Journal of Psycholinguistic Research, Vol. 25, No. 1, 1996

Coreference Processing and Levels of Analysis in Object-Relative Constructions; Demonstration of Antecedent Reactivation with the Cross-Modal Priming Paradigm

Tracy Love[1,2] and David Swinney[1,2]

This paper is concerned with two related issues in sentence processing-one methodological and one theoretical. Methodologically, it provides an unconfounded test of the ability of the cross-modal lexical priming task, when used appropriately, to provide detailed evidence about the time-course of antecedent reactivation during sentence processing. Theoretically, it provides a study of the nature of the representation that is examined when a reference-seeking element is linked to its antecedent during the processing of object-relative clause constructions. In these studies, subjects heard sentences which contained a lexical ambiguity placed in a strong biasing context. In one study this ambiguous word was the "moved" or 'fronted" object of the verb in an object-relative construction. A cross-modal lexical priming (CMLP) naming task was used to determine whether one or more of the meanings of the ambiguity are activated at three temporally distinct points during the sentence: (1) immediately after the lexical ambiguity (Study 1); (2) a later point that was 700 milliseconds before the offset of the main verb (Study 2); (3) immediately after this main verb (at the gap in this filler-gap construction) (Study 2). The probes in the CMLP task were controlled for potential confounds. The results demonstrate the following: At Test Point 1, all meanings of the ambiguity were activated; at Test Point 2, neither meaning of the ambiguity was (still) activated; at Test Point 3, only a single (context-relevant) meaning of the ambiguity was reactivated. It is concluded that an underlying (deep; non-surface-level) memorial representation of the sentence is examined during the process of linking an antecedent to a structural position requiring a referent, and that the CMLP task provides an unbiased measure of this reactivation.

The authors gratefully acknowledg the contributions of an anonymous reviewer for helpful suggestions and changes in this paper, and the support of a fellowship to the first author from grant T32 DC 00041 and NIH grant R01 DC00494 (to the second author) for this work.

[1] University of California, San Diego, La Jolla, California 92093.

[2] Address all correspondence to Tracy Love or David Swinney, Department of Psychology, UCSD, La Jolla, California 92093.

Further, it is concluded that this effect cannot be accounted for under a "compound cue" (Ratcliff & McKoon, 1994) explanation.

This paper presents evidence which speaks directly to two current issues in the sentence processing literature. The first, a theoretical issue, concerns the nature of the process by which antecedents are linked with reference-seeking elements during ongoing processing of object-relative constructions in English. The second, a methodological issue, examines the efficacy of using the cross-modal lexical priming technique to demonstrate antecedent reactivation during processing, in the light of recent challenges (McKoon & Ratcliff, 1994).

Details of the nature and time course of coreference processing have been under intense study in the past few years in an attempt to resolve fundamental questions concerning how antecedents become linked to reference-seeking elements during discourse comprehension. What has become clear from both linguistic descriptions and experimental studies is that there are quite a number of linguistic and extralinguistic information sources that could be used to aid this process for various types of coreference. Major questions exist, however, concerning which of these are actually used by the comprehension device (and when each is used) during ongoing comprehension (see, e.g., discussions by Clifton & Frazier, 1984; Clifton, Frazier, & Connine, 1984; Crain & Fodor, 1985; Fodor, 1978, 1988; McElree & Bever, 1989; Nicol & Swinney, 1989; Stowe, Tanenhaus, & Carlson, 1991; Tanenhaus, Stowe & Carlson, 1985).

This paper is concerned with one particular type of coreference processing-that involved in linking antecedent fillers to their structural gaps during ongoing comprehension, In particular, research in this paper examines such filler-gap dependencies as found in object-relative constructions (e.g., *There is the bank that the boy saw).* A number of approaches to formal linguistics (e.g., Chomsky, 1981) hold that such object-relative constructions are derived from an underlying representation something like *The boy saw the bank*. That is, the underlying representation for English sentences involves a canonical ordering of subject-verb-object (mapped to thematic roles); the surface word ordering found in the object-relative construction is thus viewed as involving "movement" of the direct object from a point following the verb (creating a gap) to a position in front of the verb. In somewhat generic psycholinguistic parlance, the moved direct object (*bank*, in the above example) becomes a filter, and the position from which it moved (following the verb) is a structural position gap.

Under this general theoretic, sentence interpretation is viewed as requiring (among other things) recovery of the underlying syntactic relationships represented by canonical word order. The issue we address here is whether there is any evidence for such recovery (i.e., for linking the structural gap-as licensed by a verb which takes a direct object-to an antecedent filler). Further, we are interested in precisely when such linkage takes place during comprehension.

