

THEMATIC ROLES AND PRONOUN COMPREHENSION¹

Rosemary J. Stevenson, Rosalind A. Crawley, Garry Wilson and David Kleinman
Department of Psychology
University of Durham
England

ABSTRACT

Two experiments tested the view that thematic role information triggers the rapid retrieval of general knowledge in pronoun comprehension. Pairs of thematic roles were contrasted as antecedents of a subsequent pronoun. The results showed that interpretation of the pronoun depended on the thematic role of the antecedent. Experiment one measured reading rates for the clause which contained the pronoun. Rates were faster when the antecedent was an Agent subject, a Patient object, a Goal, or an Experiencer. Rates were slower when the antecedent was an Agent object, a Patient subject, a Source, or a Stimulus. Experiment two required subjects to write completions to sentence fragments such as Jill like Sue and she..... and the number of references to each antecedent was recorded. The results confirmed the findings from Experiment one, although there was also an antecedent position effect (first vs. second mention) in some of the sentences. We suggest that these results are consistent with the view that thematic role information triggers the retrieval of canonical events in the real world, and may thus be responsible for the rapid retrieval of general knowledge in language comprehension.

INTRODUCTION

A striking feature of pronoun comprehension is its reliance on inferences from general knowledge (e.g. Stevenson and Vitkovitch, 1986). Some accounts propose that general knowledge may be accessed in the form of scripts (e.g. Sanford and Garrod, 1981), but scripts may not be sufficiently flexible for the fast retrieval of general knowledge (e.g. Kintsch, 1988). Others have used mental models (e.g. Garnham, 1989; Johnson-Laird, 1983). While the notion of mental models provides a useful theoretical framework, it does not explain how general knowledge is accessed when a mental model is constructed.

One possible mechanism for the fast retrieval of general knowledge when comprehending pronouns is the use of thematic role information. It has been argued that thematic roles may provide a mechanism for mediating between parsing, a discourse model and real world knowledge (e.g. Carlson and Tannenhaus, 1988; Rayner, Carlson and Frazier, 1984). However, the relationship between conceptual roles and grammatical roles is not straightforward (e.g. Finer and Roeper, 1989). Furthermore, there is no general consensus on the theoretical status of thematic roles (cf. Ladusaw and Dowty, 1988; Williams, 1989), but what is clear is that people do categorize the world in terms of agents, patients, experiencers and so on (e.g. Brown and Fish, 1983), and that these categories are also encoded in the language. Thus they may provide a convenient interface between language and general knowledge so that the rapid retrieval of general knowledge is possible, as suggested by some of the results discussed by Carlson and Tanenhaus (1988).

The hypothesis we wish to test is that there are canonical events in the world and that the lexical information about thematic roles in the verb triggers the retrieval of these canonical events. If the situation described by the sentence corresponds to the triggered canonical event, then comprehension is fast, otherwise comprehension is slowed. For example, action verbs trigger the retrieval of a canonical agent-patient sequence (e.g. Pinker, 1989). Thus, sentences containing such verbs should be comprehended more rapidly if they describe agent-patient sequences rather than patient-agent sequences.

An alternative conception of the way thematic roles might be used is the notion of a thematic hierarchy. Thematic hierarchies have been proposed by Jackendoff (1972) and Nishigauchi (1984). In Jackendoff's hierarchy, Agent is higher than Locative, Source and Goal, which in turn are higher than Theme (or Patient). Nishigauchi proposes in addition that Goal is higher than Location or Source. These hierarchies might reflect preferences in the conceptual system, and this would lead us to predict that the higher a role in the hierarchy, then the more rapidly it is retrieved during comprehension.

