

Comprehension of Deep and Surface Verbphrase Anaphors

Michael K. Tanenhaus and Greg N. Carlson

Department of Psychology, University of Rochester, New York, U.S.A.

Linguistic research on anaphora (Hankamer & Sag, 1976; Sag & Hankamer, 1984) suggests that anaphors can be divided into two classes: *Surface* anaphors that find their antecedents in some level of linguistic representation, and *deep* anaphors that find their antecedents in a discourse model or a corresponding mental representation. In three experiments, we tested the hypothesis that the syntactic form of the antecedent for a subsequent anaphor would affect the difficulty with which surface anaphors but not deep anaphors would be comprehended. In a "makes-sense" judgement task, surface anaphors were judged to make sense more often when the antecedent was introduced in a phrase that was syntactically parallel to the anaphor than when it was syntactically non-parallel. In contrast, the syntactic form of the antecedent did not affect judgements to the deep anaphors. Parallelism did, however, influence comprehension times for both types of anaphors. The results provide qualified support for the hypothesis that deep and surface anaphors access different types of representations during comprehension.

INTRODUCTION

One of the chief linguistic expressions of the general phenomenon of *anaphora* is found in the use of definite third person pronouns, such as "he" or "them". Much linguistic work on anaphora has focused on the

Requests for reprints should be addressed to Michael K. Tanenhaus, Department of Psychology, Meliora Hall, University of Rochester, Rochester, New York 14627, U.S.A.

This research was supported by NSF grants BNS 8217378 and BNS 8617738. A preliminary version of the report was completed while the first author was partially supported by a fellowship from the Max Planck Institute for Psycholinguistics, which he gratefully acknowledges. We would like to thank Kent Crowell and Mike Anderson for their assistance in conducting the experiments, Sania Hamilton and Ed Scalfina for collaborating with us on Experiment 3, and Greg Murphy and Ivan Sag for helpful discussions about the research. This article also benefited greatly from the comments of two anonymous reviewers and, especially, William Marslen-Wilson. A short report of this research appears in the *Proceedings of the eighth annual meeting of the Cognitive Science Society*, Amherst, Mass.

syntactic and semantic relations that definite pronouns can or must bear to their antecedents, and most psycholinguistic studies of anaphora have likewise focused on the comprehension of definite pronouns with antecedents in the same sentence or in prior discourse (for a review, see Garnham, 1987). As is well known, though, natural language employs a wider range of anaphoric expressions than just the definite pronouns. Some of these devices are familiar to psychologists (such as definite nounphrases with "the" as the article), whereas others are not generally known outside of linguistics (e.g. "stripping": Ross, 1967). In the end, there is a surprisingly rich array of anaphoric devices employed by English and other natural languages; many of these devices turn out to be implicitly expressed, chiefly as omitted constituents that need to be "filled in" anaphorically (as when someone says "John bought three green candies and two yellow", one understands by this two yellow *candies*, thereby "filling in" the omitted noun following the word "yellow").

Hankamer and Sag (1976) have proposed that the range of anaphoric devices can be divided into two main classes, which they label as *deep* and *surface* anaphors. Sag and Hankamer (1984) further claimed that these classes differ in terms of the level of representation that must be accessed to determine their antecedents; roughly, a deep anaphor directly accesses a conceptual level of representation in a discourse model or mental model (Garnham, 1981; Johnson-Laird, 1983; Webber, 1981), whereas surface anaphors must first access a purely linguistic level of representation. Thus, they point out, deep anaphors do not require a linguistic antecedent, because when a pronoun's interpretation is determined by non-linguistic context it is then "pragmatically controlled". "Pragmatic control" refers to situations in which an antecedent of a pronoun is presented in the context of use, but not introduced explicitly in a linguistic expression. This is common with pronouns, as when a person asks "Who is she?" at a party, gesturing to the person intended to be referred to by "she". Or, suppose two men sharing an apartment return home at the end of the day, finding the place ransacked. One can turn to the other and say "I bet the cops did it", where "it" is taken to be referring to the ransacking of the apartment, even though no words to that effect had yet been uttered.

