ANAPHOR RESOLUTION IN WRITTEN DISCOURSE:
DOES PHONOLOGY PROVIDE THE MISSING LINK?

Darla Gayle Orchard
Department of Human Communication Disorders
McGill University
Montreal, Canada

A thesis submitted to the Faculty of Graduate Studies and
Research in partial fulfilment of the requirements for the
degree of Master of Science

Darla Gayle Orchard
August, 1991
ANAPHOR RESOLUTION IN WRITTEN DISCOURSE: 
DOES PHONOLOGY PROVIDE THE MISSING LINK?

Abstract

A series of experiments was designed to investigate processes involved in written anaphor resolution. The first experiment established that lexical decisions are affected by phonological relationships between primes and targets that are presented visually. Experiment 2 explored the possibility that phonological codes are involved in processing gaps and pronouns, which are examples of surface and deep anaphors, respectively. No evidence was found to support this hypothesis. However, the complete lack of anaphor effects in Experiment 2 suggested that the lexical decision paradigm was not sensitive to anaphor processing. Experiments 3 and 4 compared the priming effects found using a probe recognition task, to those found using a lexical decision task. Results suggested that only probe recognition responses were facilitated by anaphors. Furthermore, the anaphor facilitation effect was found only for sentences containing pronouns. These results were compared to those of previous studies that have used priming methodologies.
Résumé

Cette recherche a été conçu avec le but d'examiner les processus impliqués dans la résolution d'anaphores écrites. La première expérience a établi que les décisions lexicales sont facilitées par une relation phonologique entre le mot prime et le mot cible, avec présentation visuelle des deux mots. L'étude 2 a tenté d'établir que les codes phonologiques font partie de la résolution des syntagmes nominaux supprimés, et des pronoms, qui sont respectivement exemples d'une anaphore surface, et d'une anaphore profonde. Les résultats n'ont pas soutenu cette hypothèse. Pourtant, le manque complet d'effets d'anaphores a semblé indiquer que cet paradigme n'était pas sensible aux effets prévus. Les études 3 et 4 ont comparé la sensibilité aux effets anaphoriques de deux tâches, décisions lexicales vs. reconnaissance d'une sonde. Les résultats ont suggéré que seulement la reconnaissance d'une sonde est facilitée par la présence d'une anaphore. De plus, cet effet s'est manifesté seulement si la phrase prime contenait un pronom, et non si la phrase contenait un syntagme nominal supprimé. Ces résultats ont été discuté en comparaison avec d'autres études.
First and foremost, I would like to thank my supervisor, Dr. Gloria Waters, for providing a wonderful environment in which to work. Gloria's knowledge and positive attitude have provided invaluable insights and encouragement.

There are many people who have helped with my technical aspects of this thesis. To Claudine Charpentier, Leslie de Freitas, Joyce Srebro, Heather Tomlinson, and Glenna Waters, I extend my appreciation.

I would also like to thank Dr. Shari Baum, Carol Leonard, Ken McRae, and Elizabeth Rochon for lots of lively discussion, and many helpful comments at various stages of this work. My thanks also go out to Dr. David Caplan, who reviewed an earlier draft of this thesis.

Finally, I must extend my deepest appreciation to the one person for whom this process has probably been the most painful, my fiancé, Luc Prévost. Not only did he spend hours helping me to create stimuli, but he pampered, encouraged, and loved me throughout. For this I am, and always will be, exceedingly grateful.

This research was supported by a grant from the Natural Sciences and Engineering Research Council to Dr. Gloria Waters, and by a Natural Sciences and Engineering Research Council scholarship to the author.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Résumé</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Discourse Processing and Anaphora</td>
<td>1</td>
</tr>
<tr>
<td>The Reactivation of Antecedent Information</td>
<td>3</td>
</tr>
<tr>
<td>Phonology and Antecedent Reactivation:</td>
<td></td>
</tr>
<tr>
<td>An Argument</td>
<td>11</td>
</tr>
<tr>
<td>A Complication</td>
<td>15</td>
</tr>
<tr>
<td>Evidence</td>
<td>24</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>33</td>
</tr>
<tr>
<td>Method</td>
<td>35</td>
</tr>
<tr>
<td>Results</td>
<td>41</td>
</tr>
<tr>
<td>Discussion</td>
<td>45</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>46</td>
</tr>
<tr>
<td>Method</td>
<td>47</td>
</tr>
<tr>
<td>Results</td>
<td>51</td>
</tr>
<tr>
<td>Discussion</td>
<td>62</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>70</td>
</tr>
<tr>
<td>Method</td>
<td>72</td>
</tr>
<tr>
<td>Results</td>
<td>79</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>Discussion</td>
<td>82</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>85</td>
</tr>
<tr>
<td>Method</td>
<td>85</td>
</tr>
<tr>
<td>Results</td>
<td>86</td>
</tr>
<tr>
<td>Discussion</td>
<td>100</td>
</tr>
<tr>
<td>General Discussion</td>
<td>103</td>
</tr>
<tr>
<td>References</td>
<td>111</td>
</tr>
<tr>
<td>Appendix A</td>
<td>120</td>
</tr>
<tr>
<td>Appendix B</td>
<td>124</td>
</tr>
<tr>
<td>Appendix C</td>
<td>140</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Experiment 1: Mean reaction times to targets following each of the prime conditions ........................................ 42

Figure 2: Experiment 1: Mean percentage of errors to targets following each prime condition ........................................ 44

Figure 3: Experiment 2: Mean reaction times to targets following each of the prime types in the noun gap and pronoun sentences .......... 53

Figure 4: Experiment 2: Mean percentage of errors to targets following each of the prime types in the noun gap and pronoun sentences .......... 55

Figure 5: Experiment 2: Mean reaction times to targets following identity and phonological primes in the anaphor and no anaphor sentences .......... 56

Figure 6: Experiment 2: Mean percentage of errors to targets following identity and phonological primes in the anaphor and no anaphor sentences ......................... 58
Figure 7: Experiment 2: Mean sensibility ratings of noun gap pronoun sentences with each prime type ........................................ 59

Figure 8: Experiment 2: Mean sensibility ratings of no anaphor, noun gap, and pronoun sentences with identity and phonological primes ... 61

Figure 9: Experiment 3: Mean reaction times to probes following the three sentence types ...... 80

Figure 10: Experiment 3: Mean percentage of errors to probes following the three sentence types .. 81

Figure 11: Experiment 4: Mean reaction times to targets following the three sentence types ........................................ 88

Figure 12: Experiment 4: Mean percentage of errors to targets following the three sentence types ........................................ 89

Figure 13: Experiments 3 & 4: The effect of sentence type and task on reaction time ........... 92
Figure 14: Experiments 3 & 4: The effect of sentence type and task on errors ................. 93

Figure 15: Experiments 3 & 4: The effect of sentence type and task on the amount of time taken to read the experimental sentences .... 98

Figure 16: Experiments 3 & 4: The effect of task on the mean percentage of errors made to comprehension questions ................. 99
List of Tables

Table 1: Experiment 1: Example Stimulus Set .......... 37

Table 2: Experiment 2: Example Stimulus Set .......... 50

Table 3: Experiment 2: Reaction Time Data of 50 Subjects Before and After Analysis of Covariance ......................... 63

Table 4: Experiments 3 & 4: Example Stimulus Set .... 76

Table 5: Experiments 3 & 4: Distribution of Errors ... 91

Table 6: Experiments 3 & 4: Testing the Interaction Effect on the Reaction Time Data ............ 95

Table 7: Experiments 3 & 4: Testing the Interaction Effect on the Error Data ..................... 96
Discourse Processing and Anaphora

A discourse and a set of unrelated sentences may look very similar on a page. However, the discourse has a property that no randomly strung together set of sentences will ever achieve. A discourse is distinctive because the sentences of which it is composed share meaning, so that they form a unit. Halliday and Hasan (1976) refer to this property as cohesion. They use the term cohesion to denote the relationships in meaning that occur within and across sentences in a discourse. The notion of cohesion is common to all discourse models (e.g., Kintsch & van Dijk, 1978; Grosz & Sidner, 1986).

Halliday and Hasan describe a number of methods used to accomplish cohesion in a text. Anaphora is an important linguistic device used to achieve cohesion. Furthermore, it is a process that discourse models, including those referenced above, generally attempt to account for. The definition of the term anaphor can vary somewhat from author to author. In general, the term refers to elements in a discourse that take their meaning by reference to other elements in the discourse, called antecedents. The antecedent may be specifically stated in the discourse, or
it may be implied in the context of the discourse. This is depicted in examples 1-3, below.

Some authors are very specific in their use of the term anaphor. For example, Halliday and Hasan use the term anaphora only to refer to the process involved in resolving elements, like the pronoun he in example 1, below, that refer to specifically linguistic antecedents that precede the anaphoric item. They use the term cataphora to denote the processing of elements that make reference to subsequent sentential components, as in 2, below. The term exophora is used by these authors to refer to the processing of elements that do not have linguistic antecedents, but that are understood by reference to the non-linguistic environment, as in example 3, below.

1) Sam was swimming in the pool and he began to choke.
2) As he was swimming in the pool, Sam began to choke.
3) He's choking! (The call of another swimmer who is pointing to Sam and advising the lifeguard of Sam's predicament.)

Many other authors have adopted the more general practice of using the term anaphora to define all cases where assignment of a referent is necessary for comprehension (Garnham, 1987; Hankamer and Sag, 1976; Murphy, 1985a,b; Sag and Hankamer, 1984). In this paper, the term will be used in its more general sense. Nevertheless, the reader will notice that the discussion revolves particularly around anaphors that refer to previously mentioned elements. It is important to
note the environment in which the anaphors under discussion appear, because this may be a crucial consideration for hypotheses regarding the processes involved in the resolution of these anaphors.

The study of anaphora has received an increasing amount of attention over the past two decades (see Bosch, 1983; and Garnham, 1987 for reviews). This interest can be accounted for by a number of factors. Murphy (1990) attributes it to the fact that as well as being a widely used mechanism for enhancing communicative efficiency, anaphoric inference may act as a "test case" for the inferences necessary for cohesion. Thus, while all words in a discourse must be integrated for comprehension (Marslen-Wilson & Tyler, 1980), in the case of anaphors, we have specific hypotheses regarding both the inference that must be made, and the point at which this inference becomes necessary. For instance, in example 1, above, the anaphor "he" refers to its antecedent "Sam". One might predict that a link between these items must be inferred for comprehension to take place. Furthermore, it is assumed that the inference cannot be made before the point at which the reader or listener perceives "he".

The Reactivation of Antecedent Information

A question that has been of particular interest to many researchers is whether or not anaphors that follow an antecedent, as in example 1 above, are resolved by
reactivation of the antecedent. The technique that has been used most often to study this question is the probe recognition task (Bever & McElree, 1988; Chang, 1980; Cloitre & Bever, 1988; Dell, McKoon & Ratcliff, 1983, MacDonald, 1989; MacDonald & MacWhinney, 1990; McElree & Bever, 1989; McKoon & Ratcliff, 1980; O'Brien, Duffy & Myers, 1986). In this paradigm, subjects are presented with passages either visually or auditorilly. At some point, a probe word is presented and subjects must decide if the word was in the passage they had just read/heard or not. On critical trials, the passage contains an anaphor whose antecedent is related to the probe word. For example, a passage like example 1, above, might be presented. At the end of the sentence, the word "Sam" would appear, and the correct response would be "yer" because "Sam" was in the sentence just presented. Reaction times to this type of trial would be compared to some baseline, for example, responses to "Sam" following a sentence with no anaphor, as in 4, below.

4) Sam was swimming in the pool and a girl began to choke.

Using variations on this technique, the researchers cited above have all found faster reaction times, and or less errors, to probes following passages containing anaphors than to their baseline conditions. These results have been
attributed to the reactivation of antecedent information in anaphor resolution.

Other paradigms that have been used to test for the activation of the antecedent following an anaphor include lexical decision (Cloitre & Bever, 1988; Marslen-Wilson & Tyler, 1980; Nicol, 1988; Nicol & Swinney, 1989; Tanenhaus, Carlson & Seidenberg, 1985) and naming tasks (O'Brien et al., 1986). In both of these tasks, as in the probe recognition paradigm, subjects are presented with passages followed by targets that are either related or unrelated to the antecedent. The lexical decision task requires that subjects decide if the target is a real word or not. Thus, foil trials consist of nonword targets. In the naming task, subjects must name the target. It is expected that if the anaphor has reactivated antecedent information, then there will be facilitation for the target word when it is in some way related to the antecedent.

The tasks described above have been used not only to test whether pronouns reactivate antecedent information, but also whether a similar effect is found for other anaphors, including noun anaphors and gaps. Noun anaphors are instances where the anaphoric element is either a repetition of the antecedent, as in 5a, below, or another noun which can refer to the antecedent, as in 5b.

5a) A car sped around the corner. The car hit a pedestrian.
5b) A car sped around the corner. The vehicle hit a pedestrian.

There are numerous types of gaps, but all gaps are instances of anaphors that are not physically realized. This property can be seen in example 6, below. The noun phrase gap, which refers to the car, is indicated by an asterisk (*).

6) The car that sped around the corner was sure * to hit a pedestrian.

Using the paradigms just outlined, researchers have found that responses to probes and targets are facilitated by noun anaphors (Chang, 1980; Cloitre & Bever, 1988; Dell et al., 1983; Gernsbacher, 1989; McKoon & Ratcliff, 1980; O'Brien et al. 1986; Tanenhaus et al., 1985), pronouns (Chang, 1980; Cloitre & Bever, 1988; MacDonald & MacWhinney, 1990; McElree & Bever, 1989; Nicol, 1988), and gaps (Bever & McElree, 1988; MacDonald, 1989; McElree & Bever, 1989; Nicol, 1988; Nicol & Swinney, 1989; Tanenhaus et al., 1985) when there is some relationship between the probe or target, and the antecedent. This facilitation has typically been attributed to the reactivation of antecedent information by the anaphor.

However, other interpretations of the results have been proposed. For instance, Gernsbacher's (1989) results provide evidence that repeated noun anaphors result in reactivation of their antecedents, but pronouns do not. In a probe recognition task, Gernsbacher had subjects read two clause sentences which either contained pronouns (7a, below)
or repeated noun anaphors (7b). The sentences were presented one word at a time on a computer screen. The amount of time allotted to each word was a function of both the length of the word, and the subject's average reading speed.

7a) Bill handed John some tickets to a concert but \(^{p1}\) he \(^{p2}\) took the tickets back immediately.\(^{p3}\)

7b) Bill handed John some tickets to a concert but \(^{p1}\) Bill \(^{p2}\) took the tickets back immediately.\(^{p3}\)

ANTECEDENT PROBE = BILL
NON-ANTECEDENT PROBE = JOHN

Probes were presented in one of three positions, as indicated by the superscripted "p" in the examples. Two probe conditions were tested: the antecedent, Bill, and the non-antecedent, John. Gernsbacher hypothesized that anaphors might be comprehended by two processes, enhancement of antecedent information and suppression of non-antecedent information. Activation of the antecedent was measured by calculating the difference in reaction time to the antecedent probe before (p1) and after (p2 or p3) the anaphor. Similarly, inhibition of the non-antecedent was calculated from the difference in reaction time to the non-antecedent probe before (p1) and after (p2 or p3) the anaphor.

Gernsbacher (1989) found that for the repeated noun anaphors (eg. 7b) at the immediate post-anaphor probe
position (p2), reaction times to the non-antecedent were slowed, and reaction times to the antecedent were speeded, relative to the pre-anaphor probe position (p1) measures. However, for pronouns (eg. 7a), reaction times to both the antecedent and non-antecedent probes did not differ from the pre-pronoun (p1) to the immediate post-pronoun (p2) condition. By the end of the sentences (p3) containing pronouns, the non-antecedent probe reaction times were slower than those immediately following (p2) the pronoun, but the antecedent probe reaction times had not changed. This led Gernsbacher to conclude that only repeated noun anaphors are comprehended by processes which both enhance the antecedent information and suppress the non-antecedent. The process of understanding pronouns was proposed to occur by the suppression of the non-antecedent information, which resulted in relative activation of the antecedent. However, there was no evidence that pronouns resulted in actual reactivation of antecedent information.

Gernsbacher's (1989) results contrast vividly with those of Cloitre and Bever (1988). These authors also used a probe recognition task to compare the processing involved in the resolution of written material containing pronouns and repeated noun anaphors. They presented subjects with a lead-in sentence that was followed by a sentence containing either a repeated noun anaphor, a pronoun, or no anaphor, as in examples 8(a-c), below.
Anaphor Resolution

8) Lead-in sentence:

The gangly busboy spilled soup on the famous actress.

<table>
<thead>
<tr>
<th>Anaphor Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Repeated noun anaphor:</td>
<td>A waiter ran to help the busboy.</td>
</tr>
<tr>
<td>b) Pronoun:</td>
<td>A waiter ran to help him.</td>
</tr>
<tr>
<td>c) No anaphor:</td>
<td>A waiter smothered a giggle.</td>
</tr>
</tbody>
</table>

PROBE = GANGLY

The probes presented on experimental trials were always adjectives that had been used in the antecedent noun phrase, as indicated above. Because Cloitre and Bever were not looking for inhibition effects, the non-antecedent probe manipulation was not included. Probes were always presented at the end of the second sentence.

The results indicated that probes were responded to faster following the pronoun sentences than following the repeated noun anaphor sentences. Probes following repeated noun anaphor sentences were responded to faster than probes following the no anaphor sentences. Cloitre and Bever (1988) concluded that subjects access the antecedent information following pronouns faster than following repeated noun anaphors.

The different findings of these two studies may largely be due to the different baseline measures used. MacDonald and MacWhinney (1990) have suggested that it is possible that Gernsbacher's (1989) use of the pre-anaphor baseline
may have actually increased the likelihood of finding inhibition effects, and decreased that of finding facilitation effects. They note that activation of a concept will tend to decrease as the sentence goes on. Therefore, any reactivation that occurs will have to overcome this natural decay effect. A baseline which is taken earlier in the sentence will have been subject to less of this natural decay. Comparison to such a baseline may hide activation effects, and enhance inhibition effects, at a later probe point. For this reason, MacDonald and MacWhinney suggest the use of no anaphor baselines rather than pre-anaphor baselines. Thus, the lack of an activation effect following pronouns in the Gernsbacher study may have been due to the use of an inappropriate baseline.

However, as MacDonald and MacWhinney (1990) have noted, there are also problems with the no anaphor baseline used by Cloitre and Bever (1988). Specifically, this baseline does not control for changes in discourse focus across anaphor conditions, which might affect reaction time to the probe. Using a number of baseline conditions, MacDonald and MacWhinney found that probe recognition responses to antecedents were speeded following sentences containing pronouns; whereas responses to non-antecedents were inhibited. Thus, like Gernsbacher (1989), they found evidence for the use of both enhancement and suppression processes in the resolution of anaphors. However, they also
found that pronoun resolution made use of both of these processes. Given their own findings, and the consistency of their findings with those of numerous other studies, MacDonald and MacWhinney concluded that pronouns cause reactivation of antecedent information.

