INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
THE ROLE OF MEANING IN THE SENTENCE MATCHING TASK

by

Csaba Veres

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF PSYCHOLOGY
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1997
As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Csaba Veres entitled The role of meaning in the sentence matching task and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Ken Forster
Date 10/7/97
Merrill Garrett
Date 10/7/97
Felice Bedford
Date 10/7/97
Christopher Davis
Date 10/7/97
Andrew Barss
Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Ken Forster Date 10/7/1997
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Carla Vee
ACKNOWLEDGEMENTS

First, I must acknowledge my gratitude to Ken Forster who, after having been rudely awakened by me at four in the morning so many years ago by a phone call from Australia, still had the faith to bring me to his lab. Thank you for your support, not just in academic matters, but perhaps more importantly, personal ones throughout the years. I am also grateful to Merrill Garrett whose unrelenting efforts to bring order to chaos have been an important influence throughout the years. Thank you also to my other “Arizona” committee members, Felice Bedford, Andrew Barss for your wit and insight. A special thanks to Christopher Davis who came all the way from Australia just to be on my committee, and to spot all the spelling mistakes in my thesis.

Thank you to Mark and Jonathan for their invaluable help, and just for being good guys. And all the people in the office who have been and gone, Imy and Peggy without whom this department would collapse.

And of course, thank you to all my friends who have made my life enjoyable and fulfilling. Thank you Karen and Sheryl, who came once and will again, I hope, to the land “down under”. And Ed of course, who already has his tickets. And my new friends Jack and Laleh, and many others including Anthony and Nina and Jim and Beth and Miles and Jason and Kathy and Rick and and and....
DEDICATION

I wish to dedicate this dissertation to Gittan, who has tried really really hard to keep me from becoming boring.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>8</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>10</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>2. SEMANTIC ANOMALY AND THE MATCHING TASK</td>
<td>32</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>39</td>
</tr>
<tr>
<td>Method</td>
<td>39</td>
</tr>
<tr>
<td>Results</td>
<td>43</td>
</tr>
<tr>
<td>Discussion</td>
<td>46</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>49</td>
</tr>
<tr>
<td>Method</td>
<td>49</td>
</tr>
<tr>
<td>Results</td>
<td>51</td>
</tr>
<tr>
<td>Discussion</td>
<td>53</td>
</tr>
<tr>
<td>3. SELECTIONAL RESTRICTIONS AND ANOMALY</td>
<td>55</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>63</td>
</tr>
<tr>
<td>Method</td>
<td>64</td>
</tr>
<tr>
<td>Results</td>
<td>65</td>
</tr>
<tr>
<td>Discussion</td>
<td>66</td>
</tr>
<tr>
<td>4. ANOMALY AS A LACK OF CLARITY</td>
<td>68</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>69</td>
</tr>
<tr>
<td>Method</td>
<td>70</td>
</tr>
<tr>
<td>Results</td>
<td>72</td>
</tr>
<tr>
<td>Discussion</td>
<td>74</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>76</td>
</tr>
<tr>
<td>Method</td>
<td>77</td>
</tr>
<tr>
<td>Results</td>
<td>77</td>
</tr>
<tr>
<td>Discussion</td>
<td>79</td>
</tr>
<tr>
<td>5. LEXICAL CONCEPTUAL SEMANTICS</td>
<td>82</td>
</tr>
<tr>
<td>Experiment 6</td>
<td>92</td>
</tr>
<tr>
<td>Method</td>
<td>95</td>
</tr>
<tr>
<td>Results</td>
<td>97</td>
</tr>
<tr>
<td>Discussion</td>
<td>99</td>
</tr>
<tr>
<td>6. RAPID SERIAL VISUAL PRESENTATION</td>
<td>102</td>
</tr>
<tr>
<td>Experiment 7</td>
<td>104</td>
</tr>
<tr>
<td>Method</td>
<td>104</td>
</tr>
<tr>
<td>Results</td>
<td>106</td>
</tr>
<tr>
<td>Discussion</td>
<td>107</td>
</tr>
<tr>
<td>7. GENERAL DISCUSSION</td>
<td>112</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>154</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLE 1, Mean ratings, on a seven point scale, for the items in the three conditions in experiment 1 .................................................... 44

TABLE 2, Reaction Time for Reading and Matching (and error rate for Matching) in experiment 1. Also shown is the difference between each experimental condition and the plausible baseline ........................................ 45

TABLE 3, Reaction Time for Reading and Matching (and error rate for Matching) in experiment 2. Also shown is the difference between each experimental condition and the baseline ........................................ 52

TABLE 4, The semantic feature violations observed in items used in experiment 1. .................................................................................. 62

TABLE 5, Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and its appropriate baseline ........................................ 65

TABLE 6, Mean ratings, on a seven point scale, for the items in the three conditions in experiment 4 .................................................... 72

TABLE 7, Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the plausible baseline ........................................ 73

TABLE 8, Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the plausible baseline ........................................ 78

TABLE 9, Conceptual constituent changes in experiment 6, that changed plausible sentences into anomalous .................................................. 93

TABLE 10, Mean frequency, expressed in number of words/million, for the items in the three conditions in experiment 6 .......... 94
LIST OF TABLES - Continued

TABLE 11, Mean plausibility ratings, on a seven point scale, for the items in the three conditions in experiment 6. ................................................................. 97

TABLE 12, Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the plausible baseline. ................................................................. 98

TABLE 13, Mean error rates (% correct) in the RSVP task in the three conditions, and the difference between each experimental condition and the plausible baseline. ................................................................. 107
ABSTRACT

Semantic plausibility has been shown to affect a number of sentence processing tasks, including reading, sentence matching, and RSVP (Rapid Serial Visual Presentation). In this thesis it is argued that anomaly, and not plausibility is the critical variable. Unfortunately the distinction between anomaly and implausibility has not traditionally been a clear one, and definitions can vary depending on the semantic theory that is being adopted. The experiments reported in this thesis are aimed at finding a clearly definable distinction between anomalous and implausible sentences, and to show a reliable empirical consequence of the distinction. The major emphasis is placed on the sentence matching task, which has been claimed to be sensitive to a very specific level of syntactic processing, and to be unaffected by extraneous variables. Experiments 1 and 2, however, demonstrate the very clear effect of a semantic manipulation, on the sentence matching task. Experiments 3, 4, and 5 were designed to refine our understanding of the aspect of the manipulation that was critical. Finally in experiment 6, it is argued that the violation of basic conceptual categories which have been argued to organize lexical/conceptual structure, is the only circumstance under which the effect of meaning is seen in the sentence matching task, and possibly the reading task. Experiment 7 shows that the same conclusion is true for the RSVP task. These conclusions are considered in the light of Jackendoff's (1983) and Pustejovsky's (1990) theory of semantics. It is further argued that a close symbiotic relationship can exist between these formal theories, and the empirical findings.
Chapter 1

INTRODUCTION

It is very rare in psychology that we have access to a measuring device that can give insight into a single, well defined stage of processing in mental computation. Almost always, there are several different processes which contribute to a single effect, and the source of these influences is often not clear. Conclusions about the effect of any one mental operation must therefore be tentative at best. The sentence matching paradigm introduced by Freedman & Forster (1985) is interesting because it was claimed to be a task which is affected by a very specific level of representation during sentence processing, and immune to other factors.

The subject participating in a sentence matching task (also referred to as same/different or SD matching), is asked to decide as quickly as possible if two sequences of printed words are identical or not. So, for instance, the subject might see the following display, and would be required to decide if the two strings were the same:

(1) John went to the doctor yesterday.

John went to the doctor yesterday.
Half the items in an experiment would be identical as in the above example, and half would not, typically differing by one word randomly placed in the sentence:

(2) John went to the doctor yesterday.
    John went to the grocer yesterday.

The different word is the same length as the one it is compared against, so that the source of the difference is not immediately obvious. The dependent variable is the time taken to decide that two sentences are the same. The reaction times to the "different" sentences are fairly uninteresting, since those are determined mainly by the position of the word that differs between the two sentences.

The main idea behind the task is that the speed with which sentences can be compared is determined by structure in the input strings, that subjects can make use of (Forster, 1987).

For example, consider each of the following pairs of letter strings (from Forster, 1987). Note that the response in each case is "different".

(3) \begin{verbatim}
    ATFCDHMLORTEUODGAMCH
    ATFCDHMLORBEUODGAMCH
\end{verbatim}

(4) \begin{verbatim}
    MKLP AFDM BRST KDBS HWRD
    MKLP AFDM BRKT KDBS HWRD
\end{verbatim}
It should be evident that the time taken to detect the difference tends to decrease as the structure in the strings increases. In our examples, this ranges from simple perceptual organization in (4), to the addition of phonological properties in (5), pseudo-syntactic structure in (6), and syntactic structure seen in (7).

There are several different ways in which the perceived structure might be helpful. It is possible, for instance, that short term memory is critically involved in the task, in which case the effect of structure would serve as a memory aid. A more interesting possibility suggested by Chambers & Forster (1975), is that the matching task is simultaneously being attempted at several different levels of abstraction, and that the first process to reach a successful completion determines the overall matching time. They called this the "multilevel race model". According to this model, a purely physical level comparison competes with one at the word level, phonological level, syntactic, and interpretive levels, and possibly others. The assumption of the model is that the matching
operation becomes more efficient with higher levels of abstraction. A comparison at the feature level would require hundreds of low level comparisons, whereas a comparison at a syntactic level would involve far fewer, but more intricate comparisons. At the highest level, a highly stereotyped proposition like *Butchers sell meat* might be compared as a single conceptual unit. Of course, it is very difficult to independently predict how long a comparison performed at any given level should take since our knowledge of the sorts of comparisons taking place at any level is severely limited, and we don't know exactly how long it takes to compute each successive stage in the derivation. The savings of having a high level comparison might, for instance be offset in some cases by the extra time it takes to develop the high level representation in the first place. Nevertheless, there is empirical evidence which suggests that the assumption is at least plausible, with matching times decreasing monotonically with increased structural elaboration.

Freedman & Forster (1985) point out that, in a simple same/different matching task, the pair of letter strings shown in (8a) is faster to match than the pair shown in (8b).

(8)  

<table>
<thead>
<tr>
<th></th>
<th>a. HOUSE</th>
<th>b. HSEUO</th>
<th>c. HOSUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HOUSE</td>
<td>HSEUO</td>
<td>HOSUE</td>
</tr>
</tbody>
</table>

This example shows that the time taken to discover the abstract lexical representation of the word, and to then perform a comparison on the basis of that representation, is faster than making the comparison at a letter level. The example shown in (8c) takes an amount of time intermediate between the other two, which shows that syllable structure also
benefits the comparison process, compared with a purely unstructured string. These examples clearly demonstrate that each level of abstraction that can be added to the stimulus facilitates the matching process.

These findings for single word stimuli can be extended to more complex inputs involving word sequences. In the example below the second pair shows a structurally correct, stereotyped relation, and is matched faster than the first:

(9) a. GROWL DOGS b. DOGS GROWL
   GROWL DOGS DOGS GROWL

And finally, referring back to examples (6) and (7) above, it is clear that a string of pseudo-words with syntactic structure is matched faster than a random arrangement of the same words. So, whatever the nature of the mechanisms involved in sentence matching, they clearly seem to be benefited by the presence of more elaborated structural properties in the input strings.

One intriguing possibility has been raised by Forster (personal communication), for the nature of this matching operation. In an unpublished experiment, he found that the time it takes to match a pair of words is correlated with their lengths. Now, if the earlier description of the multilevel race model were strictly correct, this should not be the case. Recall that the idea behind the model was that reaction time was controlled by the most efficient level, and the evidence suggested that more abstract levels were in fact more efficient. The reason for this added efficiency was that fewer comparisons needed to be
made at the higher levels. The reason why words are matched faster than non-word letter strings is that their representations at the lexical level can be compared directly, whereas non-words could only be compared letter-by-letter. But then word length ought to exert no influence on matching time at all, since the letter level description should have been abstracted away from. Forster's suggestion, then, is that the two high level representations are not compared directly, but that the representation of the first stimulus is used to check the identity of the second, in the following way: The lexical entry for the first word is accessed, and its contents made available. Among the information stored in the entry is a detailed specification of the word's spelling, which can then be used to see if the second stimulus had the necessary features. The increased efficiency over low-level letter-by-letter checking would then come from the fact that the features that needed to be compared are bundled into neat, easy to use packages, rather than having to be held in short term memory as individual, disconnected items. The multilevel race model would need to be altered slightly to accommodate this possibility, and say that the controlling level is the one at which information is organized and compared in the most useful way.

There is a sense, however, in which the exact details of why the more highly structured inputs are matched faster, might be irrelevant to the interpretation of the matching task. What is important is that more highly developed abstract representations are matched faster than unstructured ones, allowing us to use the task to index the availability of structural descriptions, whatever the precise mechanisms might be. Nevertheless, the reader is encouraged to keep an open mind about Forster's hypothesis,
since much of what is to come is easier to interpret if the explanation is in fact correct. However no attempt will be made to support the hypothesis directly, since its status is not of critical importance to the results reported in this thesis.

In keeping with the structural elaboration notion in SD matching, Murray (1982) showed that grossly ungrammatical word strings, which contained real words of English, were matched slower than grammatical sentences. For instance, (10a) takes longer than (10b).

(10) a. *believe police that the John Mary shot
    b. the police believe that Mary shot John

More subtle ungrammaticalities can also slow performance in the task; each of the following is slower than its appropriately matched grammatical control (Freedman & Forster, 1985).

(11) a. *Mary were writing a letter to her husband.
    b. *To go to Disneyland John wanted.
    c. *Lesley's parents are chemical engineers both.
    d. *The girl behind you the subsequent discussion.

Murray (1982) and Ratcliff (1983) went on to show that the highly abstract level of meaning interpretation could also influence the task. Thus, a pragmatically implausible sentence like (12a) takes longer to match than a plausible sentence like (12b).
(12)  a. The florist disguised the composer daily.

        b. The workman repaired the factory slowly.

This finding strongly suggests that the highest possible levels of conceptual/interpretive analysis were being recruited to help with the task. In fact, Murray (1982) reported a number of experiments where he showed the strong effect that plausibility can have on the matching task. Some examples of his implausible sentences, from various experiments, are shown below.

(13)  a. The book had clean coils.

        b. The brothers returned the station.

        c. The knights bronzed the rocks.

        d. The hydrogen had been seized.

However, while the story so far seems relatively straightforward, a number of findings reported in Freedman & Forster (1985) shed doubt on this description of the facts. Freedman & Forster (1985) showed that two types of ungrammatical sentences were matched equally quickly as their appropriately matched grammatical controls. These ungrammaticalities involved (i) the illegal movement of a wh- phrase, which is blocked by a lexically specified subject, and (ii) violations of the subjacency constraint. Sentences (14a) and (b) were no slower to match than controls matched for lexical content.

(14)  a. *What did you see Leonardo's painting of?

        b. *Who do the police believe the claim that John shot?
Forster (1987) views this as a real problem for the structural elaboration interpretation of the results obtained with the SD matching task. The point he makes is that the grammatical status of a string is determined at a level of processing which is prior to conceptual interpretation, since such interpretation depends on the structural analysis from the syntactic level. If such a structural analysis is unavailable, then this should block the availability of a conceptual level interpretation. Thus, according to the multilevel race model, if the most efficient level at which matching can occur is the interpretive one, then ungrammaticalities in the input string ought to also result in a loss of speed. In fact, experiments using sentences like (11a - d) demonstrated precisely this fact: sentence matching was sensitive to grammatical violations. So what went wrong with the sentences where movement was involved?

Forster (1987) proposed the shallow-sensitivity hypothesis to explain the apparent contradiction. Recall that the multilevel race model states that the response time in the SD matching task is determined by whichever level of representation reaches a determinate outcome first. This is called the controlling level. According to the shallow-sensitivity hypothesis, then, there is a low level structural representation at which sentences like (14) are well formed, and that the SD matching task can exploit this level, even though the sentence string is found ungrammatical at some later level in its derivation. Sentences of the type shown in (14) were claimed to be examples of overgenerated sentences. Overgeneration is a process in which some sentences are produced as an unwanted by-
product of the generative nature of linguistic rules, and subsequently need to be eliminated by various constraints. Freedman & Forster (1985) claimed that this notion had psychological reality and, that the systems involved in the SD matching task had access to the representations generated by this process.

Supporting evidence for this position came from experiments investigating the effect of binding on matching times. Recall that the puzzle to be solved is, how semantic plausibility can effect SD matching if the controlling level is a fairly shallow syntactic one. The binding experiments suggested the possibility that only certain kinds of implausibility might be able to affect the matching task. That is, some sorts of implausibility, that effected low level syntactic processes, could have an effect on the SD matching task if the effects had their locus below the controlling level of the task. This is precisely the hypothesis that was tested with the binding studies. The logic of these experiments was as follows.

