(1,59) = 6.35, p < .02$; $F2(1,59) = 5.0, p < .03$) and no other significant effects. No effects approached significance in the third region past the disambiguating item.

Combining regions D, D + 1, and D + 2 in a single analysis provided essentially the same results. In addition to a highly significant but uninteresting effect of regions, the effect of NP vs SPEC was significant ($F1(1,59) = 15.79, p < .001$; $F2(1,59) = 8.68, p < .005$). No other effect was significant beyond the .20 level.

A close examination of Table 3 suggests that the relative difficulty of SPEC readings may have been reduced in the “Prefer Inanimate/Ambiguous Her” condition. If this were a reliable effect, it could be taken as support for a reduced preference for the NP usage when her follows a verb that is frequently used with an inanimate direct object, as predicted by a frequency-based parsing account. However, the effect is not merely unreliable. A detailed examination of reading times from Experiment 1 and preferences from Questionnaires 1 and 2 provides a fuller picture of the relationship between preference and type of direct object. There was a substantial correlation between animacy preferences (Questionnaire 1) and NP usage (Questionnaire 2) in the off-line measures ($r = .749$, $p < .001$).
p < .001, for the 60 verbs that were actually used in Experiment 1; $r = .685$ for all 72 items), as can be seen in the top panel of Fig. 2, which presents the data for the 60 actually used verbs. However, there was little or no correlation between animacy preferences and the difference in reading time for SPEC and NP uses of her, summed (for each of the 180 sentences used in Experiment 1) over the disambiguating region and the two following regions, $r = .029, p > .40$; lower panel of Fig. 2). (The similar correlations for each of these three regions taken separately never exceeded .08.) The relationship between type of direct object and NP usage (top panel) shows that verb bias guided performance in these offline tasks. In contrast, the lack of relationship between preferred type of direct object and the SPEC–NP difference in reading time (lower panel) suggests no systematic effect of preference on reading time. There is substantial variance in these reading time differences, associated (we presume) with differences in plausibility and other factors, but not with the measured verb biases. Comparing the two regression analyses, it seems clear that the lexical factors that affect off-line interpretive biases do not affect on-line reading time differences.

Discussion

Experiment 1 demonstrated a consistent and persistent advantage for sentences in which the pronoun was used as an NP over sentences in which it was used as SPEC. This advantage was not related to off-line interpretive preferences.

The NP advantage: Further evidence. Before turning to an interpretation of the Experiment 1 results, we will briefly address the
termed “cumulative region reading time” by Brysbaert and Mitchell (1996), “regression path durations” by Konieczny, Hemforth, and Voelker (1994), and “go-past” or “forward pass” durations in our laboratory. This measure, which has proven in some research to be a very sensitive indicator of some form of disruption of sentence processing, is the total sum of fixation durations from first entering a region to first going past it, including time spent rereading earlier regions. Considering the region containing the pronoun and the following word, this measure averaged 722 ms for NP uses of the pronoun and 808 ms for SPEC uses ($F_{1}(1,23) = 5.13, p < .05$; $F_{2}(1,19) = 3.52, p < .08$). No effects involving ambiguity (her vs him/his) were significant, although the go-past measure showed a numerical tendency for a larger difference between SPEC and NP for the unambiguous him/his form than for her. The eyetracking results are less than robust, presumably because of the use of only one-third as many items as in the self-paced reading experiment, but do in general provide support from another methodology for the difficulty of SPEC usage.

The eyetracking results are less than robust, presumably because of the use of only one-third as many items as in the self-paced reading experiment, but do in general provide support from another methodology for the difficulty of SPEC usage observed in Experiment 1. Further evidence that the NP advantage does not simply reflect a demand effect of presenting the pronoun by itself will be presented in Experiment 2. This experiment demonstrates that the NP advantage disappears when the presence of a possible antecedent for his is manipulated, suggesting that the advantage actually reflects referential effects.