But surface anaphors, by virtue of the kind of representation required to assign them antecedents, take a linguistically introduced antecedent, and cannot be pragmatically controlled. For instance, in the ransacking example just discussed, one of the two men cannot turn to the other and ask (felicitously), "I wonder who?" On the other hand, if the other man first comments, "Somebody ransacked our apartment!", then the question "I wonder who?" would be felicitous, as under these circumstances an appropriate antecedent ("ransacked the apartment") would have been introduced linguistically. Simply having the idea in mind of ransacking the

apartment is not sufficient to license the use of such cases of surface anaphora as these; however, as we have seen, it is sufficient to support the use of a deep anaphor.

A second difference between deep and surface anaphors is that the surface anaphors require that their linguistic antecedent be presented in a suitable form (being sensitive to aspects of linguistic form), whereas deep anaphors are not sensitive to the syntactic form of the expression introducing the antecedent, when the antecedent is introduced linguistically. This difference is one of "syntactic parallelism". Consider the phenomenon of VP-ellipsis in English, where (roughly) a repeated verbphrase may be omitted, but understood anaphorically. The antecedent itself must also be a verbphrase—a type of linguistic constituent. Consider sentence (1) from Hankamer and Sag (1976):

1. Someone had to *take the oats down to the bin*.

The italicised constituent is a verbphrase. Now, consider uttering (2) immediately after (1):

2. . . . So Sandy did _ .

The blank indicates the position of the omitted verbphrase, which is understood in the context of (1), as taking the underscored verbphrase in (1) as its antecedent.¹ Now, suppose instead that in place of (1) we introduce the roughly synonymous (3):

3. The oats had to be *taken down to the bin*.

The (relevant) verbphrase in (3) is italicised. If sentence (3) instead precedes (2), the result is infelicitous:

- 3'. ??The oats had to be taken down to the bin. So Sandy did _ .

The explanation is that filling the omitted verbphrase with "taken down to the bin" results in an ill-formed structure, or at least a structure that is difficult to interpret. This indicates that VP-ellipsis is in the category of surface anaphora.

It is clear, though, that even if the message of (1) is instead encoded as (3), the concept of taking oats down to the bin remains. Thus one would

¹The parallelism requirement is actually somewhat more abstract. For example, surface anaphors can be used felicitously when the tense of the verb in the antecedent differs from that of the anaphor, as in "John nearly caught a fish yesterday. Tomorrow he is sure that he will (*catch* a fish)." Another example involves "indexical" expressions, as in "John hates to take his shower before eating breakfast, but his wife actually prefers to (*take her* shower)." Sag and Hankamer (1984) discuss these and similar examples in some detail.

expect that a deep anaphor could take this concept as its antecedent. As predicted, with a deep anaphor, (4) is a coherent discourse:

4. The oats had to be taken down to the bin. So Sandy did it.

The claim that anaphors divide into deep and surface categories is particularly striking in light of the fact that, in many cases, a sentence with a deep anaphor appears to have the same interpretation as the corresponding sentence with a surface anaphor. Consider the examples in (5):

5. We wanted Mary to paint the garage.
 a. . . . but she absolutely refused.
 b. . . . but she absolutely refused to.

Both continuations, (5a) and (5b), are equivalent, yet (5a) contains a deep anaphor ("null complement anaphora"), whereas (5b) contains a surface anaphor (VP-ellipsis). Nevertheless, the claim is that the (5b) case involves recourse to a level of linguistic structure that (5a) does not.

The most extensive experimental test of the Sag-Hankamer hypothesis was conducted by Murphy (1985a). Murphy examined the effect of three variables on the comprehension of target sentences that contained closely matched deep and surface anaphors: (1) the length of the antecedent; (2) the distance between the anaphor and its antecedent; and (3) the syntactic parallelism of the antecedent. Although each manipulation influenced reading times to the target sentences, none affected the deep and surface anaphors differentially.

