THEMATIC FIT AND SYNTACTIC AMBIGUITY RESOLUTION OF INTRANSITIVE MAIN CLAUSES

Luba Dachevski
University of Western Ontario

What do people do when reading and processing sentences similar to those presented in (1) and (2)?

(1) The cop arrested the crook.
(2) The cop arrested by the detective was guilty of taking bribes.

They quickly look to information within the sentences and to knowledge of the world in order to access the correct meanings of words and the appropriate relationships among them. One of the mechanisms used by readers is syntax, for they must construct their interpretation as the sentence unfolds in a linear fashion (Binder, Duffy & Rayner, 2001; Ferriera & Clifton, 1986; Trueswell, Tanenhaus & Garnsey, 1994). Although, there has been much research devoted to mechanisms beyond syntax that are used in sentence comprehension (Binder et al., 2001; Ferriera & Clifton, 1986; McDonald, Pearlmutter & Seidenberg, 1994; McRae, Ferretti, & Amyote, 1997; McRae, Spivey-Knowlton & Tanenhaus, 1998), there is a controversial gap in the research which requires theorists to describe the ways information is integrated by the reader. What remains to be seen is how these mechanisms, or constraints as they are sometimes called, function and the time at which they are used.

A debate between some of the researchers (Binder, Duffy, Rayner, 2001 Ferriera & Clifton, 1986; Frazier & Fodor, 1987; McDonald et al., 1994; Spivey-Knowlton, & Tanenhaus, 1998) has focused particularly on the initial source(s) readers use, especially when processing a sentence. They have looked at syntactic ambiguity resolution as a tool to examine how constraints are used in processing less frequent and less preferred sentence structures. If we look back to our examples 1 and 2, we can see that the beginning of the sentences, The cop arrested, is temporarily syntactically ambiguous because it can be completed in more than one way. It is this ambiguity that readers must resolve. The goal of certain theorists has been to posit mechanisms and models that explain how readers resolve exactly this type of ambiguity.

1.2 Dominant Approaches

There exist at least two theories that seek to provide a clear answer to the debate; a) two stage models (Binder et al., 2001; Ferriera & Clifton, 1986; Frazier & Fodor, 1987) and b) constraint based theories (Altmann, 1999; MacDonald et al., 1994; Spivey & Tanenhaus, 1998, McRae et al., 1997). They differ in claims of what kind of information is used first when processing a written sentence.

Proponents of two stage models (Ferriera & Clifton, 1986; Frazier & Fodor, 1987; Binder et al., 2001) have reported research that suggests syntax is the only initial source of information that readers employ in order to understand
the meaning of the sentence. When the sentence is syntactically ambiguous, a
garden path effect occurs. This is invoked by a sentence that suggests one
interpretation but is then resolved using another interpretation. This effect is said
to occur because readers use minimal attachment to resolve the sentence and
arrive at the appropriate interpretation (Binder et al., 2001). Minimal attachment
is the process during which readers adopt the analysis that involves the fewest
nodes in a hypothetical syntactic mental representation. For example our
sentence *The cop arrested the crook* consists of two noun phrase nodes (i.e *the
cop* and *the crook*) and a verb phrase node (i.e *arrested*). The basic assumption
here is that information from the syntax is the most important information
needed. The two stage models suggest that only after the readers parse the
sentences into nodes using this technique do they consider conceptual
information.

On the other hand, constraint based theorists believe that multiple
sources of linguistic and non-linguistic information are used simultaneously as
soon as they become available (Altmann, 1999; MacDonald et al., 1994; Spivey &
Tanenhaus, 1998, McRae et al., 1997; McRae et al., 1998). These researchers
propose that all these constraints are used together in order for the reader to
arrive at an interpretation that complies with each of the sources of
information/constraints within the sentence. Among these constraints, thematic
fit is one that has been extensively studied (McRae et al., 1997; Trueswell et al.,
1994).

