IVAN A. SAG AND JORGE HANKAMER

TOWARD A THEORY OF ANAPHORIC PROCESSING

0. Introduction

Linguistics and philosophers have long been concerned with the problem of formulating a precise theory of pronouns and other anaphoric elements or expressions in natural language. Proposed treatments have been quite diverse, but a common assumption in formal accounts of anaphora appears to be that anaphoric elements in general can be treated analogously to variables in logical languages whose semantic role is explicated by rules of interpretation which constitute a recursive definition of truth or meaning for the language.

Such definitions are highly idealized rule systems which abstract away from any particulars of the psychology of those who make use of them. In this paper, we suggest that a fundamental dichotomy among anaphoric expressions in natural language cannot be properly understood solely with reference to such an abstract semantic system. We argue that the properties of distinct types of anaphora can be understood only by distinguishing between those that reference specifically linguistic objects (and hence are essentially substitutive in nature) and those that reference more general sorts of mental representations. to understand the properties distinguishing the two classes of anaphoric expressions, one must understand that they are processed and assigned interpretations on the basis of distinct kinds of psychological objects.

1. Background

In Hankamer and Sag (1976) (henceforth HS) we argued that anaphoric processes fall into two classes, which we termed *deep* and *surface* anaphora. In support of this typology, we argued that (a) only deep anaphora can be used deictically, or in the terms of HS, can be 'pragmatically controlled'; (b) only surface anaphora required parallelism in syntactic form between anaphor and antecedent; and (c) only surface anaphora exhibit the 'missing antecedent' phenomenon (Grinder and Postal (1971), Bresnan (1971)).

Examples of 'deep' anaphora are the ordinary personal pronouns (1a), sentential *it* (1b, c), and the null complement in (1d):

Linguistics and Philosophy 7 (1984) 325–345. 0165–0157/84/0073–0325$02.10

(1) Merle smashed the Cadillac with a sledgehammer.
(a) *She* did it because *she* wanted to.
(b) She did *it* because she wanted to.
(c) *It* was good exercise.
(d) She did it whenever we let her ∅.

Examples of surface anaphora are VP ellipsis (VPE) (2a), sluicing (2b), gapping (3), and stripping (4):

(2) Morgan burned her initials on Sal's arm.
(a) At least I think she did ∅.
(b) But nobody knows why ∅.

(3) Benny will drive the car, and Max ∅ the truck.

(4) I'll take you to the movies, but ∅ not ∅ this week.

For full discussion of these constructions and the arguments for assigning them to the stated categories, see HS. Here we will simply illustrate the two major claims of the HS hypothesis.

The first claim, that 'deep' anaphora may be used deictically, but 'surface' anaphora may not, is illustrated by the differences in felicity of Sag's utterances in the following context:

(5) [Hankamer points gun offstage and fires, whereupon a blood-curdling female scream is heard]
Sag:

(a) *I wonder who ∅? [sluicing, surface]
(b) *I wonder who was ∅? [VPE, surface]
(c) I wonder who *she* was? [definite pronominalization deep]
(d) *Jorge, you shouldn't have ∅.[1] [VPE, surface]
(e) Jorge, you shouldn't have done *it* [S-*it*, deep]

The 'surface' anaphors require a linguistic antecedent, and do not sound right in the given context where nothing has been said; the 'deep' anaphors, on the other hand, can apparently be interpreted as referring directly to elements in the environment.

The second claim was that the anaphors which require a linguistic antecedent (the 'surface' anaphors) also have certain syntactic properties, most notably a requirement of parallelism in form between the antecedent expression and the anaphor. 'Deep' anaphors, even when they appear to have a linguistic antecedent in the discourse, do not require such parallelism:

(6) The children asked to be squirted with the hose, so
(a) they were Ø. [VPE, surface]
(b) *we did Ø. [VPE, surface]
(c) we did *it*. [S-*it*, deep]

In the (6) the antecedent clause is in Passive form, and VPE (a 'surface' process) is only possible when the target clause is also passive (as in (a)); The *it* of *do it*, on the other hand, is a 'deep' anaphor, and does not require such parallelism.

There have been publications (Schachter (1977), Williams (1977b)) contending that the classification proposed in HS is erroneous or illusory. Schachter suggests that there is no distinction between deep and surface anaphora regarding the possibility of deictic interpretation, and Williams proposes a different fundamental dichotomy. Hankamer (1978) counters Schachter, and the proposal of Williams is shown to be untenable in Sag (1979). We believe that the dichotomy between deep and surface anaphora as outlined in HS remains valid and nontrivial, and that any adequate theory of anaphora must account for the central observation exemplified in examples (5)–(6).[2]

While we maintain that the dichotomy of anaphoric processes which we originally noticed is still to be recognized, we now propose to revise our account of it. The view of anaphora advanced in HS was the following:

(7) (a) the processes by which interpretations are assigned to *surface anaphoric elements* must make reference to the *surface syntactic structure* of sentences in the surrounding discourse.
(b) the processes by which interpretations are assigned to *deep anaphoric* elements must make reference to either
(1) the *deep syntactic structure* (as developed in Chomsky (1965)) of sentences in the discourse or
(2) nonlinguistic elements present in the context of utterance.

Part (7a) expresses the HS theory of surface anaphora, according to which surface anaphoric elements were derived by rules of deletion under syntactic identity. The relation of syntactic identity, of course, holds only of pairs of *syntactic* objects. Deep anaphoric expressions, treated as syntactically primitive by HS, were assigned interpretation by rules of semantic interpretation, and these rules were thought to make reference sometimes to linguistic objects (deep structure constituents) and sometimes to extralinguistic objects, as outlined in (7b).