Does the NP advantage reflect a parsing preference? The main point of Experiment 1 was to test lexically based frequency accounts of parsing preferences, not to identify an overall advantage of NP or SPEC usage of her. It provided no evidence for such frequency-based accounts, and in fact appears to limit their generality. Questionnaire 2 had demonstrated that the extent to which a verb prefers a human vs an inanimate direct object sharply biases sentence completion preferences in an off-line task. If these sentence completion preferences reflect frequency of occurrence (or, more immediately, some measure of...
strength of different structures in memory), then they could have affected the resolution of the ambiguity of the word *her*. They did not. There was no sign that NP uses were favored following a human-preference verb or that SPEC uses were favored following an inanimate-preference verb.

It is important to acknowledge that the type of verb bias manipulated in Questionnaire 2 and in Experiment 1 differs from the type of verb bias manipulated by MacDonald (1994), Trueswell et al. (1993), and others. These researchers varied the frequency with which verbs occur with particular syntactic complements, for example, NP or tensed S complement. We held this variable essentially constant across our conditions. A verb usage preference questionnaire (conducted at the University of Illinois, using the procedures described in Connine, Ferreira, Jones, Clifton, & Frazier, 1984) indicated that the mean frequency of NP complement (direct object) usage was 84% for our prefer-human verbs, 76% for our no-preference verbs, and 79% for our prefer-inanimate verbs. The type of verb bias we did manipulate involved the semantic type of noun that most commonly appeared as the direct object. Finding that such a manipulation influenced subjects’ choices in Questionnaire 2 demonstrates that they do possess and can use knowledge of the semantic preferences of verbs. The failure to observe on-line effects of this knowledge in Experiment 1 disconfirms a strong lexically based position in which all lexically based information can affect initial parsing decisions. It does not disconfirm other possible positions that claim that parsing will be influenced only by certain types of information, such as information about syntactic as opposed to semantic category, or “coarse-coded” lexical information (Mitchell et al., 1995).

In addition to being inconsistent with current lexically based parsing models, the current data sit uneasily with a depth-first, single-analysis position in which a mechanism like a “thematic processor” (Frazier, 1987) serves the process of reanalyzing wrong initial analyses (cf. Fodor & Ferreira, in press). If the parser initially makes the wrong analysis of an input, by (for instance) taking a SPEC usage of *her* to be an NP usage, a mechanism that has access to information about semantics and plausibility (and perhaps frequency) presumably performs the needed reanalysis (Frazier, 1987). This mechanism could well be sensitive to the differences between the verb classes used in Experiment 1 and reanalyze in favor of SPEC usage more quickly for verbs that have stronger SPEC preferences off-line. It did not seem to do so.

We must raise the question of whether the NP advantage observed in Experiment 1 actually does reflect initial parsing decisions, as we have been assuming. There are reasons to believe that they do not. The fact that the across-the-board advantage of NP personal pronouns was seen for the unambiguous *him* (vs *his*) as well as for the ambiguous *her* strongly suggests that the advantage cannot be attributed to ambiguity-resolution processes, whether guided by verb biases or not.

Consider the possibility that the parser neither makes a single initial commitment to the syntactic analysis of the ambiguous pronoun *her* nor immediately begins the process of attempting to satisfy various constraints (e.g., frequency) on possible analyses. The parser might instead analyze the phrase-initial pronoun *her* (and possibly *his* or *him*) in an underspecified fashion. That is, some but not all possible syntactic structure involved in the grammatical analysis of the pronoun is created. The parser identifies the pronoun as a constituent of an NP and projects an NP above it. No decision is made as to whether the pronoun is the head of the NP. When the next word is a noun or begins an N’, as was the case in the SPEC conditions of Experiment 1, it is incorporated into the NP containing the pronoun, following the Late Closure strategy (Frazier, 1987). The pronoun *her* is then analyzed as SPEC and the noun (or N’) is taken as the head (or a projection of the head) of the NP. However, when the word following the pronoun is an adverbial that signals that the NP is complete, the NP is closed and the pronoun is taken as the NP. Each change is
minimal, merely the specification of the values of syntactic category features (cf. the discussion of the Minimal Revisions Strategy in Frazier, 1990).

