Title
The Time Course of Anaphora Resolution

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 7(0)

Author
Guindon, Raymonde

Publication Date
1985

Peer reviewed

The Time Course of Anaphora Resolution

Raymonde Guindon

Microelectronics and Computer Technology Corporation

Abstract

Anaphors, such as definite noun phrases and pronouns, are important contributors to discourse coherence. Anaphora resolution is the process of determining the referent of an anaphor in a discourse or dialogue. Models of discourse and sentence comprehension have made different claims about the temporal relationship between the occurence of the syntactic and semantic analyses of the sentence and the process of anaphora resolution. The **end-of-sentence** hypothesis holds that anaphora resolution occurs at the end of the sentence, after the syntactic and semantic analyses are completed. The **immediacy assumption** holds that anaphora resolution occurs as soon as an anaphor is encountered and is completed as much as possible before further words are processed. The **cognitive lag** hypothesis assumes that anaphora resolution starts when the anaphor is encountered but is completed while processing further words in the sentence. A study is described that traces the activation of a referent by its anaphor over a complete sentence. It demonstrates that anaphora resolution does not await the complete syntactic and semantic interpretations of the sentence. An anaphor starts activating its referent as soon as the anaphor is encountered and the referent stays activated until the end of the sentence. This result supports a particular version of the immediacy assumption. This is also interpreted in terms of a limited cache that stores the items currently in focus and that is updated at sentence or clause boundaries.

Introduction

Temporal Characteristics of Anaphora Resolution

An important aspect of discourse comprehension models, whether in cognitive psychology or in computational linguistics, is the assumed temporal characteristics of anaphora resolution. Two specific questions to be answered are: 1) When does anaphora resolution occur during the comprehension of a sentence? 2) When does it occur in relation to the syntactic and semantic analyses? This study provides preliminary evidence on these questions.

There are three main hypotheses regarding the temporal relationship between the syntactic and semantic analyses of the sentence and anaphora resolution. The **end-of-sentence** hypothesis holds that anaphora resolution occurs at the end of the sentence, after the syntactic and semantic analyses are completed (Bever & Hurtig, 1975). The **immediacy assumption** holds that anaphora resolution occurs as soon as an anaphor is encountered and is completed as much as possible before further words are processed (Just & Carpenter, 1980). The **cognitive lag** hypothesis assumes that anaphora resolution starts when the anaphor is encountered but is completed while processing further words in the sentence (Ehrlich & Rayner, 1983). A similar issue has arisen in designing natural language interfaces, with a tendency to make syntax and semantics work "in tandem" (e.g. Cascaded ATN, Psi-Klone (Bobrow & Webber, 1980)). In most natural language understanding systems, though, anaphora resolution occurs after the syntactic and semantic analyses of the sentence are completed (e.g. see Sidner, 1984).

1- This research was performed at the University of Colorado, Boulder, as part of the author's doctoral dissertation. Walter Kintsch, Peter Polson, Alice Healy, Richard Olson, and Andrjez Ehrenfeucht are kindly thanked for their help with this research.

Short-term memory and cache management

The design of this study and the interpretation of its results are based on a model of discourse comprehension by Kintsch and van Dijk (1978). A sketch of an enhanced version of this model is presented.

Memories The memory of the system is divided into three components: a small, very fast memory called short-term memory (STM); a relatively larger main or operating memory (OM); a vast store of general world knowledge called long-term memory (LTM).

The total **STM** contains only 7±2 chunks at any one time (Simon, 1974; Miller, 1956). STM is itself divided into two other memories to which these chunks are allocated. First, the **buffer** is used to store the incoming clause or sentence before further processing. Second, the **cache** is used to hold over, from one sentence or clause to the next, the information necessary to provide global and local coherence. It contains a subset T of the previous most topical text items and a subset R of the most recent text items. Retrieval times from short-term memory are very fast. Items stored in the cache are in focus (Guindon, 1985).

The **operating memory** is conceptually that subset of the world knowledge in long-term memory which is deemed relevant to the processing of the current part of the text. It also contains the growing structure corresponding to the text read so far. It contains the less topical and less recent information from the text. Operating memory and long-term memory can contain a very large number of entities. However, retrieval times are much longer than for short-term memory. Items not in focus are in operating memory (Guindon, 1985).

Cache Management Anaphora resolution proceeds in a number of partially concurrent steps. As a new sentence is read into the buffer and an anaphor is encountered, its referent is searched in the cache. If the referent is not found in the cache, operating memory is searched. If the referent is found in operating memory, it is reinstated in the cache and is now in focus. The cache management strategy is applied at each sentence or clause boundaries. As a consequence, once an item is in focus it will stay activated until at least the end of the sentence or clause.

