SEARCH AND INFERENCE STRATEGIES IN PRONOUN RESOLUTION: AN EXPERIMENTAL STUDY

Kate Ehrlich
Department of Psychology
University of Massachusetts
Amherst, MA 01003

The question of how people resolve pronouns has been of interest to language theorists for a long time because so much of what goes on when people find referents for pronouns seems to lie at the heart of comprehension. However, despite the relevance of pronouns for comprehension and language theory, the processes that contribute to pronoun resolution have proved notoriously difficult to pin down.

Part of the difficulty arises from the wide range of factors that can affect which antecedent noun phrase in a text is understood to be co-referential with a particular pronoun. These factors can range from simple number/gender agreement through selectional restrictions to quite complex knowledge that has been acquired from the text (see Webber,(1978) for a neatly illustrated description of many of these factors). Research in psychology, artificial intelligence and linguistics has gone a long way toward identifying some of these factors and their role in pronoun resolution. For instance, in psychology, research carried out by Caramazza and his colleagues (Caramazza et al, 1977) as well as research that I have done (Ehrlich, 1980), has demonstrated that number/gender agreement really can function to constrain the choice of referent in a way that significantly facilitates processing. Within an AI framework, there has been some very interesting work carried out by Sidner (1977) and Grosz (1977) that seeks to identify the current topic of a text and to show that knowledge of the topic can considerably simplify pronoun resolution.

It is important that people are able to select appropriate referents for pronouns and to have some basis for that decision. The research discussed so far has mentioned some of the factors that contribute to those decisions. However, part of the problem of really understanding how people resolve pronouns is knowing how the various factors combine. Certainly it is important and useful to point to a particular factor as contributing to a reference decision, but in many texts more than one of these factors will be available to a reader or listener. One problem for the theorist is then to explain which factor predominates in the decision as well as to describe the scheduling of evaluation procedures. If it could be shown that there was a strict ordering in which tests were applied, say, number/gender agreement followed by selectional restrictions followed by inference procedures, pronoun resolution may be simpler to explain. At our present level of knowledge it is difficult to discern ordering principles that have any degree of generality. For instance, for every example where the topic seems to determine choice, a similar example can often be found where the more recent antecedent is preferred over the one that forms part of the topic. Moreover, even this claim begs the question of how the topic can be identified unambiguously.

A different approach is possible. The process of assigning a referent to a pronoun can be viewed as utilizing two kinds of strategies. One strategy is concerned with selecting the best referent from amongst the candidates available. The other strategy is concerned with searching through memory for the candidates. These two types of strategy, which will be referred to mnemonically as inference and search strategies, have different kinds of characteristics. A search strategy dictates the order in which candidates are evaluated, but has no machinery for carrying out the evaluation. The inference strategy helps to set up the representation of the information in the text against which candidates can be evaluated, but has no way of finding the candidates. In the rest of this paper, the way these strategies might interact will be explored and the results of two studies will be reported that bear on

the issues.

One possible search strategy is to examine candidates serially beginning with the one mentioned most recently and working back through the text. This strategy makes some sense because, as Hobbs (1978) has pointed out, most pronouns co-refer with antecedents that were mentioned within the last few sentences. Thus, a serial search strategy provides a principled way of restricting how a text is searched. Moreover, there is some evidence from psychological research that it takes longer to resolve pronouns when the antecedent with which the pronoun co-refers is far rather than near the pronoun (e.g. Clark & Sengul, 1979; Springston, 1975). Although such distance effects have been used to argue for differences in memory retrieval, with the nearer antecedents being easier to retrieve than the further ones, none of the reported data rule out a serial search strategy.

As argued earlier, a search strategy alone cannot account for pronoun resolution because it lacks any machinery for evaluation. There are, however, many kinds of information that people can bring to bear when evaluating candidates and some of these were discussed earlier. A common method is to decide between alternative candidates on the basis of information gained through inferences. Inference is a rather ubiquitous and often ill-defined notion, and, although it is beyond the scope of this paper to clarify the concept, it is worth noting that there are (at least) two kinds of inference that play a role in anaphora generally. One kind which I will call 'lexical' inferences are drawn to establish that two different linguistic expressions refer to the same entity. For instance, in the following pair of sentences from Garrod and Sanford (1977):

(1) A bus came roaring round the corner

The vehicle nearly flattened a pedestrian

a 'lexical' inference establishes that the particular vehicle mentioned in the second sentence is in fact a bus. Inferences can also be drawn to support the selection of one referent over another. In a sentence such as:

(2) John sold a car to Fred because he needed it

a series of inferences based in part on out knowledge of selling and needing, supports the selection of Fred rather than John as referent for the pronoun "he". In the experiments to be reported, it was 'lexical' inferences rather than the other kind that were manipulated.