Underspecification accounts with greater or lesser generality have been advanced by several parsing theorists (e.g., Frazier & Clifton, 1996; Gorrell, 1995; Marcus, Hindle, & Fleck, 1983; Sturt & Crocker, 1996; Weinberg, 1993), and Frazier and Rayner (1987) have made a related proposal that the parser delays in resolving a noun/verb lexical category ambiguity (e.g., trains as noun or verb). Such proposals do seem to be a promising direction for the development of structurally based parsing theories (and it is possible that similar proposals could be developed within a lexically based parsing framework by restricting the common assumption that all pertinent information is used to activate syntactic analyses). However, we must note that the underspecification account is provisional and that we have provided no direct evidence for it.

Discourse-based accounts of SPEC difficulty. We must also note that the underspecification account does not explain why the NP analysis was easier to process than the SPEC analysis in Experiment 1. We will develop, and then test, an account of the difficulty of the SPEC analysis that appeals to discourse or referential factors.

This account makes the uncontroversial assumption that a pronoun prefers to find its antecedent within the current discourse, rather than introducing a new referent into the discourse. In the sentences used in Experiment 1, when the pronoun is SPEC (her or his), there is one syntactically possible antecedent, the sentence subject. However, in Experiment 1, the pronoun was always singular, but the subject was always plural. If the sentence subject is taken as the antecedent of the SPEC usage of the pronoun, the analysis must be given up, because its number is wrong. Constructing and then rejecting this analysis may well slow reading time. When the pronoun is NP (her or him), on the other hand, there is no syntactically acceptable within-sentence antecedent. The sentence subject is not a syntactically possible antecedent. It is in the same governing category as the pronoun, so a reflexive pronoun (herself/himself) is needed to satisfy the Binding Theory (Chomsky, 1981). If the Binding Theory blocks considering syntactically impermissible NPs as the antecedent of the pronoun, the NP condition will not suffer from the difficulty of making and then rejecting the subject NP as antecedent. We suggest that the process of considering the SPEC as antecedent and then rejecting it because of its number clash is more disruptive than the simple lack of any within-sentence antecedent for the pronoun.4

We find this account interesting and attractive. However, recent work by Badecker and Straub (1994) and Straub and Badecker (1994) provides an argument against it. These researchers used a self-paced reading task to show slow reading time following a pronoun that agreed in gender with an antecedent that is structurally unavailable according to the Binding Theory, compared with a pronoun that clashed in gender. Thus, for example, reading time in a region somewhere after the pronoun was slowed in “Bob thinks that John will give him a better cut of venison next year” compared with “Bob thinks that Joan will give him a better cut of venison next year” (the putatively unavailable noun is boldfaced for expository reasons). Badecker and Straub suggest that reading is slowed when both nouns have the same gender because gender does not immediately permit a decision between them. But this claim presupposes that both Bob and John are initially considered as potential antecedents of the pronoun him, so that a decision between them is necessary. According to this claim, the initial analysis of a pronoun violates the Binding Theory. If the claim is correct and applies beyond the particular experimental manipulations used by Badecker and Straub, it means that the subject NP in our Experiment 1 should have been considered as an antecedent for both NP and SPEC uses of the pronoun. If

4 We thank J. Radó for suggesting this account.
so, then our claim that considering a wrong-number antecedent disrupts reading cannot account for any difference between NP and SPEC usages; both would have resulted in disruption. We explore this line of reasoning in Experiment 2.

**EXPERIMENT 2**

Experiment 2 focused on referential effects rather than syntactic category ambiguity resolution. Therefore, it used the unambiguous his/him sentences from Experiment 1, modified as described below. It directly contrasted cases in which the subject NP was inappropriate as the antecedent of the pronoun (because of number, as in Experiment 1) with cases in which it was the same in number as the pronoun (see Table 4). Following our account in which the subject of a sentence is first taken and then rejected as the antecedent of a SPEC pronoun (but not of an NP pronoun), the difficulty of SPEC sentences should disappear when the subject is the same in number as the pronoun. In this case, the subject is presumably taken, both initially and finally, as the antecedent of the pronoun, and no time-consuming reanalysis is needed. On the other hand, if Badecker and Straub’s proposal (1994) is correct and the Binding Theory does not determine what phrases are considered as antecedents of a pronoun, both NP and SPEC uses should be affected by the number of the subject. If the relative difficulty of SPEC uses persists when the subject is the same in number as the pronoun, then some account other than the antecedent-clash account we have advanced must be invoked.