EXPERIMENT ONE

We tested these views by measuring reading rates for clauses containing pronouns where the thematic roles of the antecedents were varied. We used sentences consisting of two clauses. The first clause contained two noun phrases, one of which was the antecedent of a pronoun that was contained in the second clause. Our sentences were of four main types. Three contained thematic roles from Nishigauchi's hierarchy: Agent-Patient sentences, Goal-Source sentences, and Agent-Goal/Source sentences. These latter sentences were included to assess the possibility that pronoun assignment in the Goal-Source sentences is dependent on the transfer of goods, and they each contained a verb of motion in the first clause. Finally, we used Experiencer-Stimulus sentences. Examples of the four types of sentences are shown in Table one. The examples in Table one all contain linguistically ambiguous pronouns: the two individuals mentioned in the first clause are of the same gender. However, the content of the second clause biased the interpretation of the pronoun to one of the two potential antecedents. These materials were constructed and assessed by 5 judges, initially independently and then in group discussion.

On the basis of the view that thematic role information triggers the retrieval of canonical events, we predicted that pronouns would be interpreted more rapidly when their antecedents were Agents in first (subject) position and Patients in second (object) position. We had no strong predictions for the other sentence types. But a plausible view is that in Goal-Source sentences, Goals make better antecedents than Sources because the Goal is in possession of the Theme at the time of the event described by the second clause (see also Ladusaw and Dowty, 1988). In the Agent-Goal/Source sentences, the Agent is more likely to be carrying out a second event, although this prediction is confounded with the position of the antecedent. Experiencer-Stimulus sentences seem to lead people to attribute the cause of the experience to the Stimulus (e.g. Brown and Fish 1983), which suggests that people do categorize events in terms of Experiencer and Stimulus. In our Experiencer-Stimulus

sentences, the second clause was a consequence or elaboration of the experience described in the first clause (see Table one). Our general knowledge of real world situations might lead us to expect, therefore, that this additional information also concerns the Experiencer. So when the pronoun has an Experiencer antecedent it should be comprehended more rapidly than when it has a Stimulus antecedent.

Agent-Patient Sentences

Pronoun biased to Agent antecedent

Joseph hit Patrick and he made sure that it hurt.
Patrick was hit by Joseph and he made sure that it hurt.

Pronoun biased to Patient antecedent

Patrick was hit by Joseph and he began to cry very loudly.
Joseph hit Patrick and he began to cry very loudly.

Goal-Source Sentences

Pronoun biased to Goal antecedent

Sarah borrowed a record from Jenny and she listened to it that evening.
Jenny lent a record to Sarah and she listened to it that evening.

Pronoun biased to Source antecedent

Jenny lent a record to Sarah and she asked for it back again.
Jenny borrowed a record from Sarah and she asked for it back again.

Agent-Goal/Source Sentences

Pronoun biased to Agent antecedent

Phil drove towards Len and he braked at the last minute. (Agent-Goal)
Len drove away from Phil and he crashed into a brick wall. (Agent-Source)

Pronoun biased to Goal or Source antecedent

Phil drove towards Len and he jumped out of the way. (Agent-Goal)
Len drove away from Phil and he waved at the disappearing car. (Agent-Source)

Experiencer-Stimulus Sentences

Pronoun biased to Experiencer antecedent

Darren disliked Martin and he made it clear to everyone.
Martin annoyed Darren and he stormed out of the room.

Pronoun biased to Stimulus antecedent

Martin annoyed Darren and he regretted it later that evening.
Darren disliked Martin and he reciprocated the ill will entirely.

Table one: Examples of the materials used in Experiment one. The underlined noun phrase is the intended antecedent, and is mentioned either first or second in the sentence.

From Nishigauchi's hierarchy, we predicted that pronouns should be interpreted more easily when their antecedents were Agents rather than Patients, Goals rather than Sources and Agents rather than Goals or Sources. The thematic hierarchy makes no prediction about the Experiencer-Stimulus sentences.

There were four factors in the experiment: the four sentence types; the two thematic roles in each sentence; position of the antecedent in the first clause (either first or second mentioned); and type of noun phrase, either a name, as in Table one, or a definite description (e.g. the policeman). There were two sentences in each of these 32 experimental conditions.