Forster (1987) argued that the ungrammaticality of the constraint violation sentences could only be discovered if a binding operation was attempted between a wh-phrase and a trace, since it is at this stage that the illegal movement is detected. The further claim is that the controlling level for the sentence matching task is at a level at which binding had not yet taken place. It follows from this that sentences in which the implausibility can only be established as a result of a binding operation, should not show a detrimental effect on sentence matching. Stevenson (1984) carried out an experiment
which tested this hypothesis, in which he contrasted sentences like (15 a and b) with ones like (16a and b).

(15) a. What did the mayor order them to unfurl?
    b. Who did the mayor order them to unfurl?

(16) a. The mayor ordered them to unfurl it.
    b. The mayor ordered them to unfurl her.

As predicted, sentences like (15a) and (b) did not differ from each other in reaction times, but ones like (16a) and (b) did, with (16b) taking significantly longer. (Stevenson performed other experiments to rule out the possibility that non-adjacency of "... unfurl (...)" was the cause of the null effect, by showing that sentences with ungrammaticalities caused by non-adjacent words that were not moved, but base generated, like “The man who bought the cars are rich” did show effects of ungrammaticality. This shows that adjacency per se is not critical to obtain effects in this task). The experiments clearly demonstrated that an implausibility that depended on the binding of a wh- phrase with its antecedent was matched just as quickly as a plausible sentence of the same type, which is consistent with the shallow-sensitivity hypothesis. But now we have a new problem: why does (16b) take longer to match than (16a)? Surely the implausibility of the sentence depends on a high level interpretation which is developed subsequent to the controlling level for the matching task, and should therefore be irrelevant to performance.
A possible way out of the dilemma is suggested by some results reported in Freedman & Forster (1985), where it was found that sentences such as (17a) took longer to match than those like (17b).

(17)  

a. *What did the pygmies like orders to smell?  

b. *Who did the soldiers hate orders to kill?

The ungrammaticality of these sentences is due to a constraint violation which is discovered after a binding operation is attempted. Thus, by hypothesis, the SD matching task should not be sensitive to the violation. Yet, they show a plausibility effect. To account for this contradictory finding, Forster (1987) suggested that the sentence matching task was in fact sensitive to plausibility considerations at the phrasal level, and that this kind of implausibility was different from that obtained when sentences are subjected to a full sentential interpretation. Thus, the extra matching time needed for (17a) is due to the implausibility of the phrase *orders to smell*. Similarly, the difference between (16a) and (b) is due entirely to the implausibility of the phrase *unfurl her*.

In order to test this hypothesis directly, Forster (1987) constructed sentences by exchanging VPs from sentence pairs that had similar structure. For instance, (19a) and (b) were formed from (18a) and (b).

(18)  

a. His extensive cellar contained many rare wines.  

b. That part of the river is a good place to fish.
(19)  a. His extensive cellar is a good place to fish.
     b. That part of the river contains many rare wines.

Clearly, the implausibility of the sentences depends on a global interpretation. At the phrasal level, there is nothing implausible about them. Thus we can make a distinction between local (phrasal) and global (sentential) implausibility. Forster (1987) reports no significant effects of plausibility on matching times for these sentences. However, when the sentences are constructed to have local implausibilities, then a significant 78 ms effect appears. Examples (21a) and (b) illustrate two sentences, constructed from (20a) and (b), which meet this criterion. In each case, the sentences in 20 are faster to match than their corresponding pairs in 21.

(20)  a. That cheese in the refrigerator tastes like cheddar.
     b. That lady in the library writes detective novels.

(21)  a. That cheese in the library tastes detective novels.
     b. That lady in the refrigerator writes like cheddar.

Further support for the local/global distinction comes from an experiment reported in Forster (1987), who contrasted sentences of the form Adj-N, with equivalent sentences of the form N-be-Adj. The first two examples below show a plausible alternation, and the second two, an implausible one.
(22)  
a. The woman liked Peter's cotton shirt.  
b. Peter's shirt was made of cotton.

(23)  
a. The woman liked Peter's timber shirt.  
b. Peter's shirt was made of timber.

The first member of each pair defines the relation between the Adj and the N locally, within an NP. But in the second member, the Adj and the N belong to different phrases, and are thus globally defined. The important thing about the sentences is that the plausibility of a given sentence is the same in the two different constructions. In both cases the sentences describe the very same state of affairs, and hence should be equally plausible or implausible. Once again the hypothesis was supported with a significant 56 ms difference between sentences like (22a) and (23a), but a non significant 20 ms difference between (22b) and (23b).

To briefly summarize, the proposal is that the controlling level for SD matching is a syntactic level that precedes semantic interpretation, but that plausibility can still have an effect if it is locally determined.

Forster (1987) suggested a mechanism for how locally defined plausibility can have an effect on SD matching, by influencing the operations of the syntactic parser directly. He suggests that parsing is carried out in two stages. During the first phase, words are collected together into phrasal or clausal units. Then, the parser must work out how these units are related. The overall efficiency of the task is greatly influenced by how quickly the first operation can be accomplished. The suggestion is that the likelihood of
two words forming a phrasal unit can influence the speed with which those words will be clumped together in a unit. For instance in sentence (22a), it must be decided if Peter's cotton, or Peter's cotton shirt, is a phrase. Clearly, since cotton and shirt are likely to co-occur, the second analysis might be quickly adopted. In the implausible case of a timber shirt, however, an erroneous initial grouping might be considered, thereby slowing down the process. Notice that according to this claim, the effect of plausibility on sentence matching is through a "dumb" process, and not through some high level interpretive one. What seems to be important is co-occurrence information that could be stored within the lexical entry, or at least linked to the entry in some direct fashion. It must be noted, however, that this proposal was not extensively studied. An untested prediction of this view, for instance, is that sentences which were relatively similar semantically, but differed in their frequency of use, should show different effects. The phrase peel a tangerine, for instance, should lead to a slower matching time than peel an orange. On the other hand, sentences which differed wildly in their plausibility, but not in their patterns of use, should not differ. Thus, The spider crawled onto my finger, and The spider crawled into my finger should show similar matching times (Forster, 1987). However, these remain untested speculations. Nevertheless, the claim made by Forster is that a combination of the shallow-sensitivity hypothesis and the local implausibility hypothesis (my term) can account for the seemingly incompatible results in sentence matching task: The task is insensitive to violations of constraints of movement, and yet is still sensitive to plausibility, as long as the plausibility has its effect locally, on phrasal construction.
While this suggestion offers a possible resolution of the results, a number of findings question the validity of the proposal. The two main areas of concern have resulted from experiments that, i) attempted to show that the controlling level for sentence matching preceded the level at which binding took place, and ii) tried to figure out exactly what "local" meant, as far as local plausibility was concerned. The experiments mentioned thus far have suggested that the appropriate distinction between "local" and "global" effects involves the NP - VP boundary. However, some studies reported in Forster (1987), and Ratcliff (1983) suggest that this is in fact not the case, since plausibility effects across phrase boundaries can be obtained under certain circumstances. In fact, the exact definition of "local" has never been arrived at.

As for the first point, consider the following. Recall that sentences such as (15a) and (b) do not show a penalty for implausibility. This was used as evidence that the SD matching task was sensitive to a level before the binding of the wh- phrase and its trace occurred. However, a somewhat contradictory study showed that in some sentences where an ungrammaticality depends on binding, the SD task does in fact incur a cost of this ungrammaticality. Thus, (24a) and (b) take longer to match than their grammatical controls.

(24a) *John smiled for he knew herself to be innocent.

*She decided on a holiday for each other.
Forster (1987) suggests that the contradiction could be resolved if we suggest that different kinds of binding operations occur at different stages in processing. If the antecedent and the anaphoric element are both in their base-generated positions, then binding occurs at a level that is below the controlling level for the matching task. If, however, the binding involves a moved element, then the operation is carried out at a higher level, which is above the controlling level. While this story will account for the data, the need for such ad hoc assumptions weakens the proposal.

A final, more general point that needs to be made before we begin exploring an alternative proposal (which will be the starting point to the experiments in this thesis), involves the logic behind the shallow-sensitivity hypothesis. The primary point of that hypothesis was that we cannot expect global plausibility to have an effect on a task which is not even sensitive to a grammatical violation. The assumption behind this claim seems to be that it is not possible to reach a global interpretation of a sentence until its grammatical structure was fully determined. But is this necessarily the case? It is possible, for instance, that a conceptual representation was hypothesized on the basis of the main verb, and this structure would then fill its required roles with the incompletely analyzed lexical/syntactic structure. This process might be similar to one suggested by Forster (1974) who argues that a heuristic method might be employed to derive semantic interpretations from a syntactic analysis. But it is important to note that this analysis does not proceed without syntactic analysis. Instead, it is started on the basis of some minimal grammatical specification. However, syntactic analysis continues, and is essential in the final stages of
processing, where it must be ensured that the semantic analysis is compatible with the syntax. So, the suggestion is not that semantic representation can be built in the absence of a syntactic analysis, nor that semantics can influence the syntax, but simply that a great deal of semantic analysis can be performed before the syntactic analysis is completed. This could be beneficial to the processing system if, for instance, the interpretive process involved many time-consuming operations, such as establishing links with knowledge structures connected with the main concepts in the representation. Then, once syntactic processes were finished, most if not all of the work already done by the semantic system would be utilized, given that the syntactic analysis did not contradict the established semantic structure. Of course this idea needs to be specified in much greater detail, with specific proposals about how much syntax is needed to begin semantic analysis, and how the syntactic information feeds into the semantic processes as it is incrementally calculated. However, the main point is that a semantic analysis might be available in the absence of a fully resolved syntactic one.

As a preliminary observation we note that such a proposal seems possible under Jackendoff's (1983) theory of conceptual structures, in which the conceptual interpretation of a sentence is represented in a form similar to a frame, (e.g. Minsky, 1985) indexed on the basis of the main verb. The frame contains slots that become filled by the appropriate lexical items, obtained via rules which link slots in conceptual structure with particular positions in grammatical structure. It is possible that some heuristics could fill some of the slots with items on the basis of an incomplete grammatical analysis. If a
fairly accurate guess could be made before the grammatical analysis was finished, then
clearly the conceptual interpretation would be determined much faster than if the system
had to wait for the completed syntactic analysis.

It is equally important to note that this idea would not necessarily require a change
in the way we see the processes responsible for sentence parsing itself. Syntactic analysis
could proceed independently and autonomously. All that is being suggested is that the
semantic interpretive processes could be initiated before the syntactic analysis had reached
a conclusion. Once a tentative semantic representation is constructed, it could start
looking to the syntactic representation for relevant elements with which to fill any unfilled
slots, and for supporting evidence that speculatively filled slots are in fact correct.

The shallow sensitivity hypothesis, it now seems, has a number of shortcomings.
Yet the basic idea from which it springs, the multilevel race model, has much to
recommend it. It therefore seems important to consider the relative merits of the two
proposals. This thesis sets out to investigate the possibility that the race model can explain
the results of sentence matching experiments, without the need to postulate special
assumptions like shallow sensitivity. It is suggested that the controlling level for the
matching task is in fact a high level interpretive one, and that the sentences which show a
cost in the task are those that do not have useable semantic representations. The reason
why constraint violation sentences are matched just as fast as fully grammatical ones, then,
is that the constraint violation sentences do develop a perfectly well formed and useable
conceptual representation. The hypothesis is that the matching task is able to exploit
representations at the level of global interpretation which lead to the fastest matching
times. Only those sentences which for some reason are prevented from reaching such a
representation must be matched at a less elaborated level of representation which involves
a less efficient and slower procedure.

Clearly, much more needs to be said about what a "useable" semantic
representation is, and about the conditions that determine whether or not such a
representation can be generated. Indeed, much of this thesis is devoted to just that
question. However, we must note that, whatever the way we characterize semantic
factors, there are two important questions that we must be able to answer. First, we need
to give a principled explanation for why some kinds of ungrammatical sentences do not
incur a cost in matching. That is, we must determine if the hypothesis necessarily implies
that a subset of ungrammatical sentences are interpreted as if they were perfectly well
formed sentences, while others are not. If so, why? Secondly, we need to account for why,
with fully grammatical sentences describing implausible events, only some of them incur a
cost in matching. What are the conditions under which we obtain a "plausibility effect"?
The bulk of the experimental data gathered for this thesis deals with the second question,
for two main reasons. First, the principle idea being proposed here is that the conceptual
level is key to understanding the sentences matching task, yet very little existing work
addresses this issue directly. Most existing experiments have been devoted to investigating
the effects of syntactic structure. This has left us with few concrete ideas about the role of
the interpretive level, making it necessary to address this imbalance. Secondly, if indeed
conceptual structures are of such primary importance, then it will only be possible to understand the role of syntax as a part of the whole process of constructing semantic representations. But for this, we need a much more detailed understanding of the semantic factors at work.

To summarize, then, the hypothesis being considered is that the multilevel-race model can account for the results in sentence matching, with the assumption that the task is in fact sensitive to conceptual level representations: either there is, or there is not a useable structure available for a given sentence at the conceptual level. If there is, subjects will exploit it. If there is not, they will be forced to use a lower, less efficient level.

Further, we are looking for a way to characterize in a precise way the differences between sentences that do, and those that do not, have a well formed conceptual representation.
Chapter 2

SEMANTIC ANOMALY AND THE MATCHING TASK

The motivation behind the first experiment was to see if indeed there was something to the idea that the so-called “plausibility” effect in sentence matching was really only found for sentences which could not be interpreted by the normal comprehension mechanisms. If this were true, then we could perhaps go on to make a case that even effects of ungrammaticality could be attributed, at least in part, to conceptual factors.

Perhaps the most obvious starting point is to consider semantic anomaly (e.g. Lyons, 1995; Katz, 1972), and to see if this can be distinguished from implausibility. As a first approximation, perhaps implausible sentences are, roughly, perfectly interpretable sentences which describe events that are not likely to take place in the world as we know it. Anomalous sentences, on the other hand, are about things which are in some sense not possible. It is the “in some sense” which must be clarified. Unfortunately, it is difficult to define the distinction operationally, since previous discussions on the topic have not been conclusive. Lyons (1995) suggests a number of tests for anomaly, of which we consider the most relevant one presently. To illustrate, consider the following sentences.

(25) a. Andrew struggled through the tropical jungle for days.
    b. Andrew struggled through the concrete jungle for days.
    c. Andrew struggled through the surgical jungle for days.
    d. The barber disliked cutting hair with blunt scissors.
e. The barber disliked cutting hair with diamond scissors.

f. The barber disliked cutting hair with early scissors.

Intuitively, the (b) and (e) sentences, while not very plausible, seem to have meanings just as clear as (a) and (d). On the other hand, (c) and (f) have interpretations which are much less clear. The intuition we are trying to capture is reflected by Lyons (1995) who suggests verifiability as a test for anomaly. Roughly, this criterion refers to how one might ascertain the truth value of a proposition. This involves knowing what a model of the world would need to be like for the proposition to be true. It is fairly obvious how one might verify the truth of "concrete jungle", but not of "surgical jungle". Of course, the human mind being very resourceful, some interpretation can be forced on virtually any utterance after sufficient creative inference and "analogy", which makes it difficult to invoke verifiability as a robust test for anomaly. This has prompted many thinkers to abandon a qualitative distinction between anomalous and implausible sentences. For instance, Lyons (1995) quotes three famous utterances, Colorless green ideas sleep furiously (Noam Chomsky), Quadruplicity drinks procrastination (Bertrand Russell), and Thursday is in bed with Friday (Gilbert Ryle), which he points to as being "meaningless". Yet, he says that all of these are interpretable, if they are appropriately contextualized, and the meanings of their words extended beyond their normal use. If this claim is taken literally, then it would seem that the difference between anomalous and implausible sentences is merely quantitative. One of the most ambitious goals in this thesis, then, is to show that in fact a qualitative difference exists, and that anomalous sentences are
processed differently from implausible ones because they differ in the way they are represented in the mind.

Even though the literature has provided some useful hints about the sort of distinction that could be studied empirically, there were no previous proposals that bore directly on the hypothesis being investigated. Thus, the first attempt at classification of items was based on somewhat intuitive grounds, guided by the notion that anomalous and implausible sentences differed in the intuitive ease with which they could be verified as being "true" or "false", and how "impossible" the situation seemed to be. In addition, the intuition was guided by the sorts of examples various authors give of anomalous sentences, such as the ones above, and the examples given below.