**Method**

**Materials.** Thirty-two sentence sets were constructed, based on the sentences used in Experiment 1. One sample set appears in Table 4. Twenty of the sentences used the verbs in the no-bias condition of Experiment 1, while the remaining 12 used some of the least biased of the prefer-human and prefer-inanimate verbs (6 of each). These 32 experimental sentences appeared in four conditions, defined by the factorial combination of subject number (plural, as in Experiment 1, vs singular, to agree with the pronoun) and pronoun use (NP vs SPEC).

All the experimental sentences used the unambiguous pronoun his/him, since we were no longer looking for effects of syntactic category or phrase structure ambiguity resolution. However, to permit any ambiguity-resolution strategies that might have been used in Experiment 1 to operate in Experiment 1 as well, 16 more verbs from Experiment 2 (8 prefer-human and 8 prefer-inanimate) were used in filler sentences, all with the pronoun her used either as NP or SPEC and with either a singular or a plural subject. There were 36 total filler sentences, the 16 just mentioned plus 20 of a variety of forms, taken from Experiment 1.

Questions were constructed to follow 16 of the experimental items and 20 of the filler items. These required subjects to chose one of two phrases to complete a sentence. For

### Table 4

<table>
<thead>
<tr>
<th>NP usage of pronoun</th>
<th>SPEC usage of pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plural subject</td>
<td>Plural subject</td>
</tr>
<tr>
<td>The supervisors/paid/him/yesterday/to finish/typing the manuscript.</td>
<td>The supervisors/paid/his/assistant/to finish/typing the manuscript.</td>
</tr>
<tr>
<td>Singular subject</td>
<td>Singular subject</td>
</tr>
<tr>
<td>The supervisor/paid/him/yesterday/to finish/typing the manuscript.</td>
<td>The supervisor/paid/his/assistant/to finish/typing the manuscript.</td>
</tr>
</tbody>
</table>
example, the question for the sentence in Table 4 was: He/His assistant finished (a) the filing (b) the typing.

Subjects and procedure. Forty-four University of Massachusetts undergraduates were tested in individual sessions. The procedure was essentially the same as in Experiment 1. Each subject read 8 sentences in each of the four experimental conditions, and counterbalancing procedures were used so that across all 44 subjects each sentence set was seen equally often in each condition.

Results

Reading times (after eliminating times less than 100 or greater than 2000 ms, 0.8% of all responses from the verb to the end of the sentence) were averaged and appear in Fig. 3. We will consider times pooled over the regions from the pronoun (the disambiguating region, in this case) through the third following region, and then discuss the regions individually. Over the four regions beginning with the pronoun, the interaction between NP vs SPEC and plural vs singular was significant ($F_{1}(1,43) = 8.03, p < .01; F_{2}(1,31) = 4.72, p < .04$). Reading time for the plural-SPEC condition, 630 ms, was slower than in the other three conditions, averaging 590 ms and varying from 586 through 594 ms. The slow times in the singular-SPEC condition also resulted in significance for each main effect ($p < .03$ in all cases), and there was a significant but uninteresting main effect and several significant interactions involving regions.

The critical interaction of NP vs SPEC and plural vs singular appeared numerically in each region from the pronoun on, and was significant or approached significance in each region apart from the second region following the pronoun ($p$ values of .02, .12, .90, and .06 in the by-subjects analyses of the four regions).
starting with the pronoun; .06, .03, .31, and .09 in the by-items analysis). In each case, reading times for the singular-SPEC condition were fast compared with the plural-SPEC condition, while the singular and plural NP conditions did not vary. However, it is of some interest to note that the singular-SPEC condition was fast compared with all three other conditions in the region of the pronoun, while the plural-SPEC condition was slow compared with the other conditions in the remaining regions. As was the case in the unambiguous him/his conditions of Experiment 1, the penalty of the plural-SPEC condition relative to the plural-NP condition did not appear at the pronoun itself, but only in the following regions, suggesting that the effects of pronoun-antecedent-finding difficulty are somewhat delayed (cf. Ehrlich & Rayner, 1983).