In terms of the latter question, considerable recent work has been aimed at discovering whether the processor attempts to attach a potential referent to a reference-seeking element immediately upon encountering such an element in the discourse (e.g., the active filler hypothesis; Frazier, 1987), or whether the comprehension system waits until most or all information that might be relevant to the decision has been considered (e.g., at the end of the phrase, sentence, etc.). A number of clever experimental paradigms involving both auditory and visual language materials have been employed to examine this question. As simply one example, an end-of-sentence probe verification task has been used by several researchers (e.g., Bever & McElree, 1988; Corbett & Chang, 1983; Tanenhaus, Carlson & Seidenberg, 1985) to indicate that, at least by the end of the sentence, potential antecedents to a pronoun have been reactivated. This has been taken to suggest that reference-seeking elements such as pronouns trigger a search through some representational form of the sentence, resulting in reactivation of potential antecedents. However, the detailing of precisely *when* such antecedent (filler) reactivation takes place during sentence processing requires a slightly more sensitive and on-line task than the end-of-sentence probe. Two such tasks, in particular, appear to have the potential for being extremely sensitive to minute details of the time course of on-line processing: cross-modal lexical priming (CMLP) (see, e.g., Nicol & Swinney, 1989; Swinney, Ford, Frauenfelder, & Bresnan, 1987; Swinney, Onifer, Prather, & Hirshkowitz, 1979) and evoked brain potentials (see, e.g., Garnsey, Tanenhaus, & Chapman, 1989). Both of these techniques have been used to examine the processing of filler-gap constructions, and both have produced evidence demonstrating effects which strongly suggest that such antecedents are assigned to a reference-seeking element (such as a gap) immediately after it is licensed by occurrence of the verb.

We are concerned here, in particular, with evidence for such a process as revealed by the CMLP task. Antecedent reactivation has been revealed with this task across a number of variants of this technique (see, e.g., Fodor. & McKinnon study reported in Fodor, 1993; Hickok, 1993; McKee, Nicol. & McDaniel, 1993; Nagel, Shapiro, & Nawy, 1994; Nicol, 1993; Nicol & Osterhout, 1989, cited in Nicol, 1988; Nicol & Pickering, 1993; Osierhout

& Swinney, 1993; Shillock, 1982; Swinney & Osterhout, 1990; Swinney & Zurif 1995; Zurif Swinney, Prather, & Love, 1994; Zurif Swinney, Prather, Solomon, & Bushell, 1993, among others). Recently, however, McKoon and Ratcliff (1994) have argued that there is a materials confound in one of these studies involving CMLP (the very first CMLP filler-gap study ever run; Swinney et al., 1987). We will argue, below, that this potential confound does not have an effect on the CMLP task (as used standardly) and that, independent of this fact, as this potential confound does not exist in many of the above-mentioned CMLP studies which demonstrate antecedent reactivation, the reactivation effect must be real and be demonstrable via the CMLP task. However, independent of such argumentation, this paper is intended as a direct examination of the issue of reactivation of antecedent fillers at structural gap positions with materials which controll for the possible confound reported by McKoon and Ratcliff.

This paper is also aimed at examining whether an antecedent search in filler-gap constructions involves examination of a superficial memorial representation of the sentence (as found, for example, in auditory working memory) or whether it involves reference to an deep or underlying (conceptual) representation of the sentence. There are a number of existing pieces of work concerned with the boundary conditions between deeper (perhaps non-linguistic) and more superficial representations involved in antecedent assignment to certain anaphors (see, in particular, work by Cloitre and Bever, 1988; Frazier, Henstra, & Flores d'Arcais, 1995; Hankamer & Sag, 1976; Tanenhaus et al., 1985). The present study was intended to bring the CMLP on-line technique to bear on this question as it specifically relates to structural dependencies as found in object-relative constructions. This is accomplished here by employing ambiguous words as the antecedent in these constructions. It was hypothesized that if the representation that is examined during coreference processing is a deep or conceptual one, then only the single context-relevant meaning of an ambiguity (presumably chosen via frequency and context) will be reactivated at the structural gap following the verb. If, on the other hand, the representation that is consulted for coreference is a more superficial form (such as verbatim auditory short-term memory, etc.) then occurrence of a gap should result in reactivation of all meanings of the ambiguity. This latter follows from evidence that, when the form (acoustic or visual) of an ambiguous word is encountered in processing, all meanings of that word are momentarily activated (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979).

The CMLP Experimental Technique and the Study of On-Line Gap Filling

Recently, McKoon and colleagues have put forward work that questioned the efficacy of the cross-modal priming technique in revealing antecedent reactivation at gaps in filler-gap constructions. In particular, McKoon, Ratcliff, and Ward (1994) (hereafter referred to as MR&W) argued that the matched-probe variant was an inappropriate way to use CMLP, but that matched-sentence variant was appropriate-and that they found no evidence of antecedent reactivation with this technique. A reply to this paper by Nicol, Fodor, and Swinney (1994) (hereafter referred to as NF&S) argued that the MR&W reasoning was faulty-that both variants had advantages and disadvantages, that both variants had demonstrated antecedent reactivation in the existing literature (see above), and that because MR&W had indeed found no gap filling in their experiments their data could provide no evaluation of the relative efficacy of the two technique variants. Further, NF&S demonstrated that the MR&W materials differed significantly in terms of difficulty from those of prior successful demonstrations of gap filling, a difference that may have been at least part of the reason for their failure to find any such effect with any task. (It was also noted that many of the pretesting requisites standardly used with the CMLP technique were not employed by MR&W; see, e.g., the methods section of this current paper.)

In a reply to NF&S, McKoon and Ratcliff (1994) (hereafter referred to as M&R) took an entirely different tact. Rather than responding to the NF&S demonstrations and arguments, M&R attempted to demonstrate that there was a confound in the original experiment in this series (Swinney et al., 1987) that had first been used to show reactivation of fillers at a gap site in object-relative constructions. They examined the probe materials in that study and found that a number of the related probes would have been a better (more plausible) continuation of the sentence than the control probes at the point of presentation (had the sentence stopped at that point). They then chose 75% of the materials from the Swinney et al. (1987) study and ran a *non-CMLP* study (a word-by-word reading task) which demonstrated that there was a priming effect induced-even with no gap present-with this type of task. They argued that the priming effect shown in the filler-gap experiments with CMLP was caused by this confound.