In this paper we will modify both parts of (7). Assumption (7a) was already questionable at the time of writing of HS,[3] and further work on surface anaphora, Verb Phrase Ellipsis in particular, has led to the conclusion that the processes by which interpretations are assigned to elliptical verb phrases must be sensitive to scope of logical operators and variable binding. This observation has led to the development of theories of VPE which make reference to logical representations instead of purely syntactic representations. In Section 2 we discuss such theories, which have proven superior to (7a) as an account of what we called 'surface' anaphora.

Part (7b) of the HS theory has not been shown to be untenable, but there is something very dissatisfying about the disjunction it contains. It says that the basis for the interpretation of a 'deep' anaphor can be found either in the deep structure of some sentence in the surrounding discourse (an abstract syntactic object), or in the physical environment (a concrete object, event, or state of affairs). These are two very different kinds of objects, and it is difficult to imagine exactly how the interpretive principles for deep anaphors could be devised so as to treat them in a unified fashion. The alternative is to assume that there are radically different interpretation processes in the two cases, an unsavory prospect in view of the fact that we have no evidence for such a further dichotomy.

In this paper we suggest that the interpretation of a 'deep' anaphoric element is determined by reference to the *interpretation* of its antecedent (in all cases when there is one), i.e. by reference to some object in a model of the world constructed by the interpreter of the sentence of discourse; while the interpretation of a surface anaphoric element is determined by reference to a linguistic representation associated with the antecedent, specifically a propositional representation of the kind generally called *logical form*.

This means, in effect, that the interpretation of what we used to call a deep anaphor is not mediated by its relation with an antecedent expression at all; it does not, in particular, involve reference to any representation of an antecedent expression, syntactic or semantic, deep or surface. Consequently we will henceforth cease to call such elements 'deep anaphors'; we will speak of *model-interpretive anaphora* (our former 'deep' anahpora) and *ellipsis* (our former 'surface' anaphora).

Our thesis here is that the contrasts observed in HS cannot be accounted for simply in terms of assigning relations between an anaphor and its antecedent at different levels of linguistic representation. We believe that the properties of anaphora in general, and the particular differences between the two kinds of anaphora which we distinguished, can only be understood in terms of a performance model: a model of how discourses are represented, produced, and comprehended.

We will advance a theory of anaphoric processing which distinguishes model-interpretive anaphora from ellipsis in terms of the means by which an interpretation is assigned to the anaphor. The theory rests crucially on the assumption that discourse understanding involves the construction by the understander of both a propositional representation of the immediate discourse and a ‘model’ of the world evoked by the discourse.

2. Ellipsis and logical form

The view of ellipsis adopted in HS was the standard one at the time, namely that elliptical elements are derived by rules of deletion under syntactic identity. In more recent investigation (Sag (1976a, b), Williams (1977)) it has been argued that this theory should be replaced by one in which the identity conditions on elliptical processes are stated entirely on ‘logical’ representations of a particular sort.[4] Following the insight of Montague (1974) and others, standard predicate calculus representations are eschewed in favour of logical representations that are highly determined by surface syntactic structure; in particular such representations contain structural units corresponding to surface VP’s. A sentence like (8) has a logical translation like (9), where the underlined expression is a complex one-place predicate taking the translation of the subject NP as argument.[5]

(8) Robin will like Sandy.

(9) W[[*like’* (*Sandy’*)] (Robin’)]

In Sag (1976a,b) the following identity condition is proposed:

(10) Delete a VP only if its logical translation is an alphabetic variant of some expression in the logical translation of the surrounding discourse.

Two expressions are alphabetic variants if they are identical down to variable indices and they do not contain distinct free variables. In simple cases, assuming the appropriate logical translations, the identity-of-logical-form theory (ILFT) and the previously held syntactic-identity theory make identical predictions.

The advantages of ILFT argued for by Williams and by Sag include an account of data like the following:

(11) Sandy thinks someone loves everyone.
(a) Chris does ∅, too.
(b) Chris thinks someone does ∅, too.

Ignoring possible readings where either quantifier has scope over *think*, (11a) is ambiguous in a way that (11b) is not, though the examples differ only in whether the matrix or embedded VP has been deleted; and either deletion would be allowed under the syntactic-identity theory. (11a) allows two interpretations: the content of the embedded proposition may be that represented by either scope assignment of the two quantifiers (EA or AE). In (11b) the embedded proposition can only be that represented by the existential quantifier taking scope over the universal (EA).

This surprising fact is predicted by ILFT. The EA reading is represented, appealing to the lambda-operator and principles of scope assignment, as (12) ('^' is Montague's intension operator). The matrix VP-translations are singly-underlined; the embedded VP-translations doubly-underlined.

(12) $$\text{think}'\ ([\hat{}\,Ex_1\underline{[\lambda x_2 Ax_3[\text{love}'(x_3)\ (x_2)]]}\ (x_1)])(\text{Sandy}').$$

$$\text{think}'\ ([\hat{}\,Ex_4\underline{[\lambda x_5 Ax_6[\text{love}'\ (x_6)\ (x_5)]]}(x_4)])(\text{Chris}').$$

The translation of each potential VPE target is without free variables and the appropriate condition of alphabetic variance holds in both cases. Hence, according to (10), deletion of either VP is possible on this reading.

The logical translation of the AE reading is given in (13).