1.3 Thematic Fit

Thematic fit constraints function during sentence comprehension by allowing
readers to access information about the relationships between the nouns in a
sentence and the relationship(s) between the noun(s) and the verb(s) (Altmann,
1999). This information is semantic and based on world knowledge. In our
example sentences (1) and (2), thematic fit is used to understand that in the first
of these, *the cop* is the doer of the verb, in other words, he or she is the agent of
*arresting* and *the crook* is the patient (i.e. *the crook* is the recipient of the
action). However, in the second sentence, *the cop* plays the role of the patient
and the detective is the agent (i.e. *the cop* is the recipient of *arresting*, and *the
detective* is the doer). Typical agents are nouns that are commonly found to be
the doers of an action and frequently occur in the initial noun position of a main
clause structure. Typical patients are nouns that are commonly found as the
recipients of an action and frequently occur after the agents in a main clause
structure. While reading, it is also useful for a reader to recognize that cops are
usually the agents in sentences regarding arresting, this makes them good
agents, and that crooks are good patients because they are typically the ones
being arrested.

More specifically, thematic fit constraints of the sentences’ initial noun
and a particular verb can influence a reader’s expectations (McRae et al., 1997;
McRae et al., 1998; Spivey & Tanenhaus, 1998; Trueswell et al., 1994). In
addition, knowledge about a verb primes information about typical agents,
patients and instruments (Ferretti, McRae & Hatherell, 2001). As such, using a
typical patient in first noun position, where an agent would be in a main clause
structure, is another way of creating less frequent and less preferred sentences. Consider (3) and (4)

(3) The guest invited by the organizer was thrilled to attend.
(4) The host invited by the organizer was thrilled to attend.

These sentences are examples of a typical agent used as the initial noun (as in (4), which is easier to interpret) and a typical patient in initial noun position (as in (3), which requires more reading time).

In order to determine whether thematic fit influences the parsing of a sentence, McRae et al. (1997) designed a series of experiments to determine where the relationship exists and to determine the kind of information contained within thematic verb roles. The data supported the claim that the relationship does in fact exist. They also found a significant correlation between the features listed, between specific agent and patient nouns, ratings of agenthood and patienthood in relation to particular verbs, and the features for typical patients and agents for those same verbs. As a result, norming data for certain verbs and their agents’ and patients’ typicality and the features for each were established. The data supported the claim that sets of nouns and verbs contain specific and contextual information (McRae et al., 1997).

The role of thematic fit has also been demonstrated to aid the processing of spoken sentences. Using a head-mounting eye-tracking device, Altmann and Kamide (1999) have found evidence that thematic fit had an influence when participants heard the word cake within sentences that used the verbs to eat and to move. When the verb to eat preceded the noun, participants were faster to look at an image of a cake than at images of other non-edible nouns; they were slower to look at the cake when they heard the condition using the verb to move. This supports the notion that information found within the verb constrains the expectation for the noun to which the verb refers (Altmann & Kamide, 1999).

To access information regarding the mechanisms used in sentence processing, measurements of readers on-line ambiguity resolution need to be taken. One of the ways of creating syntactically ambiguous sentences is through the use of verb-tense forms. Verbs like to arrest have one form for two distinct tenses; arrested is used in the simple past tense and the past participle tense of the verb. A reader encountering arrested would experience temporary syntactic ambiguity until additional information is integrated. Other verbs do not necessarily pose the same problem. For example, the verb to choose has two distinct forms, chose and chosen.

By using verb forms like arrested, researchers can create two different sentence structures like the ones we have seen in (1) and (2) that are replicated below.

(1) The cop arrested the crook
(2) The cop arrested by the detective was guilty of taking bribes.

The first of these sentences represents a main clause structure and the second represents a reduced relative structure. Main clause constructions are formed from a subject-verb-compliment composition. Readers of English have been found to be biased towards main clause structure as it is the most frequent one
Reading times for each of these sentences have been found to be significantly different, yet the two aforementioned theories are at odds when explaining the reason for the difference. A two-stage model would claim that at first, syntax does not make sense for example 2. This causes the reader to revise the parsing of the sentence based on knowledge from other sources within the sentence (Binder et al., 2001). A constraint based model, on the other hand, suggests that the typical agent followed by the verb would shape the reader’s expectation towards a main clause structure, resulting in a conflict once the reader moves on past the verb. The parsing that the reader assumes based on thematic fit would then be recognized as incorrect, as the sentence proceeds using a reduced relative structure. As such, reading times would increase while the reader deals with the conflict between contradictory evidence by pausing or slowing down to resolve the intended meaning of the sentence (McRae et al., 1998).