(26) a. a sad book .. (where sad is interpreted as the emotion felt by a person) (McCawley, 1968)

b. The square root of Milly's desk drinks humanity. (Chierchia, 1993)
c. The fact that cheese is green skipped inadvertently. (Chierchia, 1993)
d. My toothbrush is blonde and buxom. (Chierchia, 1993)
e. My red afterimage is waterproof. (Katz, 1972)
f. Prepositions feel oily. (Katz, 1972)
g. That triangle is in love. (Katz, 1972)

As Chierchia and McConnell-Ginet (1993) point out, it "... may be that judgments of anomaly pick out a somewhat heterogeneous set of expressions, some of which are simply
contradictions ..., others of which describe situations that are bizarre because of how the world works, and others of which involve a kind of semantic incompatibility other than that of contradiction ..." Yet, as imprecise as this characterization is, the concept of anomaly seems to pick out just the kinds of sentences that are required for testing the hypothesis that only sentences without a useable semantic representation are slowed in the matching task. Katz (1972) defines anomaly as follows:

"A constituent is semantically anomalous just in case it is assigned no readings (the set of readings assigned to it is null)." (p. 49)

Thus, a set of anomalous sentences was constructed, on the basis that each sentence somehow described an impossible state of the world. On the other hand, implausible sentences described bizarre but possible events, and plausible sentences were both possible, and not unusual or odd in any way. A full list of the items appears in appendix C.

In order to test the items, the delayed sentence matching task was used. Thus far, we have discussed the same-different (SD) matching task as if the two input strings were always presented simultaneously. However, this is just one of many possible presentation methods. Many experiments use the delayed same-different matching task, in which the two strings are presented one at-a-time. The first string is typically displayed for two seconds, after which it is followed by the second string. The first string remains on the display even after the second one is presented. This method of presentation is useful
because the subject is made familiar with a string before he/she is asked to perform the actual task. Thus, any oddities due to ungrammaticalities or implausibilities involved in the actual string are encountered on the initial exposure. This minimizes the possibility that matching times for "odd" strings is due solely to the subject's surprise at being shown an unusual item. For many experiments, it does not seem to make much difference which method is adopted. However, there are other manipulations where the presentation method does affect the outcome, and it is these issues to which we now turn in order to determine the most suitable way to display the items in experiment 1.

Stevenson (1984) conducted a delayed sentence matching task, but with a presentation method that was slightly different from that used in past studies. Recall that the pre-exposure to the sentences in these experiments is supposed to ensure that the subjects have seen the item and gotten used to the fact that they might be semantically odd, before they need to make a response to it. But of course the experimenter has no real control over what the subjects are doing when they see the first sentence, since no response is required. There is no way of knowing that subjects are in fact reading the initial sentences completely, before the second one is presented. They might, for instance, learn to ignore the first sentence if they think it is not useful to read it prior to the presentation of the second sentence. Stevenson's experiment was an improvement over this situation, since subjects were required to make a response to the first sentence as well. They were told to read the first sentence quietly to themselves, and when finished, to press a response button which would bring up the second one. This had the advantage of
ensuring that the sentences were read prior to the matching task. In addition, this made it possible to collect reading times for the same sentences that were involved in matching. Stevenson (1984) ran this experiment with the item set from his earlier study, which tested the effects of subjacency constraint violations. He found, as in his earlier study, that the matching times were unaffected by the constraint violation. This was fortunate for the previous research efforts, since it showed that the change in procedure did not affect the results. However, he found a significant effect of sentence type in the reading time, with constraint violation sentences taking significantly longer to read than normal grammatical sentences. This was particularly interesting, and it was claimed to provide strong evidence for the claim that sentence matching, and only sentence matching, was controlled by a relatively shallow level of sentence processing.

In developing the idea being proposed in this thesis, we are forced to interpret Stevenson's result differently. Recall that the hypothesis is that the controlling level for sentence matching is in fact a highly developed interpretive one, and that constraint violation sentences are matched as quickly as their fully grammatical controls because they are both represented equally well at this level. But why does it take longer to read constraint violation sentences? Forster's (1987) hypothesis can account for the finding by stipulating that reading times are controlled at a higher level than the matching task. However, the current hypothesis cannot resort to this explanation since there is supposed to be no difference between the two types of sentences at this high level of representation. In order to explain the result, it is necessary to think of two different stages of processing
involved with the interpretive level of representation as it is normally used in sentence comprehension. First, it is necessary to construct the representation. Then it can be used for other purposes (like making inferences, etc.). We can then stipulate that semantic implausibility and anomaly have different consequences for these distinct processes. Semantic oddness of any sort can slow down the initial construction stage of processing. (This claim is not made without precedent. For instance, such a suggestion was put forward by Forster & Ryder (1971), in order to explain the effects of plausibility in the Rapid Serial Visual Presentation (RSVP) task, which we will be discussing in greater detail later, and by Watson, 1976). The idea is that in order to construct a valid interpretation for constraint violation sentences, as well as semantically implausible sentences, a number of operations need to be performed which are not necessary for grammatical, plausible controls. The speculative linking of a lexical item with a conceptual "slot", as suggested earlier, might be one such operation. This extra processing would increase the overall time taken to process the sentence, and would result in an increase in reading times for constraint violation sentences as found by Stevenson. But once the sentence is processed, the representation is in place, and is perfectly useable in the sentence matching task. Thus, there is no cost associated with the constraint violation sentences in SD matching because they do have a semantic interpretation.

This rather straightforward hypothesis has two very clear predictions: (1) implausible sentences should slow down all, and only, tasks which include an element of sentence interpretation (where the matching operation does not involve re-interpretation),
and (2) anomalous sentences should not only slow down these same tasks, but also those that need to make subsequent use of the representations. Recognizing the proposed distinction has very important implications for sentence matching experiments which involve semantic plausibility as a variable. It suggests that we must be very careful to treat as independent variables the factors of (a) method of presentation, and (b) the exact nature of the semantic implausibility. Any experiment which uses a particular method of presentation, or confounds the different types of semantic deviance, is at risk of being misleading.

It is these considerations which motivated the first experiment. In order to demonstrate that the current proposal has some merit it was important to show that there is indeed a difference in sentence matching performance for syntactically matched sentences that differed just in the nature of the semantic deviance. The experiment used Stevenson's method to obtain results for both matching and reading. The hypothesis is that implausible sentences will be slower to read, but not to match, than plausible controls. On the other hand, anomalous sentences should be slowed relative to their controls, in both tasks.

**Experiment 1**

**Method**

*Subjects.* Twenty seven, first year undergraduate students from the University of Arizona participated in this experiment. They received course credit for their participation.
Materials and design. The experiment consisted of six experimental conditions, three each in reading and matching. The first was a set of anomalous sentences, the second were implausible, and the third, plausible.

There were 16 items in each condition, which were constructed according to the following criteria.
Anomalous: these were sentences for which an interpretation was impossible, given our knowledge of the world. They described situations which simply could not happen in our world, as discussed above. Some examples are:

Ships entered the Socratic harbor at a slow speed.

Last night Helen wore her genetic dress to dinner.

In Chicago Alice bought a desperate camera.

His wife was annoyed at the wealthy typewriter.

After she stood on the slow nail she swore.

Implausible: these sentences described situations which were possible, but highly unlikely to actually happen. Some examples are:

Ships entered the emerald harbor at a slow speed.

Last night Helen wore her ceramic dress to dinner.

In Chicago Alice bought a salty camera.

His wife was annoyed at the concrete typewriter.

After she stood on the hairy nail she swore.
Plausible: the plausible sentences were perfectly coherent sentences which described events that were not unusual in any way. These sentences served as a control, since they had coherent and sensible semantic representations.

- Ships entered the shallow harbor at a slow speed.
- Last night Helen wore her elegant dress to dinner.
- In Chicago Alice bought an expensive camera.
- His wife was annoyed at the noisy typewriter.
- After she stood on the rusty nail she swore.

The sentences were counterbalanced across three groups, such that each sentence appeared in each condition, but any given subject saw a sentence in only one condition. Each sentence therefore served as its own control, allowing us to test the effects of plausibility independently of variations in grammatical form.

The items were given to a group of 20 subjects before the experiment, who were asked to rate them according to how “strange each sentence was, or how much sense it made” (see appendix B) for the instructions. The ratings were made on a seven point scale where 7 was “very strange” and 1 was “not at all strange”. Ratings were taken to ensure that the sentences were subjectively perceived to differ in plausibility.

In addition to the same items, an equal number of different items were constructed, by changing one word in the second member of each pair. The word which
was different was in a randomly chosen place in the sentence, and matched in length to the
word it replaced. While it is the same sentences that are theoretically interesting, the
different sentences also included all three semantic conditions, to make sure that
semantically odd sentences had the same chance of being different as of being the same.
This way, subjects had no way of predicting the correct response without actually
comparing the two sentences.

Procedure. Subjects were first given instructions about the task. They were told that they
would first see a sentence that they needed to read to themselves, and then make a
response to indicate that they had finished. This response would bring up a second
sentence. Then they would need to decide if the two sentences were identical or not, and
to register their answer with a button press. They were told to make sure that they read
the first sentence completely before pressing the response button, and that they should do
the task as quickly as possible, without making too many mistakes.

The subjects were then seated in front of a computer display in an experimental
booth. In front of the monitor was a response box with two buttons, and there was a foot
pedal on the floor. Subjects were required to press the pedal to initiate each trial. Each
trial was as follows. First, a sentence appeared on the screen. When the subject finished
reading this sentence, they were required to press the right hand side “yes” button, on the
box in front of them. The second sentence was then displayed immediately following the
button press, with the first sentence also left on the display. The subject then pressed the
“no” button if they thought the two sentences were different, and “yes” if they were the
same. Both sentences then disappeared, and the subject initiated the next trial with a press of the foot pedal.

Trials were presented in random order, and the entire experiment lasted approximately fifteen minutes.

Results

In this, as in subsequent experiments, the effect of outliers was curtailed by trimming them to the cutoffs, established at 2 standard deviations above and below the mean of each subject. The results of the experiment were analyzed in a mixed design analysis of variance with one between-subjects factor (subject group) and one within-subjects group (sentence type). The results for the reading time and the matching time were analyzed independently, with planned pairwise comparisons. A separate item analysis was also carried out, and combined with the subject analysis to give minF* values according to the procedure recommended by Clark (1973). A number of authors have criticized the minF* statistic for being too conservative (e.g. Wike & Church, 1976), but Santa, Miller, & Shaw (1979) suggested that it could be used effectively with \( \alpha = 0.1 \). We therefore adopt this criterion here, but also give the subject and item analyses (F1 and F2) separately when minF* is not significant.

The rating task was analyzed with pairwise t-tests. The results of the rating task is given in Table 1.
Table 1. Mean ratings, on a seven point scale, for the items in the three conditions in experiment 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>Last night Helen wore her genetic dress to dinner.</td>
<td>4.99</td>
</tr>
<tr>
<td>Implausible</td>
<td>Last night Helen wore her ceramic dress to dinner.</td>
<td>4.27</td>
</tr>
<tr>
<td>Plausible</td>
<td>Last night Helen wore her elegant dress to dinner.</td>
<td>2.31</td>
</tr>
</tbody>
</table>

Pairwise t-tests revealed a significant difference between the anomalous and the plausible conditions, \( t(34) = 7.85, p < 0.05 \), and between the implausible and plausible conditions, \( t(34) = 5.12, p < 0.05 \). However, there was no difference between the anomalous and the implausible conditions, \( t(34) = 1.82, p > 0.05 \).

Table 2 shows the results for reading and matching in experiment 1.
Table 2. Reaction Time for Reading and Matching (and error rate for Matching) in experiment 1. Also shown is the difference between each experimental condition and the plausible baseline.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Diff.</th>
<th>RT</th>
<th>Err.</th>
<th>Diff. (RT)</th>
<th>Diff. (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>2042</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td>Implausible</td>
<td>2059</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plausible</td>
<td>1987</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Err.</th>
<th>Diff. (RT)</th>
<th>Diff. (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>2028</td>
<td>3.9</td>
<td>87</td>
<td>1.6</td>
</tr>
<tr>
<td>Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implausible</td>
<td>1979</td>
<td>3.0</td>
<td>38</td>
<td>0.7</td>
</tr>
<tr>
<td>Plausible</td>
<td>1941</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although the numerical differences in reading time were in the predicted direction, none of them were significant by the min\(F^*\) statistic. Only the difference between the implausible and the plausible conditions showed any effect at all, in the subject analysis, with \(F(1,24) = 7.05, p < 0.05\).
Looking at matching times, the anomalous sentences differed from the plausible ones, $\text{min}F^*(1,68) = 3.42, p < 0.1$, $F1(1,24) = 8.6, p < 0.05$, $F2(1,45) = 5.68, p < 0.05$. The anomalous and implausible conditions did not differ significantly, nor did the implausible and plausible conditions.

There were no differences in the error rates.

**Discussion**

The results of the reading part of the task are suggestive, and showed a numerical difference in the expected direction, but were not statistically significant. This suggests that, contrary to the claims by Watson (1976) and Forster & Ryder (1971) the hypothesis that semantically unusual sentences take longer to process, is not true. However, the analysis by subjects in the implausible condition was significant, showing that some sub set of items might have shown the expected effect. This raises the possibility that, since the reading part of the experiment was not primary in importance, the subjects might have decided not to expand too much effort in its performance. Thus, they might have rushed through the sentences with very unusual meanings, not bothering to wait until they had completely understood it. This might have happened less with the implausible sentences than the anomalous ones, since some of them at least had a reasonable interpretation, which they tried to figure out. While we are not in a position to evaluate this possibility, we caution against the interpretation of the reading data in further experiments.

The matching results, on the other hand, are very encouraging for the theory being proposed. It clearly shows that sentences which are merely *implausible* do not lead to a
performance decrement, relative to plausible, in the sentence matching task. That is, implausible sentences do not differ significantly from plausible ones in the matching task, in spite of the significant differences between their subjectively rated level of plausibility. This is important since it shows that a within phrase implausibility does not necessarily lead to a cost in matching, contrary to the claim by Forster (1987).

However, implausible sentences show a tendency to incur a cost in reading, showing that the subjective ratings are reflected in performance with at least one on-line task. This shows that the rated difference in plausibility was strong enough to have consequences for some level of psychological processing. However as already noted, we need to show a stronger result before any definitive conclusions can be reached on this point. By hypothesis, then, the matching time is unaffected for implausible sentences because a clear and useable conceptual representation is available for them, whereas reading time is effected because it reflects the initial effort needed for interpreting the sentences. On the other hand, anomalous sentences suffer a cost for matching, because there is no useful representation that can be constructed for them.

One shortcoming of the results, however, is that the mean matching time for the implausible sentences did not differ from the anomalous ones. Yet, according to the hypothesis, the anomalous sentences should have differed from the other two by an equal amount. This would also have been important to show the dissociation between rated plausibility of sentences, and their effect on the task. One way to explain this shortcoming is that, due to the relative imprecision of the selection criterion for the sentences, some of
the implausible sentences were in fact sentences for whom no clear representation was available, or, conversely, that some of the putatively anomalous sentences were implausible.

The result that sentences which express propositions that are merely implausible in the real world, do not take longer to match that plausible sentences, has important consequences for previous hypotheses. Recall that in order to maintain the shallow sensitivity hypothesis, Forster (1987) had to explain plausibility effects in terms of a "local" phenomenon that was insensitive to global interpretive processes. Thus, he showed that a sentence which contained the implausible phrase "...timber shirt..." was matched slower than one with the plausible "...cotton shirt...", but the implausible "... shirt ... made of timber..." was no slower than "... shirt ... made of cotton ...". However, the present study shows that sentences like "... timber shirt ..." are not slower to match than their plausible controls. So, how could Forster obtain his result? It is possible that the local effect obtained by Forster was due to a sub set of items in his experiment that were in fact anomalous by accident, rather than design. This would reduce the size of the item effects, but could still yield significant results for subjects. But then we need to account for why he obtained a difference between the local and the global conditions. That is, the subset of items which were, by hypothesis, responsible for the obtained difference, should have been anomalous in both constructions, and there should have been no difference between the local and global conditions. One possibility is that the sentences were in fact more interpretable in the global condition. That is, the syntactic structure of the global sentences
might highlight the connection between the NP and its modifier more clearly than that of the local sentences, thereby facilitating the task of interpretation (a similar suggestion is made by Ratcliff, 1989). This is important because, if it could be shown that globally anomalous sentences were in fact slower to match than plausible ones, then the shallow sensitivity hypothesis would be left without an explanation for the plausibility effect.

In order to test this possibility, we need a set of sentences for which highlighting the relation would not help. That is, even with the Adj-be-NP syntax, these sentences should still have no interpretation. A set of such sentences can be constructed from those in experiment 1. In that experiment, the anomalous sentences were created by combining a noun and a modifier that simply do not belong together. Thus, it might be reasonable to expect that no amount of grammatical help would lead to an interpretation, since there is no conceptual connection between the words. If this is true, then we should in fact be able to obtain global "anomaly" effects with our sentences. This is a powerful prediction for the current hypothesis. If indeed we could obtain global effects for the anomalous sentences, which are otherwise grammatically well formed, then this would provide strong support for the hypothesis that high level interpretive processes do factor in the sentence matching task. Experiment 2 investigated this possibility.