It is worth noting three specific things about this line of reasoning. First, it has been argued that the CMLP technique, when run properly, is immune to an effect of such goodness-of-continuation between the probe and the sentence (e.g., Nicol et al., 1994). That is, in normal CMLP, at normal rates of fluent speech, subjects hear a sentence which continues on uninterruptedly, independently of the occurrence of the probe. In this, many researchers have noted that subjects largely fail to report the correct location of the probe during the sentence. It appears that, because the sentence continues, there is a tendency to *not* integrate the visual probe into the ongoing auditory sentence. This would make sense, as the subjects are required to demonstrate comprehension of the auditory sentence, and such an intrusion would make such comprehension quite difficult. Apparently the separate visual and auditory modalities allow the visual probe to *not* be integrated into the ongoing auditory sentential material. Certainly, all researchers can undoubtedly imagine a scenario in which the probe *could* be induced to be integrated into the auditory sentence-such as if there is a pause introduced into the auditory sentence at the time the probe is introduced, or if the auditory sentence is presented very slowly (as for example in the task as used by Tyler & Marslen-Wilson, 1977). However, the normal use of this task resists such integration. Note here that the task used by McKoon and Ratcliff (1994) does precisely the opposite. In their task, subjects saw a visual word-by-word presentation of the sentence with the visual probe appearing precisely following a word in the sentence (and before the next visual word of the sentence); this is a task that invites (probably, unavoidably) integration of the visual probe into the word-by-word presentation of the sentence. In sum, the first point is that while CMLP appears relatively immune from integration effects, the word-by-word task used by M&R actually is an integration task, and thus their demonstration of a "plausibility" integration effect is not surprising. However, we believe that it is largely irrelevant to claims about such effects with use of the CMLP task.

Second, it is somewhat unfortunate that M&R used only used 75% of the original Swinney et al. (1987) materials (those with the potential confound among the probes). If they had run all of the Swinney et al. (1987) materials, it would have been possible to determine whether, at least for M&R's word-by-word visual technique, this confound would demonstrate an effect similar to that found by Swinney et al. (1987) for their entire experiment. Thus, the study run by M&R was not actually capable of addressing the issue of whether their potential confound might drive results such as found by Swinney et al. (1987); M&R employed neither the technique that they were attacking nor all of the materials from the study they were attacking in their work.

Third, it needs to be reiterated that reactivation of antecedents at gaps requires two separate demonstrations-first that there is no activation at some pregap site, and second that there is activation at the gap site. Thus, any explanation based on confound-among-probes must demonstrate that that confound does *not* exist at the pregap site, but *does* exist at the gap site or else the confound cannot be responsible for the *change* in priming effect demonstrated in the CMLP task. This was not and has not been demonstrated for any filler-gap experiments (even those of M&R, much less those using the CMLP task).

While the present authors feel that, based on these crucial issues, M&R have failed to show that a confound is in effect in use of the CMLP task in prior gap-filling demonstrations, we feel that the best argument to make is an empirical one. That is, in order to best ascertain whether reactivation of the filler at the gap in object-relative filler-gap constructions is demonstrated by the CMLP technique when no potential plausible-continuation confound is present among the probes, it is most convincing to simply examine it empirically. This was the purpose of this study.

METHOD

Preface

As this paper is in part an evaluation of the efficacy of using the CMLP task to examine structural coreference processing effects, a brief methodological primer on use of this task and on certain concerns in real-time language investigation will be helpful.

First, as has been demonstrated in a number of studies (see, e.g., McKee et al., 1993; Nicol & Pickering, 1993), and as discussed in Nicol, et al. (1994), the CMLP task can be effective in demonstrating antecedent reactivation in either a matched-sentences or a matched-probes configuration (a probe is the visually presented associate to the antecedent NP). The study presented below will employ the matched-probes variation of the task. In order for this to work effectively, it is necessary to obtain related probes which are strong associates of the critical noun and control probes- which have no associative relationship to anything in the sentence. Further (and most critically) related and control probes must be matched on a number of bases, primarily on an *a priori* reaction time. Over the years approximations to such matches have typically been created by equating length and frequency (because it correlates with lexical decision times, overall). However, the best way is simply to equate these words on reaction time itself (as obtained in isolated word tests), as well as on length, frequency, category, etc. Further, experience has shown us that it is critical to run such *a priori* tests on a sample from the same population of subjects that will be used in the study. (It is most ideal to run these tests on the very same subjects, of course, but practicality typically prevents this.) We have discovered that reaction-time matches from norms created outside the local population tend not to replicate on specific local populations.

Similarly, there appear to be unfortunate variability differences among associativity ratings obtained from outside the local population, and fragile on-line effects are susceptible to such differences. Thus, ratings of strength of association between ambiguity and probe, of a priori reaction times for probes, of primary and secondary meanings for ambiguities, of potential confounds for the probes (such as goodness-of-continuation of the probe at each probe location), and of the strength of the bias in the biasing context *all* must be taken from the same population when dealing with these delicate on-line techniques. In each of the aspects mentioned, we have taken our measures as samples from the same population. Each of these pretests is described, in turn, below.