(13) $$\text{think}'\ ([\hat{}\,\text{Ax}_3\text{Ex}_1\ \underline{[\text{love}'\ (x_3)]}\ (x_1)])(\text{Sandy}').$$

$$\text{think}'\ ([\hat{}\,Ax_6Ex_4\ \underline{[\text{love}'\ (x_6)]}\ (x_4)](\text{Chris}').$$

From (13) we see why (11b) cannot have a AE interpretation. The embedded VP-translations are those doubly-underlined. They contain distinct free variables and hence are not alphabetic variants. Hence, by (10), VP-Deletion cannot delete the embedded VP on this reading. No such problems arise in the case of deletion of the matrix VP, hence the facts of (9) are predicted.[6]

Numerous other disambiguation effects are also predicted by ILFT, e.g. (14).

(14) John$_i$ said Mary hit him$_i$.
(a) Bill did ∅ too.
(b) Bill said she did ∅, too.[7]

(14a) has a purely referential reading, where the content of what Bill said is that Mary hit John, and a 'sloppy' reading (Ross (1967)), where the content of what Bill said is that Mary hit Bill. (14b) has only the purely referential reading.

The fact that (14b) does not permit a sloppy reading is predicted by ILFT once the further assumption is made (see Sag (1976a,b)) that sloppy readings arise when pronouns are analyzed as variables bound by a λ-operator, as in (15).

(15) $\lambda x_1 \text{ say}' ([{}^\wedge\underline{\text{hit}' (x_1)}\,(\text{Mary}')])(\text{John}').$

The explanation for the impossibility of deletion of the embedded VP on the sloppy reading in (14b) is thus exactly analogous to the explanation for the impossibility of the AE reading in (11b).[8]

A similar account can be given of examples like (16), once the familiar assumption is made that *wh*-expressions are analyzed as variable-binding operators.

(16) Robin knows who ate what.
(a) Leslie does ∅, too.
(b) *Leslie knows who did ∅, too.

These examples constitute only a small part of the evidence discussed in detail in the works cited above. Ellipsis processes such as VPE do not actually make reference to surface syntactic representations, as had previously been thought, but rather crucially involve representations in a suitably-structured logical language.

3. A PROBLEM FOR ILFT

It is now generally acknowledged that the ILFT is superior to the classical surface-structure theories of ellipsis (both deletion and interpretive versions). ILFT, however faces an interesting class of problems, having to do with the interpretation of elliptical anaphors whose antecedents contain indexical elements. Consider the following examples, which are similar to examples noted by Partee (1975):

(17) A: Do you think they'll like me?
B: Of course they will ∅. [∅ = like you]

As indicated, the elliptical VP in B's response must be interpreted as equivalent to the form *like you*, and not to the form *like me*. Similar effects can be observed with other indexicals:

(18) A: Are you coming over here?
B: Yes, I am $\emptyset$. [$\emptyset$ = coming over there]

The problem such examples pose for ILFT is quite simple. Under virtually any assumptions about the nature of the logical representation, the expressions *like you* and *like me* have different logical representations. Hence, under ILFT the observed interpretation appears to be barred, while the nonexistent one is predicted to be possible.[9]

One might develop a solution to the problem just noted by formulating the identity condition on VPE in model-theoretic terms, if certain further assumptions are made. For example, suppose following Kaplan (ms.) that the logical representation language is provided with a semantics which makes crucial use of the notion of a context. Recursive rules specify truth and denotation for logical expressions at a time t, a world w, with respect to a variable assignment f, *when taken in a context c*. A Kaplan-context c includes a specification of the agent of c (c_{ag}) (*the speaker, roughly*), *the addressee of* $c(c_{ad})$, the time of $c(c_T)$, the place of $c(c_p)$, and perhaps other contextual features. A non-indexical term α is assigned an intension (I_α) by the model, but indexical terms such as *me′* or *you′* are not. Denotation is specified by rules such as the following ($[\![\alpha]\!]^{cwtf}$ is read: the denotation of α at world w and time t with respect to a variable assignment function f, when taken in the context c).

(19) If α is a basic non-indexical term, then $[\![\alpha]\!]^{cwtf} = I_\alpha(w, t)$

(20) (a) $[\![\text{me}']\!]^{cwtf} = c_{ag}$
(b) $[\![\text{you}']\!]^{cwtf} = c_{ad}$

similar rules would have to be stated for other indexical expressions.[10]

Having developed a semantic theory of this sort, we would state the recoverability condition on VPE as follows:[11]

(21) Delete VP$_b$ in S_b only if
(1) c_b is the Kaplan-context of S_b
(2) c_a is the Kaplan-context of some sentence S_a not subsequent to S_a in discourse.
(3) there is some VP$_a$ in S$_a$ such that AtAwAf$[[\![\text{VP}'_b]\!]^{c_b wtf} = [\![\text{VP}'_a]\!]^{c_a wtf}]$

This semantic recoverability condition provides a correct analysis of examples like (17) because the following equivalence holds.

(22) $\mathrm{AtAwAf}[[\![\mathrm{like}'(\mathrm{me}')]\!]^{c_a wtf} = [\![\mathrm{like}'\ (\mathrm{you}')]\!]^{c_b wtf}]$

The semantic recoverability condition also deals with all the facts which were successfully analyzed in ILFT. Note in particular the deletion of embedded VP's in the examples mentioned above is correctly blocked, as the presence of distinct free variables (as in (13) above) has a consequence that there is at least one variable assignment under which the relevant VP translations have distinct denotational values. On the basis then of the apparent success of such an analysis, one might conclude that VPE and other ellipsis processes require a treatment in terms of model-theoretic evaluations of logical translations.