Overview of the On-line Activation Technique and the Study

A technique called **on-line activation** was developed to trace the activation of a referent by an anaphor during a whole sentence. The technique is *on-line* because the activation is measured at various short intervals. It is an *activation* technique because the recognition latency of an old test word is measured; the anaphor acting as a prime and the test word being the referent. This allows a **direct** measure of the activation of a **specific** referent by an anaphor.

Using the on-line activation technique, subjects read passages in which an anaphor referred to an antecedent which was not in focus, that is, the antecedent was not in the cache. The antecedent was removed from focus by introducing a topic shift. An example text is presented in Table 1. Only one of the sentences 5a and 5b was presented in a text during the study. The passages were presented using rapid serial visual presentation (RSVP), one word at a time in the center of a CRT. The presentation time was 300 msec per word.

In addition to reading each text, the participants were also given the task to recognize whether some specially marked words, presented surreptitiously within the text, had appeared before in the text or not. These specially marked words are called test words. Some of the test words were old, some were new. In each text, one of these test words was the referent of the anaphor. At some point before or after the anaphor was presented on the CRT, its referent was presented for recognition and recognition times and errors were collected. The delay between the onset of the anaphor and the onset of the test word is called the stimulus onset asynchrony (SOA). The selected SOAs spanned the whole sentence, as can be seen in Table 1. The anaphor is acting as a prime, which should activate or reinstate the referent, that is, bring the referent in the cache with its fast retrieval time. The recognition time for the referent test word indicates whether the referent is in the cache (i.e.

fast) or in operating memory (i.e. slow). In addition, there were two types of primes, as shown in sentences 5a and 5b in Table 1. The prime could be either semantically related and referential (S+R+) as in 5a, or semantically related and non-referential (S+R-) as in 5b. The S+R- condition acts as a control condition where reinstatement of the referent is not expected. The S+R- condition is not referential, because of the use of an indefinite article (e.g. "a") and an adjective incompatible with the referent, but semantic priming is the same as in S+R+

Table 1

Example of an experimental text

Antecendent/test word: thermometer Anaphor: instrument

1-- The assistant was preparing solutions for a chemistry experiment.
2-- He measured the temperature of a solution using a *thermometer*.
3-- The experiment would take at least four hours.
4-- There would then be a ten hour wait for the reaction to finish.
5a- The thin *instrument* was not giving the expected reading.
5b- A broken *instrument* was not giving the expected reading.
SOA locations ^ ^ ^ ^ ^ ^ ^ ^ ^

A schema of the procedure is shown in Table 2. The words surrounded by stars are the test words.

Table 2

Schema of the procedure for three SOAs

SOAs Time	Before anaphor	350 msec after anaphor	1250 msec after anaphor
Ti	The	The	The
Ti+1	thin	thin	thin
Ti+2	*thermometer*	instrument	instrument
Ti+3	instrument	*thermometer*	was
Ti+4	was	was	not
Ti+5	not	not	giving
Ti+6	giving	giving	*thermometer*

Given that recognition latencies are much shorter for items in the cache than in the operating memory and given that an anaphor reinstates a referent not in focus by bringing the referent in the cache from operating memory, the following pattern of recognition latencies can be derived. A pattern supporting the end-of-sentence hypothesis would be a decrease of recognition latencies for the referent observed only at the end of the sentence. A pattern supporting the immediacy assumption would be a rapid decrease of recognition latencies for the referent as soon as the anaphor is encountered and reaching its peak before the next words are processed. A pattern supporting the cognitive lag position would be a decrease in recognition latencies starting soon after the anaphor is encountered and reaching its peak a few words later.

Method

Participants The participants were 36 undergraduate psychology students.

Materials There were 60 texts: 36 experimental, 18 filler, and 6 practice. The experimental texts contained as a referent an instance of a class (e.g. thermometer) to be used later as a test word, and as an anaphor the class name (e.g. instrument). An example of the material was presented in Table 1. There were two priming conditions, S+R+ as in sentence 5a, and S+R- as in sentence 5b. During the presentation of each text, two or three test words were presented, one experimental and one or two fillers. The filler words were presented at semi-random locations in the text. There was an equal number of old and new test words in the entire experiment.

Procedure The experiment was computer controlled using real-time routines on the VAX/VMS 11/780 of the Computer Laboratory for Instruction in Psychological Research at the University of Colorado. The texts were presented using RSVP, with each word presented in the center of the screen for 300 msec. Two or three test words, surrounded by stars, were presented at different locations within each text. The participants were asked to recognize whether the test words were old or new, as fast as possible and without making mistakes. After each text, the participants were given feedback on the number of mistakes they had made.