The finding that both repeated noun anaphors and pronouns prime their antecedents is very important. It is not surprising that the noun anaphor, Bill, primes its antecedent, Bill. However, it is more interesting to find that a pronoun can prime its antecedent. It is only as a result of the relationship between the anaphor and the antecedent that the pronoun shares any of the features of the antecedent. Thus, in the process of comprehending a discourse that contains anaphors, readers and listeners must discover this relationship. Consequently, the antecedent attains a level of activation that it would not have assumed without the anaphor.

**Phonology and Antecedent Reactivation: An Argument**

Although the literature reviewed above lends strong support to the notion that antecedent information is reactivated in anaphor resolution, the processes involved in this reactivation remain to be specified. Obviously, the goal of anaphor resolution is to arrive at an integrated semantic representation which is consistent with other information provided in the discourse. Thus, the studies cited above have consistently shown that antecedents,
semantic associates of antecedents, and concepts associated with the antecedent in the discourse, are primed by anaphors. However, it is not necessarily the case that semantic codes are the only source of information available to a processor dealing with anaphors. In fact, one characteristic of anaphors is that they do not tend to inherently hold much, if any, information that specifies their antecedents. Therefore, the integration of information in discourses containing anaphors is likely to be both particularly important, and particularly difficult.

It has been postulated that under conditions of increased processing load, phonological codes may be used to maintain verbal information in working memory, even when the form of input is visual (Baddeley, 1986; Baddeley & Hitch, 1974). The exact nature of these codes has not yet been specified. They are generally referred to as speech-based, articulatory, sub-vocal, or phonological codes. These codes are not necessarily equivalent; however, the terms will be used interchangeably in this discussion.

The role of phonology in the comprehension of written material has been widely debated (eg. Baron, 1973; Kleiman, 1975; Levy, 1978; Waters, Komoda & Arbuckle, 1985; Patterson & Coltheart, 1987). Using both interference from homophones in sentence acceptability tasks (Baron, 1973; Treiman, Freyd & Baron, 1983), and priming tasks (Davidson, 1986) researchers have shown that phonological codes are computed
during reading. However, the role of this information in the comprehension of written material remains unclear.

Typically, the studies designed to show that the use of phonological information is necessary in written sentence comprehension have involved concurrent tasks that interfere with the use of speech-based codes. Results of these studies suggest that these tasks also interfere with reading. For example, Slowiaczek and Clifton (1980) found that repeated articulation of the word "cola" interfered with the comprehension of passages significantly more when input was visual than when input was auditory. However, this effect held only for texts that could be classified as "more difficult" in that they required inferencing within or across sentences. The authors concluded that for this kind of written material, the use of subvocal rehearsal aided comprehension.

These types of studies have been criticized because the conditions under which concurrent articulation impairs performance may be thought to be the most taxing. Thus, it has been proposed that the results could be interpreted as suggesting that concurrent articulation simply further depletes limited capacity processing resources, rather than specifically interfering with the use of phonological codes (Hudson & Tanenhaus, 1985; Waters et al., 1985). In fact, Waters et al. found that when the interference from concurrent articulation and a similarly process-taxing task,
like tracking tones, are compared, performance decrements were equivalent. Therefore, these authors stressed that it is necessary to control for a general reduction in processing resources when using concurrent tasks to investigate the role of specific codes in comprehension. Furthermore, they concluded that skilled readers are capable of reading material for meaning without the use of speech-based codes.

More recently, Waters, Caplan and Hildebrandt (1987) have found concurrent articulation effects above and beyond those of a control task in the time taken to judge the acceptability of written sentences. They manipulated both the syntactic complexity of, and the number of propositions in, their experimental sentences. Concurrent articulation interfered more than a control finger-tapping task only when the number of propositions was increased, and not when the syntactic complexity of the sentences was increased. Waters et al. (1987) concluded that the first-pass syntactic analysis of a sentence can be achieved without the use of articulatory codes. However, they suggest that these codes may be used to interpret readings which are derived from the syntactic structure of a sentence.

To summarize, there is support for the hypothesis that phonological codes are computed during reading, but that much of the normal reading process can be achieved without the use of phonological information. However, there is also
evidence to suggest that phonological codes are involved in the comprehension of written sentences with certain types of processing loads. Thus, one might postulate a role for phonology in the comprehension of sentences that are complicated by the necessity of interpreting anaphoric structures. This role might be, for example, the storage of a surface form of the antecedent that may trigger full antecedent activation when needed. However, before reviewing evidence that has accumulated regarding the role of phonology in anaphor resolution, it is necessary to digress slightly and introduce a proposal suggesting that different classes of anaphors exist, and that these classes differ in the processes involved in their resolution.

Phonology and Antecedent Reactivation: A Complication

Hankamer and Sag (1976) were the first to observe that there is a class of anaphors requiring antecedents of a relatively restricted type. Their hypothesis centres around the following observation: some anaphors can either take linguistic antecedents, or refer to the nonlinguistic environment, while other anaphors require specifically linguistic antecedents. In examples 9a and 9b, below, Hankamer's statement provides a linguistic antecedent for both the do it verb phrase anaphor and the verb phrase gap. In both cases, the antecedent is clearly "stuff this ball through this hoop". However, 10a and 10b do not have a specifically linguistic antecedent. Here, the context may
provide an antecedent for the anaphors, but Hankamer and Sag note that the verb gap in 10b is not quite acceptable (this is indicated by the "#" preceding the sentence). They propose that this is because verb gaps are a member of a class of anaphors that require specifically linguistic antecedents. They refer to these anaphors as surface anaphors and contrast them to deep anaphors, of which the do it anaphor is an example.

9) Hankamer: I'm going to stuff this 9" ball through this 6" hoop.

(Deep) a) Sag: It's not clear that you'll be able to do it.

(Surface) b) Sag: It's not clear that you'll be able to.

10) (Hankamer is trying to stuff a 9" ball through a 6" hoop.)

(Deep) a) Sag: It's not clear that you'll be able to do it.

# (Surface) b) Sag: It's not clear that you'll be able to.

Hankamer and Sag propose that deep anaphors do not require linguistic antecedents; therefore, deep anaphors may refer to concepts from the non-linguistic environment.

Hankamer and Sag (1976) observed another difference between deep and surface anaphors. They found that only surface anaphors require that the syntactic form of the antecedent and the anaphor be the same. This syntactic parallelism restriction can be seen in examples 11a-c, from Sag and Hankamer (1984).
11) The children asked to be squirted with the hose so...

a) they were.

# b) we did.

c) we did it.

In example 11a, the verb phrase ellipsis (a surface anaphor) results in a gap which can be filled with the verb phrase "squirted with the hose". This filler is parallel in form to the antecedent, and the surface anaphor in 11a is acceptable. In 11b, the filler would be of the form "squirt them with the hose". The filler in this case should be in active voice; the antecedent is in the passive. This inconsistency causes the surface anaphor in 11b to be an inappropriate ending to the sentence. In 11c, we see that this change of voice does not render the deep anaphor, do it, unacceptable. These observations led Hankamer and Sag to propose that surface anaphors are interpreted with reference to the surface structure of the sentence. On the other hand, deep anaphors were proposed to be interpreted with reference to either the deep structure of the sentence, or the physical environment in which the sentence occurs.

Sag and Hankamer (1984) later revised this hypothesis and changed the terms used to distinguish between these two types of anaphors. They referred to surface anaphors as elliptical expressions, having proposed that comprehension of these structures occurs by referencing propositional
representations that reside in a short term memory store. Deep anaphors were renamed model interpretive anaphoric elements, and were proposed to be interpreted with reference to the discourse model. However, most researchers (Murphy, 1985a,b, 1990; Tanenhaus, Carlson and Seidenberg, 1985; Tanenhaus and Carlson, 1990) have retained the terms deep and surface anaphor, and these are the terms that will be used to distinguish the two types here. The reader should note that if the processing of these two anaphor types is as distinctive as Sag and Hankamer have proposed, then it is possible that phonological information would be helpful only in the resolution of surface anaphors. Therefore, when looking for the use of these codes in anaphor resolution, researchers might have to be very specific about what type of anaphor they are investigating.

Several attempts have been made to find experimental evidence in support of the distinction first introduced by Hankamer and Sag (1976). Murphy (1985a) conceived of the process of referring to a representation of the antecedent in short term memory as a copying process. He suggested a number of manipulations that should affect such a copying process. Murphy argued that the length of the antecedent should affect the amount of time required for the completion of a copying process. Furthermore, he postulated that if this process is dependent on representations from a short term memory store, which are subject to decay, then
increasing the distance between the antecedent and the anaphor should interfere with copying, and therefore with anaphor resolution. As well, if only surface anaphors make use of such a process, then only surface anaphors should show these effects. However, Murphy found that his manipulations affected deep and surface anaphors equally, and he concluded that both deep and surface anaphors could be resolved by something like a copying process.

Murphy (1985b) proposed that the distinction between deep and surface anaphors is real, but that its origin is not in a processing distinction. Instead, he points out that deep anaphors, such as pronouns, typically provide more information regarding their antecedents than surface anaphors do. Murphy argued that because deep anaphors are naturally less ambiguous than surface anaphors, they can be successfully used to refer to antecedents in more ambiguous contexts. Thus, Murphy suggested that it is not the type of anaphor that determines the way in which the anaphor will be resolved. Rather, he proposed that it is the environment in which the anaphor appears that will determine whether a copying process, or reliance on what Murphy calls plausible reasoning, will be used in the process of anaphor resolution.

Whereas Murphy (1985a) found that his manipulations affected both anaphor types equally, Tanenhaus et al. (1985) found that increasing the distance between the antecedent
and the anaphor increased the amount of time taken to read the sentences containing surface anaphors only. Tanenhaus and Carlson (1990) and Murphy (1990) both provide reasonable arguments as to why they found different results using the same paradigm. Both studies involve confounds which cloud their interpretation, and so it is important to find evidence for or against the hypothesis that deep and surface anaphors are processed differently using other designs.

Tanenhaus et al. (1985) describe another paradigm that they used to investigate the possibility that deep and surface anaphors are resolved with reference to different levels of representation of the antecedent. Subjects read passages containing either deep or surface anaphors; then they were presented with a test sentence which they had to verify. The test sentence either matched the antecedent phrase in the passage, or it differed from the structural form or lexical content of the antecedent, while maintaining semantic content. The manipulation of structural form can be seen in example 14, and the manipulation of lexical content is displayed in example 15, below.

14) Structural Change Condition

Antecedent Condition

a) Matching: Jenny asked Ann's boyfriend out yesterday.

b) Mismatching: Jenny asked out Ann's boyfriend yesterday.
Anaphor Resolution

Anaphor Condition

c) Deep: Ann was furious that she did it.
d) Surface: Ann was furious that she did.

Verification Sentence

e) Jenny asked Ann's boyfriend out.

15) Lexical Change Condition

Antecedent Condition

a) Matching: My sister decided to buy a new sports car.
b) Mismatching: My sister decided to get a new sports car.

Anaphor Condition

c) Deep: Because she is conservative, I was surprised she did it.
d) Surface: Because she is conservative, I was surprised she did it.

Verification Sentence

e) My sister bought a new car.

Tanenhaus et al. (1985) predicted that if anaphors are resolved by reference to a surface representation of the antecedent, then responses to the verification sentences (14e & 15e) following passages containing the mismatching antecedents (14b & 15b) should be slower than responses following the matching antecedents (14a & 15a). According to the Sag and Hankamer (1984) hypothesis, this should be the case for the passages containing surface (14d & 15d), but not deep (14c & 15c) anaphors. Because the meaning of
the antecedent is maintained in the mismatching antecedent condition, response times to the verification sentences following deep anaphors should not be affected by this manipulation.

Tanenhaus et al. (1985) did find slower verification latencies following passages containing mismatching antecedents in the structural change condition, but only for passages containing surface anaphors. There were no significant effects of anaphor or antecedent type for the lexical change passages. These results suggest that the resolution of surface anaphors does involve referencing a linguistic representation of the antecedent; whereas, this does not seem to be the case for deep anaphors. Because the lexical change manipulation had no effect, it would appear that the linguistic representation of the antecedent that is accessed is not a literal instantiation of the antecedent. Tanenhaus et al. suggest that perhaps surface anaphors access representations of their antecedents at a level where the lexical items have been replaced by their conceptual representations.

The finding that even surface anaphors are not resolved with reference to literal codes poses a problem for the hypothesis that phonological codes might be involved in anaphor resolution. However, before abandoning this hypothesis, it should be noted that the stimuli in the lexical change condition may not have provided a fair test
of the use of literal codes. The reader should note that in the example provided, a lexical change was defined as a change from "buy" to "get", and in the matching condition, the corresponding items were "buy" and "bought". Note that in the lexical matching condition, the items differ in their phonemic and orthographic structure. Therefore, both of the conditions in this version of the experiment can be thought of as mismatching, and thus would not be expected to differ. Because of this problem in the lexical change manipulation, the possibility that surface anaphors are interpreted with reference to phonological and/or orthographic representations of the antecedent is not precluded by these results. However, this experiment does offer some support for the hypothesis that deep and surface anaphors access different levels of representation of their antecedents.

Further support for this hypothesis comes from studies by Murphy (1990) and Tanenhaus and Carlson (1990). In these experiments, subjects were required to make sentence acceptability judgements. The results of both studies suggest that manipulating the syntactic parallelism of the antecedent and the anaphor had a greater effect on judgements of sentences containing surface anaphors than sentences containing deep anaphors. Tanenhaus and Carlson concluded that these results provide strong evidence that the resolution of surface anaphors is more affected by the surface form of the antecedent than is the resolution of
deep anaphors. On the other hand, Murphy argues that the results may be a reflection of the demands of a judgement task. He maintains that the syntactic parallelism effect does not distinguish between anaphor types when subjects are simply asked to read the sentences for meaning.

Thus, while the issue of whether or not deep and surface anaphors are processed differently has not been completely resolved, the potential distinction should be kept in mind in investigations of anaphor resolution. For the purposes of investigating the use of phonological codes in antecedent reactivation, the Sag and Hankamer (1984) proposal would predict that these codes may be used in the resolution of surface, but not deep, anaphors.

**Phonology and Antecedent Reactivation: Evidence**

Recently, researchers have looked at the process of anaphor resolution with the aim of determining the types of codes involved. Black, Coltheart and Byng (1987) have proposed that the process of written anaphor resolution could make use of orthographic codes, phonological codes, and/or semantic codes. They carried out a study designed to investigate the form of linguistic representation used in the resolution of verb gaps, a type of surface anaphor. Subjects read sentences that they had to judge for semantic acceptability. On critical trials, unacceptable sentences contained either a homograph or a homophone, and if the verb gap was being filled by the orthographic or phonological
form of the antecedent, then the sentence would seem acceptable. Thus, Black et al. hypothesized that sentences containing homographs, like 16a, would be harder to reject than control sentences like 16b, below, if orthographic codes were referred to in the resolution of the verb gap.

16a) Soldiers wound enemies and watchmaker’s clocks.
16b) Soldiers wound enemies and watchmakers books.

However, if phonological codes are being referred to, then sentences like 17a should be more difficult to reject than 17b.

17a) He rights injustices and she books.
17b) He rights injustices and she the sugar.

Black et al. (1987) found no differences in either reaction time or error rate to sentences like 16a and 16b. However, subjects did make more errors in response to the homophone sentences (17a) than controls (17b). These results suggest that even in reading, subjects comprehend verb gaps by referencing a phonological representation of the antecedent. Although no evidence was found to support the hypothesis that orthographic codes are used in the resolution of verb gaps, the authors advise caution in the interpretation of this null result, as they were only able to create two items in this stimulus set.

Black et al. (1987) noted that occasionally, the surface form of the antecedent is permitted to differ from that of the gap filler. For example, sentence 18a, below,
is perfectly acceptable, despite the fact that the appropriate filler for the verb gap (play) is physically different from its antecedent (plays).

18a) Mary plays the violin and her brothers the drums.
18b) Mary plays the violin and Sue the drums.

This observation was used to investigate the possibility that semantic representations are used in the processing of verb gap sentences. The authors reasoned that if a literal representation is used to maintain the antecedent, then there should be some cost involved in processing sentences in which the surface form of the antecedent and the filler are different. On the other hand, if semantic representations are used exclusively, then subjects should have no more difficulty with sentences like 18a, above, than with 18b. Note that in 18b the surface form of the antecedent and the gap filler are identical.

Results suggested that there is a cost involved in processing sentences like 18a. However, it was noted that the results may be due to the fact that the antecedent and the gap filler differ in morphosyntactic features, as well as in physical form. Black et al. (1987) tested for an effect of parallelism in morphosyntactic features by contrasting sentences such as 19a and 19b, below.

19a) Sue polished the shoes and John the furniture.
19b) Sue polished the shoes and the adults the furniture.
Although the antecedent and the gap filler have the same physical features, morphosyntactic features are incongruous in sentence 19b. A processing cost was found in processing sentences such as 19b, suggesting that a morphosyntactic representation of the antecedent is used in processing verb gaps. Black et al. concluded that skilled readers access both phonological and morphosyntactic information in the comprehension of written material containing verb gaps.

Waters, Caplan and Leonard (in press) have suggested that the effect of phonology seen in the Black et al. (1987) study may not have been due to first-pass comprehension processes. They propose that phonology may be involved in either first-pass processing, where stored phonological representations may allow the review of the material during the initial comprehension process, or in second-pass review processes. Waters et al. indicate that first-pass processes have been conceived of as automatic sequences that are necessary and often sufficient for the comprehension of texts. However, they note that under some circumstances second-pass processes may be necessary for comprehension. It has been proposed that second-pass processing is often under conscious control, and phonological codes not required in the first-pass analysis may be activated by these review processes. Waters et al. observed that in the Black et al. paradigm, each sentence remained on the computer screen until the subject responded. Thus, it was possible for
subjects to re-read the sentences, invoking second-pass processes.

Waters et al. (in press) attempted to replicate the Black et al. (1987) findings; however, Waters et al. included two presentation conditions. In one, the sentences remained on the screen; in the other, the words of the sentences appeared one-by-one on the screen for a specified amount of time. The latter presentation mode, known as rapid serial visual presentation (RSVP), ensured that subjects could not re-read the stimuli, and therefore could not make use of second-pass processes.

Waters et al. (in press) also manipulated the type of anaphor presented. Half of their experimental stimuli consisted of verb gaps similar to those in the Black et al. (1987) studies. The other half consisted of sentences containing personal and indefinite pronouns, classified as deep anaphors (Hankamer & Sag, 1976). Therefore, the Waters et al. study included a manipulation with the potential to distinguish between the two anaphor types.

The sentences were constructed so that in some cases, the antecedent was a phonologically plausible filler for the gap or pronoun, but the sentence did not make sense. An example of the phonologically plausible verb gap (20) and pronoun (21) sentences is presented below.

20) The children sleighed in winter, and the murderer in cold blood.
21) There is a sale at the store and I have one on the boat.