Discussion

The data from the four regions beginning with the pronoun strongly suggest that the subject was considered as the antecedent of the pronoun only in the SPEC case, not in the NP case. In the SPEC case, reading was rapid when the number of the subject made it an appropriate antecedent and slow when the number of the subject made it inappropriate. In the NP case, number had essentially no effect. It appears reasonable to conclude that our readers did, at least initially, act as if their parsing decisions were constrained by the principles of the Binding Theory in that they took the subject NP as a candidate antecedent only of a pronoun in a different binding domain (the SPEC case). When they did this, reading was disrupted just in case the number of the subject was inappropriate for the pronoun. No disruption was observed when the binding Theory would have it that the subject is an inaccessible antecedent for the pronoun.

The results of Experiment 2 show that at least some of the difficulty of the SPEC sentences in Experiment 1 can be traced to the disruption caused when an accessible antecedent of a pronoun clashes with the pronoun in number. This suggests that our readers obey the principles of binding in identifying possible antecedents of pronouns, and the observed disruption suggests that they treat number as a filter on the appropriateness of the accessed antecedents.

Conclusions

The present results suggest some interesting extensions of structure-based parsing theories. Experiment 1 disconfirmed the predictions of parsing theories that claim detailed information about the frequency of usage of specific lexical items determines syntactic analyses. The preference for SPEC vs NP uses of her was shown in an off-line task (Questionnaire 2) to be dependent on the frequency with which a preceding verb preferred a human vs an inanimate direct object. However, on-line reading times did not show a similar dependency. These results demonstrate that differences in frequency of usage do not always determine initial parsing decisions, even when they influence the final interpretation of sentence fragments in, for example, an off-line sentence completion task.

On a more positive note, we proposed that the advantage of NP uses in Experiment 1 can be understood in terms of structurally guided parsing processes that can initially create underspecified syntactic structures. We suggested that when an ambiguous her is encountered, the parser creates only an NP category to contain her, and does not explicitly identify the syntactic category of her nor attach it explicitly as the SPEC vs the head of the NP. In the experimental conditions we studied, the category of the following word (assuming the operation of the structural parsing principle of Late Closure) permits easy and accurate specification of the syntactic category of her.

The difficulty of the SPEC analysis in Experiment 1 cannot, under this account, be attributed to structural parsing principles. In fact, the observation that equivalent effects were observed for the ambiguous pronoun her and the unambiguous pronouns his and him demonstrates that it should not be accounted for in terms of ambiguity resolution principles.

We suggested an account in terms of the evaluation of available antecedents for the pro-
noun. A within-clause antecedent is grammatically possible (according to the Binding Theory) for the SPEC use of her (and his) but not the NP use. However, in our sentences, the available antecedent (the sentence subject) clashed in number with the pronoun. We claimed that resolving this clash slowed reading time for SPEC uses of the pronoun. We further proposed that readers honor the Binding Theory in identifying potential antecedents of pronouns and thus did not consider any within-sentence antecedent for the NP uses of her (or him). This proposal contradicts claims made by Badecker and Straub (1994) but is consistent with data presented by Nicol and Swinney (1989) and received support from our Experiment 2. This experiment demonstrated that the initial disruption of reading the SPEC pronoun his disappeared when the sentence subject agreed in number with the pronoun, but that time to read the NP him was not initially affected by number agreement.

Pulling back from specific questions of processing personal pronouns, we suggest that the present data demonstrate that not all parsing phenomena will yield easily to accounts in terms of the frequency with which lexical items are used in particular syntactic structures. Other data gathered in our laboratory indicate that various lexical properties of verbs may not play an important role in initial decisions about the syntactic category of their complements. Kennison (1994) showed that differences among verbs like walk, force, and hurry do not affect an on-line measure of the difficulty of resolving the syntactic category ambiguity of to in a following phrase in which the word is used as a preposition (e.g., hurry the teenagers to school) vs. as an infinitive marker (e.g., hurry the teenagers to catch the bus), even though one might have expected that a verb that permitted only one usage of to would result in less difficulty than a verb that left the category of to ambiguous. Kennison (1995) showed that differences in the frequency with which verbs occurred with NP vs. S complements did not affect reading time for an unambiguous S complement (e.g., The waiter insisted [or confirmed] that the reservation was made by a woman), even though one might have expected that the preferred usage of that would have been as a demonstrative, not as a sentence complementizer, after a verb that prefers NP complements. We claim that parsing theories that propose that structural preference principles play the central role in the initial creation of syntactic analyses continue to provide the best existing account of the full range of parsing phenomena, but must be supplemented by structurally based accounts of anaphor resolution (and, one assumes, vari-