1.4 A Challenge

So far, all the reviewed literature has focused on determining if constraints beyond syntax are used when parsing a syntaxtically ambiguous sentence. The goal has been to make a less frequent and less preferred structure, reduced relative structure, easier to understand by manipulating appropriate constraints (Ferreira & Clifton, 1986; McDonald et al., 1994; Trueswell et al., 1994; McRae et al., 1997; McRae et al., 1998; Spivey & Tanenhaus, 1998). To extend the research, Frazier (1995) proposed a more rigorous test to determine whether these constraints are used. She urged researchers to examine whether a preferred and easier to process structure, main clause structure, could be made more difficult to process by manipulating constraints. By creating sentences with biased thematic fit, verb tenses, and post verbal particles, a main clause structure could be made more difficult to process due to the fact that the beginning of the sentence would appear to suggest that a reduced relative structure is emerging.

In response to this challenge, Binder et al. (2001) manipulated sentences using a main clause structure according to these factors and found no evidence that constraints influenced the reading times. They established off-line endings for a set of noun-verb combinations that were hanging at the verb. They were then interested in determining whether these biases held for on-line sentence comprehension. Using the thematic fit information from the norming task, they created two experimental conditions that were different with respect to thematic fit. Binder et al. (2001) called the first condition balanced because the off-line task found almost equal frequencies of main clause and reduced relative sentences. This is comparable to McRae et al.’s (1998) good patient condition. The second of the two experiments was the biased condition; these fragments biased the readers to a main clause interpretation based on the off-line sentence completion. This was comparable to McRae et al.’s (1998) good agent condition.

Binder et al. (2001) reported the experiment which biased thematic fit of the target sentences differently. In their experiment, participants read the
isolated target sentences. The variables that were manipulated were thematic information (based on the information from their norming study) and sentence structure (they used either main clause or reduced relative structure). The sentences differed in ambiguity (in which unambiguous sentences were used to obtain baseline reading time). In the task, they found no evidence that balanced-biased manipulation influenced reading times and therefore concluded that information regarding thematic fit did not influence ambiguity resolution and is thus not used when initially processing sentences.

There were however a number of limitations and problems with Binder et al.’s (2001) study. One of these problems is the question of whether they properly manipulated thematic fit. McRae reran the sentence fragments used in Binder et al.’s (2001) study and found that the balanced items had much fewer reduced relative completions than Binder et al.’s (2001) study reported. In addition, the biased items had fewer reduced relative completions. This may be one of the reasons why they found no influence for thematic fit.

Additionally, a second confounding variable was likely to have played a large role in the null results. Binder et al.’s (2001) sentence fragments used articles such as the, his or her after the initial verbs. These post verbal particles presumably biased the participants towards an interpretation of the main clause structure. Consider (5), (6), and (7).

(5) The car towed
(6) The car towed by
(7) The car towed the

Sentence fragment (6) biases the reader towards a reduced relative structure (i.e. one in which the agent follows the verb), whereas sentence fragment (7) biases the reader towards a main clause structure (as described above). The fragment presented in (5) is more neutral because either construction can complete it. Another way of looking at it suggests that in (5) the car can be the agent or patient. In contrast, in sentence fragment (6), the car is more likely to be the patient, suggesting a reduced relative form is necessary to complete it. In (7), we are more likely to understand the car as an agent and complete it with the main clause structure. Spivey-Knowlton et al. (1993) have demonstrated that it is not only the first noun that is the strong biasing constraint; post-verbal cues that must interact with thematic role and other contextual constraints also bias sentence interpretation.

McRae also reran these fragments and found that each of them was completed using a main clause structure. This biasing factor, along with inadequate manipulation of thematic fit most likely was the reason for Binder et al.’s (2001) experiments leading to the rejection of the hypothesis that information regarding thematic fit is used initially when processing ambiguous sentences.