Design There were 36 experimental texts and 18 experimental conditions, all within-subject. There were two priming conditions: 1) semantically related and referential (S+R+); 2) semantically related and non-referential (S+R-). The non-referentiality of the S+R- condition was achieved by the inclusion of an indefinite article (e.g. "a") followed by an adjective inconsistent with the referent. There were nine SOAs: 1) following the article at the beginning of the sentence (e.g. "the" or "a"); 2) following the adjective or, in other words, immediately before the prime; 3) 350 msec after the onset of the prime; 4) 650 msec after; 5) 1250 msec after; 6) following the third word from the end of the sentence; 7) following the second word from the end of the sentence; 8) following the last word of the sentence; 9) 600 msec after the last word of the sentence. The last four SOAs were used to test whether special integrative processes occur at the end of the sentence. The last SOA, 600 msec after the offset of the last word, was included to test whether integrative processes lag somewhat after processing the last word of the sentence (see Aaranson & Scarborough, 1976).

Results & Discussion

Separate analyses treating subjects and texts as random variables were performed. The recognition latencies for each priming conditions at each SOAs are presented in Figure 1.

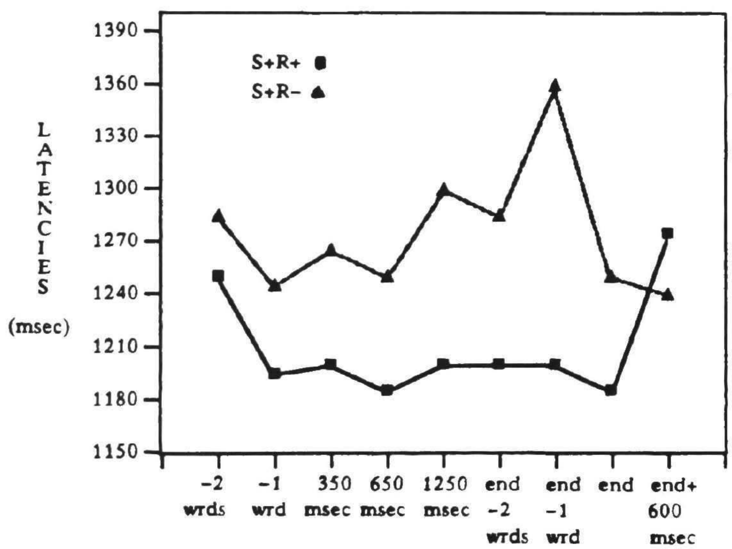

Figure 1. Recognition latencies for referent test words at each SOAs in each priming conditions.

As predicted, there was a strong effect of priming, with faster recognition latencies in the S+R+ condition than in the S+R- condition, **F(1,35)= 10.6, p= .003**, MSe= 719102 by subjects and

F(1,35)= 10.7, p= .002, MSe= 930316 by items. The referent was not reinstated in the cache from operating memory in the non-referential condition, S+R–. This indicates that in the S+R– condition the preceding syntactic cue (i.e. an indefinite article) and the preceding semantic cue (i.e. an adjective incompatible with the referent) were sufficient to quickly rule out activation of the referent. This is consistent with an on-line interactive model of discourse comprehension where syntatic, semantic, and pragmatic analyses run concurrently and synergistically.

While the interaction was not significant, the curves indicate that, in the S+R+ condition, the referent was activated early by the anaphor and stayed activated until the end of the sentence. The pattern of recognition latencies clearly does not support the end-of-sentence hypothesis and seems to support a particular version of the immediacy assumption: the referent is activated early after the anaphor is encountered and stays activated until the end of the sentence. Initial integrative processes at the discourse level – identification and search of the referent of an anaphor – are immediate. They do not await the completion of the syntactic and semantic analyses of the sentence. Again, the results support an on-line interactive system where syntax, semantics, and pragmatics cooperate concurrently. Moreover, these results support the hypothesis that cache management is applied at clause or sentence boundaries. Consequently, once an item is reinstated in the cache and is in focus, it stays activated until the end of the sentence or clause or until the next topic shift.

References

Aaranson, D. & Scarborough, H.S. Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 56 – 70.

Bever, T.G. & Hurtig, R.R. Detection of a non-linguistic stimulus is poorest at the end of a clause. *Journal of Psycholinguistic Research*, 1975, *4*, 1 – 7.

Bobrow, D.G. & Webber, B.L. *Knowledge representation for syntactic/semantic processing*. Proceedings of AAAI, Stanford University, 1980.

Ehrlich, K.E. & Rayner, K. Pronoun assignment and semantic integration during reading: Eye movements and the immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, 1983, *22*, 75 – 87.

Guindon, R. *Anaphora resolution: Short-term memory and focusing*. Proceedings of the Association for Computational Linguistics Meeting, Chicago, 1985.

Just, M.A. & Carpenter, P.A. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 1980, *87*, 329 – 354.

Kintsch, W. & van Dijk, T.A. Toward a model of text comprehension and production. *Psychological Review*, 1978, *85*, 363 – 394.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81 – 97.

Sidner, C. Focusing in the comprehension of definite anaphora. In M. Brady & R. C. Berwick (Eds.), *Computational Models of Discourse*, Cambridge: MIT Press, 1984.

Simon, H.A. How big is a chunk? *Science*, 1974, *183*, 482 – 488.