4. Discourse models and propositional representations

From the perspective of discourse understanding, the theory of surface anaphora outlined in the previous section is somewhat perplexing. The identity condition in (21), for example, does not lend itself to any particularly intuitive interpretation in terms of what object is grasped by a discourse participant who encounters a missing VP. ILFT, on the other hand, did indeed lend itself to a plausible interpretation vis-á-vis discourse understanding. Assuming that discourse comprehension involves, as an initial step, constructing a logical form for an incoming sentence, the problem of interpreting a missing VP might be thought of simply as the problem of finding an appropriate chunk of logical form within what has just been comprehended.

We would like to defend a view like this, even in the face of the difficult examples discussed in the last section. Our proposal is to separate the interpretation of indexicals from the interpretation of elliptical expressions, so that while indexicals are interpreted by reference to a constructed model of the discourse, the interpretation of ellipses remains a rather simple copying of logical form. We will argue further that the interpretation of our former "deep" anaphors involves just the same direct reference to discourse-model entities as we propose for indexicals.

We have the following conception[12] of how discourse comprehension proceeds. Discourse participants, as they process incoming sentences, synthesize models of the ongoing discourse. These models are either part of or else intimately linked to broader models of the world. Discourse comprehension consists, in part, in integrating the content of newly-produced discourse into the discourse model. Exactly how these models are structured, e.g. how 'imagistic' they might be, is a psychological question which has no clear answer, to our knowledge.

Experiments conducted by Johnson-Laird and others (reported by Johnson-Laird (1980)) suggest that as a new piece of discourse is produced, what is comprehended first is a propositional representation, which we may take to be an expression in a suitably-structured logical representation language similar to those discussed in the previous sections. Thus at any given point in a discourse, what is present in the mind of a comprehender is a pair: a model of the discourse[13] to date, and a propositional representation of (part of) the immediately-preceding discourse. As the discourse proceeds, the content of the propositional representation, which is held in short-term memory, is integrated into the model. At some point then the propositional representation is discarded, making room for new propositional representations in a presumably quite limited short-term register.

This view of discourse provides an explanation for certain often-cited experimental results. In particular there are known recency effects on the ability of subjects to perform verbatim recall tasks, whereas gross content recall is virtually unaffected by placing the relevant stimulus sentences several sentences back in discourse.[14] These observations have led researchers to conclude that there is a specific short-term register for surface syntactic structures, but they are equally compatible with the assumption that it is propositional representations such as those postulated under ILFT that are stored in such a register.

As we will show, this view of discourse understanding, assuming a tandem manipulation of logical representation and discourse models, can in addition provide the basis for an intuitive account of several peculiar properties of anaphoric processes, including the fundamental dichotomy draws by HS. Immediately we will show that once discourse models are assumed, a slight modification in assumptions about the nature of the logical representation language enables us to return to a version of ILFT in accounting for the troublesome data discussed in the previous section.

5. Anaphoric processing

5.0. We suggest that, corresponding to the two sorts of objects grasped in the process of discourse comprehension, there are two ways in which the interpretation of an anaphoric element can be recovered:

(a) by reference to the representation of propositional structure of recent discourse which the understander has just constructed;
(b) by reference to constructs of the understander's discourse model.

The 'surface' anaphora of HS and subsequent investigations (or ellipsis processes) are those whose interpretations are assigned in manner (a); 'deep anaphora' are interpreted in manner (b) (hence our decision to refer to these as model-interpretive anaphora (MIA)).

5.1. First we will outline a solution to the problem of indexicals discussed in Section 2. Indexicals, like MIA in general, will be assumed to be interpreted directly by immediate reference to the discourse model; we will further assume that this interpretation takes place *simultaneously* with the construction of propositional representations.

Consider ex. (17) (repeated here):

(17) A: Do you think they'll like me?
B: Of course they will Ø. [Ø = like you]

In this example, at the point when the null VP is to be interpreted, the register of the understander contains a representation of the immediately preceding discourse in which 'like me' has been translated into $like'(a_1)$, where a_1 is an index to the entity representing the speaker of that sentence in the hearer's discourse model. Under this interpretation, indexicals are just those elements for which there are special interpretation rules linking a constant term in the propositional representation with an entity in the discourse model according to discourse-situational conditions. When speaker B says 'Of course they will', speaker A (or any other hearer) has only to look in his record of the propositional structure of recent discourse for a VP-sized unit, not containing any free variables, and fill it in for the missing predicate. The only candidate in this particular discourse is $like'(a_1)$, which yields the desired interpretation. It is crucial, of course, that the MIA be interpreted at the time (*real* time) the missing VP is interpreted. Hence it is crucial that the two kinds of anaphoric processing proceed simultaneously.

5.2. The above illustrates the general outline of our approach to anaphoric processing, which we will now present in some detail. First we will discuss the interpretation of MIA, which we regard as involving direct reference to constructs of the understander's discourse model; then we will present a more explicit account of the interpretation of ellipsis by reference to units in the propositional register.

Consider first the paradigm case of MIA, the definite pronouns. Just as with first and second person pronouns, we assume that third person definite pronouns used referentially are interpreted by immediate reference to entities in the discourse model.[15] For example, if speaker A says

(23) The bricks on the outside of my house are not made of mud.

any hearer who understands this sentence must set up in his discourse model several entities, among them one corresponding to the set of bricks on the outside of speaker A's house. If speaker A then continues.