Experiment 2

Method

Subjects. Twenty seven, first year undergraduate students from the University of Arizona participated in this experiment.
Materials and design. The experiment consisted of six conditions, as in experiment 1. That is, three levels of semantic meaning (anomalous, implausible, and plausible) were tested in the reading and the matching components of the task.

The sentences were constructed by modifying the syntactic structure of the original sentences. That is, each sentence of the AP/NP form was converted into an NP-be-Adj form, as the examples illustrate. For many sentences, it was necessary to add lexical items to make the sentence grammatical in the new construction. However, the major lexical items were retained. Most importantly, in each case the source of the implausibility was not local. The sentences below illustrate the nature of the items in each condition:

Anomalous:

Snakes which are written are a serious hazard.
The cheese which Simon refused was very syncopated.
The tanks for rain water were quick all summer.
The jungle in Brazil is surgical and dense.
The cliff which the climbers scaled was almost transmitted.

Implausible:

Snakes which are walking are a serious hazard.
The cheese which Simon refused was very elastic.
The tanks for rain water were mirrored all summer.
The jungle in Brazil is tiny and dense.
The cliff which the climbers scaled was almost glass.
Plausible:

Snakes which are poisonous are a serious hazard.
The cheese which Simon refused was very smelly.
The tanks for rain water were empty all summer.
The jungle in Brazil is tropical and dense.
The cliff which the climbers scaled was almost vertical.

As in the previous experiment, there were 16 sentences per condition. In addition, an equivalent number of different sentences were constructed, as in experiment 1.

Procedure: The instructions given to the subjects, and the experimental method was identical to experiment 1.

Results

The table below shows the results for experiment 2.
Table 3. Reaction Time for Reading and Matching (and error rate for Matching) in experiment 2. Also shown is the difference between each experimental condition and the baseline.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>2156</td>
<td>108</td>
</tr>
<tr>
<td>Reading</td>
<td>Implausible</td>
<td>2120</td>
</tr>
<tr>
<td></td>
<td>Plausible</td>
<td>2048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Err.</th>
<th>Diff. (RT)</th>
<th>Diff. (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>1930</td>
<td>3.0</td>
<td>59</td>
<td>1.2</td>
</tr>
<tr>
<td>Matching</td>
<td>Implausible</td>
<td>1904</td>
<td>3.5</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Plausible</td>
<td>1871</td>
<td>3.5</td>
<td>33</td>
</tr>
</tbody>
</table>

The difference in reading time between anomalous and plausible conditions was significant, $\text{min}F'(1,68) = 7.35, p < 0.1$, as was the difference between implausible and plausible; $\text{min}F'(1,65) = 2.81, p < 0.1$. The difference between anomalous and implausible was not significant.

In matching, the anomalous conditions was significantly different from the plausible condition by subjects only, $F(1,24) = 7.59, p < 0.05$. The item analysis was not
significant, $F_2(1,45) = 2.99$, $p > 0.05$. The implausible sentences did not differ significantly from the plausible ones, nor did the anomalous sentences differ from the implausible ones.

There were no significant differences in error rates.

Discussion

Interestingly, the effect of both plausibility and anomaly was evident in the reading time for these items. As predicted in experiment 1, it took longer to read implausible and anomalous sentences than plausible ones. While this supports the original prediction, it raises the question of why the effect was obtained in this experiment, but not the previous one. There are a number of possible explanations, but the most straightforward one is that subjects strategies highlight the difference between the two constructions. In the AP/NP sentences, the semantic incongruity arises, on average, at the middle of the sentence. Under speeded conditions (and the fact that the reading task is only secondary), subjects might try and ignore this incongruity, hoping that the sentence would become clear by the end. However in this second experiment, the syntax of the sentences entailed that the semantic incongruity became evident only at the end of the sentence, just before the subject was required to make a response to bring up the second sentence. It is possible that under these circumstances, they hesitate a little longer before moving on, which leads to more reliable effects. As a cautionary note, then, if it is indeed the case that strategic factors enter the reading task, then once again the implication is that the data obtained with the task ought to be interpreted with care.
The results of the matching task support the hypothesis that anomalous sentences in which the anomaly arises from words that are not within the same phrase, do show a cost in this task. However, the result is not as strong as it was in Experiment 1, where both subject and item analyses were significant. The lack of a strong item analysis suggests that only a subset of the anomalous sentences in this experiment were left without an interpretation. One way to account for this, is that the NP-be-Adj syntax allows some additional factor to become involved in matching. Perhaps the general syntactic frame "... x ... was y" can be represented efficiently in a short term memory store, especially if there is some possible connection between x and y. This could enhance matching speed for some items, even though a clear semantic interpretation is not available for them.

To summarize, experiments 1 and 2 provided data which was encouraging for the hypothesis under investigation. However, a clear difference between anomalous and implausible sentences has not been demonstrated, which is problematic. In addition, it is apparent in some of the results that the subject analyses are more reliable than item analyses. The implication of this is that, while the experimental manipulation has an effect on subject performance, the items in a given condition are not all acting in the same way. Perhaps this is a consequence of the way the items were constructed. As we noted earlier, semantic anomaly seems to be a concept which covers a heterogeneous set of sentences, and the items in the experiments reflect this fact. The following experiments were designed to refine our definition of anomaly.
Chapter 3

SELECTIONAL RESTRICTIONS AND ANOMALY

The formal treatment of semantic incongruence in linguistics revolves around the notion of selectional restrictions (Chomsky, 1965; Katz, 1972), which is a way of imposing co-occurrence restrictions between words. In effect, these restrictions impose a set of conditions which ensure that only appropriate words are inserted into syntactic frames. This is important when a choice needs to be made between different readings of ambiguous words. These restrictions can therefore prevent lexical items from being inserted into inappropriate slots in a sentence. In such a case, the sentence fails to be assigned an interpretation (Katz & Fodor, 1967).

Selectional information is specified within lexical items by a set of semantic features (Chomsky, 1965; Katz, 1972), which determine what other sorts of lexical items are acceptable in combination. Theories differ on the exact details of these features. However, in order to maintain explanatory adequacy for the possible referents of words, all theories have to, at some level, capture high level distinctions like animate/inanimate, and concrete/abstract. More specific proposals include male/female, artifact/natural kind (Katz, 1972). To illustrate with an example how semantic features can be used to enforce selection restrictions, consider the sentences below:

(27)  a. The boy may frighten sincerity.

b. Colorless green ideas sleep furiously.
The deviance of (27 a) can be explained in terms of the selectional restrictions imposed by the verb *frighten*, and the semantic type of the NPs *boy* and *sincerity*. *Frighten* is specified as permitting an *abstract* subject and an *animate* object. Since *sincerity* is an *abstract* term (denoted [+Abstract]), and *boy* is [+Human], (25) is well formed. However, *boy* is [-Abstract] (concrete), and *sincerity* is not marked as [+Human], so that combination will not be permitted. The second example (28b) is similar. Note especially the restriction relation between *green* which modifies a [+Concrete] term, and *ideas*, which is [-Concrete] (or, alternatively, [+Abstract]).

Katz (1972) detailed one of the most influential semantic theories based on semantic features, in which the meaning of an expression is determined through a combination of the semantic features of its elements. For instance, the word *bachelor* might have a representation like the following:

*bachelor*: (Human) (Male) (Adult) (Unmarried) .......

Katz also proposed a number of *redundancy rules* which complete the specification of words. For instance, the following redundancy rule would assign the semantic marker *animate* to any item that contained one of the markers *human*, or *primate*, or *canine*, an so on.

(Human) v (Primate) v (Canine) v (Feline) .............. => (Animate)
So, the word *bachelor* which explicitly includes the semantic marker (Human) in its lexical entry, would be augmented with the features (Animate), (Living), and (Physical Object).

The job of combining semantic features of words whose relation to one another is given by the syntax of the sentence in which they appear, is performed by a set of *projection rules*. This task is complicated in natural languages by the fact that words can be ambiguous, so any given word could be represented by different sets of semantic markers. In establishing interpretations, it is therefore important to select the correct meaning of the word. The projection rules therefore make use of selectional restrictions specified in lexical entries, to combine the correct sets of semantic markers. The inappropriate reading is rejected because it clashes with the selectional restrictions. When there is such a clash between semantic markers, the application of the projection rules is blocked for this reading, and no interpretation is produced. One set of commonly cited examples of this type of semantic clash involve word pairs in which an attribute that is appropriate to an animate individual is *predicated of an inanimate object*. We have already seen some examples of these, in the introduction to experiment 1. For instance, the feature [+Animate] which is specified in adjectives and verbs dealing with psychological states can clash with the [-Animate] feature in words referring to inanimate things, as in the example involving the triangle which was in love. This way of looking at the process makes it clear that the role of selectional restrictions is to prevent certain lexical items from being
incorporated with the representation of certain sentences, which is precisely the sort of
tonight we are pursuing in the current set of experiments.

The sentences in experiment 1 contained a variety of semantic violations, to
produce implausible and anomalous sentences. In order to refine the experimental
manipulation, it was decided that the kinds of violation involved in any one condition need
to be controlled. The items in experiment 1 were therefore analyzed, to better categorize
the kinds of restrictions that were manipulated. The method used for this purpose involved
the enumeration of the sorts of complements each of the adjectives can legally occur with
(e.g. the plausible condition), and noting if these had a specific semantic feature in
common. If there was such a feature, then the sentences from the experiment which
contained these adjectives were looked at, to see if they did in fact violate a restriction on
the feature. This methodology is similar to Chomsky (1965), who concluded that 
frighten
but not 
admire

but not admire can take a [+Animate] subject from sentences like Sincerity frightens the
boy, The girl admires the boy and ?Sincerity admires the boy. Let us look at a few
examples.

Consider the anomalous sentences given in the Methods section of experiment 1

(28) a. Ships entered the Socratic harbor at a slow speed.

b. Last night Helen wore her genetic dress to dinner.

c. In Chicago Alice bought a desperate camera.

d. His wife was annoyed at the wealthy typewriter.

e. After she stood on the slow nail she swore.
The adjective *Socratic* in the first sentence can appear with NP's that refer to intellectual activities. That is, *Socratic argument/questions/ideals/writings/claims* are valid constructions. These are all abstract notions to do with concerns of the mind. This adjective appeared in the *anomalous* condition in the experiment, as shown in (28a), where the NP refers to a concrete geographical location. The violation is therefore one involving the [Abstract] feature. The second example is similar, with *genetic* appearing legally in *genetic code/engineering/disease/material*. Clearly these too are abstract concepts. (Perhaps *material* is slightly different in that it refers to some concrete substance. However, it still does not refer to an individual entity, but a kind of “stuff”).

The next two examples (c and d) are similar to each other, but different from a and b. Here, an adjective normally predicated of an animate individual is being used on an inanimate object. Sentence c., in particular, is odd in that it attributes psychological properties to an inanimate object. Example e. is similar again, though not as clear cut a violation. The adjective *slow* normally appears with an object capable of self propelled motion, or that motion itself. However, in sentence e., the NP *nail* is not such an object. Thus, a feature that is not relevant to *nail* is being attributed to it.

Now look at sentences from the *implausible* condition.

(29) a. Clive wanted to read the furry contract before signing.

b. Hikers believe that walking snakes are a hazard.
c. To reach the summit they scaled a glass cliff.

d. Yesterday Ruth ate a frosty apple for lunch.

e. John invited the electronic girl to the party.

In example a., a quality which normally applies to an animate entity is being predicated of an inanimate object. Yet it is different because the modifier *furry* can also appear, under certain circumstances, with inanimate objects (*e.g.* *furry coat*). So, unlike in (28c and d), the violation here seems one of a type of material rather than a mixing of concepts which apply to animate and inanimate objects. In example 29b, there is no violation as such, since *walking* is canonically predicated of animate entities, which *snake* certainly is. Instead, a physical ability which just happened not to evolve in *snakes* is attributed to them. Example c is a more straightforward case of a violation in the kind of material that something is normally made of. Sentences d and e are also similar, in that some physical quality of the NP is unusual.

Overall, both conditions seemed to be a mixed bag in terms of the sorts of semantic violations that appeared. This is consistent with the notion that the original selection criterion missed the "real" distinction. Yet, there does seem to be more similarity between items within a condition than between them, which suggests that the initial item selection at least partially reflected the correct distinction. In order to test this hypothesis, the sentences in experiment 1 were categorized according to the criterion just described. Table 4 shows how many violations of the type just described were evident in each of the conditions. Note that the abstract/concrete distinction was not always a straightforward
one. To see why, consider a sentence like *Those old tertiary songs can make Jan cry*. This sentence is difficult to interpret because the adjective selects for a very specific sort of complement, and not just an "abstract" one. For instance, *tertiary* can be used in circumstances where it makes sense to grade the noun it is modifying according to some acceptable scale, where the scale is implied by the noun itself. Thus, *tertiary education, ... of tertiary importance ...* are acceptable because the word which is being modified is able to be used in the context of a scale of some sort. That is, *education* comes at many levels, as does *importance*. On the other hand, *songs* are just songs. In order to find an interpretation, we would have to make an analogy with the other cases, and embed the phrase in a context where, perhaps, songs were being compared. However, the point is that *songs* is an abstract concept, yet is still not acceptable, showing that the abstract/concrete distinction is not restrictive enough in this context. The table lists such problematic cases as "other".
Table 4. The semantic feature violations observed in items used in experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Anomalous</th>
<th>Implausible</th>
<th>Plausible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract/Concrete</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Animate/Inanimate</td>
<td>10</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Material</td>
<td>1</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

The table shows that there is in fact a difference in the kinds of violations that are involved in the three conditions. Selectional restrictions of the type standardly considered play a much greater role in the anomalous sentences than the implausible. The converse is true of inappropriate modification of a material. This raises an interesting possibility. The implausible items in experiment 1 and 2 consisted mainly of sentences which violated a standard expectancy on the sorts of material that a physical object is normally made of, such as Forster's example of a *timber shirt*. Yet, these sentences were matched just as quickly as a plausible sentence. On the other hand, the sentences which did incur a cost in processing were primarily ones which violated a selectional restriction of the type described in linguistic theory. Perhaps this distinction reflects a difference between a
merely *pragmatic* constraint, and a set of deeper, syntactically relevant linguistic constraints. And, most interestingly, the sentence matching task might be sensitive to this difference.

**Experiment 3**

This experiment was designed to directly test the effect of selectional feature violations on the sentence matching task.

As was argued earlier, the violations in which the animate/inanimate feature is being violated seem much more clear cut than the abstract/concrete violations. As such, the initial experiment designed to investigate the effect of violating selectional restrictions concentrated on this sort of violation. The following experiment, then, was designed specifically to contrast sentences which contained the clear cases of selectional restriction violations with implausible sentences that did not violate any such restrictions. These latter sentences described implausible, but completely possible scenarios, where the implausibility came from combining an object with an unlikely material from which it was made. In this experiment, two plausible controls were used. The first of these involved an adjectival modifier of the NP, as in the previous experiments. The second was a plausible scenario produced by combining an NP with a prototypical material from which it could be made. This was thought to be a more closely matched baseline for the implausible sentences, and provided a way of checking that plausible sentences involving Adj-NP had the same reaction times as the predominantly NP-NP sentences used in the implausible condition.
Method

Subjects. Twenty eight, first year undergraduate students from the University of Arizona participated in this experiment.

Materials and design. This experiment consisted of eight conditions in all, with the four experimental conditions being tested in reading and in matching. The four conditions are shown below, together with an example for each. (A complete list of materials is given in appendix C). There were 12 sentences in each condition.

Anomalous (Selectional restriction violation)

Finally John was satisfied with his new thoughtful pillow.

Implausible (unusual construction material)

Finally John was satisfied with his new concrete pillow.

Plausible (common construction material)

Finally John was satisfied with his new feather pillow.

Plausible (adjectival modifier)

Finally John was satisfied with his new comfortable pillow.

The materials were constructed such that each sentence appeared in each condition, which controlled for the syntactic form of sentences between the conditions. Four lists were constructed and tested with four separate subjects groups. This way, any one subject saw each sentence in only one condition. In addition, an equivalent number of different sentences were constructed, as in experiment 1.
**Procedure.** The procedure for the experiment was identical to experiment 1.

**Results**

Table 5 shows the results for experiment 3.