Waters et al. (in press) argued that if the antecedent of an anaphor is reactivated in a phonological form in the first-pass analysis of a sentence, then Black et al.'s (1987) results should be replicated regardless of whether the sentence is presented as a whole, or using the RSVP paradigm. However, if phonological information is becoming activated because subjects re-read the sentence upon encountering an anaphor, then phonological effects should only be seen in the whole sentence condition. This is indeed what they found, suggesting that the effects seen in the Black et al. study were due to the activation of antecedent information when subjects re-read the sentences.

Waters et al. (in press) noticed that in their stimuli, some of the unacceptable control sentences contained anomalies that might have induced a second-pass review strategy. For example:

22) The children sleighed in the winter, and the murderer in the jar.

The second clause in example 22 contains a semantic oddity that might encourage subjects to re-read the sentence. This strategy may have then been adopted for all of the sentences in the whole sentence condition. In a second experiment, Waters et al. removed the semantically odd sentences from the stimulus set, replacing them with more neutral stimuli. They did not replicate the phonological effect found in the
first experiment. Therefore, Waters et al. concluded that phonological information is used when subjects adopt a strategy of re-reading the material, but it would appear from these experiments that phonology is not used in the first-pass processing of the sentences containing anaphors that were studied.

However, Leonard (1991) has found very different results using the same paradigm, but different stimuli. She presented subjects with sentences containing *do so* anaphora, a surface form of verb phrase ellipsis, and *do it* anaphora, a deep form. Using the RSVP technique, Leonard found that both types of anaphor resulted in a significant plausibility effect. That is, when the "antecedent" was a homophone of the appropriate filler, and thus the sentences were plausible, but not acceptable, subjects were more likely to judge them as acceptable than when the "antecedent" was not a phonologically plausible filler. Thus, the results of this experiment replicate those of Black et al. (1987) with RSVP presentation of the stimuli. In conclusion, the sentence judgement paradigm has offered inconsistent support for the proposal that both deep and surface anaphors are resolved with reference to phonological representations of the antecedent.

Using a very different paradigm, Tanenhaus et al. (1985) have found tentative evidence to suggest that the phonological form of the antecedent is reactivated in the
resolution of wh-gaps presented auditorilly. They first review results reported in Hudson and Tanenhaus (1985), which suggest that words in a sentence presented auditorilly can prime lexical decisions to rhyming targets presented visually. In this study, subjects listened to sentences such as 23a and 23b, below, and then made lexical decisions to words presented on a computer screen.

23a) Since Jane forgot to put in all the cream, the cake was dry.

23b) Since Jane forgot to put in all the milk, the cake was dry.

Hudson and Tanenhaus found that lexical decisions to words like "theme" were faster following sentences containing rhymes (23a) than following the control sentences (23b). This effect was found when up to four words intervened between the rhyme prime and the target, but was abolished when seven or more words separated the prime and the target. Tanenhaus et al. used this methodology to examine priming attributable to the presence of wh-gaps. Subjects were presented auditorilly with sentences such as 24a and 24b, below; they were then asked to make lexical decisions to words like "fear", which were presented visually 400 ms after the sentence.

24a) The man was surprised at which beer the judges awarded the first prize to.

24b) The man was surprised at which wine the judges awarded the first prize to.

Both sentences end with wh-gaps. In 24a, the appropriate
filler is "beer"; in 24b it is "wine". Tanenhaus et al. found that responses to the targets were faster following sentences in which the gap was filled with a rhyming word. Because at least seven words intervened between the antecedent and the target, the rhyme priming effect observed was attributed to the presence of the gap. Tanenhaus et al. concluded that gap-filling is a case of literal replacement. That is, the antecedent is stored in verbatim form, and this is the form used to replace the gap for comprehension. However, Tanenhaus et al. report that these results should be interpreted with caution because the rhyme priming effects found were not significant in the item analysis, only eighteen subjects were tested, and there was no manipulation in the filler-gap experiment to ensure that the antecedent was not causing the effects seen.

Although the Tanenhaus et al. (1985) results provide tentative evidence that anaphor resolution may involve the use of phonological codes, it is important to note that the sentences in this experiment were presented auditorily. Therefore, while the priming methodology may be sensitive to effects of phonology in anaphor resolution, these results provide no enlightenment with respect to the use of these codes in the comprehension of written material containing anaphors. The purpose of the studies presented here is to investigate whether the rhyme priming effect found by Tanenhaus et al. occurs with visual presentation of the
prime stimuli. The first experiment was performed to establish that lexical decisions to visually presented targets can be influenced by phonological relationships with visually presented primes. The second experiment was designed specifically to investigate the involvement of phonology in the processing of written material containing anaphors. If visually presented sentences can prime lexical decisions to visually presented targets, and this effect can be attributed to the presence of an anaphor, then we would at least have evidence that phonological representations of the antecedent are present during written anaphor resolution. This would be a first step in the line of evidence necessary to establish that phonological representations have a specific function in the comprehension of written anaphors.

Experiment 1

Because responses in a lexical decision task were to be the measure used to determine if phonological codes are present during written anaphor resolution, it was important to first ensure that lexical decisions to written targets are facilitated by phonological relationships with their primes. The results of Hudson and Tanenhaus (1985) and Tanenhaus et al. (1985) suggest that this indeed is the case; when the primes are presented auditorily. Davidson
Anaphor Resolution

(1986) has also shown that lexical decisions to visually presented targets can be primed by homophones that appear in a written text.

Although the proposal that sound-based codes mediate visual lexical decision has existed in the literature for a number of years (Rubinstein, Lewis & Rubinstein, 1971; Meyer, Schvaneveldt & Ruddy, 1974), evidence that phonological relationships between primes and targets facilitate this task has been sparse. For example, Meyer et al. concluded that the conversion from graphemes to phonemes is involved in lexical access, because they found that primes inhibited reaction times to targets that were graphemically similar but phonemically dissimilar. However, they did not find significant facilitation to targets that were both graphemically and phonologically similar to their primes. The strongest evidence to date that primes that are phonologically related to their targets can actually facilitate lexical decisions to these targets comes from Hillinger (1980). In a series of three experiments, Hillinger found that lexical decisions to targets preceded by rhyming primes were facilitated whether the graphemic relationship between the prime and target held (late-mate) or not (eight-mate).

Hillinger's (1980) results provide strong support for the proposal that lexical decisions to visually presented targets are primed by visually presented rhymes. However,
more recently, Martin and Jensen (1988) have reported repeated failures to replicate Hillinger's results. Thus, the first experiment presented here was undertaken to ensure that lexical decisions to visually presented targets could be influenced by a phonological relationship with a preceding word. The experiment was designed to mimic the conditions under which the sentence experiment would be run as closely as possible; however, single words were presented as primes rather than whole sentences. It was predicted that targets preceded by rhyming primes would be responded to faster than the same targets preceded by neutral primes. Furthermore, to ensure that any rhyme-priming effect could not be attributed to shared graphemic patterns between the primes and targets, an orthographic control was included. It was predicted that responses following these primes would not be facilitated relative to neutral primes. Finally, an identity prime condition was included. This condition was to be an important element of the sentence experiment, and so it was included here. It was expected that targets following identity primes would be responded to faster and/or more accurately than targets following neutral primes.

Method

Subjects

Fifty university students between the ages of 17 and 30 years participated in the study. All subjects were native
English speakers. All subjects in this and subsequent experiments were paid for their participation.

Materials

For Experiments 1 and 2, 80 experimental stimulus sets were prepared. Each set consisted of a single syllable target stimulus that could be matched to three other single syllable words holding one of the following relationships with the target: phonologically similar but orthographically dissimilar (phonological prime); orthographically similar but phonologically dissimilar (orthographic prime); or both phonologically and orthographically dissimilar (neutral prime). In the identity prime condition, the target and prime were the same word. Although in Experiment 1 only single word primes were presented, for the purposes of Experiment 2, each of the words from a given set had to fit into the same sentence context. Thus, 80 stimuli consisting of the four prime-target relationships depicted in Table 1 were constructed, the full set of experimental stimuli is presented in appendix A.

The design of the experiment was within both subjects and items; all subjects saw all levels of prime, and the same target in any stimulus set appeared after all prime types. In this and all subsequent experiments, a subject saw any given target in only one prime-target relationship. This was accomplished by creating as many stimulus lists as there were conditions in the experiment. For example, in
Table 1

Experiment 1: Example Stimulus Set

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Prime</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>bear</td>
<td>BEAR</td>
</tr>
<tr>
<td>Phonological</td>
<td>snare</td>
<td>BEAR</td>
</tr>
<tr>
<td>Orthographic</td>
<td>spear</td>
<td>BEAR</td>
</tr>
<tr>
<td>Neutral</td>
<td>trap</td>
<td>BEAR</td>
</tr>
</tbody>
</table>
Experiment 2, ten prime-target conditions were tested; therefore, ten stimulus lists were made. The ten conditions for a given target were distributed across these lists. Each list contained an equal number of stimuli from each experimental cell. Each subject completed one stimulus list. To keep the presentation of stimuli as similar as possible in Experiments 1 and 2, the ten lists created for Experiment 2 were used in Experiment 1, with single word rather than sentence primes. Because of the control measures necessary for the sentence reading paradigm, i.e., Experiment 2, not all prime types were presented in equal proportions (see Experiment 2, Materials, for details). The same proportions were presented in Experiment 1 as in Experiment 2. The result was that 30% of the experimental stimuli presented consisted of identity prime trials, 30% phonological prime trials, 20% orthographic prime trials, and 20% neutral prime trials.

Eighty filler stimuli were also constructed. The fillers pairs consisted of orthographically legal, single syllable nonword targets, preceded by single syllable English words. The primes words were nouns from the first noun phrase of the filler sentences that were prepared for Experiment 2. Thus, on 50% of the trials, a real word target appeared; on 50% of the trials, a non-word target appeared.
Procedure

The stimuli were presented and the data were collected by an IBM Personal System 2 computer. The word pairs were presented sequentially. Each pair was preceded by a fixation point, which consisted of an asterisk (*) appearing in the middle of the screen for 500 ms. At the offset of the fixation point, the prime word appeared in lower case print. The subjects were instructed to silently read this first word of the pair quickly but carefully. The amount of time allowed to read the prime was self-paced to mimic the presentation of the sentences that would take place in Experiment 2 as closely as possible. Subjects were instructed to press the (+) button adjacent to the number pad when they had finished reading the prime word. Immediately following this response, the target appeared in capital letters. Subjects were asked to decide if this word was a real English word as quickly and as accurately as possible. A "yes" response was indicated by pressing the (+) button adjacent to the number pad with the right thumb; a "no" response was indicated by pressing the tab button, on the opposite side of the keyboard, with the left thumb. The computer recorded the amount of time from the onset of the target until the subject's response. There was a 1000 ms interval between the offset of the target, and the fixation point that signalled the beginning of the next trial.
Subjects were told that once they had responded to a number of prime-target pairs, they would be presented with a recognition task that would require them to identify prime words that they had read on the computer screen. The recognition task was included to encourage subjects to pay attention to the prime words. One recognition list was presented after the practice trials; another was presented during a break after the 80th prime-target trial. The list following the practice trials consisted of eight words, of which four had appeared in the practice session. The other list consisted of 15 words, eight of which had been presented to the subjects.

A practice consisting of ten word pairs began the session. This practice was followed by the practice recognition list, with the aim of familiarizing subjects to this task. Subjects were encouraged to ask any questions that they may have had before beginning the experimental trials.

The stimuli were distributed equally and randomly into two blocks of 80 prime-target pairs each. Subjects were allowed as much time as they felt necessary as a rest period between blocks. During this rest period, subjects completed the recognition task designed to encourage them to read the prime words. Testing took place in one session that lasted approximately 15 minutes.
Results

Because of the precise nature of the hypotheses to be tested with this study, it was possible to use planned contrasts rather than omnibus analyses to investigate any effects. Some authors have suggested adjusting alpha levels when contrasts are used (Maxwell & Delaney, 1990), whereas others make no mention of the need for this adjustment (Ferguson & Takane, 1989). Other authors propose that it is only necessary to adjust alpha levels when a large number of contrasts is planned (Rosenthal & Rosnow, 1985). The alpha level used in this and subsequent experiments was 0.05, because the number of contrasts relative to the number of cells in the studies did not seem large.

It was hypothesized that the identity primes and the phonological primes would facilitate responses to the targets relative to neutral primes; whereas responses to targets following the orthographic primes would either not differ from the neutral condition, or would be relatively slower (Meyer et al., 1974). Thus, three contrasts were constructed to compare the effects of each of the identity, phonological, and orthographic primes to the neutral prime condition. The reaction time responses can be seen in Figure 1. Any reactions times that differed from the grand mean by more than a cutoff of three standard deviations, were replaced by this cutoff before analysis. This resulted in the replacement of 1.97% of the observed reaction times.
Figure 1. Experiment 1: Mean reaction times to targets following each of the prime conditions.
One-tailed analyses revealed that targets following identity primes were responded to significantly faster than those following neutral primes by subjects, $t(147) = 15.666$, $MSe = 2003.986$, and by items, $t(237) = 13.623$, $MSe = 5131.154$. As well, phonological primes resulted in faster reaction times by both subjects, $t(147) = 1.908$, $MSe = 2003.986$, and items, $t(237) = 2.136$, $MSe = 5131.154$, when compared to neutral primes.

It should be noted that there is a trend for facilitation following the orthographic prime condition relative to the neutral condition. This trend was not predicted. However, because this was a control condition, the trend was tested with the same powerful one-way statistics that were used to test the phonological prime effects. The analysis revealed no significant orthographic priming by subjects, one-tailed $t(147) = 0.886$, $MSe = 2003.986$, or by items, one-tailed $t(237) = 1.082$, $MSe = 5131.154$.

The error data, seen in Figure 2, reflect the same trends seen in the reaction time measures. The square-root transformation was applied to the error data before analysis. The only contrast to reach significance was that comparing the identity primes to the neutral primes, which was significant by subjects $t(147) = 7.560$, $MSe = 2.157$, and items, $t(237) = 6.313$, $MSe = 2.515$. 
Figure 2. Experiment 1: Mean percentage of errors to targets following each prime condition.
Discussion

The results of the first study confirm that the lexical decision task is sensitive to phonological priming using these stimuli, and not surprisingly, strong identity priming was also found. The phonological priming effect was not simply attributable to the orthographic similarity between the phonological primes and targets, as no orthographic priming effect was found.

It is interesting that the interference effect for orthographically similar but phonologically dissimilar primes, which was found by Meyer et al. (1974), was not replicated here. This may be because in Experiment 1, a high percentage of real word targets were preceded by word primes that held some physical relationship with the targets. On 80% of the real word trials, the prime was either identical to the target, or related to the target by phonological or orthographic similarity. This may have induced subjects to use strategies in their lexical decisions whereby they responded "yes" to the targets if they held some relationship to the prime. This type of strategy use would be a serious problem if the aim of this study had been to identify processes used in normal word recognition. However, the purpose of Experiment 1 was simply to determine if lexical decisions could be influenced by phonological relationships with prime stimuli. The results suggest that this is the case. Even if the priming
effects found are due to some sort of strategy, it appears that this strategy is particularly sensitive to phonological information, as significant priming was found with the rhyme primes, but not with the orthographic primes.

The phonological priming effect was not large; the average difference between responses to targets following phonological and neutral primes was only 17 ms. However, this difference was consistent enough to result in statistically significant priming. Thus, the results of the first experiment suggest that the lexical decision task may be a useful tool for investigating the use of phonological codes in written anaphor resolution.

Experiment 2

In Experiment 2, the primes from the first experiment were embedded in sentences that contained either noun gaps, pronouns or no anaphor. These primes functioned as the antecedents in the anaphor sentences, and were located in comparable positions in the no anaphor sentences. The noun gaps used in this experiment were surface anaphors, while the pronouns were deep anaphors (Hankamer & Sag, 1976). Given the processing distinction described in the introduction, it was hypothesized that the phonological priming effect might only be seen for surface anaphors. However, it was predicted that both anaphor types would show
identity priming, because identity primes share all codes, surface and semantic, with their targets.

So that it would be possible to establish that any priming effects found were due to the anaphors, and not simply to the presence of the antecedents, target reaction times following the identity and phonological prime conditions in the no anaphor and anaphor sentences were compared. It was predicted that targets would be responded to faster and/or more accurately following the anaphor sentences than the no anaphor sentences.

**Method**

**Subjects**

One-hundred university students fulfilling the criteria described in Experiment 1 participated in the study. None of these subjects had participated in the first experiment.

**Materials**

The stimuli in Experiment 2 consisted of the same set of 80 experimental and 80 filler targets that were used in Experiment 1. The four prime types described in the first experiment were also employed; however, the primes were embedded in one of three sentence conditions. Because of limitations on the number of stimuli that fit the priming criteria, only the identity and phonological prime conditions were used in the no anaphor sentences. This resulted in 30% of the experimental trials involving identity primes, 30% phonological primes, 20% orthographic
primes, and 20% neutral primes, as in Experiment 1.

The levels of the sentence factor were: no anaphor, gap, and pronoun. All of the sentences consisted of two clauses conjoined with "and", "but", "yet", or "because". The first noun phrase in all of the experimental sentences contained a noun from the prime types described in Experiment 1. These nouns were always at least seven words from the end of the sentence, and thus from the target. The second clause of the anaphor sentences held either a gap or a pronoun that referred back to the noun from the first clause. The anaphors were always within four words of the end of the sentence. These manipulations follow from Hudson and Tanenhaus' (1985) findings that, with auditorily presented sentences, phonological primes that occurred within four words of a target facilitated responses to the target, but this effect was not seen when seven or more words separated the prime and target. The anaphors in any stimulus set were always equally distant from the end of the sentence; where distance was defined as the number of intervening words.

As indicated above, the no anaphor condition was presented using only the two prime types expected to result in facilitation, that is, the identity and phonological primes. In the no anaphor condition, the "antecedent" was separated from the target by the same number of words as in the gap condition, but there was no anaphor in the second
clause. The ten experimental conditions are shown in Table 2. The complete list of experimental stimuli is displayed in Appendix B. The "[e]" represents the position of the gap, and the "/" divides the sentences into the clauses that were presented one at a time on the computer screen.

**Procedure**

The procedure was identical to that of Experiment 1, except for the modifications necessitated by the fact that the prime stimuli were sentences rather than single words. The sentences were presented clause-by-clause to ensure that subjects could not re-read the antecedent when the anaphor was presented. Subjects were instructed to read the first clause of each sentence, and then press the (+) button adjacent to the number pad to cause the next clause to appear. Once they had finished reading the second clause, subjects pressed the (+) button again, causing the target word to appear. They were instructed to decide if the target was a real English word as quickly and as accurately as possible. The "yes" and "no" responses were made in the same way as in the first experiment.