(24) They're made of asbestos fibres mixed with wasp spittle.

the understander interprets the pronoun *they* simply by associating it with the already established entity in his discourse model which represents the set of bricks. His propositional representation of the sentence then contains some constant term, say a_{37}, which serves as an index to the appropriate entity in the discourse model.[16]

The deictic use of definite pronouns, and of MIA elements in general, is explainable simply because the discourse model may contain entities evoked by the discourse situation as well as entities evoked by what the participants say. Aside from the obvious fact that discourse models must contain entities corresponding to the discourse participants, we can assume that they also contain a representation of any other person or object which is clearly present and perceived by the participants, if that person or object is of any importance to them; and further that any event of significance which the participants witness is incorporated into their models. Such persons, objects, and events can be referred to by deictic expressions of the proper sort or by definite pronouns, just to the degree that the user is fairly certain that his interlocutor will asociate the form with the right entity in his discourse model (which presupposes that the referent is of sufficient significance to have been evoked in the hearer's model already).[17]

From our assumption that the model constitutes (in part, at least) a record of the content of the developing discourse, certain general features follow. In particular, the discourse model must have a means of representing not only simple entities such as people and things, but relations, properties, states, actions, etc. as well. For example, an act of uttering the sentence:

(25) Merle appears to be believed by everybody to be left-handed.

evokes in the discourse model of an understander, among other things, a representation of a situation (hypothetical, in this case) which would be described simply, were it true, as

(26) Merle is left-handed.

Exactly how discourse models which are adequate for the analysis of examples like this should be represented is a question to which we do not at the moment have an exact answer.[18] Here our only claim is that the analysis of MIA is facilitated by the assumption of discourse models which, unlike the representations of 'propositional structure' constructed directly by the sentence processor, are not tied directly to the syntactic surface structure of the sentences which evoke them, even when they are evoked linguistically.

the interpretation of 'sentential *it*', as in (27) or (28):

(27) But I don't believe it,

(28) And it might be true,

uttered on the heels of (25), depends in our view on associating the anaphoric element *it* with the state of affairs in which Merle is left-handed, which, evoked by the immediately preceding discourse, is represented in the understander's discourse model. It is because this model-theoretic construct is independent of the form of the sentence originally evoking it that 'sentential' *it*, and MIA in general, appear to be indifferent to the surface form of the antecedent in discourse, when there is one; such anaphoric elements are not interpreted directly by reference to their 'antecedents', but rather by reference to entities in the discourse model which may have been evoked by previous discourse.

In order to account for the interpretation of the anaphoric expression *do it*, we assume that discourse models also may simply contain, or else allow the computation of,'actions' as abstracted from their perpetrators. If a speaker says

(29) The oats need to be taken down to the bin,

the understander's discourse model will contain a representation of the action of taking the oats down to the bin, though no perpetrator or prospective perpetrator of this action has been mentioned. If someone then says

(30) I'll do it,

We assume that the understander's sentence processor constructs a propositional representation something like

(31) $W[[\mathrm{do}'(a_{17})]\,(a_{22})]$

where a_{22} indexes the understander's discourse-model entity corresponding to the speaker of (30), and a_{17} his representation of the action of taking the oats down to the bin.

Discourse-model representations of 'actions' may apparently also abstract from such things as instruments and affected parties, since (as is well known) a *do it* may be accompanied by explicit substitutions in these roles:

(32) Paul painted Harry all over with tincture of iodine, and Mary did it to me with strawberry jam.

Such examples have long been regarded as problematic for the theory of anaphoric interpretation since there is no linguistic unit which could reasonably serve as the antecedent of the pronoun *it*. It seems clear that what is needed is a representation of the action of painting someone all over with some substance; this action is directly represented by no unit in the surface structure of the antecedent clause, but *can be represented in the understander's discourse model*. This action will be represented as a relation involving a painter, a paintee, and a substance. If the representations in the discourse model are regarded as imagistic, we must imagine that the understander can abstract the actions out of such images, perhaps by substituting different participants; if they are essentially propositional in nature, the abstraction is already done, and we can say that the interpretation of a sentence containing *do it* simply involves adding to the model another instance of some appropriate action already represented, supplying different participants as indicated in the new sentence.

5.3. Surface anaphora, on the other hand, are interpreted not by reference to elements in the discourse model, but by reference to a surface-structure-like representation of immediately surrounding discourse (differing from traditional surface structure principally in that binding and scope relations are explicitly indicated). Thus the anaphoric continuations (33a, b) are interpreted by two entirely different means:

(33) She told me to take the oats down to the bin,
(a) so I did.
(b) so I did it.

The (b) continuation is interpreted by taking *it* to represent some action in the discourse model, in this case the action of taking the oats down to the bin. The interpretation of the (a) continuation involves assigning to the null VP an interpretation equivalent to that of some full VP in the propositional representation of surrounding discourse.

Under these assumptions several differences between the two types of anaphoric processes are immediately accounted for. First, because the structural units of propositional representation must correspond to surface syntactic units, the VP anaphor requires an antecedent of appropriately parallel form:

(35) She told me the oats had to be taken down to the bin,
(a) *so I did.
(b) so I did it.
(c) and they were.

Second, the impossibility of 'deictic' use of the VP anaphor discussed at length in HS follows directly, since the record of propositional structure contains no representation of anything in the discourse environment except what has just been uttered by the participants.

Finally, note that the *do* of the (b) continuations, according to our hypotheses, is the expression. of a semantic relation between two entities (an 'actor' and an 'action') in the discourse model, while the *do* of the (a) continuations is not: and accordingly, the *do* of the (b) examples has the syntactic and semantic properties of a real verb, while the *do* in (a) is syntactically and semantically just a dummy.

