Processing Verb Phrase Anaphors[1]

Michael K. Tanenhaus
University of Rochester

and

Greg N. Carlson
University of Iowa

In this paper we present three experiments which investigate the hypothesis proposed by Hankamer and Sag (1976) that there are two distinct kinds of anaphors in natural language--"deep" and "surface" anaphors. "Deep" anaphors in English include among other things definite pronouns, "One"-pronominals, and Null Component anaphora, exemplified in (1) below.

1. a. John left. He was angry. (Definite pronoun)
   b. Mary bought a green car. Frank bought a red one. (One pronominal)
   c. Mary knew who was guilty. But she wouldn't tell___. (Null Component)

"Surface" anaphors, on the other hand, include examples of Verb Phrase Ellipsis, Gapping, and Sluicing.

2. a. Sander built a new house. Max did___, too. (Verb Phrase Ellipsis)
   b. William caught a barracuda, and Harry,___, a shark.
   c. Someone just called you. But I don't know who___.

All anaphors, we assume, fall into one of these two classes.

One of the primary differences between the two categories is that deep but not surface anaphors may find their antecedents in the general context of use (e.g. something pointed at or otherwise made salient). Surface anaphors, unlike deep anaphors, require the presence of a linguistically-expressed antecedent. Consider, for instance, a situation in which two people are watching a fisherman reel in a fish. Under those circumstances, one of the bystanders could turn to the other and say, "Do you think he'll eat it?" (using a deep anaphor), but not "And Bill's nephew,___a rainbow trout" (infelicitiously using a surface anaphor, meaning Bill's nephew caught a rainbow trout). Thus, deep anaphors take antecedents which may or may not be linguistically-introduced antecedents. We will assume that deep anaphors find their antecedents in some non-linguistic form of representation, while surface anaphors seek antecedents among linguistic representations.

Moreover, we follow Sag and Hankamer (1984) in assuming that the level of linguistic structure in which surface anaphors find their antecedents is more abstract than surface structure; in fact, a level of logical form seems to be the most appropriate level at the present time (this is a level of linguistic representation in which scope of operators, such as quantified NP's, is unambiguously represented). Although multiple levels of linguistic representation are assumed by many formal linguistic and AI theories, little if any processing evidence supports the need for such abstract levels. Deep and surface anaphors offer an ideal contrast for studying the processing of these representations, because in the same context of interpretation, they may index different aspects of mental representations on the way to the same final interpretation.

One of the more compelling arguments for a deep-surface distinction is that surface anaphors seem to require that their antecedents be constituents at the appropriate level of linguistic representation, whereas deep anaphors do not. Consider the examples in (3).

3. a. Someone has to take out the garbage.
   b. The garbage has to be taken out.
   c. But Bill refused to___.
   d. But Bill refused___.

Sentence (3a) may be felicitously followed by either (3c) or (3d) equally well, though (3c) is a case of surface VP-Ellipsis, whereas (3d) is an example of a deep Null Complement anaphor. However, if both are preceded in a discourse by (3b) instead, (3c)--the surface anaphor--becomes infelicitous, whereas (3d)--the deep anaphor--remains perfectly acceptable and easily interpretable. This is because in the logical form of (3a) the verb phrase ("take out the garbage") is a constituent, as it is on the surface, whereas the logical form of (3b) has no constituent assigned the meaning "take out the garbage," again the same as surface form in which "the garbage" and "(be) taken out" do not form a single constituent. Following Hankamer and Sag, we will refer to this as the "parallelism" requirement of surface anaphors, reflecting the intuition that the antecedent of a surface anaphor requires structure parallel to that required at the site of the anaphor itself; in the case of (3c) a VP, missing at the site of the anaphor, is required as an antecedent. On the other hand, deep anaphors find antecedents based on knowledge of situations and other conceptual phenomena, where the linguistic notions of category and constituency do not come directly into play. Hence there is no "parallelism" requirement for deep anaphors, like (3d).