**APPENDIX**

**Verbs Selected in Questionnaire 1 and used in Questionnaire 2 and Experiment 1**

All of these verbs were used in Questionnaire 2. Those listed below the line of asterisks were omitted from the self-paced experiment. Listed with each verb is the probability of occurring with a human direct object.

<table>
<thead>
<tr>
<th>Prefer-human direct object</th>
<th>Prefer-inanimate direct object</th>
<th>Equi-biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.88 convinced 0.04 needed 0.46 paid</td>
<td>0.88 helped 0.00 retrieved 0.63 questioned</td>
<td>0.96 complimented 0.00 mentioned 0.42 criticized</td>
</tr>
<tr>
<td>0.96 helped 0.00 burned 0.71 confused</td>
<td>0.92 disapproved 0.04 prepared 0.42 hurried</td>
<td>0.84 amused 0.00 lost 0.25 obeyed</td>
</tr>
<tr>
<td>0.92 harassed 0.04 studied 0.17 analyzed</td>
<td>0.92 terrified 0.00 inspected 0.79 ridiculed</td>
<td>0.92 deceived 0.13 hunted 0.34 dressed</td>
</tr>
<tr>
<td>1.00 frustrated 0.00 publicized 0.42 introduced</td>
<td>0.96 invited 0.00 searched 0.54 disturbed</td>
<td>0.84 embarrassed 0.00 painted 0.46 monitored</td>
</tr>
<tr>
<td>0.96 invited 0.00 publicized 0.42 introduced</td>
<td>0.96 assaulted 0.00 searched 0.54 disturbed</td>
<td>0.84 flattered 0.00 viewed 0.67 influenced</td>
</tr>
<tr>
<td>1.00 frustrated 0.00 studied 0.17 analyzed</td>
<td>1.00 interviewed 0.04 measured 0.50 visited</td>
<td>0.92 threatened 0.00 inspected 0.29 documented</td>
</tr>
<tr>
<td>0.96 assaulted 0.00 searched 0.54 disturbed</td>
<td>1.00 interviewed 0.04 measured 0.50 visited</td>
<td>0.88 impressed 0.00 submitted 0.45 recommended</td>
</tr>
<tr>
<td>1.00 frustrated 0.00 studied 0.17 analyzed</td>
<td>0.84 amazed 0.04 enhanced 0.42 satisfied</td>
<td>0.84 amazed 0.04 enhanced 0.42 satisfied</td>
</tr>
<tr>
<td>0.92 complimented 0.00 mentioned 0.42 criticized</td>
<td>1.00 astonished 0.04 signed 0.59 admired</td>
<td>Mean 92.80 2.25 43.95</td>
</tr>
<tr>
<td>0.96 complimented 0.00 mentioned 0.42 criticized</td>
<td>Range 0.84–1.00 0.00–0.13 0.13–0.79</td>
<td>Mean 92.79 2.42 40.83</td>
</tr>
<tr>
<td>0.96 complimented 0.00 mentioned 0.42 criticized</td>
<td>Range 0.79–1.00 0.00–0.13 0.13–0.79</td>
<td>Range 0.79–1.00 0.00–0.13 0.13–0.79</td>
</tr>
</tbody>
</table>

Note: The verb “complimented” inadvertently appeared twice in Questionnaire 2 and the verb “inspected” appeared twice in the self-paced experiment.
ous other theories) to account for the full range of sentence-comprehension phenomena.

REFERENCES


(Received August 11, 1995)

(Revision received August 29, 1996)