Murphy reasoned that the length of the antecedent should affect comprehension of surface anaphors, on the assumptions that interpreting a surface anaphor requires replacing the anaphor with a "copy" of the antecedent and copying a long phrase should take longer than copying a short phrase. Murphy also assumed that accessing a conceptual representation of the antecedent should not be affected by the length of the phrase introducing the antecedent. Thus he argued that antecedent length should not affect comprehension times to a deep anaphor if, in fact, deep anaphors are interpreted using conceptual representations.

Examples of the short and long antecedents are given in (6a) and (6b), respectively, for the surface and deep anaphors given in (6c) and (6d):

- 6a. Jimmy swept the tile floor behind the chair.
 6b. Jimmy swept the tile floor behind the chair free of hair and cigarettes.
 6c. Later, his uncle did too.
 6d. Later, his uncle did it too.

There was a main effect of length, with target sentences being read faster with the short antecedents, but there was no interaction between length

and type of anaphor. Moreover, the effects of length were markedly reduced when the sentence introducing the antecedent and the sentence containing the anaphor were separated by an intervening sentence. Murphy concluded that when the verbatim form of the antecedent was available in memory, both deep and surface anaphors are interpreted by a "copying" process, whereas when the form of the antecedent is not available, "plausible reasoning" is used to construct an interpretation for the anaphor.

However, the manipulation of antecedent length also introduced potential scope and attachment ambiguities which could have complicated comprehension of both types of anaphors. Thus, the length effect alone does not provide convincing evidence for a copying process. The fact that the length effect disappears with intervening material is consistent with Murphy's hypothesis but, as we will see, the interpretation of the distance effect, itself, is problematic. We should also note that copying is not the mechanism by which a surface anaphor would be associated with its antecedent in most current linguistic theories. The more standard assumption is that anaphors are co-indexed with their antecedents, i.e. linked by pointer to the antecedent. On an indexing or pointer account, there is no reason to expect that the length of the antecedent should have different effects on deep and surface anaphors.

Murphy also varied the distance between the antecedent and the anaphor. He contrasted a "near" antecedent condition in which the sentence containing the anaphor immediately followed the sentence containing the antecedent with a "far" antecedent condition in which a sentence intervened between the anaphor and its antecedent. If surface anaphors require access to the linguistic form of their antecedents, comprehending surface anaphors in the far condition should be especially difficult, because memory for linguistic form decays rapidly, whereas memory for conceptual information is much more robust (Sachs, 1967; Wanner, 1974). Thus the amount of material that intervenes between an anaphor and its antecedent should have smaller effects on the comprehension of deep anaphors than the comprehension of surface anaphors (see Malt, 1985, and Garnham and Oakhill, 1987, for studies examining the effects of distance on interpreting surface anaphors). Murphy found that comprehension times to both surface and deep anaphors were longer in the far antecedent condition than in the near antecedent condition. Although the interaction with distance was not significant, the difference between the near and far antecedents was numerically larger for the surface anaphors than the deep anaphors.

The lack of a significant interaction between distance and type of anaphor might have been partially due to the intervening sentence having introduced material that removed the antecedent from focus. For pro-

nouns and definite nounphrase anaphora, the distance between an antecedent and its anaphor has little effect on comprehension times for the sentence containing the anaphor as long as the antecedent remains in focus (Anderson, Garrod, & Sanford, 1983; Garnham, 1987; Lesgold, Roth, & Curtis, 1979). Thus, the fact that distance effects obtained with the deep anaphors in Murphy's experiments raises the possibility that the intervening material introduced topic shifts. Tanenhaus, Carlson, and Seidenberg (1985) report an experiment in which an intervening sentence increased comprehension times to surface—but not deep—anaphors. The intervening sentences in this experiment were usually stative sentences, which as a rule do not introduce topic shifts or focus shifts. The issue of how intervening material affects the interpretation of deep and surface anaphors remains unresolved, and further research is needed to separate clearly focus and topic shifts from distance effects.