In three experiments we manipulated the parallelism of the antecedent. Using a "makes sense" judgment task (henceforth, "the judgment task"), we asked subjects to read a context sentence and then decide whether a subsequently presented target sentence made sense given the context. This task was chosen because it provides both judgment and reaction time data, and because it requires the subjects to integrate the anaphor with preceding discourse in order to make the judgment. All of the context and target sentences were grammatical sentences in English, but some of the filler targets did not make sense given the context (e.g. "Bill won first prize. He was glad that he didn't___," or "Tom took out the garbage willingly. He objected to doing it.")

In our first experiment, we created non-parallel antecedents by changing active sentences into passives (e.g. "Somebody had to take out the garbage" vs. "The garbage had to be taken out.") The passive creates a non-parallel antecedent because the VP in the active (the only reasonable antecedent in the context) is no longer a constituent in the passive. We reasoned that if deep anaphors find their antecedents in conceptual representations, they should be equally comprehensible with both parallel and non-parallel antecedents as the context sentences should give rise to the same or nearly the same conceptual representations (though see the general discussion below).

Sample materials for this experiment are illustrated in (4). The parallel antecedent is represented in context sentence (4a), the non-parallel antecedent in context sentence (4b). The surface and deep anaphors (with a definite pronoun) were presented as exemplified in target sentences (4c) and (4d), respectively.

4. a. Someone had to take out the garbage.
   b. The garbage had to be taken out.
   c. But Bill refused to___.
   d. But Bill refused to do it.

Twenty sets of materials, similar to (4), with fillers, were counterbalanced across four presentation lists. The results for 32 subjects are presented in Table 1. An interaction was obtained between parallelism and type of anaphor in the judgment data, with parallelism affecting judgments to surface but not deep anaphors, and in the reaction time data. However, contrary to our expectations, parallelism did significantly affect comprehension times to deep anaphors as well as surface anaphors.

In the second experiment, we created non-parallel antecedents in a different way, by presenting antecedents in a nominalized form. This presents the non-parallel antecedent as a constituent (unlike in experiment 1), but a constituent of the wrong category for the surface anaphor--an *N* instead of a *VP*. Sample materials are presented in (5). The parallel and non-parallel antecedents are presented in the context sentences (5a) and (5b), respectively, and the surface and deep anaphors in target examples (5c) and (5d), respectively.

5. a. It always annoys Sally when anyone mentions her sister's name.
   b. The mention of her sister's name always annoys Sally.
   c. However, Tom did anyway out of spite.
   d. However, Tom did it anyway out of spite.

The results for 28 subjects are presented in Table 2. Again, parallelism interacted with type of anaphor in the judgment task, with parallelism having no effects on the proportion of deep anaphors judged to make sense, but strong effects on the surface anaphors. There was a main effect of parallelism in the reaction time data and no interaction with type of anaphor.

Summarizing the results of the first two experiments, we find that parallelism did not affect judgments to deep anaphors, whereas it had robust effects on judgments to surface anaphors. In contrast, reaction time (when the anaphor was judged to make sense) was affected by parallelism for both types of anaphors. Why should non-parallelism increase reaction times to comprehend deep anaphors, but have no effect on judgments? One explanation, proposed by Murphy (1982), is that deep anaphors in some way contain more clues to the nature of the antecedent, and are thus less dependent on surface form than surface anaphors. The partial dependence on surface forms accounts for the increased reaction times, while the additional "clues" facilitate judgment.

In order to test this hypothesis, as well as to test the more general observation that the critical factor in our experiments could be the contrast between null anaphors and phonologically-realized anaphors (e.g. "it"), we conducted an experiment that contrasted Null Complement anaphors (e.g. "Bill refused ___" as in (3d)) with Verb Phrase Ellipsis (e.g. "Bill refused to___" as in (3c)). Both anaphors are null. On Murphy's analysis, though, Null Complement anaphors contain fewer cues than VP-Ellipsis anaphors, although Null Complement anaphors fall into the category of deep, rather than surface, anaphors.

Twelve sets of materials were constructed in which parallel and non-parallel antecedents preceded either a Null Complement anaphor or a VP-Ellipsis anaphor, as exemplified above in (3). The materials were counterbalanced across four presentation lists, and 40 subjects were tested using the judgment task. Only twelve sets of materials were used because the number of English verbs that allow the construction of a Null Complement/VP-Ellipsis contrast is limited. The results are presented in Table 3. Again we see a robust interaction between type of anaphor and parallelism of the antecedent in the percentage of sentences judged to make sense, and a main effect of parallelism in the comprehension time data.