Table 5. Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and its appropriate baseline.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection</td>
<td>2399</td>
<td>151</td>
</tr>
<tr>
<td>restriction</td>
<td>2292</td>
<td>63</td>
</tr>
<tr>
<td>Implausible</td>
<td>2223</td>
<td></td>
</tr>
<tr>
<td>material</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plausible</td>
<td>2223</td>
<td>63</td>
</tr>
<tr>
<td>material</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plausible</td>
<td>2248</td>
<td></td>
</tr>
<tr>
<td>adjective</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Matching</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection</td>
<td>2101</td>
<td>3.0</td>
</tr>
<tr>
<td>restriction</td>
<td>2077</td>
<td>2.1</td>
</tr>
<tr>
<td>Implausible</td>
<td>2041</td>
<td>3.9</td>
</tr>
<tr>
<td>material</td>
<td>2076</td>
<td>4.2</td>
</tr>
<tr>
<td>Plausible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjective</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In the reading time, the anomalous (selection restriction violation) sentences were significantly slower to read than the plausible AP/NP sentences, $\text{min}F'(1,67) = 8.16$, $p < 0.05$. The implausible and plausible construction material conditions were significantly different by subjects, $F1(1,24) = 5.36$, $p < 0.05$, but not by items, $F2(1,44) = 2.54$, $p > 0.05$. The two plausible baseline conditions did not differ.

In matching, neither the anomalous, nor the implausible sentences took significantly longer than their plausible counterparts. None of the error differences were significant.

Discussion

The reading times show that the selectional restriction violation sentences, and to a lesser extent the implausible material sentences, were processed more slowly than the plausible baseline. This shows that the plausibility manipulation was sufficiently strong to be detected in this experimental paradigm.

The matching times, on the other hand, showed no significant effects at all. The experiment failed to support the hypothesis that violations involving the animate/inanimate distinction can produce sentences which show a cost in the sentence matching task.

To further study the effects of selectional restriction, then, it seems that we must return to considering the more problematic distinction between abstract/concrete entities. We have already noted that this was a somewhat problematic distinction, because it often involves violations of very specific and idiosyncratic conditions. But consider what this might entail. Such violations often involve the contradiction of a condition which is
fundamental to the definition of the words in question, or at least incompatible with it, as in *colorless green*, *tertiary song*, and *Socratic harbor*. Perhaps these cases are special not because they violate certain semantic requirements, but because the way in which the words are meant to be connected is unclear.
Chapter 4

ANOMALY AS A LACK OF CLARITY

An experiment reported by Ratcliff (1989) argues along lines that are very similar to the previous suggestion. She contrasted a set of sentences she called *clear* with another set she called *unclear*. She constructed plausible and implausible versions of both of these sentence types, giving her a four way contrast. Perhaps the best way to appreciate the distinctions is with the following examples. The first two show the plausibility contrast for *clear* items, and the second two for *unclear* ones.

(30) a. He was given a lace handkerchief.
    b. He was given a brass handkerchief.

(31) a. Fruity syrup splashed onto the table.
    b. Architectural syrup splashed onto the table.

The relevant contrasts were (30a) vs. (b), and (31a) vs. (b). She found a highly significant effect in the *unclear* condition, with implausible sentences taking 77ms longer to match than plausible ones. However, in the *clear* condition, there was a non-significant 5ms trend in the wrong direction. These items, then, appear to demonstrate the clear difference between two different types of implausible sentences, which may reflect the implausible/anomalous distinction which is under investigation in this thesis.

Unfortunately, her experimental design prevented the clear and unclear sentences from being directly compared. Her items were constructed such that the any given sentence
participated only in the clear or the unclear condition, but not both. That is, the clear implausible sentences were compared with a set of matched plausible sentences, and the unclear implausible ones were compared against a different set of plausible sentences. As the above examples illustrate, there is no suitable direct comparison between unclear and clear sentences, as would have been the case if the set included the sentence *He was given an architectural handkerchief*. This design also raises the possibility that the obtained effects were entirely due to the differences between the two plausible sentences used as baseline. Thus, the experiment needed to be replicated, with a fully counterbalanced design in which every sentence appeared in each condition. This way, item specific effects can be factored out fully. In order to do this, however, some new adjectives had to be found so that any given sentence would contain an adjective-noun pair which made it either plausible-clear, implausible-clear, or unclear. To maintain compatibility with the terminology of the earlier experiments, these conditions will be referred to as plausible, implausible, and anomalous

**Experiment 4**

Experiment 4 was a modified version of Ratcliff’s (1989) experiment which demonstrated that implausible sentences that had a clear interpretation were matched as efficiently as plausible sentences. In this experiment we also hope to demonstrate a strong difference between the clear and unclear sentences, by using an improved design.
Method

Subjects. Twenty seven, first year undergraduate students from the University of Arizona participated in this experiment.

Materials and design. The items were taken from Ratcliff (1989), and modified to allow a fully counterbalanced design. That is, each sentence appeared in each of the conditions, in different files. This necessitated the construction of some new AP/NP combinations. For instance the pair of sentences below, taken from Ratcliff’s items,

He was given a lace handkerchief.
He was given a brass handkerchief.

had to be complemented with a third sentence that was implausible and unclear. For some sentences this involved a shuffling of the original items, but for some, new adjectives were found. Altogether, ten new adjectives were used in each file.

There were, then, six conditions in all, with three in each of the tasks. The conditions were called anomalous, implausible, and plausible, to maintain the terminology with the earlier experiments. As in previous experiments, three sets of items were constructed and counterbalanced across three files. The sentences below illustrate each of the conditions. The full set of items appears in appendix C. There were ten sentences in each condition. In the majority of cases, the sentences were taken from Ratcliff (1989). When it was necessary, an adjective was made up for a sentence, attempting to adhere to the notions of clear and unclear as closely as possible.
Anomalous (unclear):

She watched the hyphenated leopard feed the ravenous cubs.
At the theater the audience applauded the syllogistic dancer.
The misaligned fuel stained the concrete floor in the garage.
The existential milk was removed by the health department.

Implausible (clear):

She watched the skating leopard feed the ravenous cubs.
At the theater the audience applauded the purring dancer.
The curdled fuel stained the concrete floor in the garage.
The flaming milk was removed by the health department.

Plausible:

She watched the purring leopard feed the ravenous cubs.
At the theater the audience applauded the skating dancer.
The flaming fuel stained the concrete floor in the garage.
The curdled milk was removed by the health department.

The sentences were also designed to control for word frequency differences that might exist between the different conditions. Thus, a word which appeared in the anomalous condition with one sentence, would appear with the implausible version of another, and
the plausible version of yet another. In this way, the set of words which differed between
the three versions of each sentence, was also counterbalanced across the three files.

In addition, an equivalent number of different sentences were constructed, as in
experiment 1.

Since the clarity of meaning manipulation was new, subjects were asked for
subjective ratings as in experiment 1.

Procedure. The experimental procedure was identical to experiment 1.

Results

Table 6 shows the mean ratings assigned by subjects for the items in each
condition.

Table 6. Mean ratings, on a seven point scale, for the items in the three
conditions in experiment 4.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>Last night at Albertson's, subtracted hinges were on sale.</td>
<td>4.61</td>
</tr>
<tr>
<td>(Unclear)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implausible</td>
<td>Last night at Albertson's, lace hinges were on sale.</td>
<td>3.97</td>
</tr>
<tr>
<td>(Clear)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plausible</td>
<td>Last night at Albertson's, brass hinges were on sale.</td>
<td>1.99</td>
</tr>
</tbody>
</table>
Pairwise t-tests revealed a significant difference between the anomalous and the plausible conditions, $t(34) = 8.1$, $p < 0.05$, and between the implausible and plausible conditions, $t(34) = 5.24$, $p < 0.05$. However, there was no difference between the anomalous and the implausible conditions, $t(34) = 1.65$, $p > 0.05$.

The table below shows the Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the control.

Table 7. Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the plausible baseline.

<table>
<thead>
<tr>
<th>Reading</th>
<th>RT</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous (Unclear)</td>
<td>2382</td>
<td>80</td>
</tr>
<tr>
<td>Implausible (Clear)</td>
<td>2296</td>
<td>-6</td>
</tr>
<tr>
<td>Plausible</td>
<td>2302</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matching</th>
<th>RT</th>
<th>Err.</th>
<th>Diff. (RT)</th>
<th>Diff. (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous (Unclear)</td>
<td>2214</td>
<td>11.9</td>
<td>60</td>
<td>5.2</td>
</tr>
<tr>
<td>Implausible (Clear)</td>
<td>2204</td>
<td>4.8</td>
<td>50</td>
<td>-1.9</td>
</tr>
<tr>
<td>Plausible</td>
<td>2154</td>
<td>6.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
None of the reading time differences were significant with the minF* statistic. The anomalous condition did differ, however, from both plausible and implausible conditions, by items: $F_2(1,27) = 5.25, p < 0.05,$ and $F_2(1,27) = 6.33, p < 0.05,$ respectively.

None of the planned matching time comparisons was significant. However, for the first time, the error difference between the anomalous and the other two sentence types was significant, but by subjects only. The difference between anomalous and implausible yielded $F_1(1,24) = 13.01, p < 0.05,$ and the anomalous, plausible difference, $F_1(1,24) = 10.05, p < 0.05.$

**Discussion**

The results from the subjective ratings was comparable to the ratings in experiment 1. Subjects considered both the clear, and unclear conditions to be equally implausible, compared to the plausible sentences. However, the reading times contrasted with the general pattern of results obtained so far. The implausible sentences did not differ, even numerically, from the plausible ones. The 80ms difference between the anomalous and plausible sentences was not significant by subjects, which could indicate once again that subject variability, possibly due to strategic factors, makes reading time data unreliable.

Unfortunately, we failed to replicate Ratcliff's very compelling result in the matching task. While it took 60ms. longer to match implausible sentences than plausible ones, this difference is 27ms. less than the comparable comparison in experiment 1, and not significant. The difference in error rates is interesting, in that it might reflect the previously suggested strategic effect in reading. That is, if subjects don’t read the
anomalous sentences sufficiently carefully, then they might be more likely to make errors on the classification. That is, since they have not familiarized themselves with the sentence, its anomaly might influence their performance, biasing them to respond "different". However, since this is the first time such an effect has been observed, this observation must be taken as tentative.

In order to determine why we did not replicate Ratcliff’s (1989) result, we must consider the differences between the experiments. There are two obvious differences. The first is the change in the item set, perhaps due to the matching of the controls. This would imply that Ratcliff’s result was unreliable. The second difference was the way in which items were presented. Ratcliff used a fixed delay sentence matching paradigm, in which the first sentence of a pair is displayed for two seconds prior to the second. In the present study, we used a response contingent procedure as before. As previously discussed, this is supposed to give subjects the opportunity to fully process the first sentence, as well as providing data about the time it takes to process the sentences in each of the conditions. However, we have already noted the possibility of strategic effects in the reading time. In this study in particular, the experimenter noticed that several subjects did not read the first sentence at all, and pressed the response button to bring up the second sentence as soon as the first one appeared. Apparently they decided that it was a waste of time to read through the first sentence, since it did not seem so important to the matching task. While every effort was made to dissuade subjects from this strategy, there is no guarantee that the efforts were successful. In fact, the non-significant differences obtained in the reading task
support the claim that in fact subjects did not carefully read the sentences. Clearly this is a problem. First, it lessens the confidence we can have in the results obtained from the reading task. But perhaps more importantly, it might be the reason we did not get strong effects in this experiment. If the first sentence was not processed before the presentation of the second, then the time taken to interpret the sentence would be combined in some unknown way with the time taken to match them. It was therefore decided that the second possibility would be investigated with a change in methodology was called for.

**Experiment 5**

In the introduction, we mentioned several possible psychological mechanisms for how sentence matching is performed. The most favored possibility was that a high level interpretation of the first stimulus is used to check for a match in identity of the second one. It is this checking process which is facilitated by the presence of structure in the stimuli, in one or both of two ways. First, it might enable the retention of a larger "chunk" of the first stimulus in some working memory buffer, which would allow for a more efficient check of the second stimulus. Second, the structure might help because, as previously stated in the multilevel race model, more elaborated levels of interpretation could involve fewer comparisons, resulting in a faster overall process. Of course, structure might be helpful in both of these ways. In any case, it is critical for evaluating the proposed hypothesis to ensure that subjects extract whatever structure they can from the first stimulus before moving to the second. It was therefore decided that a change in methodology was needed to dissuade subjects from neglecting to read the first sentence,
by removing it from the screen as soon as the subject has responded to it. This would force them to finish reading the first sentence if they wanted to be able to do the matching task at all.

Method

Subjects. Twenty four, first year undergraduate students from the University of Arizona participated in this experiment.

Materials and design. This experiment was identical in this respect to experiment 4.

Procedure. Once again, the procedure was identical to the previous experiments, except for one detail. As soon as the subject made their response to indicate that they had finished reading the initially presented sentence, that sentence was taken off the screen, and a second sentence displayed. This sentence was displaced such that it would have appeared beneath the first sentence, had it been there. This was to maintain compatibility with the earlier experiments.

Results

Table 8 shows the results for experiment 5.
Table 8. Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the plausible baseline.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anomalous (Unclear)</td>
<td>2799</td>
<td>65</td>
</tr>
<tr>
<td>Implausible (Clear)</td>
<td>2715</td>
<td>-19</td>
</tr>
<tr>
<td>Plausible</td>
<td>2734</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Err. (RT)</th>
<th>Diff. (RT)</th>
<th>Diff. (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anomalous (Unclear)</td>
<td>2000</td>
<td>4.6</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>Implausible (Clear)</td>
<td>1927</td>
<td>9.17</td>
<td>-34</td>
<td>4.57</td>
</tr>
<tr>
<td>Plausible</td>
<td>1961</td>
<td>4.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 65ms difference in reading time between the anomalous and plausible conditions was not significant. The difference between anomalous and implausible, however, was significant, $\min F'(1,43) = 3.17$, $p < 0.1$. 
In the matching time data, the anomalous and implausible conditions differed by subjects only, $F(1,21) = 5.72$, $p < 0.05$. None of the other planned comparisons differed significantly.

None of the comparisons in the error data differed significantly.

Discussion

The change in presentation conditions eliminated the effect in error rates that was noted in experiment 4. This suggests that subject strategies might have been responsible for that effect, as suggested. Interestingly, the overall error rates are similar to those obtained in earlier experiments, which supports the claim that high level representations of some sort are involved in the matching task, because the visual stimulus was absent.

Surprisingly, however, the change in methodology did not lead to the expected result in matching, since the anomalous sentences did not differ from the plausible ones. However, for the first time in the series, we did obtain a difference (by subjects) between anomalous and implausible sentences in the matching task. This is important for the hypothesis, which stipulates a qualitative difference between the two kinds of sentence. But, we need a more reliable result before any definite conclusions can be reached.

Unfortunately, then, Ratcliff’s items were not as useful as hoped in the quest for a criterion by which implausible and anomalous sentences could be distinguished. The experiments do, nevertheless, show that implausibility does not slow down sentence matching when compared to a plausible baseline. This recurring finding is important, since it contradicts many claims in the literature.
But now we are in a strange position, in that we have shown time and again that highly implausible sentences do not show an effect in sentence matching. Yet there are many experiments in which sentences which seem no more implausible, do show strong effects of plausibility (Ratcliff, 1983; Murray, 1982). There are obviously many differences in the particulars of these experiments, but one general difference stands out. Namely, very few of those experiments involved a modifier/NP construction as used in the experiments in this thesis. Perhaps this particular construction is not the best choice for evaluating the present hypothesis. But why?

Recall that the hypothesis under investigation is that anomalous sentences are not represented at some level of conceptual structure, and therefore are not stored in a way that would facilitate efficient comparison with the second stimulus. The aim thus far has been to find AP/NP combinations which would prevent such a representation. However, perhaps this particular syntactic construction makes it difficult to find such examples. In support of this, Ryder (1994) in her study of NP/NP constructions has suggested a number of alternative strategies which can be used to interpret novel combinations. She points out that the strategies are so powerful that they rarely fail to come up with some interpretation. One of the strategies she proposes involves comparing against a “template” based on an existing representation. A similar operation might take place in the AP/NP sentences used in these experiments, making the selection of items particularly difficult.