6. On the nature of the 'propositional representation'

We have assumed that the interpretation of surface anaphora depends on reference to a representation of propositional structure which is constructed as the discourse progresses by some kind of parsing mechanism, and which fades rapidly as the content of what has been said is integrated into a more permanent discourse model. Most experimental investigations into the psychological processes involved in discourse understanding have taken these minimal assumptions as a starting point. They have differed, however, on the nature of the parsing mechanism and the nature of the representations that it constructs.[19]

In order to account for the Sag-Williams observations regarding the constraints on ellipsis, we are committed to certain assumptions about the propositional representation constructed by the parser, which we will briefly outline and justify here.

Theories of human sentence processing going back as far as Woods (1970) and Kimball (1973) have assumed that the human parsing mechanism, in addition to providing some representation of syntactic surface structure, includes devices which in effect assign binding relations between dislocated constituents (topics, fronted WH elements, controllers of unbounded deletions) and the vacant positions to which they are related. For more recent, and more explicit, accounts of such devices, see Cowper (1976), Marcus (1977), Wanner and Maratsos (1978).

Some of these treatments (Kimball, for example) merely hint at the parser's construction of binding relations; others, especially the more recent ones, are very explicit. There is no disagreement on the necessity for some representation of binding relations, since it is clear that the understanding of a sentence containing dislocated constituents must involve associating such constituents with the appropriate grammatical roles normally associated with elements in the 'trace' positions. Since the general abandonment of the idea that the parser constructs classical surface structures, and that some additional mechanism works through transformational derivations in reverse to arrive at a representation corresponding to classical deep structure, every parsing theory has incorporated, at least implicitly, the assumption that the representations produced by the parser include binding relations between dislocated elements and their trace locations indicated directly.

The resulting representations consequently are not traditional surface structures, but rather something more like the 'logical forms' of Sag and Williams. If it is only assumed that such binding devices also come into play in the parsing of sentences involving quantifiers and similar operators, then the representations produced by the parser are *exactly* the surface-determined propositional representations we require.

Given these assumptions, we have an obvious and quite compelling explanation for the surprising discovery by Sag and Williams that surface anaphora are not determined by classical surface structure. The reason is simply that an understander *never constructs a representation equivalent to classical surface structure* – the first output of this parser is a propositional representation complete with indications of scope and binding relations.

This claim should be experimentally testable, given the recent development of techniques sensitive to very short-term properties of the parsing mechanism (Neely (1977); Tanenhaus, Leiman, and Seidenberg (1979); Merrill, Sperber, and McCauley (1981)). One study (Hudson, Tanenhaus, and Carlson (1982)) already has yielded results which appear to substantiate the claim. This study shows that priming effects from dislocated WH expressions indicate that they are associated with the location of their traces within 200 msec. after the trace location is reached in input. Similar experiments, according to our theory, should show the same kinds of effects for other constructions involving binding in the propositional representation, including those involving quantification without the equivalent of (classical) movement or deletion.

7. Conclusion

We have argued that the distinction between two kinds of anaphoric processes observed by HS can be accounted for in terms of a theory of discourse understanding the postulates two sorts of psychological objects: propositional representations (of the sentences of the immediately prior discourse) and discourse models (of the broader discourse context). Elliptical elements (HS's 'surface' anaphora) are interpreted in terms of propositional representations; model-interpretive anaphoric elements (HS's "deep" anaphora) are interpreted in terms of discourse models. These arguments provide support for the theory of language understanding, involving exactly these two kinds of objects, that is defended empirically on entirely independent grounds by Johnson-Laird (1980).

We conclude by pointing out one immediate prediction which follows from the theory of anaphoric processing that we have advanced. If elliptical expressions are interpreted with reference to propositional representations which lead a temporary life in a limited short-term memory register, it follows that the interpretability of elliptical expressions should be affected by the amount of discourse that intervenes between anaphor and antecedent. Too much intervening discourse, on this theory, should cause the propositional representation of a given sentence to 'fall off' the short-term register, so to speak, rendering the units of these propositional representations no longer available as the basis for the interpretation of a subsequent elliptical expression. Model-interpretive anaphoric elements, whose interpretation does not involve propositional representations, should be subject to no such recency effect.

That MIA is not subject to a short-term recency effect is a common enough observation. To appreciate this point, one only need consider the following dialogue from Grosz (1977, p. 23).

(37) E: Good morning. I would like for you to reassemble the compressor . . .
E: I suggest you begin by attaching the pump to the platform . . . (other subtasks).
A: All right. I assume the hold in the housing cover opens to the pump pulley rather than to the motor pulley.

E: Yes, that is correct. The pump pulley also acts as a fan to cool the pump.
A: Fine. Thank you.
A: All right, the belt housing cover is on and tightened down. (30 minutes and 60 utterances after beginning.)
E: Fine, Now let's see if *it* works.

Here *it* refers to the compressor, an entity in the discourse model, which was evoked by the first sentence in (37), and which is 'in focus' (to use Grosz's terminology) at the end of the dialogue, and hence may be referred to by a (MIA) anaphoric element like *it*.

Consider, however, the result of replacing the final sentence of this dialogue with either of the following,

(38) (a) *Fine. I knew you would be able to Ø. [Ø = reassemble the compressor] (VPE).
(b) *Fine. Now you know how Ø.
[Ø = to reassemble the compressor] (Sluicing).

Neither of these elliptical expressions, we observe, can be assigned an interpretation on the basis of the first sentence in Grosz's dialogue. This of course follows from the theory we have suggested, as the propostional representation of that sentence has long since fallen off of the short-term register of either participant.