Murphy also varied the syntactic parallelism of the antecedent for deep and surface anaphors, crossing parallelism with distance and with length of the antecedent. In the near antecedent condition, both types of anaphors took longer to comprehend when their antecedents were not syntactically parallel; however, when a sentence intervened between the antecedent and the anaphor, there were no effects of parallelism.

The finding that both deep and surface anaphors are equally dependent upon the syntactic parallelism of their antecedents is clearly incompatible with any form of Sag and Hankamer's (1984) hypothesis. This result is also somewhat surprising in light of the strong linguistic intuitions that parallelism is important only for surface anaphors. In addition, the results are inconsistent with results we obtained in a study that found syntactic parallelism effects only for surface anaphors (Tanenhaus & Carlson, 1984).

The present experiments manipulated syntactic parallelism in similar ways to Murphy's experiments; however, the materials and the task were somewhat different than his. Murphy embedded his context and target sentences in short paragraphs and used a reading time paradigm. We used pairs of stimuli in which a context sentence introduced an antecedent for an anaphor in a following "target" sentence, and we used a "makes sense" judgement task in which subjects were instructed to decide as quickly as possible whether or not the target sentence made sense given the context sentence. The measures in this task are the percentage of sentences judged to make sense and latencies to those sentences that are judged to make sense. We have adopted the judgement task in recent studies (e.g. Tanenhaus & Carlson, 1984) because it requires the subject to take the antecedent into account when reading the sentence with the anaphor and because in a number of experiments in our lab more stable data is obtained with this task than with simple reading time or comprehension time tasks.

To see why this might be the case, it is important to keep in mind that

comprehension time and reading tasks are in fact variants of signal detection experiments. In the instructions to a standard comprehension time experiment, subjects are typically told to press a button when they have "understood" a sentence, leaving what is meant by "understood" up to the subject. In effect, this leaves the subject free to set his or her own criterion. We find that subjects are often confused by what we mean by "understand", and it is common lore among researchers who conduct comprehension time experiments that the type of questions that are included to make subjects pay attention during the experiment have a great deal of influence on overall comprehension times. Presumably, this is because subjects use the questions to set their criterion for comprehension; if they can answer the question, they have understood the material. The makes sense task allows the subject to set a criterion using filler sentences that do not make sense. The criterion set by the subject will determine whether differences will be reflected in judgements or latencies or both, although we have not, to date, conducted parametric studies that manipulate the subject's criterion.

We conducted three experiments investigating the effects of parallelism on the comprehension of deep and surface anaphors. The first experiment manipulated parallelism by introducing the antecedent in either an active or a passive sentence (e.g. "Someone has to take out the garbage" vs "The garbage has to be taken out"). In the active version, there is a linguistic constituent that can serve as the antecedent for the anaphor, whereas the passive version does not contain an appropriate constituent.

The second experiment adopted a different approach to manipulating parallelism. Antecedents for verbphrase anaphors were either introduced in a verbphrase or in a nounphrase using a nominalised verb. The use of a nominalised verb allows an event, which is the right type of *conceptual* object for a verbphrase anaphor, to be introduced in a nounphrase, which is the wrong type of *linguistic* category for a verbphrase anaphor. The Sag-Hankamer hypothesis predicts that only conceptual-level representation should matter for deep anaphors and thus the nominalisation manipulation should affect the surface but not the deep anaphors.

The third experiment was conducted to eliminate a confound between type of anaphor in the Hankamer-Sag taxonomy and phonological explicitness of the anaphor. In most surface anaphors, the anaphoric element is a deleted phrase and thus is not realised phonologically. In contrast, for most deep anaphors, there is an explicit anaphoric element. However, there are some types of deep anaphors in which the anaphoric element is unrealised phonologically. One such type is *null complement anaphora*. In null complement anaphora, the missing infinitive or sentential complement of some verbs can function as an anaphor, as was illustrated earlier in (5). The same verbs that allow null complement anaphora also can be used as