Subjects

In order to insure homogeneity of subject populations, subjects were required to be native English speakers, with no brain injuries, learning disabilities, or abnormal mental behavior evidenced at any time in their lives. This was determined through extensive screening procedures, which were employed for each of the pretest and experimental conditions described below. All subjects were University of California, San Diego (UCSD) undergraduates who participated for course credit, if they passed our screening procedures. The actual numbers of subjects used in each pretest or experimental condition are given in each of the appropriate sections, below.

Pretest for Ambiguity Bias

Thirty UCSD undergraduate students were presented with 165 lexical ambiguities which had at least two noun meanings. Presentation was oral/auditory (as the words often have different auditory and visual biases). For each word, subjects were asked to provide the first word they thought of (an associate, most often). Following that, subjects were asked to provide an associate to another meaning of the word, if they could think of one. Tallies of numbers of first and second choices for each meaning of each of the ambiguous words were made. From this, for the present study, 40 lexical ambiguities were chosen which had clear preferences for a primary interpretation; this chosen, the primary meaning had to have been rated in first place (rather than second, etc.) 75% of the time and, additionally, the secondary meaning had to have been rated as second place at least 75% of the time.

Pretest-a priori Equated Reaction Times for Related and Control Probes

Probes for the primary and secondary meanings of the ambiguity were chosen by utilizing one of the three most frequently provided associates for each meaning (tallied over all subjects). A large number of words equated to the related associates on the basis of frequency, length, form class, and animacy were included (with the related words) in an isolated lexical-naming

task (the same task to be used as part of the experiment). This naming pretest was performed on a separate groups of 30 UCSD undergraduate students. From this, a matched control word was chosen for each related associate for each meaning of the ambiguity. Overall, the mean reaction time for the related probes was 483 msec and for the matched control probes was 484 msec. This difference was not significant ($t_{79} = .396$, $p = .70$).

Creation and Pretest of Object-Relative Construction Sentential Materials

Each ambiguous word was placed in a sentence biased toward its primary, dominant meaning, in the structural frame provided in the example below, wherein the ambiguous word was the filler for the gap (represented by e_i where the gap is an implicit structural entity created by movement of the direct object from its canonical position following the verb). The gap and its filler were underlined and coindexed in the example. The biasing contexts were created following the Tabossi (1988) criteria for bias toward a strong aspect of the meaning of the ambiguity (a minimum of 75% of 12 judges agreed on the intended aspect of meaning of the ambiguous word in the sentence). The following is an example of the sentential materials: *The professor insisted that the exam be completed in ink, so Jimmy used the new* pen$_i^{*1}$ *that his mother-in-law recently*^{*2} *purchased*^{*3}e_i *because the multiple colors allowed for more creativity*.

Asterisks were employed to indicate where the three probe positions would be in the experiment. (These were points during the auditorily presented sentence at which the visual probe words occurred.) Probe Position 1 immediately follows the ambiguity. Probe Position 2 occurred at a later point, 700 msec prior to the offset of the verb in the object-relative construction (that point is approximated in the above example). Probe Position 3 is at the offset of the verb involved in the critical object-relative construction (where the gap is licensed). As is detailed further below, while the actual sentence used to examine the critical Probe Positions 2 and 3 were identical to the above example (i.e., an object-relative construction), the sentence used to test Probe Position 1 was identical to the above sample sentence only up to the point just after Probe Position 1 occurred; following that test point, the sentence was continued in structurally different, but semantically similar, fashion (see Study 1, below, for more details).

The probes used with this example (in both studies) were as follows:

Primary meaning: related = PENCIL
Primary meaning: control = JACKET
Secondary meaning: related = JAIL
Secondary meaning: control = TALE

Pretests of Plausible Continuation of Probes with the Sentence

In order to make certain that the related probes were not better/more plausible continuations of the sentence than the control probes, at each point at which they were to occur during the sentence, two separate pretests were run.

A. Goodness-of-Fit into Sentence Pretest. In the first of these, subjects were provided with visual representation of the experimental sentence fragments up to the point where probe activation was to be tested. Thus, there were three such fragments for each sentence, each ending with the last word occurring before the probe point. Subjects were told to rate (on a 5-point scale, where 1 = *poor continuation* and 5 = *good continuation)* the degree to which each probe word constituted a good or plausible continuation of the sentence at that point. This was run on 10 subjects. The results were as follows (where PR-related to primary meaning of ambiguity; PC = control for primary meaning; SR = related to secondary meaning of ambiguity; and SC = control for secondary meaning):

	PR	PC	SR	SC
Probe Point 1:	1.2	1.1	1.0	1.0
Probe Point 2:	1.0	1.0	1.0	1.0
Probe Point 3:	1.0	1.0	1.0	1.0

There were no significant differences between related and control probes at any of these positions. (And, to anticipate the results, this pattern of good continuation ratings did not predict or correlate with the pattern of the reaction time means in the study presented below.)

B. Relatedness of Probe to Sentence Fragment (overall). A second set of ratings was taken in which subjects rated how much the probe word was related to the sentence fragment as a whole-at each of the test points. (In this, of course, the primary probe should be rated very highly, as it is supposed to relate to the context and the contextually biased meaning of the ambiguity.) The results, for 60 subjects, were as follows (where a rating of 1 was *not related* and a rating of 5 was *highly related* and where PR = related to primary meaning of ambiguity; PC = control for primary; SR = related to secondary meaning of ambiguity; and SC = control for secondary):

	PR	PC	SR	SC
Probe Point 1:	4.7	1.3	1.3	1.1
Probe Point 2:	4.7	1.3	1.3	1.1
Probe Point 3:	4.7	1.3	1.2	1.1

The only significant differences involving any of these values were between the PR category (mean = 4.7) and each of the other values. More importantly, and, again, to anticipate, the pattern of relatedness to sentence fragment as a whole did *not* correspond to the pattern of reaction-time priming found in the study below.