Subjects in the experiment were asked to read texts such as the one given below:

(3) Fred was outside all day

John was inside all day

a) He had a sleep inside after lunch

b) He had a sleep in his room after lunch

and then immediately after, answer a question such as "Who had a sleep after lunch?" that was designed to elicit the referent of the pronoun in the last sentence. Two factors were independently varied. The antecedent could be near or far from the pronoun, the latter effected by switching the order of the first two sentences. The second factor was whether a 'bridging' inference had to be drawn to establish co-reference between part of the predicate of the last sentence and the target sentence. The two versions, (a) no inference and (b) inference, are shown as alternative third sentences in example (3) above. The principal measures were the time to answer the question and the accuracy of the response.

The experiment addresses two critical issues. One is whether the 'lexical' inference is drawn as part of the evaluation procedure, or, whether it is drawn independently of that process. The other issue concerns the search strategy itself: do subjects examine candidates serially, and, if so, do they still use other criteria to reject the first candidate and choose the second? Two distinct models of processing can be constructed from a consideration of these issues. In the case where inferences are triggered by the need to

evaluate a candidate, any effect due to extra processing should be unaffected by whether the antecedent is near or far from the pronoun. In either case the inference will be drawn in response to the need to decide on the acceptability of the candidate. In the second model, the inference is triggered by the anaphoric expression, e.g. "in his room" in the third sentence, and the need to relate that expression to the location "inside" mentioned in a previous sentence. The inference is expected to take a certain amount of time to be drawn (cf. Kintsch, 1974). According to the second model, one would expect that in cases where the antecedent is near the pronoun, there will be some effect due to inference because the process may not be completed in time to answer the question. When the antecedent is far from the pronoun, however, the inference process will be completed and hence no effect of inference should still be detected. The two models assume rationality on the part of the subjects; that is, they assume that subjects will accurately select the further antecedent where appropriate even though recency would predict selection of the first candidate that is evaluated. If this assumption is valid, subjects should select the far antecedent where appropriate more often than the (erroneous) near candidate.

The results of the experiment, shown in Table 1, support the second model; 'lexical' inferences are drawn only once and in response to an anaphoric expression. The data also provide evidence of a serial search strategy by showing that there are more errors and longer latencies associated with far rather than near antecedents. The data further show that even when the correct choice is far from the pronoun, subjects will choose it in preference to the nearer candidate, thus demonstrating that a serial search strategy alone cannot predict the choice of referent.

The inferences that subjects had to draw in this experiment concerned simple lexical relations. The increase in latency due to having drawn such an inference supports the results of earlier studies, particularly those of Garrod and Sanford (1977). What the present study fails to do, however, is to determine whether that inference is drawn spontaneously, while reading. Previous research (e.g., Kintsch, 1974, Garrod and Sanford, 1977) has shown that inferences are more likely to be drawn while reading than at a response stage. It was thus of some interest to know when the lexical inferences in the present study were drawn. This issue was examined by modifying the previous experiment to include both an additional measure of reading time and a 1.5s delay between presentation and test. The latter modification is important since if subjects are drawing inferences while reading, the process may not be completed by the time the question is asked immediately after presentation. The introduction of a delay also allows for a further test of the two processing modeled outlined earlier. If indeed 'lexical' inferences are drawn to establish co-reference between anaphoric expressions rather than to determine pronominal reference, as the previous experiment indicated, then there should be an effect of inference on reading time but not at response when there is a delay, because by response the inference should have been drawn. The data were consistent with this hypothesis. However, what also emerged from the second study was that only some of the passages seemed to elicit inferences at reading; the number of passages was increased in the second experiment to counter possible repetition effects. In fact, for half the passages subjects responded by saying there was no answer. An example of such a passage is given below:

(4) Jill had a newspaper in the living-room
Ann had a book in the living-room
She read some chemistry in the evening