General Discussion

Summarizing the results of the three experiments, we find that the parallelism did not affect judgments to deep anaphors, while it had a robust effect on judgments to surface anaphors. In contrast, reaction time (when the anaphors were judged to make sense) was affected by parallelism for both types of anaphors. Our hypotheses account for the difference in judgments, but do not explain why parallelism should not differentially affect reaction times. In particular, why should lack of parallelism increase reaction time to deep anaphors? There would appear to be two possible accounts consistent with our hypotheses. One is that subjects were making grammaticality judgments. In checking for grammaticality, attention is paid to the linguistic form antecedents are expressed in. The non-consistency of the antecedents of the passives requires extra steps in checking plausibility, and the nominalized forms represent the less usual sort of antecedent for pronouns standing for activities, verbal phrases being the more expected type of antecedent. But this hypothesis seems implausible given that all the sentences given subjects were in fact grammatical, and that subjects were basing judgments on understanding the sentences in context instead of attending to their formal properties. A more likely possibility, which we plan to pursue in further research, is that the manipulations we used to create non-parallel antecedents at the level of linguistic representation also had effects at the conceptual or discourse model level. More specifically, passivizing a sentence shifts focus to the underlying direct object (the surface subject), now in contrast to the backgrounded verb and possibly other remaining VP material emphasizing disunity, unlike the active counterpart. Similarly, the nominalized forms used in the second experiment present the actions as if presupposed or otherwise backgrounded, in contrast to the verbal forms (Kiparsky and Kiparsky (1971)). In both cases, using a (deep) VP anaphor requires a search which requires a shift of focus. These focus shifts should take time, but result in antecedents that are completely comprehensible. Thus, the focus hypothesis promises an account of the complete dissociation between the effects of non-parallelism on reaction time, and the effects on comprehensibility as indexed by the proportion of sentences judged to make sense.

In summary, our experiments support a processing difference between deep and surface anaphors. They further suggest that surface anaphors take linguistic antecedents. These results do not, however, address the question of what level or aspects of linguistic representation are important for the online interpretation of surface anaphors. In work in progress we are addressing this issue by manipulating parallelism in different ways in order to see which aspects of representation are important for the comprehension of surface anaphors.

Table 1

Results of the Passive Experiment

	Antecedent Type			
	Active (Parallel)		Passive Non-Parallel)	
Anaphor	% judged to make sense	(RT)	% judged to make sense	(RT)
Deep	94%	2181 msec	91%	2381 msec
Surface	89%	2165 msec	70%	2848 msec

Note: Judgments are to the sentence with the anaphor and reaction times are to those sentences judged to make sense.

Table 2

Results of the Nominalization Study

	Antecedent Type			
	Active (Parallel)		Passive (Non-Parallel)	
Anaphor	% judged to make sense	(RT)	% judged to make sense	(RT)
Deep	87%	2686 msec	87%	2952 msec
Surface	89%	2557 msec	71%	2923 msec

Note: Judgments are to the sentence with the anaphor and reaction times are to those sentences judged to make sense.

Table 3

Results of Null Complement Verb Phrase Ellipsis Experiment

	Antecedent Type			
	Parallel		Non-Parallel	
Anaphor	% judged to make sense	RT(msec)	% judged to make sense	RT(msec)
Null Comp. (Deep)	93%	18250	89%	21525
VPE	95%	20153	77%	21208

Note: Judgments are to the sentence with the anaphor. Reaction time data are to sentences judged to make sense.

References

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

Kiparsky, P., & Kiparsky, C. (1971) Fact. In D. Steinberg & L. Jakobovits, Semantics. Cambridge, England: Cambridge University Press, 345-369.

Murphy, G. (1982) Understanding anaphora. Stanford University doctoral dissertation.

Sag, I., & Hankamer, J. (1984) Toward a theory of anaphoric processing. Linguistics & Philosophy, 7, 325-345.

[1]This research was supported by Grant BNS-8217378 from the National Science Foundation. We thank Ed Sclafani and Sania Hamilton for collaborating with us on Experiment 3.