A self-paced reading time task was used. Sentences appeared one clause at a time on a computer screen and subjects pressed the space bar of the computer keyboard when they had read and understood each clause. The time taken to read the second clause of each sentence was measured. Questions were asked after approximately one third of the sentences to ensure that they were read for comprehension. Sixty four filler sentences were also included which tested hypotheses about temporal expressions. The results are shown in Table two. Since there was no effect of type of noun phrase and nothing interacted with this factor, the data have been combined for this factor.

Position of antecedent in the sentence			
Thematic role of antecedent	First Mentioned	Second Mentioned	Overall Means
Agent	5.07	4.71	4.89
Patient	4.77	5.09	4.93
Goal	5.65	5.68	5.66
Source	5.00	5.33	5.16
Agent-Goal	5.16	5.16	5.16
Agent-Source	5.24	4.77	5.01
Experiencer	5.38	5.12	5.20
Stimulus	4.87	4.70	4.78

TABLE TWO: Mean reading rates (in words per second) for the second clause in each antecedent condition for each sentence type.

The main findings were as follows: the clause containing the pronoun was read more quickly when the antecedent was Agent in first position and Patient in second position. This is consistent with the retrieval of a canonical Agent-Patient event, but inconsistent with the thematic hierarchy which predicted that the clause should be read more quickly when the antecedent was Agent rather than Patient, regardless of its sentence position. In the Goal-Source sentences, the clause was read more quickly when the antecedent

was Goal rather than Source, as both the hierarchy and the canonical event hypotheses would predict. In the Agent-Source sentences, the clause was read more quickly when the antecedent was the Agent rather than Source. This is only partially consistent with the two hypotheses, which both predict a similar preference in Agent-Goal sentences. In the Experiencer-Stimulus sentences, the clause was read more rapidly when the antecedent was Experiencer rather than Stimulus, consistent with the view that the verbs trigger the retrieval of canonical events. The thematic hierarchy makes no predictions for such sentences.

The details of the analyses are as follows: Agent-Patient sentences showed an interaction between thematic role and position of the antecedent ($F_1=7.00$, $df=1,31$, $p<.05$; $F_2=5.31$, $df=1,15$, $p<.05$). Reading rates were facilitated when the Agent was mentioned first and when the patient was mentioned second. Goal-Source sentences showed a main effect of thematic role ($F_1=18.45$, $df=1,31$; $p<.01$; $F_2=15.62$, $df=1,15$, $p<.01$). Goal antecedents were retrieved more quickly than Source antecedents. Agent-Goal/Source sentences showed a main effect of position of antecedent ($F_1=6.89$, $df=1,31$, $p<.05$; $F_2=3.28$, $df=1,15$, $p<.1$). Reading rates were facilitated when the antecedent was the Agent, and hence mentioned first. There was also a significant interaction between thematic role and antecedent position in the F_1 analysis only ($F_1=4.47$, $df=1,31$, $p<.05$; $F_2=2.02$, $df=1,15$,). Facilitation for the Agent was only apparent when the Agent occurred with a Source. Experiencer-Stimulus sentences showed a main effect of thematic role ($F_1=11.17$, $df=1,31$, $p<.01$; $F_2=7.94$, $df=1,15$, $p<.05$). Pronouns were interpreted more quickly when the antecedent was Experiencer rather than Stimulus. There was also a main effect of position of the antecedent, in the F_1 analysis only ($F_1=4.33$, $df=1,31$; $p<.05$; $F_2=2.76$, $df=1,15$). Reading rates were facilitated when the antecedent was mentioned first rather than second in the sentence.

These results were replicated in an additional experiment using modified materials, which included linguistically unambiguous pronouns: the two individuals mentioned in the first clause were different genders. The results confirmed the main findings above with two exceptions: (1) the facilitating effect of the Agent appeared in both Agent-Goal and Agent-Source sentences, and (2) the facilitating effect of the Experiencer antecedent only appeared when it was the first mentioned antecedent. These results held for linguistically unambiguous pronouns as well as ambiguous ones.