Materials and Design for the Experiments:

The 40 experimental materials were combined with 40 filler sentences (of the same approximate structure as the experimental sentences). The latter were paired with real words which were unrelated to anything in the sentence. These filler probes were matched in length and form class to the probes for the experimental materials.

The 40 experimental sentences and 40 filler sentences were pseudorandomly assigned to positions in a script, such that no more than three of either type ever appeared in a row. In addition, 10 practice sentences and probes were created, with the same structure and constraints as for the fillers. These 10 practice sentences, followed by the pseudorandomly intermixed 80 experimental and filler sentences, were recorded by a male speaker of English, at a normal speaking rate. This tape was digitized at a 48-kHz sample rate. Probe positions were then assigned to each tape by placing a 1-kHz square-wave-onset tone digitally on a parallel channel to that containing the speech (a channel not audible to subjects). This tone was used by the computer program RTLAB to initiate presentation of visual word probe (for the naming task) and to begin response timing (which was terminated by the subject's naming response). Placement of the probe points was done by the use of the digital speech wave editing routine Qdisplay (for use with the PAS 16 digital audio computer board).

It is a standard procedure in psycholinguistic studies of this type that no subject hear more than a single experimental sentence paired with more than a single visual probe in the study (to eliminate repetition-bias effects of any sort). Given this, and in order to maximize the number of exemplars of each condition that each subject would experience, the overall experiment was divided into two separate studies.

Study 1 was designed to serve as a basic replication of a well-established effect-the activation of more than one interpretation of a lexical ambiguity (even following the occurrence of a strongly biasing context) (e.g., Swinney, 1979; Onifer & Swinney, 1981; Tanenhaus, et al., 1979). Thus, in this study, subjects heard the entire script of 90 sentences, and experimental and control probes for the primary and secondary meanings of the ambiguity were presented at the points immediately following occurrences of the ambiguity in the sentences (Probe Position 1). However, this study was also part of another, larger, experiment (presented elsewhere), and there was one difference between the sentences presented in Study I and those used in Study 2. This difference occurred in the sentence *following* the occurrence of the ambiguity (i.e., sentences in the first and second studies were essentially identical up through and including the occurrence of the critical ambiguous word); sentences in Study 1 were concluded so as to allow a conjoined sentence to be attached. As Study 1 only examined processing up to the point of the ambiguous word, the effects discovered therein should be essentially no different than if Probe Position 1 were run on the sentences of Study 2, and hence the data from this first study are presented here to complete an overall picture of processing for these words. It is important to note here that all critical issues in this paper hinge totally on effects found in Study 2. However, Study 1 set up the issue nicely, and hence is briefly discussed here.

There were four visual probe conditions in Study 1 (created by assigning the four probe types to four separate lists, such that only one probe type from any one sentence was present in any list, and all four probe types were equally represented across the four lists). Thus, there were four experimental subjects' groups conditions, corresponding to the four lists, and 12 subjects were assigned to each group condition (one extra subject was run in three of the groups, for a total of 51 subjects).

Study 2 was comprised of the 90 object-relative construction sentences precisely as described in the section titled Creation and Pretest of Object-Relative Construction Sentence Materials, above. This study contained two Probe Positions 2 and 3 factored into the design. In this study, two tapes were created which were identical, except that Probe Points 2 and 3 were equally counterbalanced across the tapes. These two audio tape conditions were combined with four visual probe list conditions (identical to those described in the previous paragraph for Study 1). Thus, Study 2 contained eight between-subjects groups, created by the 2 Tapes X 4 Lists conditions. This design allows each subject to experience all of the experimental conditions (probe types and probe positions), distributed across different items. (Eleven subjects were assigned to each of the eight experimental groups in this second study, for a total of 88 subjects.)

Procedure for Studies 1 and 2

Subjects in both studies in the experiment experienced the same procedure. Subjects were comfortably seated in front of a computer monitor in a sound attenuated booth, and were told that they had two tasks. The first was to listen to and understand sentences that they would hear over headphones. They were told that, at certain times during the experiment, they would be tested on the content of the sentence they heard. Their second task was to watch the computer screen, and, whenever they saw a word appear on the screen, they were to name it aloud as rapidly as they could. As in all CMLP studies of this type, the auditory sentence continued on, uninterruptedly, even while the visual probe was presented.

Each subject's attention was focused on the middle of a screen (where the word would appear) by means of a black mask that allowed only a small word-sized area of the screen to be visible. The subject's voiced response was recorded via microphone by a tape recorder and by the computer. Onset of the response stopped timing which had initiated with onset of the visual probe word. The RTLAB program controlled for all of the usual timing problems in video display (raster position, etc.), so timing was accurate to less than 1 msec. This program also recorded all reaction-time responses. A four-track reel-to-reel TEAC tape deck was used to present the auditory sentences to the subjects.

The experiment was paused five times for each subject (each pause following a filler sentence and prior to another filler sentence) and subjects were asked a multiple-choice question about the content of the sentence they had just heard. Following each such pause, subjects were reminded of the task.