The task used to encourage the subjects to read the sentences carefully was similar to that used in Experiment 1. Subjects were presented with lists of sentences, and had to decide if they had read each sentence or not. These lists were presented following the practice trials, and during the break allowed after the 80th sentence-target
Table 2

Experiment 2: Example Stimulus Set

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Prime Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Anaphor</td>
<td>Identity</td>
<td>The bear proved equal to the task/ and the prey died quickly.</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>The snare proved equal to the task/ and the prey died quickly.</td>
</tr>
<tr>
<td>Noun Gap</td>
<td>Identity</td>
<td>The bear proved equal to the task/ and [e] killed the unlucky prey.</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>The snare proved equal to the task/ and [e] killed the unlucky prey.</td>
</tr>
<tr>
<td></td>
<td>Orthographic</td>
<td>The spear proved equal to the task/ and [e] killed the unlucky prey.</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>The trap proved equal to the task/ and [e] killed the unlucky prey.</td>
</tr>
<tr>
<td>Pronoun</td>
<td>Identity</td>
<td>The bear proved equal to the task/ and it killed the unlucky prey.</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>The snare proved equal to the task/ and it killed the unlucky prey.</td>
</tr>
<tr>
<td></td>
<td>Orthographic</td>
<td>The spear proved equal to the task/ and it killed the unlucky prey.</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>The trap proved equal to the task/ and it killed the unlucky prey.</td>
</tr>
</tbody>
</table>

Note: The target that followed all ten conditions shown above was BEAR.
pair. Subjects were again allowed to pause as long as they felt necessary during the break.

Although efforts were made to keep the sentences corresponding to any one target as similar as possible, they obviously differed across experimental conditions. For this reason, 50 of the subjects were presented with a rating scale designed to determine if the sentences used as priming stimuli were equivalent with respect to how sensible they were. This rating scale was always presented at the end of the experimental session. For each of the 50 subjects, the rating scale consisted of the same 80 experimental sentences that they had read in the lexical decision task. Subjects were asked to rate these sentences from 1 to 5; "1" indicated a rating of "makes no sense", and "5" indicated "makes perfect sense". It was hoped that this measure would be sensitive to any consistent differences in the stimuli that might arise across the experimental conditions.

Results

The cutoff replacement procedure described in Experiment 1 was used to replace reaction times that were more than three standard deviations from the grand mean. This resulted in replacement of 1.51% of the reaction time data. Six planned contrasts were used to analyze the reaction time and the error data. These were comprised of three contrasts to test the predicted prime effects, and three to test the hypothesized anaphor effects. Because the
orthographic condition showed a trend toward priming in the first experiment, the phonological prime effects were tested by contrasting the phonological condition with the combination of the orthographic and neutral prime conditions. This contrast was applied to the gap and pronoun sentences individually. The identity prime condition was contrasted to the neutral condition alone, and both anaphor types were combined in this contrast. This was done because our predictions specified that only phonological priming might be different for the two anaphor types.

The contrasts designed to investigate the anaphor effects compared the gap and pronoun conditions individually to the no anaphor condition for phonological primes. The gap and pronoun conditions were combined and compared to the no anaphor condition for the identity primes.

The reaction time data for gap and pronoun sentences across all four prime types is pictured in Figure 3. The only prime contrast to reach significance was that for the identity primes in both anaphor conditions. Reaction times following identity primes were significantly faster than those following the neutral primes by both subjects, one-tailed $t(297) = 7.822$, $MSE = 9814.179$, and items, one-tailed $t(237) = 7.154$, $MSE = 11619.303$. However, the phonological priming effects were not tested, as Figure 3 reveals that reaction times following phonological primes were actually
Figure 3. Experiment 2: Mean reaction times to targets following each of the prime types in the noun gap and pronoun sentences.
slower than those following the neutral primes for both gaps and pronouns.

The error data for these sentence types is displayed in Figure 4. Analyses following the square root transformation revealed significant identity priming by subjects, one-tailed $t(297) = 3.468, \text{MSe} = 0.295$, and items, one-tailed $t(237) = 4.485, \text{MSe} = 0.200$. The effect of phonological priming for the sentences containing gaps was not tested, because the effect was in the wrong direction. For the sentences containing pronouns, the phonological priming effect was not significant by subjects, one-tailed $t(297) = 0.904, \text{MSe} = 0.218$, or items, one-tailed $t(237) = 1.44, \text{MSe} = 0.183$.

The effect of the presence of the anaphor on reaction times can be seen in Figure 5. Analyses revealed that responses to the identity primes were not significantly faster in the gap and pronoun conditions than in the no anaphor condition by subjects, one-tailed $t(198) = 1.05, \text{MSe} = 5343.52$, or by items, one-tailed $t(158) = 1.20, \text{MSe} = 8107.539$. For the phonological primes, targets following the gap sentences were responded to faster than those following the no anaphor sentences by subjects, one-tailed $t(198) = 1.76, \text{MSe} = 5343.52$, but not by items, one-tailed $t(158) = 0.432, \text{MSe} = 8107.539$. The phonological effect for the pronoun sentences was not tested, because reaction times were in the opposite direction from that predicted.
Figure 4. Experiment 2: Mean percentage of errors to targets following each of the prime types in the noun gap and pronoun sentences.
Figure 5. Experiment 2: Mean reaction times to targets following identity and phonological primes in the anaphor and no anaphor sentences.
Figure 6 presents the error data for these conditions. The square-root transformation was applied before the analysis. For the identity primes, no more errors were made following the no anaphor sentences than following the combination of the noun gap and pronoun sentences by subjects, one-tailed $t(198) = 0.94$, MSE = 0.186, or by items, one-tailed $t(158) = 0.98$, MSE = 0.133. For the phonological primes, the effect of gaps over no anaphor was not tested, because the effect found was in the opposite direction from that predicted. The effect of pronouns for the phonological primes was not significant by subjects, one-tailed $t(198) = 0.88$, MSE = 0.186, or by items, one-tailed $t(158) = 0.57$, MSE = 0.133.

The rating scale data was analyzed with two analyses of variance. One 2(anaphor type) X 4(prime type) analysis included the two anaphor levels of the sentence factor and all four levels of the prime factor. The other 3(anaphor type) X 2(prime type) analysis tested all levels of the sentence factor and only the identity and phonological prime types.

The 2 X 4 analysis revealed a main effect of sentence type, by subjects $F(1,49) = 4.985$, but not by items, $F(1,79) = 1.474$. The main effect of prime type was significant by subjects $F(3,147) = 17.044$, and by items, $F(3,237) = 6.225$. The interaction was not significant by subjects, $F(3,147) = 1.042$, or by items, $F(3,237) = 0.822$. Figure 7 shows that
Figure 6. Experiment 2: Mean percentage of errors to targets following identity and phonological primes in the anaphor and no anaphor sentences.
Figure 7. Experiment 2: Mean sensibility ratings of noun gap and pronoun sentences with each prime type.
the main effect of sentence type in the subject analysis can be attributed to the fact that gaps were rated as making more sense than pronouns. Post-hoc analysis using Tukey's Honestly Significant Difference procedure revealed that the main effect of prime type was attributable to the fact that the sentences containing neutral primes were rated as more acceptable than the sentences containing all other primes types, by the subject analysis, all $p$'s $< 0.01$, and the neutral primes were rated as more acceptable than the orthographic and phonological primes in the item analysis, all $p$'s $< 0.01$. Furthermore, the subject analysis revealed that the sentences containing identity primes were rated as more sensible than the sentences containing orthographic primes, $p < 0.05$, but this effect was not replicated in the item analysis.

Figure 8 shows the rating data relevant to the 3 X 2 analysis. There was no significant interaction between the prime type and sentence type factors by subjects, $F(2,98) = 0.779$, or by items, $F(2,158) = 0.186$. The analysis revealed a main effect of sentence type by subjects, $F(2,98) = 50.46$, and by items, $F(2,158) = 29.783$. Tukey's Honestly Significant Difference procedure revealed that this effect was attributable to the fact that both the gap and pronoun sentences were rated as more sensible than the no anaphor sentences in the subject and item analyses (all $p$'s $< 0.01$). The main effect of prime type was also significant by
Figure 8. Experiment 2: Mean sensibility ratings of no anaphor, noun gap, and pronoun sentences with identity and phonological primes.
subjects, \( F(1,49) = 11.683 \), and was marginally significant by items, \( F(1,79) = 3.755, \ p = 0.0562 \). Sentences containing identity primes were rated as more acceptable than sentences containing phonological primes.

Because the sensibility ratings differed across prime types and sentence types, a 2 X 4 analysis of covariance was performed on the reaction time data of the 50 subjects who had completed the rating scale, with sensibility rating as the covariate. The significance tests that resulted from this analysis are not interpretable because the number of subjects is only half of those actually sampled. However, the marginal means for these 50 subjects over the eight experimental conditions before and after adjustment by the covariate were compared. Table 3 reveals that including the covariate had very little influence on the mean reaction times.

**Discussion**

The purpose of Experiment 2 was to investigate the hypothesis that phonological representations of the antecedent are present during written anaphor resolution. It was predicted that lexical decisions to target words would be primed by sentences that contained anaphors referring to antecedents that rhymed with these targets. The results of Experiment 2 failed to provide any support for this hypothesis. Not only were the expected phonological priming effects nonexistent, but there was no
### Table 3

**Experiment 2: Reaction Time Data of 50 Subjects Before and After Analysis of Covariance**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Marginal Cell Means (ms)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence Type</td>
<td>Prime Type</td>
<td>Non-Adjusted</td>
</tr>
<tr>
<td>Noun gap</td>
<td>Identity</td>
<td>776.948</td>
<td>776.796</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>863.713</td>
<td>864.033</td>
</tr>
<tr>
<td></td>
<td>Orthographic</td>
<td>859.707</td>
<td>860.362</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>833.205</td>
<td>830.362</td>
</tr>
<tr>
<td>Pronoun</td>
<td>Identity</td>
<td>775.655</td>
<td>775.353</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>896.314</td>
<td>897.960</td>
</tr>
<tr>
<td></td>
<td>Orthographic</td>
<td>867.042</td>
<td>868.509</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>848.037</td>
<td>846.713</td>
</tr>
</tbody>
</table>
evidence to suggest that anaphors prime targets under any of the prime conditions tested.

The lack of both the phonological priming effect and the anaphor priming effect will be discussed in turn. The information provided by the sensibility ratings will also be discussed, as these ratings gave strong indications that the experimental manipulations changed the sentences in unexpected ways.

Sensibility Ratings

The sensibility ratings provide interesting information regarding the sentences constructed in this experiment. It was hoped that consistent differences between the ratings of the various experimental conditions would not be found. However, it is important to note that the ratings of the two control conditions, neutral and no anaphor, were significantly different from the ratings of the other levels of their respective factors.

These differences are of potential importance if they affect the subjects' responses to the targets. It seems reasonable to suppose that sentences that make less sense are more difficult to comprehend. Despite the fact that the presentation of the sentences was self-paced, it is possible that the processing of the sentence was still going on when the target was presented. The processing of less sensible sentences may have slowed the subjects' responses to the targets. Thus, any prime type effects in this study may
have been masked by the fact that sentences containing phonological and orthographic primes were rated as being less sensible than sentences containing neutral primes. Furthermore, anaphor effects could have been falsely enhanced by the fact that the no anaphor sentences were rated as less sensible than both the noun gap and pronoun sentences. It is important to keep these potential confounds in mind throughout the discussion of the phonological priming and anaphor effects which follows.

As informative as the sensibility ratings are, they do not provide us with a mechanism for controlling for the differences across the experimental conditions. The analysis of covariance carried out for the 50 subjects who had completed both the lexical decision and the rating tasks suggested that the covariate resulted in very little adjustment of the cell means. This may be taken as an indication that the sensibility of the sentence did not much affect reaction times to the targets.

Phonological Priming

The results of Experiment 2 did not replicate Tanenhaus et al.'s (1985) findings, which indicated that anaphors were able to prime lexical decisions to targets that rhymed with their antecedents. However, the studies differed in a number of respects which may account for the inconsistent results.
Firstly, Tanenhaus et al. (1985) tested the effects of wh-gaps, rather than noun gaps and pronouns. It is possible that wh-gaps are not processed the same way as other anaphor types (Nicol & Swinney, 1989). In sentences containing wh-gaps, there is generally a wh-word preceding the gap. This word may serve as a cue to the reader or listener that a gap will eventually occur in the sentence. Readers or listeners who are alerted to the fact that they will have to fill a gap may attempt to retain items from the sentence more than they normally would. Phonological codes may be used to aid this retention, causing wh-gaps to exhibit rhyme-priming effects. If this is the case, then the use of phonological representations would be specific to situations in which the parser can predict that an anaphor will occur. This explanation seems tenable, but it is inconsistent with the findings of researchers who have reported phonological effects for other types of anaphors, using other paradigms (Black et al., 1987; Leonard, 1991).

Another difference between the Tanenhaus et al. (1985) study and Experiment 2, was the interval between the presentation of the anaphor and the target. The wh-gaps in the Tanenhaus et al. experiment occurred at the end of the sentences, 400 ms before the targets were presented. In Experiment 2, there were up to four words between the noun gaps or pronouns and the end of the sentence, which was immediately followed by the target. Hudson and Tanenhaus's
(1985) results suggest that words that occur within four words of the end of a sentence can prime targets that are presented immediately after the sentence. However, these results were found using nouns, rather than gaps. It is possible that any representations activated during anaphor resolution decay faster than representations that follow actual nouns. Furthermore, the sentences in the Hudson and Tanenhaus study were presented auditorily, which may have resulted in longer lasting phonological representations of the prime words.

Tanenhaus et al. (1985) also presented their sentences auditorily. In Experiment 2, the sentences were presented visually so that conclusions could be made regarding the processing of written material containing anaphors. The failure to find phonological priming effects under these conditions suggests either that the paradigm used was not sensitive to the presence of these representations, or that when sentences are presented visually, phonological representations of the antecedent are not available at the end of the sentences, even when an anaphor occurs within four words of this point.

In Experiment 1, we found that lexical decisions to visually presented targets were influenced by phonological relationships to visually presented primes. This suggests that the most appropriate interpretation of the null effect found in Experiment 2 is that there were no phonological
representations of the antecedent available at the end of the sentences to prime the lexical decisions to the targets. However, we have failed to rule out the possibility that these phonological representations were available at some point between the presentation of the anaphor and the end of the sentence. In Experiment 1, there was no delay between the presentation of the prime word and the presentation of the target. In Experiment 2, there was no delay between the presentation of the second clause and the presentation of the target. However, the words that intervened between the anaphor and the end of the sentence were likely to have caused some unknown delay.

**Anaphor Priming**

Strong identity priming was seen in both the reaction time and the error data. However, this effect could not be attributed to the presence of an anaphor, because responses to the targets following sentences containing identity primes did not differ across the three levels of sentence type. Responses following sentences with no anaphors were just as fast and accurate as responses following sentences with noun gaps or pronouns. Thus, it appears that the identity priming found is due to the antecedent, and not the anaphors.

This finding is particularly troublesome, as we have no evidence from Experiment 2 that anaphors result in any reactivation of their antecedents. There were no
differences between the anaphor and no anaphor conditions that were consistent in both the subject and item analyses. A great deal of research has indicated that anaphors prime targets that hold some relationship with their antecedents, but most of this research has made use of the probe recognition task. Furthermore, in almost all of the studies reported which have investigated these effects using the lexical decision task, the prime stimuli were presented auditorily (Marslen-Wilson & Tyler, 1980; Nicol, 1988; Nicol & Swinney, 1989; Tanenhaus et al., 1985).

One notable exception to this generalization is a lexical decision experiment performed by Cloitre and Bever (1988). With visual presentation of their prime and target stimuli, they found that lexical decisions to targets that followed prime sentences were facilitated by repeated noun anaphors and pronouns relative to a no anaphor control condition. However, in the lexical decision version of this experiment, facilitation was greater for the repeated noun anaphors than for the pronouns. The same stimuli were used with a probe recognition paradigm, and greater facilitation was found for pronouns than for repeated noun anaphors. Thus, it may be that lexical decisions are not particularly sensitive to the processes involved in the resolution of written material containing the types of anaphors investigated in Experiment 2. The probe recognition task
may provide a much more sensitive measure of the reactivation of antecedent information by noun gaps and pronouns. Experiments 3 and 4 were designed to investigate this possibility.

Experiment 3

Some of the most robust evidence to suggest that both gaps and pronouns result in reactivation of antecedent information has come from the work of Bever and his colleagues (Bever & McElree, 1988; Cloitre & Bever, 1988; McElree & Bever, 1989). In particular, McElree and Bever have shown that probe recognition responses are faster following sentences containing pronouns, or a variety of gaps, than following matched control sentences with no anaphors. They presented subjects with sentences containing either PRO gaps, NP-raising gaps, NP-tough gaps, or pronouns, and contrasted these to control sentences. Probes were presented at two positions, either shortly after the gap/pronoun, or at the end of the sentence. The sentences were presented visually, in phrases that subjects read at a self-paced rate. One stimulus set can be seen in examples 25a-e. The phrases are indicated by slashes, the two probe positions are indicated by a superscripted "p1" or "p2", and the gaps are indicated by "[e]".
25a) PRO gap: The stern judge/ who met with the defense/ adamantly refused [e] to/ \(p_1\)/ argue about the appeal./ \(p_2\)

25b) NP-raising gap: The stern judge/ who met with the defense/ is sure [e] to/ \(p_1\)/ argue about the appeal./ \(p_2\)

25c) NP-tough gap: The stern judge/ was difficult for the defense/ to argue with [e] about/ \(p_1\)/ the pending appeal./ \(p_2\)

25d) Pronoun: The stern judge/ who met with the defense/ thought he should/ \(p_1\)/ argue about the appeal./ \(p_2\)

25e) No anaphor: The stern judge/ who met with the defense/ flatly rejected the/ \(p_1\)/ arguments for an appeal./ \(p_2\)

PROBE: STERN

McElree and Bever (1989) compared each anaphor type individually to the no anaphor control. At the first probe position, \(p_1\), no differences in probe recognition latencies were found. However, at the probe position at the end of the sentence, \(p_2\), recognition latencies were faster following each anaphor type, than they were following the no anaphor control. This led McElree and Bever to conclude that both the gap constructions and the pronouns that they had tested access antecedent information by the end of the sentence.

In Experiment 3, sentences containing noun phrase-raising gaps and sentences containing pronouns were contrasted to no anaphor control sentences in their ability to facilitate the recognition of probes that were nouns from
the antecedent noun phrase. Given the findings of McElree and Bever (1989), it was predicted that both anaphor types would be shown to facilitate responses to the probes when compared to the no anaphor control condition.

**Method**

**Subjects**

Thirty right-handed college and university students between the ages of 17 and 30 years participated in the study. All subjects were native English speakers and were paid for their participation.

**Materials**

The stimuli in Experiments 3 and 4 were modelled after those of McElree and Bever, Experiment 1 (1989), with some modifications. Only one factor was tested, sentence type, with three levels: no anaphor; NP-raising gap; and pronoun. Because McElree and Bever only found priming at the end of the sentences, that was the only probe position tested in Experiments 3 and 4.