As a final example, consider the result of replacing the last sentence of (37) with either of the following.

(39) (a) Fine. I knew you would be able to do *it*. (Sentential *it*)
(b) Fine. You've succeeded Ø. (Null Complement)

Here, the anaphoric elements are instances of MIA, and they may indeed receive interpretations that appear to be based on the first sentence of the dialogue. That is, *do it* in (39a) may be assigend the interpretation of *reassemble the compressor*, and the null anaphor in (39b) may be interpreted as 'in reassembling the compressor'. These interpretations are possible, however, only because the relevant property is prominent in the discourse model at the end of the dialogue, as predicted by our theory.

Notes

[1] On the fact that (d) is acceptable in the given context just in case Sag desires the person offstage shot, and regards the shooting as a welcome but unexpected favor, see Hankamer (1978).

[2] Because of the relative delicacy of the crucial judgments, and because of disagreement

between the two authors of this paper over their significance, we will not discuss the missing antecedent phenomenon in this paper. For some discussion see Sag (1976b, Chapter 4), Williams (1977b), and Sag (1979, p. 155 Note 2).

[3] It has been known since Lees (1960) that a superficial syntactic identity condition would not be sufficient, since elements which are identical in surface structure but differ at deeper levels cannot give rise to ellipsis. See Sag (1976b) for some discussion.

[4] Sag (1976a, b) regarded ellipses as syntactically involving deletion, while Williams (1977) advocated in interpretive approach in which there was no deletion. In this section we are not interested in this difference, but in the common claim that the relevant level of structure for the determination of the interpretation of an elliptical expression is not classical surface structure but some richer propositional representation.

[5] Here 'α' denotes the logical translation of any expression 'α' and W is the future tense operator [*like'* (*Sandy'*)] should be thought of as denoting a function from individuals to truth values, and *like'* as denoting a function from individuals to functions from individuals to truth values.

[6] It might be thought that the possibility of EA interpretations for examples like (i) are problematic for the claim being made:

(i) Lonnie thinks everyone likes a friend of mine from Texas.
Connie thinks everyone does ∅, too.

Fodor and Sag (1982) argue, however, that indefinite NPs may be analyzed as referring expressions, as well as quantifiers of a familiar sort. Hence examples like (i) do not necessarily involve wide scope existential quantification, even on the EA interpretation. See Fodor and Sag (1982) for further discussion of this point.

[7] Coindexing here indicates (intended) coreference between the expressions *John* and *him*.

[8] The explanation, of course, requires appeal to a further convention on variable indexing to prevent accidental identity of variable indices.

[9] These facts are also, of course, problematic for a deletion analysis of verb phrase ellipsis which deletes under identity of surface structure, as in the classical standard-theory conception. The problem of superficial non-identity of pronouns in antecedents and targets of VP deletion were recognized in Ross (1967), though no satisfactory account of them was ever developed in the classical framework.

[10] We present this sketch primarily for expository purposes, to illustrate how the problem of indexicals in ellipses might be formally treated in a model-theoretic semantics. We do not necessarily espouse this particular treatment, nor Kaplan's approach in general. It is not in fact clear how Kaplan's approach would extend to the full range of examples that have been discussed by linguists.

[11] This formulation draws on a similar formulation by Ladusaw (1980).

[12] This conception owes much to Webber (1978) and Johnson-Laird (1980); see also Johnson-Laird and Garnham (1980). A similar conception underlies the approach outlined in Marslen-Wilson and Tyler (1980). These are just some of the more recent and more explicit statements of the position; something similar is widely accepted in both psychological and artificial-intelligence work on information processing. Winograd (1972) and Anderson and Bower (1973) present important early developments of the idea.

[13] By 'model of the discourse' we mean a partial model of the world as conceived by the person, containing some representation of that part of the world which is of current interest. This partial model is not presumed to contain a representation of only that information which has been explicitly gleaned from immediately prior discourse. Our 'worlds' here of course includes imaginary subworlds.

[14] See Jarvella (1971), Sachs (1967). For a survey of relevant research in this area, see Clark and Clark (1977, Chapter 4).

[15] Of course there are other uses of definite pronouns which require different principles of interpretation, most notably where the pronoun corresponds to a variable in propositional

representation. The ambiguities resulting from this double function of pronouns are now well known. We have nothing to say about how an understander chooses from among alternative possible interpretations. Nor will we address here the problem of so-called 'pronouns of laziness', but for an account compatible with our general views, see Kamp (ms.).

[16] Since a discourse model may contain hypothetical entities, it is perfectly straightforward that pronouns can 'refer' to such hypothetical entities as well as to real ones. How the model distinguishes between real and hypothetical entities is of course an interesting question but not one we will address here.

[17] For the beginning of a theory of how entities move in and out of 'focus' as dialogues progress, see Grosz (1977).

[18] One possibility is that the understander sets up and keeps track of potentially different 'belief worlds' for each sentient being in the model; these belief worlds would then have the same sort of representation as the main world.

[19] They have also differed regarding assumptions about the nature of the memory representations incorporated into the model and the nature of the interaction between the two kinds of representations. Earlier investigations tended to adopt the simple assumption that the propositional representation was constructed prior to and independently of the incorporation of newly acquired information into the model. This assumption has been shown in recent work to be too simplistic; studies such as Marslen-Wilson and Tyler (1980) have shown that information present in the model representation is available to the parsing mechanism. These results are consistent with our assumption that there are two distinct representations, which are constructed in parallel.