RESULTS

Data from both studies were treated similarly. In all cases, data for each subject were screened for errors on an individual basis (errors were defined as speaking an incorrect word, failure to respond, or failure of computer to register a response). In addition, each subject's data were screened for data points lying 3 standard deviations above or below his/her mean reaction time. Any subject with more than 10% errors was dropped from the analysis. Similarly, any subject who missed more than one of the comprehension questions was dropped from the analysis. Only data for remaining subjects are described in the analyses below. For the remaining subjects, screens resulted in loss/dropping of an average of 2 data points per subject from the analysis (less than 5% of the data). In both studies, means were calculated for each subject for each experimental condition. These data were then submitted to planned comparisons and overall analyses of variance, as described below.

Additionally, following completion of the study it was discovered that five of the sentences used had technically failed to completely meet the Tabossi (1988) 75% agreement about context criteria (and our related completion test criteria for the context, in one case), and these five items were dropped from any further analysis, leaving 35 exemplars in the study.

Recall that Study I examined of activation for each meaning of the ambiguous word immediately following its occurrence in the sentence. (Again, because this study is being presented elsewhere, data analysis for this study is merely summarized here in order to connect meaningfully with data from the Study 2.) Means for each of the experimental conditions, calculated across all 51 subjects, are provided in Column 1 (Probe Position a) of Table I. As can be seen in this table, priming occurred for naming times to the related probe (as compared to the control probe) for both the primary and secondary meanings of the ambiguity. There were no significant effects for the between-subjects materials counterbalancing factors (lists) in this study. There was, however, a main effect for probe type (related vs. control) [$F(1, 47) = 4.844$, $p = .001$]. This effect did not interact significantly with the ambiguity meaning factor (primary vs. secondary). Planned a priori comparisons were performed on the related versus control probes for each of the ambiguity meanings. There was a significant priming effect in each of these cases: $t_{50} = 2.242$, $p = .015$ for the primary meaning and $t_{50} = 1.805$, $p = .038$ for the secondary meaning. Thus, this study replicated a number of prior studies which demonstrated that both meanings of an ambiguous word are activated immediately following occurrence of that word in a sentence, even a sentence with a strong prior biasing context of the form used by Tabossi (1988).

Study 2 examined precisely which meanings of the ambiguity, if any, are activated at two later pointsin the sentence: Probe Position 2-a point 700 msec before the offset of the verb which licenses a gap (effectively, before the verb which licenses a gap is processed)-and Probe Position 3-a point just after that gap is licensed (immediately following the verb).

Table I. Mean Reaction Times (in msec) for Related and Control Probe Naming Times for Both Primary and Secondary Meanings of the Ambiguity at All Three Probe Positions

Meaning of Ambiguity	Probe type	Probe Position I	Probe Position 2	Probe Position 3
Primary	Related	521[a]	510(n.s.)	496[a]
	Control	533	513	514
Secondary	Related	529[a]	510(n.s.)	507(n.s.)
	Control	537	515	509

[a]Significant difference between related and control contrast; n.s. = nonsignificant difference for related-control contrast. See text for details.

Means for each of the experimental conditions (related vs. control for both the primary and secondary meanings of the ambiguity, at both Probe Positions 2 and 3) were calculated across all subjects and are presented in the second and third columns of Table I.

The overall analysis of variance (ANOVA) was run on individual subjects' data employing tapes 2 and lists (4) as between-subjects factors (materials counterbalancing factors) and Probe Positions 2 and 3, ambiguity meaning (primary and secondary), and probe type (related and control) as within-subjects factors. Overall significant main effects of probe position [$F(1, 83) = 6.493$, $p = .01$], probe type [$F(1, 83) = 7.267$, $p = .008$], and lists [$F(3, 83) = 2.843$, $p = .042$] were found. Additionally, a significant effect was found for the Ambiguity Meaning x Probe Type x Probe Position interaction [$F(1, 83) = 3.768$, $p = .055$].

These effects were further examined via a priori planned (one-tailed) comparisons of the probe type condition (related vs. control probes) for each of the ambiguity meaning conditions (primary and secondary) at each of Probe Positions 2 and 3. At Probe Position 2, there was no significant priming for related (compared to control) probes for either the primary and secondary meanings of the ambiguity ($t_{87} = .49$, $p = .31$; $t_{87} = 1.03$, $p = .15$, respectively). At Probe Position 3, we had a different pattern of results. There was a significant priming effect for related (compared to control) probes for the primary meaning of the ambiguity ($t_{87} = 3.26$, $p = .00005$, with subjects as a random factor; $t_{34} = 1.6$, $p = .057$ with items as the random factor). However there was no significant effect for related (compared to control) probes for the secondary meaning ($t_{87} = .56$, $p = .29$)

Finally, in order to determine whether the nonsignificant priming effect for the primary meaning of the ambiguity at Probe Position 2 differed from the significant priming effect found for this same meaning at Probe Position 3, an ANOVA was run for each of the ambiguity meanings involving just probe position and probe type (related/control) as within-subject factors. For the ANOVA involving only the primary meaning of the ambiguity, there were main effects for both probe type [$F(1, 83) = 7.945$, $p = .006$] and probe position [$F(1, 83) = 4.322$, $p = .04$] and, critically, an interaction of Probe Type x Probe Position [$F(1, 83) = 4.697$, $p = .033$]. Thus, not only is there a significant activation of the primary meaning of the ambiguity at the gap (Probe Position 3), but as demonstrated by the significant interaction (i.e., when Probe Position 3 effects are compared to the nonsignificant activation shown at Probe Position 2), this represents a significant reactivation of the meaning. For the ANOVA involving only the secondary meaning of the ambiguity, there were no significant main effects nor was there a significant interaction of Probe Type x Probe Position [$F(1, 83) = .094$, $p = .76$].