The first noun phrase of the experimental sentences was the antecedent in the anaphor conditions, and held the noun that would act as the probe for all three sentence types. These nouns were chosen so that they could eventually be matched to other words fulfilling the four prime conditions described in Experiments 1 and 2; however, the prime type manipulation was not included in Experiments 3 and 4. The result of restricting the choice of antecedent nouns in this
way was that many of the nouns were inanimate objects. Therefore, while the no anaphor and anaphor sentences in this experiment contained relative clauses, as in McElree and Bever (1989), the pronoun condition consisted of sentences formed by conjoining two sentences with "and", "but", "yet", or "because". This change was necessitated by the fact that the pronoun sentences with relative clauses in the McElree and Bever study required animate antecedents.

McElree and Bever (1989) presented their sentences phrase by phrase. It was not clear from their study how a phrase was defined; however, their frequent references to Cloitre's unpublished dissertation (cited in McElree & Bever, 1989) suggested that it might have been defined as in Cloitre and Bever (1988). In this paper, a phrase was defined according to the criteria of Janus and Bever (1985). This resulted in phrases that were between two and five words in length, divided into the major sentential constituents. The sentences for Experiments 3 and 4 were divided following these guidelines, with some added restrictions, outlined below. This resulted in sentences consisting of four to six phrases. Each phrase was two to five words long. All sentences had to fit the following pattern:

A B C D

where:

A = the first noun phrase of the sentence, i.e. the
antecedent in the two anaphor conditions.

B₁ - B₃ = phrases forming the relative clause, or the corresponding elements in the pronoun sentences.
C = the verb phrase, which included the anaphor in the gap and pronoun conditions.
D = the object, or prepositional phrase following C.

For any given stimulus set, the number of words between the antecedent and the end of the sentence was the same across sentence types. For the two anaphor conditions, the number of words between the anaphor and the end of the sentence was also equated, and was never more than five words, as in Bever and McElree (1988). Furthermore, the number of phrases was the same, and the "D" phrase was identical, across the experimental conditions within a given stimulus set. These measures assured that the sentences were as similar as possible following the anaphor.

It should be noted that in McElree and Bever's (1989) stimuli, the last phrase of the sentences contained a verb associated with the antecedent in the anaphor conditions, but not in the no anaphor condition. For example, in 25b and 25d, the verb "argue" appears in the last phrase, however it is replaced by the noun "argument" in 25e. This resulted from the fact that the no anaphor controls were created by changing the verb in the final clause of the sentence to its nominal form. In order to eliminate this
potential confound, the control stimuli in Experiments 3 and 4 were created so that this verb was present in all three sentence types, and it occurred in the penultimate phrase across experimental conditions. Table 4 shows one stimulus set. The phrases are marked by slashes, and the gaps are marked by "[e]". The complete list of stimuli is presented in appendix C.

Three-hundred and fifteen filler sentences were also created. These fillers included passages that were two sentences in length, to discourage the anticipation of probes. As well, words from positions other than the first noun phrase were probed. The fillers also included probes that were not in the sentences, requiring a "no" response. These negative probes were always semantically related to a word in the sentence, so that it was not possible for subjects to reject probes based on their lack of relatedness to the content of the sentences. Fillers included sentences containing NP-raising gaps, pronouns, and no anaphor. In total, subjects were presented with 360 passages, one or two sentences long, of which 50% were followed with a probe from the passage (i.e., a "yes" probe). Only 12.5% of the passages were experimental items. It is not unusual to find that most of the stimuli in such experiments are fillers (eg., Cloitre & Bever, 1988; MacDonald, 1989; McElree & Bever, 1989).
### Table 4

**Experiments 3 and 4: Example Stimulus Set**

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Anaphor</td>
<td>The nurse/ who remained at his bedside/ day and night/ obviously loved/ the soldier./</td>
</tr>
<tr>
<td>NP-Raising Gap</td>
<td>The nurse/ who remained at his bedside/ all night/ seemed [e] to love/ the soldier./</td>
</tr>
<tr>
<td>Pronoun</td>
<td>The nurse/ remained at his bedside/ all night/ because she obviously loved/ the soldier./</td>
</tr>
</tbody>
</table>
Comprehension questions were prepared for 25% of the stimuli to encourage subjects to read the sentences, as in McElree and Bever (1989). The questions were always presented after the probes, and always followed filler passages. The comprehension questions required a yes/no response.

Procedure

Each session began with a practice consisting of ten passage-probe pairs with comprehension questions following three of the passages. Subjects were encouraged to ask any questions they had before continuing.

The presentation of the stimuli was similar to that in Experiment 2. Stimuli were presented in the middle of the computer screen. All passages were preceded by a fixation point, which consisted of an asterisk, that appeared on the screen for 250 ms, and was followed immediately by the first phrase of the passage. The interval between the last response (to the probe or comprehension question) of one trial, and the fixation point signalling the beginning of the next trial, was 1000 ms.

Subjects were instructed to press the (+) key adjacent to the number pad with their left index finger when they had finished reading a phrase. This response caused the next phrase to appear. The sentences were presented with normal upper and lower case characters. Probes were presented in upper case print. Subjects were instructed to respond "yes"
to probes that had appeared in the previous passage, and
"no" to probes that had not. These responses were made on a
Micro Experimental Laboratory response box attached to an
IBM Personal System 2 computer. They responded "yes" by
pressing the right-most key of the four-key response box
with the middle finger of their right hand. "No" responses
were made by pressing the key immediately to the left of
this one with the index finger of their right hand.

Subjects were advised that once in a while a
comprehension question would appear following the probes.
The questions appeared surrounded by a box, so that they
could be distinguished from the passages. Subjects were
told that they would not know when a question would appear;
therefore they were to read all passages well enough to be
able to answer questions about them. Subjects were
instructed to answer these yes/no questions with the
response keys that they used to respond to the probes.

Testing took place over two sessions, which lasted
approximately 40 minutes each. These sessions took place on
different days. Each session was broken up into three
blocks with 60 passages per block. The stimuli were
randomly divided into blocks, with the restriction that each
block contained seven or eight experimental stimuli and
fifteen comprehension questions. Within a block, stimuli
were presented in a fixed, random order.
Results

The replacement procedure described in the previous experiments resulted in the replacement of 1.41% of the reaction time data. As in McElree and Bever, each anaphor type was compared individually to the no anaphor control. Two planned comparisons were used for this purpose. The mean reaction times to probes following the three sentence types can be seen in Figure 9. Although a trend towards facilitation following both types of anaphor is evident, these effects were not significant for the gap sentences by subjects, one-tailed $t(58) = 0.66$, $MSe = 4151.282$, or by items, one-tailed $t(88) = 0.35$, $MSe = 8385.099$.

Furthermore, the effect was not significant for the pronoun sentences by subjects, one-tailed $t(58) = 1.64$, $MSe = 4151.282$, or by items, one-tailed $t(88) = 1.28$, $MSe = 8385.099$.

The percentage of mean errors for each sentence type is depicted in Figure 10. The square-root transformation was applied to the error data before analysis. Again, the effect of the presence of the NP-raising gap is not significant by subjects, one-tailed $t(58) = 0.736$, $MSe = 0.308$, or by items, one-tailed $t(88) = 1.12$, $MSe = 0.221$. However, probes were responded to with significantly fewer errors following pronoun sentences, than following no anaphor sentences by subjects, one-tailed $t(58) = 2.19$, $MSe = 0.308$, and by items, one-tailed $t(88) = 2.78$, $MSe = 0.221$. 
Figure 9. Experiment 3: Mean reaction times to probes following the three sentence types.
Figure 10. Experiment 3: Mean percentage of errors to probes following the three sentence types.
Anaphor Resolution

Discussion

Experiment 3 provides the first evidence in this series that at least one type of anaphor results in the reactivation of antecedent information. Significantly fewer errors were made in the recognition of probes following pronoun sentences than following no anaphor control sentences. However, McElree and Bever's (1989) results were not entirely replicated, because the NP-raising gaps in Experiment 3 did not result in any significant facilitation of the antecedent probes. Furthermore, McElree and Bever's results were seen in the reaction time data, while the error data was not compared across the anaphor and no anaphor conditions in their study.

One difference between the two studies which may account for the weaker effects found in this experiment is the different way in which the sentences were parsed into phrases. As discussed above, the parsing of the McElree and Bever (1989) stimuli resulted in systematic differences between the two anaphor sentence types under discussion, and the no anaphor sentences. This resulted in the last phrase of the NP-raising gap and pronoun conditions containing a verb whose subject was the antecedent, while the last phrase of the no anaphor sentences did not contain such a verb. In the present experiment, the sentences were divided so that the corresponding verb occurred in the penultimate phrase across all sentence types, and the last phrase was identical
across sentence types. These measures result in an arguably better control condition. The presence of the verb in the last phrase of the anaphor sentences, but not the control sentences, could confound interpretation of the anaphor effects. For example, the verb may affect the reader's discourse focus such that the antecedent is more strongly in focus because of the verb, rather than because of the presence of the anaphor. Thus, it is possible that McElree and Bever's anaphor effects were inflated by this confound.

Another difference between the two experiments was in the choice of the word to be probed. McElree and Bever (1989) used an adjective that modified the antecedent noun rather than the antecedent noun itself. Although they provided no justification for this choice, it is conceivable that stronger effects would be seen with the adjective probe. If the activation of the antecedent noun has not decayed sufficiently by the end of the sentence, regardless of the sentence type, then it is possible that the anaphors would not add significantly to a level of activation that is still near ceiling. However, the modifier may decay much faster than the noun itself. Thus, reactivation by the anaphor would have the potential to change the activation level of the modifier sufficiently to be evident in the response measures. If this is the case, then one might expect to find stronger reactivation effects for antecedent
noun probes with increased distance between the antecedent and the anaphor.

The results of this experiment suggest that only pronouns result in reactivation of antecedent information, relative to a no anaphor control. The pronouns used in Experiment 3 are deep anaphors, and the NP-raising gaps are surface anaphors (Hankamer & Sag, 1976). Although differences were found for the two types of anaphor, they were not predicted. It was expected that both anaphor types would result in priming of the antecedent targets, because these identity targets shared all information, "deep" and "surface", with the antecedents.

The fact that in Experiment 3, significant facilitation resulted only following the pronoun sentences, and not the gap sentences, poses somewhat of a problem. Because the structure of the pronoun and no anaphor sentences was quite different in this experiment, the results may be due to different processing loads posed by the different sentence structures. Thus, it may not be appropriate to attribute this effect to the reactivation of the antecedent by the anaphor. Nevertheless, Experiment 4 was undertaken to investigate whether this effect could be replicated using the lexical decision paradigm. It was predicted that the facilitation of responses to probes by pronouns that was found with the probe recognition task would not be replicated using the lexical decision task.
Experiment 4

Method

Subjects

Thirty college and university students fulfilling the same requirements as those outlined in Experiment 3 participated in the study. None of the subjects had participated in any previous experiments. All subjects were paid for their participation.

Materials

The experimental passages and targets were the same as those used in Experiment 3. The filler stimuli had to be slightly changed for the lexical decision task. The 180 "no" trials from Experiment 3 consisted of real word probes that were not in the passages. Ninety of these probes were changed to nonword targets for Experiment 4 by changing just one letter in the word. This left 90 targets that were real words, but were not in the preceding passage. Ninety more filler passages were created. These were followed by nonwords formed by changing one letter of a word that had occurred in the passage. This assured that visual similarity to words in the passages was not predictive of word status. Five of the targets in the practice trials were also changed to nonwords. All nonword targets were orthographically legal in English. These changes resulted in a total of 450 passages, excluding the practice trials,
of which 60% were followed by real English words, and 40% by nonwords. Experimental trials accounted for 10% of the passages.

Twenty-two new comprehension questions were also created, so that 25% of the trials were followed by questions. The new stimuli were randomly added to the three versions of the experimental stimuli created for Experiment 3, with the restriction that each block of these versions contained 75 passages, and approximately equivalent numbers of comprehension questions.

Procedure

The procedure was almost identical to that of Experiment 3. Rather than deciding if the target was in the previous passage or not, subjects were told to decide if the target was a real English word or not. This was the only change made in the instructions to the subjects. The only other change in the procedure was that instead of six blocks of 60 stimuli each, the passages were presented in six blocks of 75 stimuli each. In all other respects, the procedures for the two experiments were the same.

Results

The data were first analyzed using the same contrasts that were used in Experiment 3; then analyses combining the data from Experiments 3 and 4 were performed. The results of these analyses follow.
Experiment 4

The replacement procedure resulted in the replacement of 1.79% of the reaction time data. The contrasts performed in Experiment 3 were planned for the reaction time and error data from the lexical decision task of Experiment 4, as well. Mean reaction times to the experimental targets following the three sentence types are presented in Figure 11. Reaction times were actually slower following the two anaphor conditions, than following the no anaphor condition. Because the differences were predicted to be in the opposite direction, the effects were not tested statistically.

The error data are depicted in Figure 12. The square-root transformation was again applied before analysis. Only the difference between the NP-raising gap and no anaphor conditions was tested, as there was obviously no difference between the pronoun and no anaphor sentences. The number of errors made to targets following sentences containing the gap construction was not significantly different from that following sentences with no anaphor by subjects, one-tailed $t(58) = 0.287$, $MSe = 0.07$, or by items, one-tailed $t(88) = 0.153$, $MSe = 0.072$.

Comparison of Data from Experiments 3 and 4

It should be noted that in Experiment 4, only seven out of the 30 subjects made any errors at all to the experimental stimuli. In Experiment 3, 21 of the 30 subjects made errors. Because so few false negative errors
Figure 11. Experiment 4: Mean reaction times to targets following the three sentence types.
Figure 12. Experiment 4: Mean percentage of errors to targets following the three sentence types.
were made, particularly in Experiment 4, it was decided to observe the number of false positive errors to the filler targets. This would ensure that the subjects were not exhibiting a response bias. Table 5 shows the false positive errors to the filler stimuli and the false negative errors to the experimental stimuli for both the probe recognition data from Experiment 3, and the lexical decision data from Experiment 4. While there is a tendency for subjects to make more false positive errors in both experiments, it appears that in general, very few errors are made. Thus, the very low percentage of errors made in Experiment 4 probably results from the fact that the lexical decision task is easy, rather than from a response bias.

Planned contrasts, designed to compare the interaction effect of the probe recognition and lexical decision tasks on probe/target responses, were tested on the residuals of the cell means with the row effects, the column effects, and the grand mean removed (Rosenthal & Rosnow, 1985). The residual means for the reaction time data and the error data can be seen in Figures 13 and 14, respectively. The item analyses using these planned contrasts are relatively straightforward, because both the sentence type and task factors are within items manipulations. However, the subject analyses are complicated by the fact that the task factor is a between subjects manipulation. Rosenthal and Rosnow suggest that to test planned contrasts involving
### Table 5

Experiments 3 and 4: Distribution of Errors

<table>
<thead>
<tr>
<th>Experiment</th>
<th>% False Positive Errors</th>
<th>% False Negative Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Filler Stimuli)</td>
<td>(Experimental Stimuli)</td>
</tr>
<tr>
<td>3</td>
<td>9.63</td>
<td>4.888</td>
</tr>
<tr>
<td>4</td>
<td>2.35</td>
<td>0.814</td>
</tr>
</tbody>
</table>
Figure 13. Experiments 3 and 4: The effect of sentence type and task on reaction time.
Figure 14. Experiments 3 and 4: The effect of sentence type and task on errors.
interaction effects with mixed sources of variance, the $M_{se}$ values for all effects that are involved in the contrast must be taken into account. Because the $M_{se}$ values of the effects involved in the subject analysis are too divergent to be aggregated, analyses were performed using the smallest and largest $M_{se}$ values, as suggested by Rosenthal and Rosnow. The results of these analyses on the reaction time data are displayed in Table 6. It should be noted that using the smallest $M_{se}$ value results in an $F$ statistic that is overestimated, whereas using the largest $M_{se}$ value results in an $F$ statistic that is underestimated. Because the two estimates give inconsistent interpretations of the data, one cannot conclude whether or not the interaction effect is statistically significant by subjects. However, the item analysis is not significant.

Table 7 shows the same analyses performed on the transformed error data. In this analysis, neither of the subject analyses are significant, nor is the item analysis significant. In fact, the low $F$ values suggest a problem in the data, which is likely to be the very small number of errors found with the lexical decision task.

As well as the reaction time and error data to the targets, the amount of time required to read the sentences, and the accuracy of the responses to the comprehension questions were collected in Experiments 3 and 4. This data was analyzed in the hopes that it would help to clarify
Table 6

Experiments 3 and 4: Testing the Interaction Effect on the Reaction Time Data

<table>
<thead>
<tr>
<th>Analysis</th>
<th>MSE</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Smallest = 3780.475</td>
<td>$F(1, 58) = 4.48, \ p &lt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>Largest = 78701.848</td>
<td>$F(1, 58) = 0.215, \ p &gt; 0.05$</td>
</tr>
<tr>
<td>Item</td>
<td>5786.549</td>
<td>$F(1, 88) = 2.928, \ p &gt; 0.05$</td>
</tr>
</tbody>
</table>
Table 7

Experiments 3 and 4: Testing the Interaction Effect on the Error Data

<table>
<thead>
<tr>
<th>Analysis</th>
<th>MSE</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Smallest = 0.07</td>
<td>$F(1, 58) = 0.046$</td>
</tr>
<tr>
<td></td>
<td>Largest = 0.200</td>
<td>$F(1, 58) = 0.016$</td>
</tr>
<tr>
<td>Items</td>
<td>0.128</td>
<td>$F(1, 88) = 0.025$</td>
</tr>
</tbody>
</table>
differences in the task demands. The amount of time required to read the experimental sentences is shown in Figure 15. An analysis of variance performed on this data revealed a main effect of sentence type both by subjects, \( F(2, 116) = 20.791, \text{MSE} = 947.824 \), and by items, \( F(2, 88) = 3.286, \text{MSE} = 9038.613 \). Post-hoc analysis using Tukey's Honestly Significant Difference procedure revealed that in the subject analysis, the sentences containing NP-raising gaps were read faster than sentences containing pronouns, which were read faster than sentences containing no anaphor. However, the item analysis revealed only a significant difference in reading time between the NP-raising gaps and the no anaphor sentences. The main effect of task was significant by items, \( F(1, 44) = 138.992, \text{MSE} = 2744.079 \), but not by subjects, \( F(1, 58) = 2.467, \text{MSE} = 103263.849 \). The interaction of sentence type and task was not significant by subjects, \( F(2, 116) = 0.679, \text{MSE} = 947.824 \), or by items, \( F(2, 88) = 0.486, \text{MSE} = 2257.680 \). Thus, the only consistent effect on reading time found, was that sentences containing NP-raising gaps were read faster than sentences with no anaphor.

The errors made on the comprehension questions are displayed in Figure 16. The sentence type factor is not included in this analysis because the comprehension questions only followed filler trials, and so they differed
Figure 15. Experiments 3 and 4: The effect of sentence type and task on the amount of time taken to read the experimental sentences.
Figure 16. Experiments 3 and 4: The effect of task on the mean percentage of errors made to comprehension questions.
across the sentence types. Analysis\(^1\) revealed no significant differences between the two tasks in the number of errors made to comprehension questions by subjects, \(t(58) = 0.230, \text{MSe} = 0.904\), or by items, \(t(200) = 1.173, \text{MSe} = 0.886\).