References

Anderson, J. and G. Bower: 1973, *Human Associative Memory* (Winston/Wiley, Washington, D.C.).

Bresnan, J: 1971, 'A Note on the Notion "Identity of Sense Anaphora"', *Linguistic Inquiry* **2**, 589–597.

Clark, H. H. and E. V. Clark: 1977, *Psychology and Language*, (Harcourt Brace Jovanovich, New York).

Cowper, E.: 1976, *Constraints on Sentence Complexity: A Model for Syntactic Processing* (Ph.D. dissertation, Brown University).

Fodor, J. D. and I. A. Sag: 1982, 'Referential and Quantificational Indefinites', *Linguistics and Philosophy* **5**, 355–398.

Grinder, J. and P. Postal: 1971, 'Missing Antecedents', *Linguistic Inquiry* **2**, 269–312.

Grosz, B.: 1977, *The Representation and Use of Focus in Dialogue Understanding* (Technical Note 151, SRI International, Menlo Park, California).

Hankamer, J. and I. Sag: 1967, 'Deep and Surface Anaphora', *Linguistic Inquiry* **7** 391–426.

Hankamer, J.: 1978, 'On the Nontransformational Derivation of Some Null VP Anaphors', *Linguistic Inquiry* **9**, 66–74.

Hudson, S., M. Tanenhaus, and G. Carlson: 1982, 'Phonological Codes and Parsing' (unpublished paper presented at the 1982 summer meeting of the Linguistic Society of America, College Park, Maryland, July 1982).

Jarvella, R. J.: 1971, 'Syntactic Processing of Connected Speech', *Journal of Verbal Learning and Verbal Behavior* **10**, 409–416.

Johnson-Laird, P: 1980, 'Lectures on Mental Models (Stanford University).

Johnson-Laird, P. and A. Garnham: 1980, 'Descriptions and Discourse Models', *Linguistics and Philosophy* **3**, 371–394.

Kamp, J. A. W.: (ms.) 'Disjoint Reference' (unpublished. University of Massachusetts. Amherst).

Kaplan, D.: (ms.) 'Demonstratives: An Essay on the Semantics, Other Indexicals' (unpublished, University of California, Los Angeles).

Kimball, J.: 1973, 'Seven Principles of Surface Structure Parsing in Natural Language', *Cognition* **2**, 15–47.

Ludusaw, W.: 1980, *Polarity Sensitivity as Inherent Scope Relations* (Garland Publishing, New York).

Lees, R.: 1960, *The Grammar of English Nominalizations* (Mouton, The Hague).

Marcus, M.: 1977, *A Theory of Syntactic Recognition for Natural Language* (unpublished Ph.D. dissertation, MIT).

Marslen-Wilson, W. and L. Tyler: 1980, 'Towards a Psychological Basis for a Theory of anaphora', in J. Kreiman, and A. Ojeda (eds.), *Papers from the Parasession on Pronouns and Anaphora* (Chicago Linguistic Society).

Merrill, E., R. Sperber, and C. McCauley: 1981, 'Differences in Semantic Coding as a function of Reading Comprehension Skill', *Memory and Cognition* **9**, 618–624.

Montague, R.: 1974, *Formal Philosophy* (R. Thomason, ed.) (Yale University Press, New Haven).

Neely, J.: 1977, 'Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention', *Journal of Experimental Psychology* **106**, 226–254.

Partee, B.: 1975, 'Deletion and Variable Binding', in E. L. Keenan (ed.), *Formal Semantics of Natural Language*, Cambridge University Press, Cambridge.

Ross, J.: 1967, *Constraints on Variables in Syntax*, (unpublished Ph.D. dissertation, MIT).

Ross, J.: 1969, 'Guess Who' CLS **6**.

Sag, I. A.: 1976a, 'A Logical Theory of Verb Phrase Deletion', CLS **12**.

Sag, I. A.: 1976b, *Deletion and Logical Form* (Ph.D. dissertation, MIT). [Published 1980 by Garland Publishing, New York].

Sag, I. A.: 1979, 'The Nonunity of Anaphora', *Linguistic Inquiry* **10**, 152–164.

Sachs, J. S.: 1967, 'Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse', *Perception and Psycholinguistics* **2**, 437–442.

Schachter, P.: 1977, 'Does She or Doesn't She?' *Linguistic Inquiry* **8**, 763–767.

Tanenhaus, M., J. Leiman, and M. Seidenberg: 1979, 'Evidence for Multiple Stages in the Processing of Ambiguous Words in Syntactic Contexts', *Journal of Verbal Learning and Verbal Behavior* **18**,427–441.

Wanner, E. and M. Maratsos: 1978, 'An ATN Approach to Comprehension', in Halle, Bresnan, and Miller (eds.), *Linguistic Theory and Psychological Reality* (MIT Press).

Webber, B. L.: 1978, *A Formal Approach to Discourse Anaphora* Ph.D. dissertation, Harvard University) [Published 1979 by Garland Publishing, New York].

Williams, E.: 1977a, 'Discourse and Logical Form', *Linguistic Inquiry* **9**, 138–141.

Williams, E.: 1977b, 'On Deep and Surface Anaphora', *Linguistic Inquiry* **8**, 692–696.

Winograd, T.: 1972, 'Understanding Natural Language', *Cognitive Psychology* **3**, 1–191.

Woods, W.: 1970, 'Transition Network Grammars of Natural Language analysis', *communications of the ACM* **13**, 591–606.

Dept. of Linguistics, Stanford University,
CA 94305, *U.S.A.*