EXPERIMENT TWO

Experiment two extended the generality of the findings by using a sentence continuation task. Sentences from the replication experiment were used. Subjects were presented with with sentence fragments up to and including the pronoun (e.g. Jill loathed Susan and she....), and were instructed to complete the sentence. The number of times the completions referred to either of the two potential antecedents was recorded. The results are shown in Table three. The main findings of Experiment one were confirmed. But the sentence continuation task produced additional antecedent position effects in the Agent-Patient and Experiencer-Stimulus sentences.

Position of antecedent in the sentence			
Thematic role of antecedent	First Mentioned	Second Mentioned	Overall Means
Agent	3.00	0.63	1.81
Patient	7.34	5.00	6.17
Goal	6.66	5.72	6.19
Source	2.25	1.34	1.80
Agent-Goal	5.63	2.22	3.92
Agent-Source	5.75	2.34	4.05
Experiencer	6.66	5.50	6.08
Stimulus	2.47	1.31	1.89

TABLE THREE: Mean number of completions (out of 8) containing references to each antecedent in Experiment two.

Details of the analyses are as follows: Agent-Patient sentences showed a main effect of thematic role ($F_1=93.27$, $df=1,31$, $p<.01$; $F_2=59.24$, $df=1,15$, $p<.01$), and a main effect of antecedent position ($F_1=22.44$, $df=1,31$, $p<.01$; $F_2=50.42$, $df=1,15$, $p<.01$). Continuations referred to the Patient rather than the Agent and to the first mentioned antecedent rather than the second. Goal-Source sentences showed a main effect of thematic role ($F_1=129.76$, $df=1,31$, $p<.01$; $F_2=115.56$, $df=1,15$, $p<.01$). Continuations referred to the Goal. Agent-Goal/Source Sentences showed a main effect of antecedent position ($F_1=13.16$, $df=1,31$, $p<.01$; $F_2=15.71$, $df=1,15$, $p<.01$): Continuations referred to the first mentioned antecedent (the Agent) rather than to the second. Experiencer-Stimulus sentences showed a main effect of thematic role ($F_1=135.50$, $df=1,31$, $p<.01$; $F_2=166.26$, $df=1,15$, $p<.01$), and a main effect of antecedent position ($F_1=4.03$, $df=1,31$, $p<.06$; $F_2=14.93$, $df=1,15$, $p<.01$). Continuations contained more references to the Experiencer than to the Stimulus and more references to the first mentioned antecedent than to the second.

Thus the sentence continuation task also reveals an effect of thematic roles. There was a preference to refer to (1) Patients rather than Agents, (2) Goals rather than Sources, (3) Agents rather than Goals or Sources and (4) Experiencer antecedents rather than Stimulus antecedents. In addition there was a preference to refer to the first mentioned antecedent in all sentence types except Goal-Source.

DISCUSSION

The clearest evidence for the proposal that verbs retrieve canonical real world events comes from the Agent-Patient sentences in Experiment one: clauses containing pronouns were read more rapidly when the antecedent was an Agent

subject (mentioned first) or a Patient object (mentioned second). At first glance, the results of the continuation task seem to contradict this hypothesis because the continuations referred to Patients rather than Agents in both antecedent positions. However, there was also a first mention effect, which suggests that the continuations may be sensitive to topicalization in the passive form. That is, over and above a preference for Agent subjects and Patient objects, there is an additional preference for references to first mentioned noun phrases in the passives, i.e. to Patient subjects. However, the results for Agent-Patient sentences from both experiments are inconsistent with the hierarchy hypothesis, which predicts an overall preference for the Agent rather than the Patient. There is an alternative interpretation of the Agent-Patient results of Experiment one. This is that the reading rates are due to a lack of parallelism in the sentences with passive first clauses (see e.g. Frazier, Taft, Roeper and Clifton, 1984). Unfortunately it is difficult to construct Agent-Patient sentences where the canonical event hypothesis is not confounded with a parallelism hypothesis. However, inspection of the continuations from Experiment two reveals that people frequently produce non-parallel structures when the Patient is mentioned first.