**Discussion**

The results of Experiment 4 support our prediction that the lexical decision task does not provide a sensitive measure of the antecedent reactivation effect under the circumstances outlined above. Although Experiment 3 revealed that sentences containing pronouns facilitated probe recognition responses relative to a no anaphor control, this effect was not replicated for lexical decision responses to the same stimuli. However, it should be noted that the planned contrast to test for an interaction between the sentence type and task factors did not reveal a significant interaction for either the reaction time or the error data. This contrast consisted of weights chosen to test the hypothesis that recognition of probes would be facilitated by the pronoun sentences relative to the no anaphor sentences, but that lexical decisions to targets would not differ following the pronoun and no anaphor

---

\(^1\)The lexical decision stimuli included 22 more comprehension questions than the probe recognition stimuli to keep the proportion of stimuli followed by comprehension questions at 25%. Because of the unequal number of comprehension questions in the two experiments, task was treated as a between groups factor in the item analysis.
sentences. Consideration of the residual means plotted for both the reaction time and error data, displayed in Figures 13 and 14, respectively, suggests that a better characterization of the interaction would be that probe responses are facilitated following the pronoun sentences while target responses are inhibited following pronoun sentences.

The possible inhibition of lexical decision responses to targets following sentences containing pronouns is quite puzzling. However, if one were to assume that the lexical decision measures are indeed not sensitive to antecedent reactivation occurring at an earlier point in the sentence, then it is possible that this inhibition is a reflection of some other aspect of the processing of the pronoun sentences relative to the no anaphor control sentences. If this is the case, then it would appear that the pronoun sentences have a general tendency to inhibit responses relative to the no anaphor control sentences used in these experiments. Thus, our concern that the facilitation effect found in Experiment 3 was due to processing factors other than the presence of the anaphor may be unfounded. In fact, it is possible that the observed facilitation of probe responses following pronouns found in Experiment 3 was an underestimate of the actual effect.

A number of points are noteworthy with respect to the lexical decision and probe recognition tasks. Firstly,
subjects made very few errors on the lexical decision task. It was possible that the very low number of false negative responses on the experimental items may have been due to a response bias. However, the number of false positive errors to the filler items was also extremely low, suggesting that a response bias was not occurring (see Table 5). It appears that lexical decision is simply an easy task. Furthermore, subjects tended to make more false negative and false positive errors in the probe recognition experiment than in the lexical decision experiment, suggesting that probe recognition is a more difficult task than lexical decision.

Secondly, because probe recognition requires that subjects pay attention to the words in the sentences, whereas lexical decision does not, it is possible that subjects read the sentences differently in Experiments 3 and 4, depending on which task they were assigned. Therefore, both the amount of time that it took subjects to read the experimental sentences, and the errors that subjects made on the comprehension questions which followed 25% of the passages, were analyzed. Although there was a trend for subjects to take longer to read the sentences in the probe recognition experiment, this effect was not significant in the subject analysis. The only robust effect found in the reading time measures was an effect of sentence type which was attributable to the fact that subjects read the NP-raising gap sentences more quickly than the no anaphor
sentences. This may be due to the fact that although the number of words in the sentence types was equated, the no anaphor sentences tended to contain more content words than the gap sentences. Eye-fixation measures have suggested that normal readers fixate longer on content words than on function words (Just, Carpenter & Wooley, 1982).

Comprehension questions were answered with equal accuracy in the two experiments. If subjects were not reading the sentences as carefully in the lexical decision study, then it might have been predicted that they would make more errors on the comprehension questions with this task. However, this was not the case. The reading time and comprehension question measures do not suggest that subjects read the sentences differently depending upon the task requirements.

**General Discussion**

This series of experiments was undertaken with the aim of investigating the hypothesized role of phonology in the resolution of two types of anaphor. The notion that speech-based codes are involved in the comprehension of written discourse has been both supported (eg. Slowiaczek & Clifton, 1980; Waters et al., 1987) and refuted (eg. Waters et al., 1985) in the literature.
Previous studies have tended to look at the comprehension of entire written passages, and have generally not focused on the comprehension of specific areas in a text that might be under particular processing load constraints. It was posited that given the evidence that anaphor resolution involves the reactivation of previously mentioned elements in a text, phonological codes might be of use in the processing of these types of constructions. Furthermore, some studies that have looked for evidence of phonological code activation in the processing of sentences containing anaphors have provided support for this hypothesis (Black et al., 1987; Leonard, 1991).

Priming paradigms have been used to investigate the potential role of phonology in the resolution of anaphors in spoken discourse (Tanenhaus et al., 1985). It was hoped that this methodology would prove useful in investigations of written discourse. However, despite the fact that phonological priming was established using a lexical decision paradigm in Experiment 1, Experiment 2 provided no evidence that either noun gaps or pronouns primed targets that were phonologically related to the antecedents of these anaphors. Nevertheless, before concluding that phonological representations of the antecedent are not used in anaphor resolution, it should be noted that the anaphor conditions in Experiment 2 did not even facilitate responses to the actual antecedent relative to the no anaphor condition.
This lack of an anaphor effect under the identity prime condition requires one of two conclusions: either anaphors do not result in the activation of antecedent information, or the paradigm we used was not sensitive to these effects.

Given the large number of studies outlined in the introduction that have found evidence for the activation of antecedent information in the processing of written material containing anaphors, we decided to explore the possibility that the lexical decision task is not a particularly sensitive measure of antecedent activation. Experiments 3 and 4 were designed to compare the facilitation attributable to the presence of an anaphor using a probe recognition task versus a lexical decision task. This facilitation was only found for probes following sentences containing pronouns, and only with the probe recognition paradigm.

The results of Experiments 3 and 4 suggest that indeed, the lexical decision task does not provide a good measure for the investigation of antecedent reactivation effects under the conditions in which the stimuli in these experiments were presented. However, a number of studies have found priming effects due to anaphors, for targets which hold some relationship with the antecedents, using the lexical decision paradigm. Most of the studies that have found these effects have involved cross-modal presentation of the stimulus materials; the passages were presented auditorily (Nicol, 1988; Nicol & Swinney, 1989; Tanenhaus
et al., 1985). However, Cloitre and Bever (1988) have found that lexical decisions to targets from the antecedent noun phrase are primed by pronouns and noun anaphors when both the priming passages and the targets are presented visually.

The most consistent difference between the studies that have found priming using the lexical decision paradigm, and our Experiments 2 and 4, which did not, is the interval separating the presentation of the anaphor and the presentation of the target. Some studies have found effects when the target is presented at the end of the sentence, as in Experiments 2 and 4. However, in these studies (Tanenhaus et al., 1985; Cloitre & Bever, 1988) the stimuli were constructed so that the anaphor occurred at the end of the sentence. In our experiments, the anaphor usually occurred three to five words before the end of the sentence. In fact, in Experiment 4, a phrase which was two to three words long separated the phrase containing the anaphor, and the end of the sentence. In other studies where the lexical decision paradigm was used to explore these effects (Nicol, 1988; Nicol & Swinney, 1989), the targets are generally presented soon after the anaphor, and before the presentation of the sentence has been completed.

How important is the inter-stimulus interval? Tanenhaus et al. (1985) report one experiment in which they presented targets at three different time intervals following the anaphor: 0 ms, 600 ms, and 900 ms. The
results indicated that the antecedent information was increasing in activation at the 600 ms position, but by 900 ms after the anaphor, this effect was decreasing. Thus, it appears that the lexical decision task is only influenced by the activation of antecedent information for a very short period after the presentation of the anaphor.

Therefore, it may be that the point at which we presented the target was simply too far downstream from the anaphor to show any effect using the lexical decision task. However, the alternative possibility, that anaphors do not reactivate their antecedents in such a way as to result in significant priming, is also worthy of discussion. Much of the literature that has shown that passages containing anaphors prime probes or targets related to the antecedents, has been criticized for not providing adequate control measures (MacDonald & MacWhinney, 1990). Thus, while anaphor priming effects have been widely reported in the literature, they cannot always be conclusively attributed to the anaphors.

For example, in the McElree and Bever (1989) study, the anaphor and control stimuli differed with respect to the presence of a verb in the final phrase of the sentences. When we attempted to control for this confound, we did not replicate McElree and Bever's results for NP-raising gaps. Admittedly, the experiments differed in other respects that may have caused the inconsistent findings. Probing with the
actual antecedent noun, rather than a modifier from the noun phrase, may have been one important change, as discussed earlier. However, the stimuli in the no anaphor and NP-raising conditions of Experiments 3 and 4 were very similar to those of McElree and Bever, and our inability to replicate their priming effects is troublesome. It is also important to note that anaphor priming effects are often small, and inconsistent across subject and item analyses (MacDonald, 1989; Nicol, 1988; Tanenhaus et al., 1985).

Researchers who plan to make use of the lexical decision task to investigate anaphor resolution should be aware of the possibility that any effects will only be seen shortly after the anaphor. The short duration of these effects do not make their investigation simple. Although the results of these experiments suggest that it might be worthwhile to replicate Experiment 2 with targets presented closer to the anaphor, it is not clear how this should be done when the sentences are presented visually. One option that has been used is to devise the stimuli so that the anaphor occurs at the end of the sentence. Another would be to present the words of the sentence one-by-one, and to present the target before the sentence is complete (Davidson, 1986). These modifications would result in a paradigm which one might expect to be more sensitive to the postulated effects.
Another option available to the researcher intent on investigating potential effects of phonology in written anaphor resolution using a priming paradigm, would be to make use of other tasks. Seidenberg, Waters, Barnes and Tanenhaus (1984) have found that the naming of targets may be more affected by phonological relationships with prime stimuli than lexical decisions to these targets are. Furthermore, the probe recognition task may be useful for investigating phonological effects. Responses to probes which are phonologically related to antecedents might be slowed relative to a number of control conditions. If a phonological interference effect were found, it would suggest that phonological representations of the antecedent were available. Using one of these methodologies, the inter-stimulus interval may remain an important factor. Any role that phonology might play in anaphor resolution may be of very short duration. Thus, time course studies using paradigms similar to those outlined above may provide valuable information with respect to the use of speech-based codes in the processing of written material containing anaphors.

Finally, it is important to note that task differences may influence the way that subjects process stimulus materials. The probe recognition task requires subjects to identify words that have been presented in a text. Therefore, it may induce subjects to attempt to memorize
words from the text, when they would not do so if the objective was simply to comprehend the material, as in "normal" reading. Levy (1978) has shown that the instructions given to subjects will influence whether or not phonology plays a role in reading. The use of speech-based codes seems to improve subjects' performance when they are instructed to retain the verbatim form of written passages. However, when reading for meaning, these codes did not seem to play an important role.

Thus, tasks, like probe recognition, that emphasize verbatim recall, may be more likely to show phonological effects. There was no robust evidence from Experiments 3 and 4 to suggest that subjects read the sentences differently in the probe recognition and lexical decision studies. Nonetheless, task demands will always be an important consideration in the investigation of the processes involved in the comprehension of written material.
References


## Appendix A: Experimental Stimuli for Experiment 1

<table>
<thead>
<tr>
<th>Target/ID PRIME</th>
<th>PHON PRIME</th>
<th>ORTH PRIME</th>
<th>NEUT PRIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) squad</td>
<td>fraud</td>
<td>grad</td>
<td>man</td>
</tr>
<tr>
<td>2) fad</td>
<td>plaid</td>
<td>squad</td>
<td>song</td>
</tr>
<tr>
<td>3) braid</td>
<td>shade</td>
<td>plaid</td>
<td>dress</td>
</tr>
<tr>
<td>4) aid</td>
<td>jade</td>
<td>plaid</td>
<td>plant</td>
</tr>
<tr>
<td>5) swan</td>
<td>fawn</td>
<td>man</td>
<td>boy</td>
</tr>
<tr>
<td>6) wan</td>
<td>fawn</td>
<td>plan</td>
<td>sick</td>
</tr>
<tr>
<td>7) swap</td>
<td>drop</td>
<td>trap</td>
<td>crime</td>
</tr>
<tr>
<td>8) swarm</td>
<td>storm</td>
<td>charm</td>
<td>gang</td>
</tr>
<tr>
<td>9) wart</td>
<td>fort</td>
<td>art</td>
<td>chair</td>
</tr>
<tr>
<td>10) quart</td>
<td>snort</td>
<td>tart</td>
<td>glass</td>
</tr>
<tr>
<td>11) chase</td>
<td>race</td>
<td>phase</td>
<td>trial</td>
</tr>
<tr>
<td>12) squash</td>
<td>slosh</td>
<td>mash</td>
<td>corn</td>
</tr>
<tr>
<td>13) swat</td>
<td>pot</td>
<td>cat</td>
<td>broom</td>
</tr>
<tr>
<td>14) bead</td>
<td>seed</td>
<td>head</td>
<td>rock</td>
</tr>
<tr>
<td>15) thread</td>
<td>sled</td>
<td>bead</td>
<td>glass</td>
</tr>
<tr>
<td>16) mead</td>
<td>feed</td>
<td>bread</td>
<td>oats</td>
</tr>
<tr>
<td>17) beak</td>
<td>sheik</td>
<td>steak</td>
<td>boat</td>
</tr>
<tr>
<td>18) break</td>
<td>lake</td>
<td>leak</td>
<td>hole</td>
</tr>
<tr>
<td>19) freak</td>
<td>sheik</td>
<td>steak</td>
<td>witch</td>
</tr>
<tr>
<td>20) creak</td>
<td>shriek</td>
<td>break</td>
<td>crash</td>
</tr>
<tr>
<td>21) bear</td>
<td>snare</td>
<td>spear</td>
<td>trap</td>
</tr>
<tr>
<td>22) spear</td>
<td>deer</td>
<td>bear</td>
<td>rock</td>
</tr>
<tr>
<td>23)</td>
<td>fear</td>
<td>deer</td>
<td>bear</td>
</tr>
<tr>
<td>24)</td>
<td>tease</td>
<td>fees</td>
<td>grease</td>
</tr>
<tr>
<td>25)</td>
<td>please</td>
<td>freeze</td>
<td>cease</td>
</tr>
<tr>
<td>26)</td>
<td>meat</td>
<td>sheet</td>
<td>sweat</td>
</tr>
<tr>
<td>27)</td>
<td>heat</td>
<td>sleet</td>
<td>threat</td>
</tr>
<tr>
<td>28)</td>
<td>wheat</td>
<td>beet</td>
<td>sweat</td>
</tr>
<tr>
<td>29)</td>
<td>threat</td>
<td>bet</td>
<td>cheat</td>
</tr>
<tr>
<td>30)</td>
<td>great</td>
<td>late</td>
<td>meat</td>
</tr>
<tr>
<td>31)</td>
<td>wreath</td>
<td>teeth</td>
<td>breath</td>
</tr>
<tr>
<td>32)</td>
<td>sheath</td>
<td>teeth</td>
<td>death</td>
</tr>
<tr>
<td>33)</td>
<td>height</td>
<td>might</td>
<td>weight</td>
</tr>
<tr>
<td>34)</td>
<td>weight</td>
<td>gait</td>
<td>height</td>
</tr>
<tr>
<td>35)</td>
<td>where</td>
<td>share</td>
<td>here</td>
</tr>
<tr>
<td>36)</td>
<td>brew</td>
<td>clue</td>
<td>pew</td>
</tr>
<tr>
<td>37)</td>
<td>screw</td>
<td>glue</td>
<td>pew</td>
</tr>
<tr>
<td>38)</td>
<td>crew</td>
<td>glue</td>
<td>few</td>
</tr>
<tr>
<td>39)</td>
<td>two</td>
<td>crew</td>
<td>pro</td>
</tr>
<tr>
<td>40)</td>
<td>pro</td>
<td>foe</td>
<td>two</td>
</tr>
<tr>
<td>41)</td>
<td>hoe</td>
<td>throw</td>
<td>shoe</td>
</tr>
<tr>
<td>42)</td>
<td>toe</td>
<td>blow</td>
<td>shoe</td>
</tr>
<tr>
<td>43)</td>
<td>doe</td>
<td>crow</td>
<td>shoe</td>
</tr>
<tr>
<td>44)</td>
<td>shoe</td>
<td>glue</td>
<td>hoe</td>
</tr>
<tr>
<td>45)</td>
<td>tomb</td>
<td>broom</td>
<td>bomb</td>
</tr>
<tr>
<td>46)</td>
<td>womb</td>
<td>room</td>
<td>bomb</td>
</tr>
<tr>
<td>47)</td>
<td>bombs</td>
<td>psalms</td>
<td>tombs</td>
</tr>
<tr>
<td>48)</td>
<td>gnome</td>
<td>foam</td>
<td>some</td>
</tr>
<tr>
<td></td>
<td>Anaphor Resolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49)</td>
<td>some drum chrome ring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50)</td>
<td>don fawn son chimp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51)</td>
<td>son nun con thief</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52)</td>
<td>blood mud hood mask</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53)</td>
<td>food nude flood mail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54)</td>
<td>mood dude flood job</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55)</td>
<td>brood dude blood song</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56)</td>
<td>tool jewel wool blade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57)</td>
<td>stool rule wool cup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58)</td>
<td>wool bull pool dance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59)</td>
<td>boot suit foot glove</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60)</td>
<td>forms swarms worms pests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61)</td>
<td>pose crows dose dog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62)</td>
<td>ghost boast cost clown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63)</td>
<td>host coast frost inn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64)</td>
<td>post roast cost car</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65)</td>
<td>both oath moth boys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66)</td>
<td>bough prow tough horse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67)</td>
<td>dough hoe plough toy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68)</td>
<td>four store tour band</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69)</td>
<td>stove mauve glove suit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70)</td>
<td>crow pro cow lamb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71)</td>
<td>cow bough crow branch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72)</td>
<td>throw pro vow man</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73)</td>
<td>show foe cow pie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74)</td>
<td>known clone town book</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clue</td>
<td>crew</td>
<td>cue</td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>75</td>
<td>bull</td>
<td>wool</td>
<td>gull</td>
</tr>
<tr>
<td>76</td>
<td>lure</td>
<td>poor</td>
<td>cure</td>
</tr>
<tr>
<td>77</td>
<td>fuse</td>
<td>cues</td>
<td>ruse</td>
</tr>
<tr>
<td>78</td>
<td>brute</td>
<td>toot</td>
<td>mute</td>
</tr>
<tr>
<td>79</td>
<td>flute</td>
<td>fruit</td>
<td>mute</td>
</tr>
</tbody>
</table>
Appendix B: Experimental Stimuli for Experiment 2

The following stimuli are presented with each prime condition in brackets in the following order: identity, phonological, orthographic, and neutral. Each sentence type is presented individually.

1) Target = SQUAD
   No anaphor: The (squad, fraud) fooled everybody/ and the team saved the day.
   Gap: The (squad, fraud, grad, man) fooled the other team/ and [e] saved the day.
   Pronoun: The (squad, fraud, grad, man) fooled the other team/ and it/he saved the day.