For instance, Forster (personal communication) argues that the sentence containing the phrase *surgical jungle* is not anomalous, because it invokes in him an image of a surgery
in which tubes and wires and things holding up human bodies are hanging from the ceiling, much as the leaves and vines from trees hang in a jungle. Perhaps this scene is inspired by the television show M*A*S*H, where doctors run around in the jungles of Korea. In any case, the interpretation seems to be based on a mental model that was acquired through experience with the world. But if the interpretation allows the sentence to have a high level conceptual representation, and if many sentences in the anomalous condition were of this type, then this could eliminate the sentence matching effect altogether. What we need is a set of sentences which minimizes this possibility. Once again, we look to linguistic theory for help.
Chapter 5

LEXICAL CONCEPTUAL SEMANTICS

The semantic relations between various syntactic constituents of a sentence can be looked at in terms of Jackendoff's (1983, 1990) theory of lexical/conceptual semantics, which is a system of conceptual elements and rules that attempt to explain interactions between syntax and semantics. Ultimately, he hopes to explain many grammatical facts that are not easily handled by syntactic analysis, in terms of a semantic/conceptual one. In order to achieve his goal, he proposes that conceptual structure is divided into a number of primary conceptual constituents, and that the rules of conceptual combination are phrased in terms of these constituents. These conceptual constituents form the backbone of conceptual structure by defining the major ontological categories of concepts. A tentative list of these major constituents is: [Thing], [Event], [State], [Action], [Place], [Path], [Property], and [Amount]. These are supposed to pick out all the different sorts of referents in the world that a word could have, and the identities of these constituents is supposed to be inferred directly from linguistic evidence. However, it is clear that the above list is tentative and incomplete since, for instance, Jackendoff suggests that there might be other constituents like [Smell] and [Sound]. Nevertheless, the important claim is that some (small) set of fundamental constituents does exist. It is these constituents which form the conceptual "parts of speech", which parallel the role of syntactic categories, in that they form the basic vocabulary on which the recursive rules of combination operate. The claim is that every permissible concept is generated by the rules which combine these
constituents in various ways. But while the major constituents form the backbone of the system, finer grained descriptions of concepts must be given by more idiosyncratic features, as we will see.

The attempt to identify a relatively small set of primitives which, when combined in a principled way, can account for the full range of expressible concepts, is of course not new. Many attempts have been made to identify a suitable set of primitives (see Fodor, 1983, for an excellent and highly critical summary of this approach). In spite of the as yet unsolved difficulties with the approach, the practical solution of many problems seem to require such a reduction. Schank & Rieger (1985), for instance, outline a system whose aim is to determine the conceptual content of natural language sentences. In order to accomplish this, and to support the inferences the system was required to make, they made use of a set of six conceptual categories in the translation of the sentences. These were: PP (real world objects); ACT (real world actions); PA (attributes of objects), AA (attributes of actions); T (times); and LOC (locations). Of course these, as in Jackendoff's system, are further elaborated to describe the specific situation. Miller and Fellbaum (1992) suggest a set of 26 unique concepts, one of which is supposed to form the most general descriptor of every English noun. These are \{act, action, activity\}, \{animal, fauna\}, \{artifact\}, \{attribute, property\}, \{body, corpus\}, \{cognition, ideation\}, \{natural object\}, \{natural phenomenon\}, \{person, human being\}, \{plant, flora\}, \{possession, property\}, \{process\}, \{communication\}, \{event, happening\}, \{feeling, emotion\}, \{food\}, \{group, collection\}, \{location, place\}, \{motive\}, \{quantity, amount\}, \{relation\}, \{shape\}. 
{society}, {state, condition}, {substance}, and {time}. Each of these heads a hierarchy of concepts which become more specific with each successive level. Every noun, then, has a place in one of these hierarchies. Clearly the three proposals differ in the exact identity of the primitives, and in the generality of each of them. Which is the most successful proposal is of course an empirical matter. However, there is also considerable agreement between the primitives. All three give priority to object, event, property/attribute and action, for instance.

In Jackendoff’s theory, while on the one hand the constituents pick out non-overlapping and independent sets of words, on the other hand, they share some properties which result in some useful computational properties. Formal interpretive processes across the set of constituents can to a large extent be standardized, such that the same process can apply in the same way to any constituent. Jackendoff (1990) ( p.22-41) provides a summary of the properties of the constituents, and the interested reader is referred there. What is important here, is that very strong linguistic reasons exist for postulating the existence of these constituents in conceptual structure, and the role that these constituents play in sentence interpretation.

The conceptual structure of a lexical item can be viewed as a structure with zero or more open argument places that are filled by lexical items from a sentence. Sometimes the arguments are filled with default assignments that can be overridden if necessary. Consider the verb drink. One can use the word with or without specifying the thing that was drunk, as in the sentences below.
(32)  a. John drank the wine.
     
     b. John drank too much last night.

We can represent the lexical entry for *drink* in the following way.

Notice that, if there is no explicit mention in the sentence of what was being drunk, then this information is implicitly filled in by the default assignment *liquid*.

Once an appropriate representation is found for a sentence, the lexical items need to be mapped into their corresponding conceptual constituents. This is performed in a number of steps. Consider the conceptual representation of the sentence *John ran into the room*. The illustration below shows how the final product, after the syntactic constituents are used to fill in the open arguments of the conceptual representation of *run*.

(33)  [Event GO ([Thing JOHN], [Path TO ([Place IN ([Thing ROOM)])]]))]}
Notice that, in addition to conceptual constituents, there are argument functions that are involved with conceptual structures. Some of these functions are GO, TO, and IN. Notice that in the examples given so far, the function GO was associated with an [Event], the function TO with a [Path], and IN with a [Place]. While some functions are specific to a given constituent type, others can be used across types. In fact, this is the mechanism which allows some words to be used in several different semantic fields. Thus, John went from Chicago to Boston, and The meeting went from one to two can have the same basic conceptual structure, differing only in the conceptual constituent type, and the specific argument assignments.

Let us see in more detail how words in an utterance are mapped into the correct argument slots in conceptual structure. Consider once again example (33). The sentence John ran into the room describes an [Event] in conceptual structure. The verb corresponds to the Event-function GO, which has two arguments as shown below:

\[(34) \quad \text{[Event]} \Rightarrow \text{[Event GO ([thing], [Path])]}\]

The subject of the sentence is linked to the first argument of GO, and the PP to the second argument. But the [Path] constituent is itself decomposed into the Path-function TO and its argument as shown in (35).

\[(35) \quad \text{[Path]} \Rightarrow \text{[Path TO/FROM/TOWARD/VIA... ([thing]/[Place])]}\]

In turn, the [Place] constituent is described in (36).
(36)  [Place] => [Place ON/UNDER/ABOVE ... ([Thing ])]

Finally, the object of the sentence is mapped onto the [Thing] constituent of the argument of the Place-function. The whole structure is put together recursively, to give the representation in (33). The one thing to note is that syntactic constituents must be mapped onto the appropriate conceptual ones. It would not do any good, for instance, to map *room* into the second argument of GO. This must be filled with a [Path], and not a [Thing].

Of course as we have already seen, there are cases in which not all the slots can be filled because they are not explicitly given in the sentence. In this case they are assigned default interpretations. The interesting case, though, is when the sentence contains information which clashes with the default assignment. So, in example (32), the default value of *liquid* is assumed as a minimal condition on what was drunk. We could not say *John drank too much steak last night*, since the word *steak* does not refer to a liquid. Thus, the linking rules have to be designed to take note not only of appropriate syntax to concept mappings, but also of *selectional restrictions* on the conceptual type of a word which is proposed to fill a given slot. But note that the role of selectional restrictions is not simply to ensure that lexical items are not inserted into the wrong place, or to pick the correct member of a set of interpretations for an ambiguous word (as in Katz, 1972, for instance). Selectional restrictions can be seen as default assignments which are essential to the conceptual interpretations for sentences, when those sentences do not explicitly provide explicit information about some aspect of the conceptual representation.
which is constitutive of its meaning. Thus, selectional restrictions can be seen as an integral part of a theory of conceptual representations, and not simply a mechanism for preventing certain word combinations. But most importantly, Jackendoff's theory allows us to postulate two different sorts of restrictions which might be operating.

The first type of selectional restriction can be seen from the *drinking* example above. That is, a specific detail of a candidate word does not match some special requirement of the empty slot, as with the verb *drink*. So, one can drink *water*, but not *table salt*, or *computers*. This is similar to the types of restriction tested in experiment 3. However, the second possibility is new. Jackendoff (1983, 1990), as we have seen, proposed that there is some relatively small set of conceptual elements which can act as maximal superordinates of all words in the lexicon (or concepts). (As we saw already, this is a proposal which has been made by other researchers, who claimed special priority for a small set of essential primitives). In Jackendoff's formalism, then, it is possible for a candidate word to violate the requirement for a particular, major conceptual category in a conceptual slot, as in *John drank the color*. In this example the word *color*, which is a [Property], is the syntactical candidate to fill the slot meant for a [Thing]. But this is a violation in the category of conceptual constituents, and can be contrasted with a more idiosyncratic, within-category violation, as in the earlier examples. Jackendoff (1990) does not distinguish between these two possible types of violation, since his aim is to account for all cases of violations of selectional information with the same formal mechanisms: the rule which fuses lexical items with the conceptual slots with which they are linked. The
formal statement of the principle of fusion does not need to distinguish between the sorts of violation that can prevent the operation. However, the claim being made here is that the distinction might have consequences for the psychological mechanisms which implement the formal theory. Let us see how this distinction might impact on the current investigation.

Jackendoff (1990) discusses constraints on restrictive modification, which refers to cases where a constituent modifies another. The prototypical case of such modification is an AP modifier of an NP, as was the case for the sentences used in all the experiments reported thus far in this thesis. But other cases might involve a PP modifier of an NP as in house in the woods; a PP modifier of a VP as in John went home at 6:00; or an AP modifier of a PP as in high on the hill. In each case, the modifier is integrated with the conceptual constituent of the phrase it is modifying, as in the following example, red hat:

\[
\begin{array}{c}
\text{HAT} \\
\text{Thing} [\text{Property RED}]
\end{array}
\]

But let us now consider how we might formulate restrictions involving category violations, in the case of restrictive modifiers. Jackendoff suggests that there are indeed some restrictions on the kinds of constituents that might be combined in this relation. So,
he finds that a [Thing] or an [Event] cannot modify a [Manner]. In general, however, it is not as easy as before to state conditions on this kind of modification. The biggest problem with Jackendoff’s theory in this regard is that he does not offer a precise explanation for how the relation between the modifier and the head can be calculated, so it is difficult to stipulate rules for finding interpretations in a given word pair. The sentences used in the experiments in this thesis almost all involved AP/NP constructions, whose interpretation is most readily seen as a [Property] modifying a [Thing]. In the plausible and implausible conditions, the way in which the modifier can be construed as a property is fairly clear (e.g. walking snakes, simmering tribes). However, this is not intended to be so in the anomalous condition (e.g. Socratic harbor, genetic dress). What would happen in these cases? Let us say that it is assumed by interpretive processes that the adjective describes some [Property], since this is a canonical role. But there is no way to figure how the AP can be construed to be a [Property] of the NP. Perhaps the conceptual representation would be left incomplete with the adjective being designated as a property, but without a specific proposal for how the AP is a property. In this case, however, the representation would still be well formed. On the other hand, a different kind of relation might be tried, perhaps with equally little success. In this case, the representation would be left ill formed. The problem is that we are left with the possibility that the conceptual structure would be well formed in some cases but not others, and most importantly, we have no formal way for deciding which is true for a given sentence. This would leave us with a mixed bag of interpretable and un-interpretable sentences in the anomalous condition. Perhaps this is
the same point that was made above, that it is difficult to specify exactly when such a combination is not interpretable.

Fortunately, Jackendoff's theory suggests a simple but powerful alternative to the kinds of sentences we have been using, if we focus on a more broad range of syntactic positions. Consider sentences like *John drank the color*. It is clear that the constituent [Thing] required by *drink* is not satisfied by the [Property] *color*. Unlike in the questionable AP/NP cases, there is not even a superficial agreement here between the ontological type that is needed by the verb, and the type supplied by the syntactic object. There is a very clear and easy to define violation that is occurring. Similarly, we can clearly define violations where only idiosyncratic aspects of the restrictions are being violated. Thus, *John drank the powder* does not involve violation of ontological type, but instead there is a violation of a specific restriction on the kind of [Thing] that is required. Notice that this forces us to use sentences which are more similar in structure to the ones commonly employed in the matching task, and not like the ones used thus far, as suggested above.

The distinction between the two kinds of violation, then, can be stated in formal terms which are useful for generating experimental items. Jackendoff's proposed set of constituents seems useful in that they are based on elaborate linguistic analyses, and tend to overlap fairly widely with other similar proposals. The following experiment, then, tests the difference between sentences in which lexical items violate a basic ontological kind,
and those that violate only idiosyncratic semantic restrictions. The former of these are, by hypothesis, anomalous and the latter, implausible.

**Experiment 6**

This experiment was intended to test sentences where the type of the conceptual constituent that appeared in various syntactic positions, was clearly violated. In order to accomplish this, sentences were created by changing the words in plausible sentences. The changed word was either an implausible word (given the context) from the same constituent, or a different constituent altogether. Importantly, since the mapping from syntactic to conceptual constituents is not one-to-one (e.g. a [Property] or a [Thing] can be an NP), it was possible to keep the syntactic form of the sentences identical across the three conditions.

Table 9 shows a tabulation of the changes to conceptual constituent type that were made in order to change plausible sentences into anomalous ones.
Table 9. Conceptual constituent changes in experiment 6, that changed plausible sentences into anomalous.

<table>
<thead>
<tr>
<th>Changed elements</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td>[THING] ⇒ [PROPERTY]</td>
<td>22</td>
</tr>
<tr>
<td>[THING] ⇒ [EVENT]</td>
<td>12</td>
</tr>
<tr>
<td>[THING] ⇒ [STATE]</td>
<td>13</td>
</tr>
<tr>
<td>[THING] ⇒ [ACTION]</td>
<td>10</td>
</tr>
<tr>
<td>[THING] ⇒ [DIRECTION]</td>
<td>1</td>
</tr>
<tr>
<td>[THING] ⇒ [TIME/AMOUNT]</td>
<td>2</td>
</tr>
<tr>
<td>[ACTION] ⇒ [THING/STATE]</td>
<td>2</td>
</tr>
<tr>
<td>[EVENT] ⇒ [THING]</td>
<td>1</td>
</tr>
<tr>
<td>[TIME] ⇒ [PROPERTY]</td>
<td>2</td>
</tr>
<tr>
<td>[TIME] ⇒ [THING]</td>
<td>2</td>
</tr>
<tr>
<td>[PROPERTY] ⇒ [EVENT]</td>
<td>1</td>
</tr>
<tr>
<td>[PROPERTY] ⇒ [STATE]</td>
<td>2</td>
</tr>
<tr>
<td>[MANNER] ⇒ [ACTION]</td>
<td>1</td>
</tr>
</tbody>
</table>

It is evident that most of the changes involved replacing a [Thing] with some other constituent. This reflected a practical problem with designing items which were not like this, whose theoretical implications will be discussed later. For the present purposes, however, the concern was that this might introduce a systematic difference in the average frequency of words used in the different conditions. Word frequency might affect processing in many ways, since it has a large effect on the time it takes to recognize individual words (e.g. Forster & Chambers, 1973) This potential confound was of greater
concern in this experiment than previous ones because a larger proportion of the words differed for each sentence in the different conditions.

In order to ascertain the potential seriousness of this problem, the average frequency of the words which differed from sentence to sentence, was calculated for each condition. Table 10 shows the results of this tabulation.

Table 10. Mean frequency, expressed in number of words/million, for the items in the three conditions in experiment 6.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>The problem ran around under the accident all night long.</td>
<td>67.8</td>
</tr>
<tr>
<td>Implausible</td>
<td>The computer ran around under the plate all night long.</td>
<td>57.8</td>
</tr>
<tr>
<td>Plausible</td>
<td>The mouse ran around under the table all night long.</td>
<td>117.24</td>
</tr>
</tbody>
</table>

Pairwise t-tests revealed a significant difference between the anomalous and the plausible conditions, t(136) = 2.27, p < 0.05, and between the implausible and plausible conditions, t(130) = 2.21, p < 0.05. However, there was no difference between the anomalous and the implausible conditions, t(134) = 0.25, p > 0.05.

Clearly, the average frequencies of the words used in the different conditions was not matched, since the words in the anomalous and implausible conditions were less frequent than in the plausible condition. However, this finding does not invalidate the
experiment, since it works against the hypothesis under investigation. That is, the hypothesis predicts that the implausible sentences should be just as fast to match as the plausible ones, and only the anomalous ones should be slower. However, if word frequency has any effect on the task, then the implausible sentences would be slowed down to the same extent as the anomalous ones, making them differ from the plausible ones. The difference in frequency, if it was an important factor, would bias the experiment against the predicted outcome.

Finally, even though the reading task has given somewhat inconsistent and therefore inconclusive results, the following experiment adheres to the procedure used in experiments 1 through 5. However, more care was taken to ensure that subjects did not skip over the first sentence, by monitoring their performance carefully from the control room. If their reaction times for the first sentence dropped below one second for a number of successive trials, the experiment was halted and the instructions re-administered.

Method

Subjects. Thirty six, first year undergraduate students from the University of Arizona participated in this experiment.