McRae, Spivey & Tanenhaus (2004) designed an experiment that improved on Binder et al. (2001). One of the corrections made by McRae et al. (2004) was the use of noun-verb sets that were shown to be related in thematic fit. This information regarding thematic fit was based on three norming studies. The first of these studies were role/typicality norms similar to (8) and (9).
(8) How common is it for each of the following to invite someone?
(9) How common is it for each of the following to be invited by someone?

By offering the participants a list of nouns, participants provided ratings of agenthood for questions like that in (8), whereas ratings of patienthood were offered for questions resembling (9). The second and third norming studies asked participants to complete sentences departing at the verb and prepositional phrase, such as those in (10) and (11), respectively.

(10) The host invited ______________________
(11) The host invited to the luncheon ______________________

Based on the results of the norming studies, they created twenty target sentences using main clause structure and manipulated for thematic fit and ambiguity, such as in (12) and (13).

(12) The host (had) invited to the luncheon three old buddies who really had no business being there.
(13) The guest (had) invited to the luncheon three old buddies who really had no business being there.

These sentences differ in thematic fit between the first noun and the verb. A host is a good patient for an inviting event, and a guest is a good patient for an inviting event. By including a good patient in initial noun position, the participants are steered towards a reduced relative structure rather than the main clause structure that is actually used, thus complying with Frazier’s (1995) suggestions. The use of the had particle is similar to the post verbal particle used by Binder et al. (2001) in that it is a strong cue for a main clause continuation, meant to be used for baseline comparisons.

The results supported the constraint based theory of sentence processing (McRae et al., 2004). The thematic fit by ambiguity interaction at the prepositional phrase (to the luncheon) was significant. In the good patients condition, reading times were shorter for condition without had than condition with had. This suggests that the presence of had causes additional competition. Furthermore, in the good patients condition, reading times were longer for condition without had than for condition with had. In both cases, evidence for main clause structure was stronger than evidence for reduced relative structure. These findings contradict Binder et al.’s (2001) findings and lend themselves to the constraint based theory of sentence processing.

One possible criticism of McRae et al. (2004) is that sentences with shifted noun phrases may come across as odd in English. One way around that potential criticism is to use intransitive sentences. Intransitive main clause sentence structure is different from the transitive sentences used in the examples above in that they do not have objects. These sentences would not have shifted patients and thus may be easier for readers to interpret. Sentences (14) and (15) below are examples of intransitive main clauses that will be used in the present investigation.
The hunter trapped near the river because the game warden rarely checked there.
The fugitive trapped near the river because the game warden rarely checked there.

By introducing intransitive main clauses, this experiment will continue in the vein of previous literature by exploring whether information regarding thematic fit is used in these forms as well. I will manipulate thematic fit of initial noun such that good patients and good agents begin the sentence. There will also be sentences with the auxiliary verb *had* prior to the first verb, thus biasing readers to a main clause structure, in order to use these sentences as baseline comparison items.

2. Method

2.1 Participants
Research participants were volunteers recruited by posters from the Psychology Research Participation Pool at the University of Western Ontario and other on-campus sites. Age and gender of the participants were not measured, as they were irrelevant. Scores for the participants who did not display native-like fluency of English were not used. A total of 48 participants were tested, with 12 participants per list of stimuli items.

2.2 Materials

2.2.1 Norming Studies.
The results of three off-line norming studies were used to create sixteen sets of experimental. These norming studies were identical to those used by McRae et al. (2004) (role/filler typicality ratings, completions at the verb and completions at the prepositional phrase, as described above) except for the verbs used in the typicality ratings and the fragmented sentences used for the completion tasks.

For each of the 42 verbs presented in the role/filler typicality ratings, a pair of sentences was constructed. In each pair, one sentence included a good agent in first noun position (i.e. with high ratings as the doer of the action), and one sentence included a good patient (i.e. with high ratings as the receiver of the action) in first noun position. One member from each pair was used in one of two sentence completion studies.