DISCUSSION

The results of these studies support four basic conclusions. First, there is an immediate reactivation of a structurally marked antecedent filler once a structural gap is licensed during language processing (by a verb requiring a direct object) in object-relative constructions. [Note that the filler in these constructions is structurally marked by the immediately following relativizer (e.g., *that, which*).] This conclusion is supported by the significant interaction of probe type (related vs. control) for the primary meaning of the ambiguity across Probe Position 2 (prior to verb) and Probe Position 3 (at the gap). Given that there was no significant activation for this meaning at Probe Position 2, but there was significant activation for this meaning at Probe Position 3 (and that this change in activation was significant between these two positions), the conclusion to be drawn is that the comprehension device's need for recovery of the canonical order of sentence elements resulted in the immediate reactivation of the antecedent at the position where it was required (immediately following the verb). This points directly to the immediate and on-line use of structural knowledge in coreference processing of the type examined here.

Second, these results demonstrate that, when the putative confound of goodness/plausibility-of-fit of the probe type materials with the sentence (as discussed in McKoon & Ratcliff, 1994) is controlled for, evidence for reactivation of the antecedent is *still* obtained. That is, given that the pattern of priming effects across the three test points did not correspond at all to either of the patterns of plausible-continuation and general-relatedness ratings obtained for these materials, the effects cannot be argued to have been a function of such a confound. In fact, examination of the continuation/fit ratings reveals that (while any differences were likely too small to cause any difference in processing in any event) many of the reaction time (priming) effects went in the opposite direction to that which would have been predicted by the McKoon and Ratcliff confound argument. Thus, there are two possible conclusions. The first is that effects from a standardly and appropriately employed CMLP task are not amenable to intrusion by plausible/good fit or implausible/poor fit probes; the ongoing auditory processing required by this task does not encourage continuation or plausible continuation to be performed. This is not to say that some tasks will not demonstrate this effect-there are many processing tasks (including that used by McKoon and Ratcliff) which will encourage or even force integration of probes into ongoing processing of other material. The CMLP technique, when used normally, does not appear to be one of these. The second conclusion is that these data argue that data from prior CMLP studies are likely to be valid indicators of antecedent reactivation (as was argued, for example, by Nicol et al., 1994).

It also appears that a deep or conceptual level representation of the sentence is examined by the sentence comprehension device in an effort to determine the antecedent to a structural gap. That is, while there is evidence from the first study that both meanings of the ambiguity were momentarily activated when the lexical ambiguity was first encountered, the second study revealed that the reactivation of that ambiguity at the gap only involved reactivation of the contextually appropriate meaning (the primary meaning, in this study). Context is known to allow rapid choice of a single appropriate meaning for an ambiguous word (e.g., Swinney, 1979; Tanenhaus et al., 1979) during sentence processing. This meaning becomes part of the underlying representation of the sentence, and it is only *this* meaning that was reactivated by the reference-seeking element (the gap) in the present study. Hence, these data support a view that coreference (at least syntactically constrained coreference) is carried out across a deep (nonechoic or acoustic) memory representation of the sentence.

Finally, these data argue against the validity of a compound-cue hypothesis of priming. Ratcliff and McKoon (1988, 1994) have argued that all priming effects can best be accounted for by a compound-cue concept. This is a formal notion, built on the Gillund and Shiffrin (1984) model of memory recall and recognition, which proposes that, in priming, the prime and target form a compound cue, in which the compound or joint occurrence familiarity strengths between two lexical elements [the cue (the prime) and target] are built up in memory. Connection between a cue and a target (and everything else) is obtained upon the appearance of an explicit cue, and retrieval of a target is facilitated by the existence of strong cue-target compound familiarity strengths in memory. In this work, it is argued that structural factors are not relevant to coreference processing, as revealed by priming.

There were a number of facts in the current data that this theory cannot account for. First, there was priming (at Probe Position 3) for a word, when there is no cue available to prime the reaction time response to the probe. That is, it was a gap (a missing element) in a structural analysis that was causing the reactivation of an antecedent that was presumably causing priming for the related probe. There was no explicit cue in the sentence to cause priming at the gap. However, there was priming being obtained, without a cue. The compound-cue hypothesis has no explanation for such an event. If, for example, the argument should be attempted that it was the entire context (biased toward the primary meaning) that was causing priming at Probe Position 3, then there is no explanation in this theory as to why there was no similar priming for that meaning at Probe Position 2. After all, the biasing context in the sentence occurred before Probe Position 2, and, in fact, Position 2 was closer to such context than Position 3. In all, the compound cue hypothesis fails to account for these data.

Further, recall that this task is a naming task. As McNamara (1992) has pointed out (and as Ratcliff and McKoon, 1994, in their reply to McNamara did not dispute), the compound-cue model cannot account for priming with naming data at all. This follows, as they stated, from the fact that while lexical decision tasks involve a binary decision, naming responses involve the retrieval of one of a very large number of names from the lexicon. The compound-cue retrieval model is comfortably built to the characteristics of retrieval as involved in the binary decision (retrieval/non-retrieval), but nothing more. Thus, while some type of metaphoric extension might be argued rhetorically in the case of naming data such as these, the formal model itself holds no explanation for the data.