Materials and design. There were six conditions in this experiment, three each in the reading and the matching task. These were anomalous, implausible, and plausible. Each sentence appeared in each condition. The sentences were counter balanced across three files. There were ten sentences in each condition.
In order to have each sentence appear in each condition, key lexical items were changed. The number of differing items varied across sentences, but the average number was around three. The three conditions were constructed as follows:

Anomalous: These were sentences in which the words substituted for the ones appearing in the plausible version, were of incompatible conceptual types.

Implausible: In these, the substituted words were very unusual examples of the correct conceptual type.

Plausible: Sentences which described plausible events.

The following examples demonstrate the kinds of substitutions that were made to form the three conditions:

It is always a pleasure to eat anomaly[STATE] at a nice allegation[ACTION].

It is always a pleasure to eat lobster[THING] at a nice restaurant[THING].

It is always a pleasure to eat charcoal[THING] at a nice mortuary[THING].

Yesterday we saw the predicament[STATE] jump over the dream[STATE], and land with a thump.

Yesterday we saw the kangaroo[THING] jump over fence[THING], and land with a thump.

Yesterday we saw the automobile[THING] jump over the moon[THING], and land with a thump.
The items were given to a group of 20 subjects to ensure that the rated plausibility differences were comparable to the earlier experiments.

In addition, an equivalent number of different sentences were constructed, as in experiment 1.

Procedure: The procedure was identical to experiment 1.

Results

Table 11 shows the ratings that subjects assigned to the items in experiment 6.

Table 11. Mean plausibility ratings, on a seven point scale, for the items in the three conditions in experiment 6.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>The problem ran around under the accident all night long.</td>
<td>5.32</td>
</tr>
<tr>
<td>Implausible</td>
<td>The computer ran around under the plate all night long.</td>
<td>5.30</td>
</tr>
<tr>
<td>Plausible</td>
<td>The mouse ran around under the table all night long.</td>
<td>1.86</td>
</tr>
</tbody>
</table>

Note that numerically, the difference between the plausible and the other two conditions is larger than it was in the other experiments, and the difference between the anomalous and implausible conditions is negligible. Pairwise t-tests revealed a significant
difference between the anomalous and the plausible conditions, \( t(34) = 23.3, p < 0.05 \), and between the implausible and plausible conditions, \( t(34) = 22.05, p < 0.05 \). However, there was no difference between the anomalous and the implausible conditions, \( t(34) = 0.12, p > 0.05 \).

Table 12 shows the reading and matching task results in experiment 6.

Table 12. Reaction Time for Reading and Matching (and error rate for Matching), as well as the difference between each experimental condition and the plausible baseline.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>2654</td>
<td>191</td>
</tr>
<tr>
<td>Reading</td>
<td>Implausible</td>
<td>2486</td>
</tr>
<tr>
<td>Plausible</td>
<td></td>
<td>2463</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>Err.</th>
<th>Diff. (RT)</th>
<th>Diff. (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>2347</td>
<td>5.8</td>
<td>133</td>
<td>2.2</td>
</tr>
<tr>
<td>Matching</td>
<td>Implausible</td>
<td>2270</td>
<td>4.7</td>
<td>56</td>
</tr>
<tr>
<td>Plausible</td>
<td></td>
<td>2214</td>
<td>3.6</td>
<td></td>
</tr>
</tbody>
</table>
Considering first the reading time, the 191ms difference between anomalous and plausible sentences was significant, min\(F'(1,49) = 7.92, \ p < 0.1\), as was the 168ms difference between the anomalous and implausible conditions, min\(F'(1,53) = 7.37, \ p < 0.1\). On the other hand, the implausible and plausible conditions did not differ.

The anomalous and plausible conditions also differed in matching, min\(F'(1,56) = 4.63, \ p < 0.1\). The anomalous and implausible sentences differed by subjects, but not items, F1(1,33) = 5.04, p < 0.05, F2(1,27) = 3.52, p > 0.05. The implausible and plausible conditions did not differ.

There were no differences in error rates.

Discussion

The reading times showed that subjects were spending considerably more time with the anomalous condition than the other two, which did not differ from each other. An interesting pattern is starting to emerge from these experiments, namely, that implausible sentences seem to have no effect on reading time. This contradicts the initial hypothesis that implausible sentences take longer to interpret, even when it is possible to interpret them. However, as discussed, this conclusion must be considered cautiously.

In the matching task, there is a very strong and reliable difference between anomalous and plausible sentences, as expected. There is also a significant difference (by subjects) between the anomalous and implausible sentences, and no difference between implausible and plausible. These results are almost exactly what is predicted by the
hypothesis. The only blemish is the lack of a significant difference in the item analysis between the anomalous and implausible sentences. However, in this experiment, there were fewer items per condition than in the other experiments. It is possible that the item analysis simply lacked the power to reveal a significant difference.

The implication from both the reading and matching tasks in this experiment is that implausible sentences are not processed any differently from plausible ones. Yet, as the subjective ratings indicate, at some level of interpretation, the two are clearly distinguished. This leads to an interesting possibility, namely, that reading and matching times are sensitive to a level of semantic representation that can be distinguished from a high level conceptual representation that is sensitive to constraints of pragmatic knowledge. Thus, these experiments could be used to contribute to the debate about whether or not some such two levels should be distinguished (e.g. Jackendoff, 1983; Pinker, 1989; Katz & Fodor, 1967).

The difficulty with the interpretation of the experiment is that the reading time data might be unreliable due to strategic factors as previously discussed. This particular task can not be taken to reliably indicate the effects of the manipulated variable. Yet, this is the only other task that we have converging evidence from at this stage. It is possible that some other on line task that involves interpretation, is affected by the meaning manipulation. If so, then the implication is that the sentence matching task is sensitive to a semantic level that precedes pragmatic interpretation, but other tasks are sensitive to a higher level representation. If this were true, then a revised version of the shallow
sensitivity hypothesis would be favored. In fact Ratcliff (1989) proposed just such a hypothesis, claiming that the task is sensitive to a level of representation at which thematic roles are assigned to lexical items. In order to test this possibility, we decided to employ the Rapid Serial Visual Presentation (RSVP) task, which has been claimed to be sensitive to plausibility (Forster, 1974; Forster & Ryder, 1971).
Chapter 6

RAPID SERIAL VISUAL PRESENTATION

In the RSVP task, subjects are presented with word strings, one word at a time, in rapid succession (usually about 16 words/sec). Each word is superimposed on the previous one, preventing a cumulative sensory storage of the input. Typically, subjects are required to report as many words as they can remember from the sentence. Usually the presentation rate is such that, on average, four or five words are successfully reported from a seven word sentence. Forster (1974), and Forster & Ryder (1971) showed that performance on the RSVP task was affected by syntactic and semantic factors. In particular, they showed that semantically anomalous and implausible sentences were more poorly reported than semantically plausible ones. Now, their notion of anomaly and implausibility was not precisely defined. Yet, the list of materials they provide suggest that some items, at least, meet the currently suggested criteria.

The explanation for why the RSVP task should be sensitive to violations of syntactic or semantic well formedness is relatively straightforward (though see Forster, 1974, for a defense of this interpretation against a number of potential challenges). The idea is that the assimilation of each word in the string into a short term memory buffer is limited by the rapid presentation rate of the stimuli. Each word, it is argued, is perceived at the time of presentation, but then lost from memory if it is not successfully integrated into memory. The presence of structure, then, helps because it assists with this process of assimilation. The reason that semantically implausible sentences are reported less
accurately than plausible ones is that the perceived meaning is incoherent, and therefore not beneficial for the encoding of the lexical items.

One important confound in the study of semantic factors with the RSVP task is the possible role of guessing. That is, if subjects are asked to report words they saw in the sentence, then they are likely to guess on some occasions that they really did not remember a word. If the words are plausible in the context, then this might help them to guess correctly. To avoid this difficulty, Forster (1974) used a variant of the RSVP task in which subjects were not asked to report the entire sentence after they finished a trial. Instead, at the end of each trial they were asked to perform a two alternative forced choice task on two visually presented candidates for words which appeared in the sentence. Consider the following example, which shows that four sentences are required to test the effect of a single source of implausibility. One of these sentences would be given to each subject.

(37) a. The conductor asked the violins to play louder.
    b. The conductor asked the trumpets to play louder.

(38) a. The customers asked the violins to lean steadily.
    d. The customers asked the trumpets to lean steadily.

After each sentence, the subjects were given a choice between violins and trumpets. Either word is approximately equally plausible in the sentences in (37) and implausible in (38). Thus, in both the plausible and the implausible sentences, the effect of context on guessing
is negligible since this will not favor either word. In addition, any perceptual advantage that one of the words might have is factored out since both words appear in both conditions, and can thus be used as their own controls.

The RSVP task is a particularly interesting one to test the hypothesis in this thesis, because the rapid rate of presentation seems to make it particularly sensitive to semantic deviations. That is, if an interpretation cannot be rapidly constructed, then there will be no benefit to the assimilation of words into short term memory. Thus, it is possible that the relatively unconstrained reading task in the sentence matching task was insensitive to deviations caused by implausibility, and that the RSVP task is more sensitive since rapid interpretation is of key importance.

**Experiment 7**

The results of experiment 6 suggest that implausible sentences do not affect the processing of linguistic stimuli. However, this conclusion contradicts many findings in the literature, and therefore has potentially important theoretical ramifications. Unfortunately, part of the evidence was gained through the analysis of reading times, which are potentially unreliable, as suggested in earlier experiments. Therefore, it is important to test the items which showed a clear pattern of results in sentence matching, in another task which has been claimed to be sensitive to meaning. One such task is RSVP.

**Method**

*Subjects.* Forty two first year undergraduate students from the University of Arizona participated in this experiment.
Materials and design. The sentences were taken from experiment 5, but were adapted for use in the RSVP task. It is very important in the forced choice RSVP task that the items which are probed are carefully counterbalanced across the conditions, as shown in examples (37) and (38) above. Thus, two separate lists need to be constructed for controlling the probe word in a single condition. For instance, in the example above, the plausible sentence contained a choice between violins and trumpets. It is possible, however, that one of these words is easier to perceive than the other, which would bias the results. Let us suppose that trumpets is in fact easier to recognize. Then, if the test item did in fact contain trumpet, then the accuracy of report would be higher than if it contained violins. Thus, separate files were constructed such that in one of them trumpets appears, and in the other, violins appears. In addition to this, however, it must also be ensured that the same pair of words appears in each condition. This is to ensure that the every property of the probe words is controlled, and the only variable is the semantic context in which it is found. Thus in examples (37) and (38), the same pair of words appear in sentences where they are plausible and implausible.

In order to achieve such counterbalancing for items in experiment 6, six separate files were constructed. In effect, these were three pairs of files, with each pair being identical except for which of the two probe words appeared in the sentence. Since there were three conditions, the items in these were counterbalanced across the three sets of items. In order to achieve this level of control, some of the sentences needed to be altered,
such that the same word could appear in each of the three experimental conditions. The full set of items appears in appendix C.

The three experimental groups were as before. That is, the conditions were (a) anomalous, (b) implausible, and (c) plausible.

Procedure. Subjects were given instructions and seated in an experimental booth, as before. During the experiment, each time the subject pressed the foot pedal, a sentence was presented on the computer screen. The sentence was presented one word at a time, each word being displayed for 56.8ms (= 17.6 words/sec). Each word was superimposed on top of the one that preceded it. After the final word, a row of hash marks (###) was displayed for the same duration, to eliminate visual persistence of the last word in the sentence. Immediately following the hash marks, the two test items appeared, one on each side of the screen. Subjects were required to press a response button to signal which word they thought they saw in the sentence. The two choices remained on the screen until the subjects made their response.

Results

Since the point of interest in the RSVP task concerns the accuracy of report under the different conditions, only error rates were recorded. These are shown in table 13.
Table 13. Mean error rates (% correct) in the RSVP task in the three conditions, and the difference between each experimental condition and the plausible baseline.

<table>
<thead>
<tr>
<th></th>
<th>Err.</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomalous</td>
<td>46.2</td>
<td>7.9</td>
</tr>
<tr>
<td>Implausible</td>
<td>39.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Plausible</td>
<td>38.3</td>
<td></td>
</tr>
</tbody>
</table>

The difference in error rates between the anomalous and plausible conditions was significant, \( \min F'(1,62) = 2.95, p < 0.1 \). The anomalous and implausible conditions differed by subjects and items but not \( \min F', \) \( F_1(1,36) = 4.4, p < 0.05, F_2(1,27) = 6.18, p < 0.05 \).

**Discussion**

The results show that implausible sentences do not show a decrement in performance in the RSVP task, when compared to plausible sentences. However, anomalous sentences are processed much less efficiently. This pattern of results corresponds exactly to the results obtained with the matching and reading tasks. The lack of an effect for the implausible sentences in both sentence matching and RSVP, and in reading, could be interpreted in several ways. First, it is possible that the level of semantic representation that is used by these tasks is identical, and this level is unaffected by
plausibility. This is interesting, because it implies that the construction of this
representation is not affected by pragmatic facts. The implication is that there is a level of
representation which we can call the semantic level, which encodes the meaning of
linguistic entities, and is insensitive to real world pragmatic knowledge. This level of
representation seems to contain the same features which are involved at the highest level
of conceptual representation, yet differs from that representation. In contrast, Jackendoff
(1983), who provided the basic framework for the experiment, argues that high level
conceptual analysis is the same as the semantic level. However, we differentiate the
conceptual level from the semantic one on the basis of the dissociation between empirical
data and subjective plausibility ratings. Simply, at some level of interpretation, we clearly
do appreciate that there is something very odd about implausible sentences. The semantic
level which controls performance on these experiments seems completely oblivious to
these implausibilities, however.

An alternative interpretation of the results is that plausibility does affect
processing, but was not found in the experiments only because the difference between the
plausible and implausible sentences was not sufficiently large. Perhaps the implausible
sentences were not implausible enough. The large difference in subjective ratings between
the plausible and implausible sentences mitigates against this interpretation to some extent,
but it could always be argued that the rating task is not a good measure of the kind of
plausibility that is relevant for the on line processing task, and is thus irrelevant. Or,
perhaps, the ratings are affected by the overall composition of the sentences because
subjects feel that they should make use of the entire scale, and the presence of some really implausible sentences forces them to rate any sentence which is not ridiculously weird as plausible.

The possibility that the plausible sentences were too similar to the implausible ones, is difficult to refute. One way to do it would be to find an online task which did differentiate between the items in the two conditions. But let us tackle the question a different way. Consider what the claim that the plausible sentences were not plausible enough, entails. One possibility is that the plausible sentences should really describe highly stereotypical scenarios like The maid polished the silver (from Ratcliff, 1983) In fact, Ratcliff found, in several experiments, that implausible sentences did not differ from non-stereotyped plausible ones (e.g. The clerk dissolved the jacket vs. The clerk borrowed the jacket), but that a difference was only apparent when stereotyped sentences were compared to anything else. (However, in other experiments she did find an effect between non-stereotyped plausible, and implausible sentences, so her conclusions cannot be accepted without qualification). It is therefore possible that a strong effect is seen only if the plausible sentences describe events which have a high probability of occurring, whereas the ones in the experiment merely described situations which could happen, but actually happened with low frequency. (How often is one told at the breakfast table that The mouse ran around the table all night?). Thus, maybe only stereotyped sentences are processed faster than implausible ones. But even if this were true, would it have an impact on the theoretical position taken in this thesis in a significant way?
According to the kinds of theories we have been discussing, the way in which implausibility would effect processing is by slowing down the process of interpretation, because the time to retrieve information about implausibly connected concepts would be greater than the time for plausibly connected ones. Now, by hypothesis again, the plausible sentences in the experiment would not have been hindered because there were supposedly no semantic incongruencies in the sentence. However, the implausible items should have been slowed because they did encounter such incongruencies. The finding that there was no difference in performance between the two conditions was therefore interpreted as showing that implausible sentences did not take longer to process. The criticism of the experiment, then, is that this characterization is not correct, since the processing of plausible sentences is slowed to the same extent as implausible ones because they were not "plausible enough". But what is "plausible enough"? The suggestion was that the plausible sentences should really have described highly stereotypical events which occurred with a high frequency in the subjects' personal experiences. But this raises an objection to the objection. That is, how can we rule out the possibility that highly stereotypical sentences are processed faster only because some other system becomes involved. Perhaps some sort of episodic system might realize that it already has the relevant interpretation, and interject in the normal process. Thus, introducing highly stereotyped sentences might contaminate the data by introducing the consequences of mechanisms which are not a part of the ordinary interpretive processes. The experiment,
as it stands, compares the interpretation of novel but plausible sentences with novel implausible sentences which, it seems, is as it should be.