Twenty participants completed each of the last two norming studies, where each participant encountered each verb only once. One hundred and twelve fillers were randomized within each list. At least one filler item separated each target item. Participants’ responses were scored as to how many intransitive main clause (IMC), reduced relative (RR), noun phrase shift (NPS), self, nonsense, and blank completions there were for each verbal condition. See Table 1 for agenthood and patienthood ratings and the means of each of these in the Good Agent and Good Patient conditions.
Table 1
Mean Agenthood and Mean Patienthood Ratings and Sentence Fragment Completion Data from Norming Studies

<table>
<thead>
<tr>
<th>Variable</th>
<th>Good Agent Mean</th>
<th>Good Patient Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agenthood Rating</td>
<td>6.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Patienthood Rating</td>
<td>3.0</td>
<td>5.7</td>
</tr>
<tr>
<td>MC completions at the verb</td>
<td>99.4</td>
<td>74.1</td>
</tr>
<tr>
<td>RR completions at the verb</td>
<td>0.0</td>
<td>16.5</td>
</tr>
<tr>
<td>Self completions at the verb</td>
<td>0.3</td>
<td>8.2</td>
</tr>
<tr>
<td>RR completions at the PP</td>
<td>32.4</td>
<td>77.6</td>
</tr>
<tr>
<td>NPS completions at the PP</td>
<td>28.1</td>
<td>6.0</td>
</tr>
<tr>
<td>Intransitive MC completions at the PP</td>
<td>37.2</td>
<td>10.7</td>
</tr>
</tbody>
</table>

Note: Less than 1% of responses for each sentence completion tasks were left blank.

2.2.2. Self Paced Reading Materials.

Sixteen intransitive main clause experimental sentences were created based on the data gathered in the three norming studies. They were chosen on the basis that they were intransitive and had highly polarized ratings for agents and patients. Each of the sentences were modified so as to create four conditions; good agent or good patient in the first noun position for the first two conditions, as in (16) and (17), and ambiguous or unambiguous for the second two conditions, as shown by (18) and (19).

(16) (good agent condition)
The hunter trapped near the river because the game warden rarely checked there.

(17) (good patient condition)
The fugitive trapped near the river because the game warden rarely checked there.

(18) (ambiguous condition)
The hunter trapped near the river because the game warden rarely checked there.

(19) (unambiguous condition)
The hunter had trapped near the river because the game warden rarely checked there.
As can be seen by these examples, the good agent condition used a noun that is very likely to be the doer of the action, whereas the good patient condition used a noun that is very likely to be the recipient of the action. Additionally, the difference between the ambiguous and unambiguous conditions lies in the inclusion or exclusion of the word *had*. As it is a strong cue for main clause sentence structure, *had* is found in the unambiguous condition. Furthermore, each verb phrase was followed by a three-word prepositional phrase (e.g. *near the river*) that began the compliment portion of each main clause.

Only one condition of the target sentence appeared per list, such that the participants saw each verb only once. Each target sentence was followed by a continuation sentence in order to provide additional contextual information. The continuation sentence that followed each condition of (16) through (19) was *No one ever found out where he was hiding*. These questions were constructed so that half of the questions should yield a ‘yes’ response, and half the questions should yield a ‘no’ response. Twenty practice items were also included in the self-paced reading materials.

2.3 Procedure

The experimental sessions each began with an oral and written description of the task and distribution of the informed consent form. Twenty practice items were presented, followed by a random presentation of 16 target items among 60 filler items. A comprehension question was presented to measure each participant’s focus on the task. Each sentence was presented in a one-word-at-a-time self-paced moving window format, as described in Just, Carpenter and Woolley (1982).

At the beginning of each trial, a series of dashes appeared on the screen. Each dash represented a non-space character of each sentence pair. As each participant pressed the green button on the button box, a word was revealed. With each successive button press, a new word appeared as the previous word became concealed by dashes. After each sentence pair was revealed and obscured, the comprehension question appeared on the computer screen where the sentences were presented. Participants were instructed to press the red button on the button box if they believed the answer to be ‘no’ and the yellow button on the button box if they believed the answer to be ‘yes’. Upon completion of each third of the experiment, participants were given the opportunity to take a break by delaying the beginning of the next sentence pair. Once all sentences and questions were completed, participants were debriefed. A Macintosh computer operating with OS 8.5 and PsyScope presented each sentence and recorded button press latencies. A CMU button box measured word by word reading times and rounded them to the nearest ms.

3. Results

The target sentences were divided and analyzed by four critical regions of the experimental sentences. These regions were the initial NP, the VP, and the three
(or four or in one case five) word prepositional phrase (PP), and the first three words of the complement as can be seen in (20) below.