2) Target = FAD
   No anaphor: The (fad, plaid) was considered vulgar/ and the store was unappealing to her.
   Gap: The (fad, plaid, squad, song) was considered vulgar by some/ and [e] was unappealing to her.
   Pronoun: The (fad, plaid, squad, song) was considered vulgar by some/ and it was unappealing to her.

3) Target = BRAID
   No anaphor, identity: That braid always looked terrible/ and a pony tail was never worn.
   No anaphor, phonological: That shade always looked terrible/ and the other colour was never worn.
   Gap: That (braid, shade, plaid, dress) always looked terrible on the child/ and [e] was never worn.
   Pronoun: That (braid, shade, plaid, dress) always looked terrible on the child/ and it was never worn.

4) Target = AID
   No anaphor: The (aid, jade) was offered with great kindness/ but she felt indebted.
   Gap: The (aid, jade, plaid, plant) was offered with kindness/ but [e] made her feel indebted.
   Pronoun: The (aid, jade, plaid, plant) was offered with kindness/ but it made her feel indebted.

5) Target = SWAN
   No anaphor: The (swan, fawn) sat by the lake/ and the keeper slept peacefully.
   Gap: The (swan, fawn, man, boy) sat by the perfectly clear lake/ and [e] slept peacefully.
   Pronoun: The (swan, fawn, man, boy) sat by the perfectly clear lake/ and it/he slept peacefully.
6) Target = WAN
No anaphor: The (wan, fawn) worried the specialist/ and the staff monitored everything.
Gap: The (wan, fawn, plan, sick) worried the specialist/ and [e] was/were monitored carefully.
Pronoun: The (wan, fawn, plan, sick) worried the specialist/ and they/it was/were monitored carefully.

7) Target = SWAP
No anaphor: The (swap, drop) went unnoticed/ and the public went about their business without incident.
Gap: The (swap, drop, trap, crime) went unnoticed by the rookie detective/ and [e] was completed without incident.
Pronoun: The (swap, drop, trap, crime) went unnoticed by the rookie detective/ and it was completed without incident.

8) Target = SWARM
No anaphor: The (swarm, storm) was highly destructive/ and the lack of food caused many problems.
Gap: The (swarm, storm, (Her) charm, gang) was highly destructive in the small town/ and [e] caused many problems.
Pronoun: The (swarm, storm, (Her) charm, gang) was highly destructive in the small town/ and it caused many problems.

9) Target = WART
No anaphor, identity: The wart was very ugly/ but the skin could not be removed.
No anaphor, phonological: The wart was very ugly/ but the site could not be removed.
Gap: The (wart, fort, art, chair) was ugly in her opinion/ but [e] could not be removed.
Pronoun: The (wart, fort, art, chair) was ugly in her opinion/ but it could not be removed.

10) Target = QUART
No anaphor: A (quart, snort) seemed sufficient/ but George was not satisfied.
Gap: A (quart, snort, tart, glass) seemed sufficient/ but [e] did not satisfy him.
Pronoun: A (quart, snort, tart, glass) seemed sufficient/ but it did not satisfy him.
11) Target = CHASE
No anaphor: The (chase, race) lasted too long/ and the hunters were annoyed.
Gap: The (chase, race, phase, trial) lasted much too long/ and [e] was very annoying.
Pronoun: The (chase, race, phase, trial) lasted much too long/ and it was very annoying.

12) Target = SQUASH
No anaphor: The (squash, slosh) was fed to the pigs/ but the farmers worried about the animals' nutrition.
Gap: The (squash, slosh, mash, corn) was fed to all of the farmer's small pigs/ but [e] was not nutritive.
Pronoun: The (squash, slosh, mash, corn) was fed to all of the farmer's small pigs/ but it was not nutritive.

13) Target = SWAT
No anaphor: The (swat, pot) stunned the housefly/ and Mary killed the pest.
Gap: The (swat, pot, cat, broom) hit the annoying housefly/ and [e] killed the pest.
Pronoun: The (swat, pot, cat, broom) hit the annoying housefly/ and it killed the pest.

14) Target = BEAD
No anaphor, identity: The head was very smooth/ and the egg seemed perfectly round.
No anaphor, phonological: The seed was very smooth/ and the pod seemed perfectly round.
Gap: The (bead, seed, head, rock) was smooth to the touch/ and [e] seemed perfectly round.
Pronoun: The (bead, seed, head, rock) was smooth to the touch/ and it seemed perfectly round.

15) Target = THREAD
No anaphor, identity: The thread was fragile/ and the stitch broke immediately.
No anaphor, phonological: The sled was fragile/ and the seat broke immediately.
Gap: The (thread, sled, bead, glass) was very fragile/ and [e] broke almost immediately.
Pronoun: The (thread, sled, bead, glass) was very fragile/ and it broke almost immediately.
16) Target = MEAD
No anaphor: The (mead, feed) was given to the cows/ and they were sick.
Gap: The (mead, feed, bread, oats) was/were given to the cows and [e] made them sick.
Pronoun: The (mead, feed, bread, oats) was/were given to the cows and it/they made them sick.

17) Target = BEAK
No anaphor: The (beak, sheik) was large/ and John's curiosity was aroused.
Gap: The (beak, sheik, steak, boat) was very large/ and [e] aroused John's curiosity.
Pronoun: The (beak, sheik, steak, boat) was very large/ and it/he aroused John's curiosity.

18) Target = BREAK
No anaphor: The (break, lake) was very small/ but Sue was worried anyhow.
Gap: The (break, lake, leak, hole) was very small/ but [e] caused some concern.
Pronoun: The (break, lake, leak, hole) was very small/ but it caused some concern.

19) Target = FREAK
No anaphor: The (freak, sheik) won first prize at the costume party/ and the host was amusing.
Gap: The (freak, sheik, steak, witch) won first prize at the costume party/ and [e] amused the other guests.
Pronoun: The (freak, sheik, steak, witch) won first prize at the costume party/ and he/she amused the other guests.

20) Target = CREAK
No anaphor: The (creak, shriek) sounded very loud in the still night/ and she woke suddenly.
Gap: The (creak, shriek, break, crash) sounded very loud in the still night/ and [e] woke her suddenly.
Pronoun: The (creak, shriek, break, crash) sounded very loud in the still night/ and it woke her suddenly.

21) Target = BEAR
No anaphor: The (bear, snare) proved equal to the task/ and the prey died quickly.
Gap: The (bear, snare, spear, trap) proved equal to the task/ and [e] killed the unlucky prey.
Pronoun: The (bear, snare, spear, trap) proved equal to the task/ and it killed the unlucky prey.
22) Target = SPEAR
No anaphor, identity: The spear fell heavily/ and the prey landed in the lake.
No anaphor, phonological: The deer fell heavily/ and the arrow landed in the lake.
Gap: The (spear, deer, bear, rock) fell from the ledge/ and [e] landed in the lake.
Pronoun: The (spear, deer, bear, rock) fell from the ledge/ and it landed in the lake.

23) Target = FEAR
No anaphor, identity: The Indian chief's fear left him as the battle began/ and his courage returned.
No anaphor, phonological: The Indian chief's deer left him as the battle began/ and his wolf returned.
Gap: The Indian chief's (fear, deer, bear, son) left him just as the battle began/ and [e] never returned.
Pronoun: The Indian chief's (fear, deer, bear, son) left him just as the battle began/ and it/he never returned.

24) Target = TEASE
No anaphor: The (tease, fees) disgusted the man, and the brute annoyed the woman.
Gap: The (tease, fees, grease, road) disgusted the elderly man/ and [e] annoyed the young woman.
Pronoun: The (tease, fees, grease, road) disgusted the elderly man/ and she/it annoyed the young woman.

25) Target = PLEASE
No anaphor: "(Please, Freeze)" was said authoritatively/ but the young girl's voice had no effect.
Gap: "(Please, Freeze, Cease, Stop)" was said authoritatively by the young girl/ but [e] had no effect.
Pronoun: "(Please, Freeze, Cease, Stop)" was said authoritatively by the young girl/ but it had no effect.

26) Target = MEAT
No anaphor: The (meat, sheet) was foul-smelling/ and she felt very ill.
Gap: The (meat, sheet, sweat, room) was foul-smelling/ and [e] made her feel ill.
Pronoun: The (meat, sheet, sweat, room) was foul-smelling/ and it made her feel ill.
27) Target = HEAT
No anaphor: The (heat, sleet) discouraged the guests/ and the staff became alarmed.
Gap: The (heat, sleet, threat, storm) discouraged the guests/ and [e] alarmed the staff.
Pronoun: The (heat, sleet, threat, storm) discouraged the guests/ and it alarmed the staff.

28) Target = WHEAT
No anaphor: The (wheat, beet) tasted bitter/ and she had to gasp.
Gap: The (wheat, beet, sweat, gin) tasted bitter/ and [e] caused her to gasp.
Pronoun: The (wheat, beet, sweat, gin) tasted bitter/ and it caused her to gasp.

29) Target = THREAT
No anaphor: The (threat, bet) pleased the mafia/ but the police were angry.
Gap: The (threat, bet, cheat, crime) pleased the mafia leader/ but [e] angered the police.
Pronoun: The (threat, bet, cheat, crime) pleased the mafia leader/ but it angered the police.

30) Target = GREAT
No anaphor: The (great, late) entrée did not please the difficult customer/ and he complained again.
Gap: The (great, late, meat, large) entrée did not please the difficult customer/ and [e] caused further complaint.
Pronoun: The (great, late, meat, large) entrée did not please the difficult customer/ and it caused further complaint.

31) Target = WREATH
No anaphor, identity: The wreath was rotten/ and the room smelled bad.
No anaphor, phonological: The teeth were rotten/ and the mouth smelled bad.
Gap: The (wreath, teeth, breath, milk) was/were quite rotten/ and [e] smelled very bad.
Pronoun: The (wreath, teeth, breath, milk) was/were quite rotten/ and it/they smelled very bad.
32) Target = SHEATH
   No anaphor: The (sheath, teeth) frightened the cowardly man/ and he began to panic.
   Gap: The (sheath, teeth, death, cry) frightened the cowardly man/ and [e] caused him to panic.
   Pronoun: The (sheath, teeth, death, cry) frightened the cowardly man/ and it/they caused him to panic.

33) Target = HEIGHT
   No anaphor: His (height, might) is rather unusual/ and he inspires respect.
   Gap: His (height, might, weight, size) is greater than usual/ and [e] inspires respect.
   Pronoun: His (height, might, weight, size) is greater than usual/ and it inspires respect.

34) Target = WEIGHT
   No anaphor: His (weight, gait) is the greatest ever recorded in the sport/ and he frightens his opponents.
   Gap: His (weight, gait, height, jump) is the greatest ever recorded in the sports world/ and [e] frightens his opponents.
   Pronoun: His (weight, gait, height, jump) is the greatest ever recorded in the sports world/ and it frightens his opponents.

35) Target = WHERE
   No anaphor: "(Where, Share)?" was asked but he did not answer.
   Gap: "(Where, Share, Here, How)?" was asked diplomatically/ but [e] was not answered.
   Pronoun: "(Where, Share, Here, How)?" was asked diplomatically/ but it was not answered.

36) Target = BREW
   No anaphor: The (brew, clue) was hidden in the church/ and the priest was never found.
   Gap: The (brew, clue, pew, stash) was well hidden in the church basement/ and [e] was never found.
   Pronoun: The (brew, clue, pew, stash) was well hidden in the church basement/ and it was never found.

37) Target = SCREW
   No anaphor: The (screw, glue) was too old/ and the whole kit was thrown away.
   Gap: The (screw, glue, pew, bench) was felt to be too old/ and [e] was thrown away.
   Pronoun: The (screw, glue, pew, bench) was felt to be too old/ and it was thrown away.
38) Target = CREW
No anaphor: The (crew, glue) worked well/ and the captain repaired the radio.
Gap: The (crew, glue, few, team) worked well/ and [e] repaired the ship's radio.
Pronoun: The (crew, glue, few, team) worked well/ and it/they repaired the ship's radio.

39) Target = TWO
No anaphor: The (two, crew) did not win the race/ but the crowd cheered.
Gap: The (two, crew, pro, girl) did not win the race/ but [e] received many cheers.
Pronoun: The (two, crew, pro, girl) did not win the race/ but they/it/he/she received many cheers.

40) Target = PRO
No anaphor: The (pro, foe) entered the building quickly/ and the police went to work.
Gap: The (pro, foe, two, men) entered the poorly guarded building quickly/ and [e] went to work.
Pronoun: The (pro, foe, two, men) entered the poorly guarded building quickly/ and he/they went to work.

41) Target = HOE
No anaphor: The (hoe, throw) was too long/ and he felt awkward.
Gap: The (hoe, throw, shoe, bat) was much too long/ and [e] felt awkward.
Pronoun: The (hoe, throw, shoe, bat) was much too long/ and it felt awkward.

42) Target = TOE
No anaphor: His (toe, blow) touched the button accidentally/ and he triggered the alarm.
Gap: His (toe, blow, shoe, hand) touched the button by mistake/ and [e] triggered the alarm.
Pronoun: His (toe, blow, shoe, hand) touched the button by mistake/ and it triggered the alarm.

43) Target = DOE
No anaphor: The (doe, crow) was placed in a box/ and the box was taken to Bob.
Gap: The (doe, crow, shoe, purse) was carefully placed in a special box/ and [e] was given to Bob.
Pronoun: The (doe, crow, shoe, purse) was carefully placed in a special box/ and it was given to Bob.
44) Target = SHOE
No anaphor: The (shoe, glue) fell/ and the overloaded shelf dropped to the floor.
Gap: The (shoe, glue, hoe, vase) fell from the overloaded shelf/ and [e] dropped to the floor.
Pronoun: The (shoe, glue, hoe, vase) fell from the overloaded shelf/ and it dropped to the floor.

45) Target = TOMB
No anaphor: The (tomb, broom) should have frightened the children/ but only their parents reacted.
Gap: The (tomb, broom, bomb, witch) should have really frightened the small children/ but [e] did not.
Pronoun: The (tomb, broom, bomb, witch) should have really frightened the small children/ but it/she did not.

46) Target = WOMB
No anaphor: The (womb, room) provided security/ but the baby's growth was not sufficient.
Gap: The (womb, room, bomb, cloak) provided security for some time/ but [e] did not suffice.
Pronoun: The (womb, room, bomb, cloak) provided security for some time/ but it did not suffice.

47) Target = BOMBS
No anaphor: The (bombs, psalms) were displayed at the museum/ and the visitors were awed.
Gap: The (bombs, psalms, tombs, works) were displayed at the local museum/ and [e] awed the visitors.
Pronoun: The (bombs, psalms, tombs, works) were displayed at the local museum/ and they awed the visitors.

48) Target = GNOME
No anaphor: The (gnome, foam) jumped up against the waves/ but the sea swept everything aside.
Gap: (The gnome, The foam, Some, The duck) jumped up against the sea's pounding waves/ but [e] was/were swept aside.
Pronoun: (The gnome, The foam, Some, The duck) jumped up against the sea's pounding waves/ but he/it/ they was/were swept aside.
49) Target = SOME
No anaphor: (Some, One drum) shone in the light of the store/ but the others seemed dull.  
Pronoun: (Some, The drum, The chrome, The ring) shone in the light of the department store/ but they/it seemed dull outside.

50) Target = DON
No anaphor: The (don, fawn) ran through the courtyard/ and the ladies were amused.  
Gap: The (don, fawn, son, chimp) ran through the Spanish courtyard/ and [e] amused the ladies.  
Pronoun: The (don, fawn, son, chimp) ran through the Spanish courtyard/ and he/it amused the ladies.

51) Target = SON
No anaphor: The (son, nun) hid/ and the angry kidnapper ran away.  
Gap: The (son, nun, con, thief) hid until dark/ and then [e] ran away.  
Pronoun: The (son, nun, con, thief) hid until dark/ and then he/she ran away.

52) Target = BLOOD
No anaphor: The (blood, mud) covered his face/ and he concealed his identity.  
Gap: The (blood, mud, hood, mask) covered his scarred face/ and [e] concealed his identity.  
Pronoun: The (blood, mud, hood, mask) covered his scarred face/ and it concealed his identity.

53) Target = FOOD
No anaphor, identity: The food was not expected/ and the caterers surprised the crowd.  
No anaphor, phonological: The nude was not expected/ and the artists surprised the crowd.  
Gap: The (food, nude, flood, mail) was not expected at all/ and [e] surprised the crowd.  
Pronoun: The (food, nude, flood, mail) was not expected at all/ and it surprised the crowd.
54) Target = MOOD
No anaphor: The (mood, dude) increased her agitation/ and she had a heart attack.
Gap: The (mood, dude, flood, job) further increased her agitation/ and [e] caused her heart attack.
Pronoun: The (mood, dude, flood, job) further increased her agitation/ and it/he caused her heart attack.

55) Target = BROOD
No anaphor: The (brood, dude) intrigued the children/ but the others were bored.
Gap: The (brood, dude, blood, song) intrigued the young children/ but [e] bored the others.
Pronoun: The (brood, dude, blood, song) intrigued the young children/ but it/he bored the others.

56) Target = TOOL
No anaphor: The (tool, jewel) had survived the passage of time/ and the archaeologist was astonished.
Gap: The (tool, jewel, wool, blade) had survived the passage of countless years/ and [e] astonished the archaeologist.
Pronoun: The (tool, jewel, wool, blade) had survived the passage of countless years/ and it astonished the archaeologist.

57) Target = STOOL
No anaphor, identity: The stool was broken/ and all of the furniture was discarded.
No anaphor, phonological: The rule was broken/ and all of the orders were discarded.
Gap: The (stool, rule, wool, cup) was broken by the children/ and [e] was therefore discarded.
Pronoun: The (stool, rule, wool, cup) was broken by the children/ and it was therefore discarded.

58) Target = WOOL
No anaphor: The (wool, bull) amused most of the children/ but Fred was not entertained.
Gap: The (wool, bull, pool, dance) amused most of the children/ but [e] did not entertain Fred.
Pronoun: The (wool, bull, pool, dance) amused most of the children/ but it did not entertain Fred.
59) Target = BOOT
No anaphor, identity: The boot was too large/ and the shoe was ugly.
No anaphor, phonological: The suit was too large/ and the jacket was ugly.
Gap: The (boot, suit, foot, glove) was much too large/ and [e] was very ugly.
Pronoun: The (boot, suit, foot, glove) was much too large/ and it was very ugly.

60) Target = FORMS
No anaphor: The (forms, swarms) slithered through the trees/ and the farmers were worried.
Gap: The (forms, swarms, worms, pests) slithered through the young trees/ and [e] worried the farmers.
Pronoun: The (forms, swarms, worms, pests) slithered through the young trees/ and they worried the farmers.

61) Target = POSE
No anaphor: The (pose, crows) frightened the little girl/ and she cried.
Gap: The (pose, crows, dose, dog) frightened the girl/ and [e] made her cry.
Pronoun: The (pose, crows, dose, dog) frightened the girl/ and it/they made her cry.