It was also the case for these passages that the inferences did not seem to be drawn while reading but rather in response to the question. There is some doubt here about cause and effect, nevertheless, the

observation raises some interesting questions concerning what triggers an inference to be drawn. One answer, supplied by Garrod & Sanford in their experiments, is that a relation between expressions must somehow be perceived before an inference is drawn to determine the nature of the relation. In other words, people do not draw inferences randomly to relate linguistic expressions. Thus, whereas Garrod & Sanford found that subjects would infer co-reference between "bus" and "vehicle" in example (1), they failed to make that connection, quite rightly, in a similar passage shown below:

(5) A bus came roaring round the corner
It nearly smashed some vehicles

What kinds of strategies do readers adopt when they search their memory to find plausible referents for pronouns? Results of the experiments reported here point to a strategy in which entities are examined serially from the pronoun. The purpose of a serial search strategy is to provide a principled way in which readers can examine those entities they have stored in memory, for their appropriateness as the referent of a particular pronoun. The strategy is thus unnecessary when there is only one entity in memory by virtue of simple criteria such as number and gender agreement with the pronoun. What constitutes 'simple' criteria is, of course, an open question; the answer, however, will materially affect the applicability of the search strategy.

The most important part of reference resolution is, however, deciding on the referent. A serial search strategy has no machinery for evaluating candidates, it can only direct the order in which candidates are examined. The process of selecting a plausible referent depends on the inferences a reader has drawn while the text is read. Thus, when subjects found it hard to select a referent at all they also failed to draw many inferences while they read the text. Moreover, because the inferences for these passages did seem to be drawn in response to a question eliciting the referent, the implication is that inferences for the clearer material are generally drawn spontaneously and before a specific need for the information arises. One can conjecture from these data that the selection of plausible referents is dependent on how well a reader has understood the preceding text. If inferences are not drawn until a specific need arises, such as finding a referent, then it may be too late, to select a referent easily or accurately. Thus, reference can also be viewed in terms of what a text makes available for anaphoric reference (cf. Webber, 1978).

The picture of pronoun resolution that emerges from the studies reported here, is one in which effects of distance between the pronoun and its antecedent may play some role, not as a predicator of pronominal reference as has often been thought, but as part of a search strategy. There certainly are cases where nearer antecedents seem to be preferred over ones further back in the text; however, it is more profitable to look to concepts such as foregrounding (cf. Chafe, 1974) rather than simple recency for explanations of the preference. It is also of some interest to have shown that inferences may contribute to pronoun resolution but drawn for other reasons.

REFERENCES

Caramazza, A., Grober, E., Garvey, C. and Yates, J. (1977). Comprehension of anaphoric pronouns. Journal of Verbal Learning and Verbal Behavior, 16, 601-9.

Chafe, W.L. (1974). Language and consciousness. Language, 50, 111-133.

Clark, H.H., and Sengul, C.J. (1979). In search of referents for nouns and pronouns. Memory and Cognition, 7, 35-41.

Ehrlich, K. (1980). Comprehension of pronouns. Quarterly Journal of Experimental Psychology, 32, 247-

Garrod, S. and Sanford,A.J. (1977). Interpreting ana-

photic relations: the integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 16, 77-90.

Grosz, B.J. (1977). The representation and use of focus in a system for understanding dialogs. In Proceeding of the Fifth International Joint Conference on Artificial Intelligence. Cambridge: MIT.

Hobbs, J.R. (1978). Resolving pronoun references. Lingua, 44, 311-338.

Kintsch, W. (1974). The representation of meaning in memory. Potomac, Md: Erlbaum.

Sidner, C. (1977). Levels of complexity in discourse for anaphora disambiguation and speech act interpretation. In Proceedings of the Fifth International Joint Conference on Artificial Intelligence. Cambridge: MIT.

Springston, F.J. (1975). Some cognitive aspects of presupposed coreferential anaphora. Unpublished doctoral dissertation, Stanford University.

Webber, B.L. (1978). A formal approach to discourse anaphora. BBN report no. 3761. Cambridge, Mass: Bolt, Beranek and Newman, Inc.

TABLE 1

Percent correct responses (P.C.) and mean response times (R.T.).

	Inference condition			
Distance	No inference		Inference	
	R.T.	P.C.	R.T.	P.C.
Near	1.32	95%	1.42	87%
Far	1.56	72%	1.56	70%