(20) The hunter / (had) trapped / near the river / because the game…

Analyses of variance were conducted using reading times as the dependent variable for each of the three critical regions of the experimental sentences. The independent variables were thematic fit (good agent condition vs. good patient condition) and ambiguity (had condition vs. no had condition). Both independent variables were within subjects.

The purpose of this investigation was to demonstrate the effect of thematic fit on the intransitive main clause sentences. Figure 1 demonstrates the reading times for each region of the text at each of the four variants of the stimuli.

Constraint-based theory predicts an interaction of thematic fit and ambiguity to be at the prepositional phrase or the compliment region of the sentence. The interaction was not significant for the prepositional phrase, $F(1, 47)= 0.014$, n.s., and the compliment, $F(1, 47)=0.356$, n.s.

It was also predicted that there would be a difference in reaction time between the sentences that included the word ‘had’ and those that excluded it as an auxiliary verb. The analyses found a significant main effect for ambiguity at the prepositional phrase, $F(1, 47)= 5.29$, $p<0.05$, with the sentence including ‘had’ taking an average of 20ms longer to read than the sentence excluding ‘had’. No significant main effect for ambiguity was found at the compliment, $F(1, 47)=0.28$, n.s.

Consistent with previous research, evidence for thematic fit alone was found in one of the two post-verbal regions of the sentence. A significant main effect was found for thematic fit at the first three words of the compliment, $F(1, 47)= 9.14$, $p<0.005$, with the sentence featuring a good agent in first noun position taking 26ms shorter to read than the sentence featuring a good patient in first noun position. This effect was not found at the prepositional phrase, $F(1, 47)= 1.34$, n.s.
4. Discussion

Although there was no interaction between thematic fit (good agent or patient in first noun position) and ambiguity (inclusion or exclusion of the auxiliary verb *had* before the verb) at the prepositional phrase, the main effect permits an explanation beyond the original theory.

As proposed by Trueswell (1996), multiple completions may be computed simultaneously until the one constrained by context allows the reader to assume one structure over another. The fastest reading time being found for the good agent condition without *had* suggests that the reader assumed the dominant, or main clause structure, and proceeded without much difficulty. It may be the case that the version without *had* was faster than the version including it because the verbs chosen for this task were more frequently occurring in simple past form rather than past participle (which would require the *had*). As such, readers’ knowledge of frequency for the verbs in particular and the sentence in general guided their reading times for this condition.

The second and third fastest reading time are found for stimuli items with the good patient without *had* and good agent with *had* respectively. It may be the case that competition between the reduced relative completion and main clause completion as cued by the prepositional phrase and the presence of *had* respectively explains this unpredicted result (Trueswell, 1996). This competition would cause the reader to slow down for a longer period during the prepositional phrase and delay the presence of the next region to compute which is more frequent. Having come to the conclusion that the verbs used in this experiment typically (or more frequently) come in simple past form as opposed to past participle form, as mentioned above, the reader proceeds faster in the condition that appears to suggest a reduced relative structure as cued by the prepositional phrase and lack of *had* in the verb phrase.

The slowest reading time at the prepositional phrase is found in the condition featuring a good patient in first noun position and exclusion of the auxiliary *had* before the verb. This is consistent with the previously explained data, which suggests that the verbs used in this experiment may be more frequently occurring as simple past forms rather than past participle. Having sensed this oddness, readers take longer when the verb is preceded by *had*.

In the case of the compliment portion of the sentence, the findings are more consistent with the constraint based theories. Readers are faster for conditions featuring good agents than good patients in first noun position due to the fact that they become aware that the sentence is intransitive and no longer expect a transitive reduced relative form (featuring a patient) at this point.

To ensure that this explanation suffices for these findings, a measurement of how frequently each verb appears in simple past form versus past participle form would be useful. In order to find evidence of an interaction, a counterbalancing of both kinds of verbs would be required. It is also suggested...
that the verbs used in future paradigms be truly intransitive, as opposed to those that are transitive and can be used intransitively. Any further examination of thematic fit and ambiguity would require a more rigorous examination into which verbs are used in order to demonstrate an interaction.

References