62) Target = GHOST
No anaphor: The (ghost, boast) amused the men/ but the ladies were alarmed.
Gap: The (ghost, boast, cost, clown) amused the placid men/ but [e] alarmed the ladies.
Pronoun: The (ghost, boast, cost, clown) amused the placid men/ but he/it alarmed the ladies.

63) Target = HOST
No anaphor: The (host, coast) was uninviting/ and the guests were dismayed.
Gap: The (host, coast, frost, inn) was uninvitingly cold/ and [e] caused some dismay.
Pronoun: The (host, coast, frost, inn) was uninvitingly cold/ and he/it caused some dismay.
64) Target = POST
   No anaphor, identity: The post was acceptable/ but the workers were disappointed.
   No anaphor, phonological: The roast was acceptable/ but the diners were disappointed.
   Gap: The (post, roast, cost, car) was considered acceptable/ but [e] disappointed them nonetheless.
   Pronoun: The (post, roast, cost, car) was considered acceptable/ but it disappointed them nonetheless.

65) Target = BOTH
   No anaphor: (Both, The oath) annoyed the elderly woman/ and she left.
   Gap: (Both, the oath, the moth, the boys) annoyed the woman/ and [e] made her leave.
   Pronoun: (Both, the oath, the moth, the boys) annoyed the woman/ and [e] made her leave.

66) Target = BOUGH
   No anaphor, identity: The bough is strong/ but the tree can be broken.
   No anaphor, phonological: The prow is strong/ but the ship can be broken.
   Gap: The (bough, prow, tough, horse) may be strong/ but [e] may still be broken.
   Pronoun: The (bough, prow, tough, horse) may be strong/ but it/they may still be broken.

67) Target = DOUGH
   No anaphor, identity: The dough was tainted/ and the bread was discarded.
   No anaphor, phonological: The hoe was painted/ and the axe was discarded.
   Gap: The (dough, hoe, plough, toy) was no good/ and [e] was quickly discarded.
   Pronoun: The (dough, hoe, plough, toy) was no good/ and it was quickly discarded.

68) Target = FOUR
   No anaphor: The (four, store) were/was very exciting/ and a sensation ensued.
   Gap: The (four, store, tour, band) were/was very exciting/ and [e] caused a sensation.
   Pronoun: The (four, store, tour, band) were/was very exciting/ and they/it caused a sensation.
69) **Target = STOVE**
No anaphor, identity: The stove was the exact shade that she wanted/ and the fridge fit perfectly.
No anaphor, phonological: The mauve was the exact shade that she wanted/ and the suit fit perfectly.
Gap: The (stove, mauve, glove, suit) was the exact shade that she was looking for/ and it fit perfectly.
Pronoun: The (stove, mauve, glove, suit) was the exact shade that she was looking for/ and it fit perfectly.

70) **Target = CROW**
No anaphor, identity: The crow cleared the fence/ and feathers fluttered to the ground.
No anaphor, phonological: The pro cleared the fence/ and flowers fluttered to the ground.
Gap: The (crow, pro, cow, lamb) barely cleared the fence/ and it/he landed on the ground.
Pronoun: The (crow, pro, cow, lamb) barely cleared the fence/ and it/he landed on the ground.

71) **Target = COW**
No anaphor, identity: The cow was very weak/ and the ground held no nourishment.
No anaphor, phonological: The bough was very weak/ and the tree held no nourishment.
Gap: The (cow, bough, crow, branch) was weak with age/ and it fell to the ground.
Pronoun: The (cow, bough, crow, branch) was weak with age/ and it fell to the ground.

72) **Target = THROW**
No anaphor: The (throw, pro) impressed his fiancée/ but his competitors were discouraged.
Gap: The (throw, pro, vow, man) impressed his lovely fiancée/ but [e] discouraged his competitors.
Pronoun: The (throw, pro, vow, man) impressed his lovely fiancée/ but it/he discouraged his competitors.

73) **Target = SHOW**
No anaphor: His (show, foe) won/ but then the judge's decision was disqualified.
Gap: His (show, foe, cow, pie) initially won the contest/ but [e] was then disqualified.
Pronoun: His (show, foe, cow, pie) initially won the contest/ but it was then disqualified.
74) Target = KNOWN
No anaphor: The (known, clone) holds no surprises/ and he is not impressed.
Gap: The (known, clone, town, book) holds no surprises/ and [e] does not impress him.
Pronoun: The (known, clone, town, book) holds no surprises/ and it does not impress him.

75) Target = CLUE
No anaphor: The (clue, crew) was mysterious/ but he found his home.
Gap: The (clue, crew, cue, sign) was very mysterious/ but [e] led him home.
Pronoun: The (clue, crew, cue, sign) was very mysterious/ but it led him home.

76) Target = BULL
No anaphor: The (bull, wool) was wet/ and the barn smelled bad.
Gap: The (bull, wool, gull, box) was soaking wet/ and [e] smelled very bad.
Pronoun: The (bull, wool, gull, box) was soaking wet/ and it smelled very bad.

77) Target = LURE
No anaphor, identity: The lure was interesting/ and the fish paid attention.
No anaphor, phonological: The poor were interesting/ and the rich paid attention.
Gap: The (lure, poor, cure, town) was/were of interest/ and [e] attracted much attention.
Pronoun: The (lure, poor, cure, town) was/were of interest/ and it/they attracted much attention.

78) Target = FUSE
No anaphor, identity: The fuse worked during testing/ but later the lights failed.
No anaphor, phonological: The cues worked during testing/ but later the timing failed.
Gap: The (fuse, cues, ruse, bulb) worked during the test/ but [e] failed later on.
Pronoun: The (fuse, cues, ruse, bulb) worked during the test/ but it/they failed later on.
79) Target = BRUTE
No anaphor: The (brute, toot) was loud/ and the teacher was startled.
Gap: The (brute, toot, mute, child) was surprisingly loud/ and [e] startled the teacher.
Pronoun: The (brute, toot, mute, child) was surprisingly loud/ and he/it startled the teacher.

80) Target = FLUTE
No anaphor: The (flute, fruit) was praised by the professor/ but the students were not appreciative.
Gap: The (flute, fruit, mute, speech) was praised by the young professor/ but otherwise [e] was not appreciated.
Pronoun: The (flute, fruit, mute, speech) was praised by the young professor/ but otherwise it/he was not appreciated.
Appendix C: Experimental Stimuli for Experiments 3 & 4

a = No Anaphor  
b = NP-Raising Gap  
c = Pronoun

1) PROBE = NURSE  
a) The nurse/who remained at his bedside/day and night/obviously loved/the soldier.  
b) The nurse/who remained at his bedside/all night/seemed [e] to love/the soldier.  
c) The nurse/ remained at his bedside/all night/because she obviously loved/the soldier.

2) PROBE = SWAP  
a) The swap/that involved fifty kilos/of South American cocaine/certainly caused/some trouble.  
b) The swap/that involved fifty kilos/of cocaine/was certain [e] to cause/some trouble.  
c) The swap/involved over fifty kilos/of cocaine/and it certainly caused/some trouble.

3) PROBE = WOMB  
a) The womb/that appears in all/of the artist's paintings/might embarrass/many potential buyers.  
b) The womb/that appears in all/of his paintings/seems [e] to embarrass/many potential buyers.  
c) The womb/appears in all/of his paintings/but it might embarrass/many potential buyers.

4) PROBE = FLUTE  
a) The flute/that had made/the old musician famous/always held/ the audience's attention.  
b) The flute/that made/the musician famous/was sure [e] to hold/ the audience's attention.  
c) The flute/had made/the musician famous/and it always held/ the audience's attention.

5) PROBE = MEAD  
a) The mead/that caused/the captain's drunken stupor/was really not fit/for consumption.  
b) The mead/that caused/the captain's drunken stupor/appeared [e] to be unfit/for consumption.  
c) The mead/had caused/the captain's drunkenness/but it was not fit/for consumption.
6) PROBE = BEAK
a) The beak/that stuck through/the crack in the egg/reached out/for help.
b) The beak/that stuck through/the cracked egg/seemed [e] to reach out/for help.
c) The beak/stuck through/the cracked egg/and it reached out anxiously/for help.

7) PROBE = CROW
a) The crow/that squawked loudly/in the old dilapidated barn/was disturbing/the herd.
b) The crow/that squawked/in the dilapidated barn/was certain [e] to disturb/the herd.
c) The crow/squawked loudly/in the dilapidated barn/and it was disturbing/the herd.

8) PROBE = QUART
a) The last quart/that left the gasoline pump/in a big rush/messily spilled/onto his shoes.
b) The last quart/that left the pump/in a rush/was sure [e] to spill/onto his shoes.
c) The last quart/left the gasoline pump/in a rush/and it messily spilled/onto his shoes.

9) PROBE = WREATH
a) The wreath/that hung/on the gaily decorated door/merrily added/ the final touch.
b) The wreath/that hung/on the gay door/appeared [e] to add/the final touch.
c) The wreath/gaily hung/on the door/and it merrily added/the final touch.

10) PROBE = SHEATH
a) The sheath/that had belonged/to the bravest knight/in the old kingdom/proudly bore/many ornate trimmings.
b) The sheath/that belonged/to the bravest knight/in the kingdom/was likely [e] to bear/many ornate trimmings.
c) The sheath/had belonged/to the bravest knight/in the kingdom/and it proudly bore/many ornate trimmings.

11) PROBE = FREAK
a) The freak/who worked/with the large travelling circus/was always frightening/the young children.
b) The freak/who worked/with the travelling circus/was certain [e] to frighten/the young children.
c) The freak/worked well/with the travelling circus/because he always frightened/the young children.
12) PROBE = TEASE
a) The tease/who was flirting/with the young men/at the Christmas party/was annoying/the young women.
b) The tease/who was flirting/with the men/at the party/was sure [e] to annoy/the young women.
c) The tease/was flirting/with the young men/at the party/and she was annoying/the young women.

13) PROBE = STOOL
a) The stool/that strangely fell/out of the moving van/at the intersection/was disrupting/traffic flow.
b) The stool/that fell/out of the van/at the intersection/was certain [e] to disrupt/traffic flow.
c) The stool/strangely fell/out of the van/at the intersection/and it was disrupting/traffic flow.

14) PROBE = FUSE
a) The fuse/that will quickly blow/after a short circuit/in the old wiring/will prevent/a fire.
b) The fuse/that will blow/after a short/in the wiring/will be certain [e] to prevent/a fire.
c) The fuse/will quickly blow/after a short circuit/in the wiring/and it will prevent/a fire.

15) PROBE = TOE
a) The toe/that broke when she fell/off of the large swing/quickly turned/shades of blue.
b) The toe/that broke when she fell/off of the swing/seemed [e] to turn/shades of blue.
c) The toe/broke when she fell/off of the swing/and it quickly turned/shades of blue.

16) PROBE = POSE
a) The pose/that best suited/the young female model/had required/ lots of stamina.
b) The pose/that best suited/the young model/appeared [e] to require/ lots of stamina.
c) The pose/well-suited/the young model/but it had required/ lots of stamina.

17) PROBE = SHOE
a) The shoe/that finally fit/her tiny foot perfectly/did not match/ her budget.
b) The shoe/that finally fit/her foot/was likely [e] to not match/ her budget.
c) The shoe/finally fit/her tiny foot/but it did not match/her budget.
18) PROBE = BOOT
a) The boot/that was sitting/on the bedroom window sill/apparently fell/through the opening.
b) The boot/that was sitting/on the window sill/appeared [e] to fall/through the opening.
c) The boot/was sitting/on the window sill/and it apparently fell/through the opening.

19) PROBE = STOVE
a) The old stove/that heats up/their small cottage/is also/the only light.
b) The old stove/that heats up/their cottage/seems [el to be/the only light.
c) The old stove/heats up/their cottage/and it is also/the only light.

20) PROBE = CLUE
a) The clue/that prompted the pirates' search/for buried treasure/initially led/to an island.
b) The clue/that prompted the pirates' search/for treasure/appeared [e] to lead/to an island.
c) The clue/prompted the pirates' search/for treasure/and it initially led/to an island.

21) PROBE = ACHE
a) The ache/that pounded at the back/of her head/unfortunately lasted/for days.
b) The ache/that pounded the back/of her head/seemed [e] to last/for days.
c) The ache/pounded the back/of her head/and it unfortunately lasted/for days.

22) PROBE = SNEAK
a) The sneak/who easily took advantage/of the company chairman's trust/slyly disappeared/with the money.
b) The sneak/who took advantage/of the chairman's trust/was likely [e] to disappear/with the money.
c) The sneak/easily took advantage/of the chairman's trust/and he slyly disappeared/with the money.

23) PROBE = STEAK
a) The steak/that fills the room/with such delicious spicy odours/should satisfy/her hunger.
b) The steak/that fills the room/with spicy odours/is certain [e] to satisfy/her hunger.
c) The steak/fills the room/with delicious spicy odours/and it should satisfy/her hunger.
24) PROBE = NIECE
a) The niece/who always worked/around her uncle's stables/in the summertime/would inherit/the prize horses.
b) The niece/who worked/around the stables/in the summertime/was certain [e] to inherit/the prize horses.
c) The niece/always worked/around the stables/in the summertime/and she would inherit/the prize horses.

25) PROBE = FEAT
a) The feat/that proved impossible/for the king's imperial guards/simply bored/the unknown archer.
b) The feat/that proved impossible/for the king's guards/seemed [e] to bore/the unknown archer.
c) The feat/proved impossible/for the king's guards/but it simply bored/the unknown archer.

26) PROBE = RAID
a) The raid/that will lead/to the downfall/of the evil old tyrant/will result/in many deaths.
b) The raid/that will lead/to the downfall/of the tyrant/is certain [e] to result/in many deaths.
c) The raid/will lead/to the downfall/of the evil tyrant/but it will result/in many deaths.

27) PROBE = MARE
a) The mare/that pulled the poor family/and their few belongings/certainly looked/old and feeble.
b) The mare/that pulled the family/and their few belongings/appeared [e] to be/old and feeble.
c) The mare/pulled the family/and their few belongings/but it certainly looked/old and feeble.

28) PROBE = YACHT
a) The yacht/that carried the veteran crew/to the championship race/proudly sits/in the museum.
b) The yacht/that carried the crew/to the championship/is sure [e] to sit/in the museum.
c) The yacht/carried the veteran crew/to the championship/and it proudly sits/in the museum.

29) PROBE = GLOW
a) The glow/that emanated from the eyes/of the young woman/surely revealed/her hidden feelings.
b) The glow/that emanated from the eyes/of the woman/seemed [e] to reveal/her hidden feelings.
c) The glow/emanated from the eyes/of the woman/and it surely revealed/her hidden feelings.
30) PROBE = SCREWS
a) The screws/that held the phone/to the kitchen wall/were becoming/old and rusty.
b) The screws/that held the phone/to the wall/appeared [e] to be/old and rusty.
c) The screws/held the phone/to the wall/but they were becoming/old and rusty.

31) PROBE = FLOOD
a) The flood/that left so many hundreds/without a home/unexpectedly lasted/over two weeks.
b) The flood/that left so many/without a home/seemed [e] to last/over two weeks.
c) The flood/left so many/without a home/and it unexpectedly lasted/over two weeks.

32) PROBE = ROAR
a) The roar/that caused the rapid dispersal/of all of the mammals/truly frightened/the hunters.
b) The roar/that caused the rapid dispersal/of all the mammals/seemed [e] to frighten/the hunters.
c) The roar/caused the dispersal/of all of the mammals/and it truly frightened/the hunters.

33) PROBE = GAINS
a) The gains/that further improved/the corporation's financial status/inadvertently fostered/another takeover bid.
b) The gains/that improved/the corporation's financial status/appeared [e] to foster/another takeover bid.
c) The gains/further improved/the corporation's status/and they inadvertently fostered/another takeover bid.

34) PROBE = TOAST
a) The toast/that embarrassed the groom/during the wedding reception/fortunately made/everyone else laugh.
b) The toast/that embarrassed the groom/during the reception/seemed [e] to make/everyone else laugh.
c) The toast/embarrassed the groom/during the reception/but it fortunately made/everyone else laugh.

35) PROBE = FAME
a) The fame/that overwhelmed the young actress/at the very beginning/is now simply/part of life.
b) The fame/that overwhelmed the young actress/in the beginning/now appears [e] to be/part of life.
c) The fame/completely overwhelmed the young actress/in the beginning/but it is now/part of life.
Anaphor Resolution

36) PROBE = SQUASH
a) The squash/that grew huge/in old farmer Brown's garden/should win/first prize.
b) The big squash/that grew/in farmer Brown's garden/is sure [e] to win/first prize.
c) The squash/grew huge/in farmer Brown's garden/and it should win/first prize.

37) PROBE = JEEP
a) The jeep/that won the famous race/across the Sahara desert/sadly crashed/after the finish.
b) The jeep/that won the race/across the Sahara desert/appeared [e] to crash/after the finish.
c) The jeep/won the race/across the Sahara desert/but it sadly crashed/after the finish.

38) PROBE = STRAP
a) The strap/that held the heavy boxes/onto the car/suddenly broke/under the weight.
b) The strap/that held the boxes/onto the car/appeared [e] to break/under the weight.
c) The strap/held the boxes/onto the car/but it suddenly broke/under the weight.

39) PROBE = LAD
a) The lad/who raced past the gate/on the spirited stallion/will fall off/before long.
b) The lad/who raced past/on the spirited stallion/is certain [e] to fall off/before long.
c) The lad/raced past the gate/on the stallion/but he will fall off/before long.

40) PROBE = MOTH
a) The moth/that carelessly flew/towards the bright flame/was unaware/of the consequences.
b) The moth/that flew/towards the flame/was unlikely [e] to know/of the consequences.
c) The moth/carelessly flew/towards the flame/but it was unaware/of the consequences.

41) PROBE = CHART
a) The chart/that displays the migration path/of the Canada goose/also indicates/the nesting grounds.
b) The chart/that displays the migration/of the Canada goose/appears [e] to indicate/the nesting grounds.
c) The chart/displays the migration/of the Canada goose/and it also indicates/the nesting grounds.
42) PROBE = BEAST
a) The beast/that still lived/within her tortured soul/repeatedly tormented her/every night.
b) The beast/that lived/within her tortured soul/appeared [e] to torment her/every night.
c) The beast/still lived/within her soul/and it relentlessly tormented her/every night.

43) PROBE = BAIT
a) The bait/that had attracted/the typical dim-witted consumer/was not successful/with others.
b) The bait/that attracted/the dim-witted consumers/was unlikely [e] to be successful/with others.
c) The bait/had attracted/the dim-witted consumers/but it was not successful/with others.

44) PROBE = FUEL
a) The fuel/that gradually leaked/from the rusty old engine/badly stained/the garage floor.
b) The fuel/that gradually leaked/from the engine/was likely [e] to stain/the garage floor.
c) The fuel/gradually leaked/from the old engine/and it badly stained/the garage floor.

45) PROBE = KITE
a) The kite/that flew through the blue skies/above the park/certainly amused/the children.
b) The kite/that flew through the skies/above the park/seemed [e] to amuse/the children.
c) The kite/flew through the skies/above the park/and it certainly amused/the children.