The results from the Goal-Source sentences endorse the importance of thematic roles without distinguishing between canonical events and a thematic hierarchy. Reading rates are fast when the sentences describe situations where the person in possession of the goods is also the actor of the second event. In the continuations, the person who possesses the goods becomes the actor of the second event. The results from the Agent-Goal/Source sentences are also consistent with both hypotheses: the Agent being preferred to both Goal and Source. Finally, the results from the Experiencer-Stimulus sentences again endorse the importance of thematic roles and are consistent with the canonical event hypothesis. The preference for Experiencer antecedents in both experiments can only be explained by reference to real world events.

Thus the canonical event hypothesis explains more of the data than does the hierarchy hypothesis. We suggest, therefore, that one way in which general knowledge is rapidly retrieved for the comprehension of pronouns is by the lexical information in the verb triggering the retrieval of canonical events in the real world. The more closely the situation described by the sentence maps onto the canonical event, the more speedily will the pronoun be interpreted. These suggestions are not conclusive, but they do indicate how thematic roles might be used for the rapid retrieval of general knowledge. This use of thematic roles is likely to be modified by other information in the sentence, as shown by the topicalizing effect of the passive form. In addition it is likely that the verb information is used in conjunction with the connective. In these experiments we used and. If we had used because the precise pattern of results would probably be different, particularly for the Experiencer-Stimulus sentences (e.g. Garvey, Caramazza and Yates, 1976). Nevertheless, if the canonical event hypothesis is correct, then it should contribute to a general theory of the way general knowledge is retrieved in comprehension and not just apply to pronouns.

REFERENCES

- Brown, R. & Fish, D. (1983). The psychological causality implicit in language. Cognition, 14, 237-273.
- Carlson, G.N. & Tanenhaus, M.K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), Syntax and Semantics, Volume 21. San Diego: Academic Press.
- Finer, D. & Roeper, T. (1989). From cognition to thematic roles: the projection principle as an acquisition mechanism. In R.J. Mathews & W. Demopoulos (Eds.), Learnability and Linguistic Theory. Dordrecht: Kluwer Academic Publishers.
- Frazier, L., Taft, L. Roeper, T. & Clifton, C. (1984). Parallel structure: a source of facilitation in sentence comprehension. Memory and Cognition, 12, 421-430.
- Garnham, A. (1989). Mental Models as Representations of Discourse and Text. Chichester, England: Ellis Horwood/John Wiley and Sons.
- Garvey, C., Caramazza, A. & Yates, J. (1976). Factors influencing assignment of pronoun antecedents. Cognition, 3(3), 227-243.
- Jackendoff, R.S. (1972). Semantics in Generative Grammar. Cambridge, Mass.: MIT Press.
- Johnson-Laird, P.N. (1983). Mental Models. Cambridge, Mass.: Harvard University Press.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. Psychological Review, 95, 163-182.
- Ladusaw, W.A. & Dowty, D.R. (1988). Toward a nongrammatical account of thematic roles. In W. Wilkins (Ed.), Syntax and Semantics, Volume 21. San Diego: Academic Press.
- Nishigauchi, T. (1984). Control and the thematic domain. Language, 60, 215-250.
- Pinker, S. (1989). Learnability and Cognition: The Acquisition of Argument Structure. Cambridge, Mass., London, England: Bradford/MIT Press.
- Rayner, K., Carlson, M., & Frazier, L. (1984). The interaction of syntax and semantics in sentence processing: eye movements in the analysis of semantically biased sentences. Journal of Verbal Learning and Verbal Behaviour, 22, 358-374.
- Sanford, A.J. & Garrod, S.C. (1981). Understanding Written Language. Chichester, England: John Wiley & Sons.
- Stevenson, R.J. & Vitkovitch, M. (1986). The comprehension of anaphoric relations. Language and Speech, 29, 335-360.
- Williams, E. (1989). The anaphoric nature of O-roles. Linguistic Inquiry, 20, 425-456.

¹ This research was supported by Grant No. RC0023 2441 from The Economic and Social Research Council, Great Britain to the first two authors, and by the Human Communication Research Centre of the ESRC, Great Britain at the Universities of Edinburgh, Glasgow and Durham.