Overall then, this study argues for the active role of structural processing in the reactivation of antecedents at gap sites, for the use of a deep or conceptual representation for determination of which the appropriate antecedent is, and for the efficacy of the use of the cross-modal lexical priming task in detailing these processes.

REFERENCES

Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 19,* 35-43.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, The Netherlands: Foris.

Clifton, C. & Frazier, L. (1986). The use of syntactic information in filling gaps. *Journal of Psycholinguistic Research, 15,* 209-224.

Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectation in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23,* 696-708.

Cloitre, M., & Bever, T. (1988). Linguistic anaphors, levels of representation and discourse. *Language and Cognitive Processes, 3,* 293-322.

Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11,* 283-294.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language processing: Psychological, computational, and theoretical perspectives.* New York: Cambridge University Press.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9,* 427-473.

Fodor, J. D. (1988). On modularity in syntactic processing. *Journal of Psycholinguistic Research, 17,* 125-168.

Fodor, J. D. (1993). Empty categories in sentence processing: A question of visibility. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing' The Sperlonga Meeting I1.* Hove, England: Erlbaum.

Frazier, L. (1987). Processing syntactic structures: Evidence from Dutch. *Natural Language and Linguistic Theory, 5,* 519-559.

Frazier, L., Henstra, J., & Flores d'Arcais, G. B. (1995). Finding candidate antecedents: Phrases or conceptual entities. Unpublished manuscript, University of Massachusetts.

Gamsey, S., Tanenhaus, M., & Chapman, R. (1989). Evoked potentials and the study of sentence comprehension. Sentence Processing [Special issue]. *Journal of Psycholinguistic Research, 18,* 51-60.

Gillund, G. & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91, 1-67.*

Hankamer, J., & Sag, 1. (1976). Deep and surface anaphora. *Linguistic Inquiry, 7,* 391-426.

Hickok, G. (1993). Parallel parsing: Evidence from reactivation in garden path sentences. *Journal of Psycholinguistic Research, 22,* 239-250.

McElree, B., & Bever, T. (1989). The psychological reality of linguistically defined gaps. Sentence Processing [Special issue]. *Journal of Psycholinguistic Research, 18,* 21-35.

McKee, C., Nicol, J., & McDaniel, D. (1993). Children's application of binding during sentence processing. *Language and Cognitive Processes, 8,* 265-290.

McKeon, G., & Ratcliff, R. (1994). Sentential context and on-line lexical decision tasks. *Journal of Experimental Psychology: Language, Memory and Cognition, 20,* 1239-1243.

McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Language, Memory and Cognition, 20,* 1219-1228.

McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99,* 650-662.

Nagel, H. N., Shapiro, L. P., & Nawy, R. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research, 23,* 473-486.

Nicol, J. (1988). *Coreference processing during sentence comprehension.* Unpublished doctoral dissertation, MIT.

Nicol, J. (1993). Processing syntactic dependencies: Task specific effects. Paper presented at the Sixth Annual CUNY Conference on Sentence Processing, Amherst. MA.

Nicol, J., Fodor, J. D., & Swinney, D. A. (1994). Using cross modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Language, Memory and Cognition, 20,* 1229-1238.

Nicol, J., & Swinney, D. (1989). The role of structure and coreference assignment during sentence comprehension. Sentence Processing [Special issue]. *Journal of'Psycholinguistic Research, 18,* 5-19.

Nicol, J. L., & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence for semantic priming. *Journal of Psycholinguistic Research, 22,* 207-237.

Onifer, W., & Swinney, D. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9,* 225-236.

Osterhout, L., & Swinney, D. (1993) On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research, 22,* 273-286.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385-408.

Ratcliff R., & McKoon, G. (1994). Retrieving information from memory: Spreading activation theories versus compound cue theories. *Psychological Review, 101, 177-*184.

Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech, 25,* 385--401.

Stowe, L., Tanenhaus, M., & Carlson (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech, 34,* 319-340.

Swinney, D., & Osterhout, L. (1990). Inference generation during auditory language comprehension. *The Psychology of Learning and Motivation, 25,* 17-33

Swinney, D., & Zurif, E. (1995). Syntactic processing in aphasia. *Brain and Language,* 50, 225-239.

Swinney, D., Ford, M., Frauenfelder, U., & Bresnan, J. (1987). *The time course of co-indexation during sentence comprehension.* Paper presented at the Psychonomic Society Meeting, Seattle.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645-660.

Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition, 7,* 159-165.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27,* 324-340.

Tanenhaus, M. K., Carlson, G., & Seidenberg, M. S. (1985). Do listeners compute linguistic representations? In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives.* New York: Cambridge University Press.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427-440.

Tanenhaus, M. K., Stowe, L. A., & Carlson, G. (1985). *The interaction of lexical expectation and pragmatics in parsing filler-gap constructions.* In Proceedings of the Seventh Annual Cognitive Science Society Meetings, pp. 361-365.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic Processing, *Journal of Verbal Learning and Verbal Behavior, 16,* 683-692.

Zurif, E., Swinney, D., Prather, P., & Love, T. (1994). Functional localization in the brain with respect to syntactic processing. *Journal of Psycholinguistic Research 23,* 487-498.

Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language,* 45,448-464.