The other possibility is that there is another level of plausibility between highly stereotyped sentences, and the ones that were used in the plausible condition of the last two experiments. But it is difficult to know what this level would be, especially if we cannot rely on subjective ratings. The only possible way to investigate this would be to change the sentences until, eventually, one of these tasks at least would show a difference between the implausible and the plausible conditions. But then we would also need to compare these new plausible sentences against highly stereotyped sentences, to show that they too differed. However, given the size of the effects we can obtain in these tasks, this seems like a difficult goal to accomplish. So for the present, we can conclude only that some sentences which subjects typically rate as having very clear and sensible interpretations, are not processed faster in the three tasks we looked at than certain other sentences which subjects rate as being very unusual and implausible. Future research will need to address the question of whether or not various levels of plausibility can be distinguished at the level of semantic representation that these tasks are sensitive to. This is an important question to answer in order to maintain the strong form of the theory proposed in this thesis, which is that the only distinction is between strings that are anomalous and those that are not. Pragmatic factors ought to have no effect at all on processing.
Chapter 7

GENERAL DISCUSSION

The experiments reported in this thesis set out to find evidence for the hypothesis that performance on the sentence matching task was sensitive to the meaning of the sentences in the task. This is contrary to claims in the literature, that the task is sensitive to a very precise level of syntactic analysis, and that so called “plausibility” effects were really due to locally defined lexical co-occurrence frequencies, which affected the speed with which a sentence could be parsed. Experiment 1 showed that locally implausible sentences were in fact not matched slower than plausible ones, but anomalous ones were. Experiment 2 showed that even globally anomalous sentences could result in a decrement in matching speed. Reading times for these sentences, however, suggested that both implausible and anomalous sentences took longer to read than plausible ones. Experiments 3 to 6 attempted to clarify the distinction between “anomaly” and “implausibility”. Of these experiments, 1, and 3 to 5 used similar sentences, in which the semantic plausibility was varied by manipulating the identity of a modifier in an AP/NP construction. Experiment 3 ruled out the possibility that selectional restrictions of the type typically discussed in the linguistic literature, were the correct way to describe the distinction. Experiments 4 and 5 failed to find the distinction in terms of Ratcliff’s (1989) proposed criterion of clarity of interpretation. However these experiments, (5 in particular), showed that the reading times for implausible sentences did not differ from the reading of plausible ones. This contradicts the initial hypothesis that plausibility can affect the speed with
which a semantic interpretation of a sentence can be established. In experiment 6, the nature of the semantic manipulation was changed, such that the plausibility variation occurred between words in different syntactic positions. This experiment clearly demonstrated that implausible sentences did not have an effect on reading or matching times, but that anomalous sentences did. Furthermore, we now had a tentative definition of anomaly, in terms of the violation of conceptual constituents. Finally, experiment 7 suggested that the matching task was not unique in the level of semantic representation that affected it, but that potentially a large number of tasks which measured sentence processing, were unaffected by plausibility.

The implication from these conclusions is that there is a level of semantic representation at which the meaning of linguistic input is represented, and that this representation is important for many experimental tasks which measure aspects of sentence processing. Furthermore, this level of representation is not sensitive to facts of a pragmatic nature in the real world. Instead, they are sensitive to fundamental beliefs about the basic ontological structure of the world. This has interesting parallels with recent claims in the cognitive developmental literature, about the way in which fundamental beliefs about the world can direct the emergence of mental life (Keil, 1989; Carey, 1985). The claim also supports Pinker (1989) who argues for a distinction between semantic and conceptual representations, and contradicts Jackendoff (1983) who argues against it. Perhaps one way to characterize the process of interpretation that is suggested by these experiments is that a semantic interpretation is automatically extracted from the input, and
that various default assumptions are made in the construction of this representation on the basis of the rules of semantic structure. Then, this representation is elaborated into a mental model, which integrates different knowledge bases to give a representation which can support various inferences (Johnson-Laird, 1983). It is only the latter of these that is sensitive to real world knowledge and pragmatic considerations. However, it is the former that controls performance in the majority of language processing experiments.

If these claims are true, then a large number of prior claims in the sentence matching task need to be reevaluated. For instance, we already saw that the defense of the shallow sensitivity hypothesis depended on the local/global plausibility distinction. However, we have claimed that no such distinction can be made, and that the controlling level for the matching task is not one that is prior to semantic interpretation. An alternative explanation for the effects obtained with constraint violation sentences must be sought.

This issue, however, raises an important challenge to the present proposal. Recall that sentences which violate a constraint on movement do not take longer to match than grammatical controls, but those which violate agreement rules, do. If the proposal that interpretive processes determine matching time is correct, then there must be some way of explaining why only some kinds of ungrammaticality have an effect on this level of representation, by blocking an interpretation.

Consider the subjacency constraint violation sentence from example (12b): *Who do the police believe the claim that John shot?*. The problem is to explain why this
sentence takes no longer to match than a control, whereas *John smiled for he knew herself to be innocent does take longer to match. How do these strings relate to their possible interpretations?

We have already suggested in the introduction that it might be possible that semantic interpretation can begin before the syntactic analysis of a sentence was completed. We suggested that semantic interpretation could begin after some minimal amount of syntactic processing had occurred, but we could not specify the details of this process. While we are still not in a position to give a detailed explanation, we can at least consider how such a process might be possible in terms of, say, Jackendoff's (1983) theory of conceptual semantics. Recall that in his theory, the constituent type of the interpretation of the sentence as a whole is determined by the verb. Thus, John walked into the room is an [Event], and John is tall is a [State]. It is possible, then, that the frame for the interpretation can be established from identifying the main verb alone. Following this, according to the theory, the linking rules map lexical items from syntactically appropriate positions into the unfilled slots in conceptual structure. However, it is not inconceivable that the semantic representation itself begins to claim lexical items from the input string, before an exhaustive syntactic analysis is completed. For instance in English where word order is critical, a fairly good guess can be made about the semantic role that many words play in a sentence, based on a rudimentary syntactic analysis and the position in the string. For instance, it is not too difficult to guess the significance of each word in the following incomplete string
John ...his...... went ....... store .....bought....... yesterday.

The first step, then, involves establishing a conceptual frame for the interpretation. Next, the frame is filled in as completely as possible, based on the existing syntactic analysis, and other slots are tentatively filled by a heuristic process. This would allow time consuming knowledge retrieval and linking processes to commence before the syntactic analysis was completed. The tentatively filled slots would ultimately need to be confirmed by the completed syntactic analysis, but by this time much interpretive work would have been completed. Of course, it is also possible that some slots cannot be filled by the heuristic processes operating at the semantic level. These would be left unfilled until the grammatical structure is fully determined.

Recall also that slots do not always have to be filled with words that appear explicitly in the sentence. Conceptual representations themselves can supply assumed, default assignments to certain aspects of interpretation (e.g. Jackendoff, 1990; Pustejovsky, 1995). Thus, in *John drank a lot last night*, it is assumed that he drank some sort of liquid. Similarly, in *John buttered the bread*, the verb specifies not only that the bread was “covered with something”, but that the “something” was butter. Because of this assumed assignment, it seems bad to say *John buttered the bread with butter*. However, it is possible to say *John buttered the bread with stale butter*, since this overrides the default assignment (see Pustejovsky, 1995). Consider once again the case of *John buttered the bread with butter*. There is a redundancy of information, which leads to the oddness of the sentence. However, there is no contradiction or conflict which could lead
to a semantic confusion, as in *John buttered the bread with water. Thus, information which is filled in by default leads to a well formed conceptual representation, information which is supplied redundantly in the sentence leads to a well formed but odd representation, but a clash in information leads to an ill formed representation. With this in hand, let us return to the constraint violation sentence (39a), which is based on the well formed (39b).

(39)   a. *Who do the police believe the claim that John shot?

       b. Who do the police believe that John shot?

Grammatically, it is the extraction of the Wh- phrase from the object position of shot that is blocked by the claim, and which results in the ill formed sentence. But does the sentence have a possible interpretation? In sentence (39a), the grammar of the string cannot supply a well formed syntactic analysis which can be used to link the words to the conceptual representation. Nevertheless, it might be possible to establish a representation as suggested above, based on a tentative analysis. The main verb is believe, so a belief state frame is established in conceptual representation. Now, when we say in a grammatical sentence that *x believes y, what we really mean is that there is some proposition, or state of affairs, that is to do with y, and that x believes this proposition or state of affairs to be TRUE (see Pustejovsky, 1995). In the grammatically correct sentence that the example was derived from (Who do the police believe that John shot?), the state that is being believed is that the event where John shot someone is true, and the Wh- phrase is trying to
identify the effected entity in that event. Notice that it is necessary to presuppose that the event did take place, to answer the question in the affirmative. If, on the other hand it was revealed to the police that, in fact, John did not shoot anyone, then the answer to the sentence Who do the police believe John shot? would highlight the assumption that the event was no longer believed. For instance, an answer could be “no one... in fact, the police know that John did not shoot anyone” To reflect these elements in the meaning, the conceptual representation of the sentence would include an [Event] (John shooting someone), and a [State] (the police believing that the event took place, and that there was a victim).

The representation of the ungrammatical example would include the same elements as the sentence from which it was derived. But now, the additional component the claim has to be incorporated. Since believe can take propositions as the object of the belief, this constituent could be accommodated. Thus, the state now involves two components: 1) John shooting someone and 2) a claim being made. The problem is in identifying the exact relationship between the two. Since this relationship is not obvious, the sentence becomes difficult to interpret. But the important point is that all of the words in the sentence find a slot in the conceptual representation that they can fill, without violating any constraints on binding or conceptual type. Consequently, the conceptual representation forms an appropriate organizing structure for the sentence. (If this example seems a little strained, the reader might wish to think through a few more examples from Stevenson (1984): Which yacht has Al heard the rumor that Ed sailed?, What did the
'cook like the plan to curry?') Now contrast the example with *John smiled for he knew herself to be innocent, which is an example of an ungrammatical sentence that does lead to a cost in matching. In this sentence John, he, and herself have to co-refer. When the conceptual representation for this sentence is established, it includes an element for John. This element is indexed according to Jackendoff (1990), so that any co-refering expressions in the sentence can be co-indexed in the conceptual representation. Thus, he is assigned the same index as John. But herself, which is syntactically marked as co-referring, cannot be co-indexed because of the mismatching semantic feature. Thus, herself is left without an interpretation in conceptual structure. The clash between the interpretation and the words in the sentence leads to an ill formed representation. Obviously, the intended interpretation seems clear, so it should be possible to adjust the elements in the representation to reflect this meaning. However, this would become a representation of a different sentence - not the one which was presented.

An interesting consequence of this explanation is that the current hypothesis might be remarkably similar to the one proposed by Forster & Stevenson (1987), at least in the way it explains effects with ungrammatical sentences. That is, the reason why some ungrammatical sentences can be represented at a conceptual structure might well be that those are the “overgenerated” sentences, which have reasonable interpretations before the movement of the Wh- phrase. For instance, the police believe the claim that John shot who? makes the semantic relations clear. The advantage of the present proposal is that the same data structure, namely conceptual representation, is exploited by all sentences when
possible. Syntactic factors can be subsumed because of the consequences they have on interpretation. In this way, syntactic and semantic effects can be explained by the same mechanism.

It must be pointed out that many problems remain for this hypothesis. For instance, it is not clear that the difference between examples (15b) and (16b), (where *Who did the mayor order them to unfurl?* does not have a cost in matching but *The mayor ordered them to unfurl her,* does), can be explained in this way. Perhaps it could be that in the interrogative form, the initial implausibility forces one to re-construe the meaning of the sentence, such that *them* and the thing being *unfurled* now refers to different things. To make this more clear, consider *Who did the mayor order them to die?* Perhaps *them* becomes a redundant pronoun referring to the people that the *mayor* ordered to do something (*die, or unfurl something*). It is also difficult to explain why jumbling words, as in *Who the shot believe police ....*, slows down the matching operation. Taken to the extreme, the hypothesis would predict that the conceptual representation could be constructed on this basis. Clearly, the syntactic structure must maintain some correspondence with the conceptual one even though, by hypothesis, it is possible to construct elements of the interpretation before syntactic processing is completed. The problem is with specifying how much syntax is needed, and how the two levels interact. While this is a matter for further research, the general principle that is suggested for the way we can look at psychological theories is clear.
The way in which a computational system like the human mind implements a formal theory of rules and representations, is often left undefined by the formal theory. For instance, suppose that humans really did perform deductive reasoning by following the rules of first order predicate logic. According to that formalism, a conclusion is valid if it is derivable from the premises, by legally applying a sequence of rules to the premises, until the conclusion is reached. But the formal system says nothing about which rule should be applied at any given stage, or how one should decide that a given rule is a good one to use. All it can do is to say if a given rule can be applied legally. It is possible, therefore, that the mind performs its proofs by simply testing each possible rule against the sentence at a given stage in the proof, and using the first one that legally applies. Or, perhaps, a separate chain of reasoning is initiated for each legal rule. This is the difference between a depth first and a breadth first search strategy. Or, perhaps, rules are inserted to some arbitrary depth according to some heuristic, and then the chain is tested to see if every application of a rule was legal. In any case, the description of how one legal sentence leads to another would be exactly the same according to the rules of the formal system. So, what is being suggested in the present hypothesis is that, even though the linking rules between syntax and semantics can be expressed in a formal theory (eventually), the process which constructs the representations which are consistent with these rules, does not necessarily follow a strictly serial derivation. Instead, it is more like the process suggested above, where elements in the conceptual structure are inserted according to some heuristic, and the results checked against the rules of the formal system.
Clearly, this would be more sensible in a constrained representational language like Jackendoff’s than in first order logic.

Another important theoretical question we faced in this thesis, is whether or not we should draw a sharp distinction between anomalous and implausible utterances on theoretical grounds. As we have already suggested, the prevailing view in the literature is that such a distinction is one of degree rather than kind (e.g. Lyons, 1995), since the definition of anomaly tends to cover a relatively heterogeneous group. However, the results of these experiments show the opposite. In particular, we were able to demonstrate a strong dissociation between the rated implausibility of a sentence, and the way that sentence might affect psychological processes. While anomalous and implausible sentences were judged to be equally meaningless, only the former had an effect on the time it took to read, or to match a given sentence. This is an important result because it shows that at some level, there is a difference in the way the different types of sentences are interpreted, and that possibly, these differences are obscured at a higher level in processing.

The experiments suggest that there is an automatic, fast, informationally encapsulated process which derives a high level semantic representation of linguistic input. These processes are insensitive to acquired, “encyclopedic” knowledge, but are sensitive to fundamental ontological distinctions in conceptual space. This supports empirically those theories in which some aspects of lexical semantic interpretation are kept separate from general conceptual structures (e.g. Pinker, 1989; Katz, 1972), and contradicts theories which conflate semantic and conceptual processes (e.g. Jackendoff, 1983).
Semantic anomaly, then, is defined as a lack of a well formed structure at the semantic level. Utterances which are anomalous, therefore, can only be interpreted if extralinguistic processes intervene, and conjure up a context in which the utterance becomes meaningful. Often, as Lyons (1995) points out, this involves an extension of the way in which particular words are interpreted. For instance, in the earlier example *Thursday is in bed with Friday*, Lyons suggests that *Thursday* and *Friday* might be taken to refer to people, instead of days of the week. What the limits of this process are, is an open question. It is not clear, for instance, that they could be applied to a sentence like *It is always a pleasure to eat anomaly at a nice allegation*.

The claim is that there is a real difference between word combinations that violate major ontological constituents, and those that violate only "pragmatic" constraints, and that this fact characterizes the difference between anomalous and implausible sentences. The finding that there is a privileged position for knowledge to do with basic ontological categories is interesting, because it accords with conclusions which have been reached independently by researchers in developmental psychology (e.g. Keil, 1989; Carey, 1985), whose main concern is with identifying innate constraints on the acquisition of knowledge. Basically, they argue that some basic beliefs about ontological categories of the sort that we have been discussing are innate, and necessary for the acquisition of knowledge. For instance, Keil (1989) shows that children are not concerned about hypothetical transformations between different kinds of animals, but they don't accept a transformation between, say, an animal [+Animate] and a plant [-Animate]. If the basic categories which
are used to interpret linguistic strings by a semantic system are the same that can be inferred from developmental studies, then it might prove interesting to determine if in fact the same knowledge base is being tapped by the diverse methodologies.

As a word of caution for this line of reasoning, however, the construction of the experimental items did highlight a difficulty with the effort to identify basic categories by which words can be classified. In designing sentences for experiment 6, it was often difficult to categorize words according to the list of conceptual constituents given by Jackendoff. For instance, consider the following short list of problematic items: problem, dream, theory, and story. In order to help categorize these words George Miller’s “Wordnet”, which specifies the taxonomic hierarchy into which words of English fall, was consulted. Wordnet is claimed to be a psychologically realistic model of lexical knowledge in which words are categorized in a manner most suited to their syntactic type. Nouns, which are of primary concern here, are categorized according to the primitives outlined in Miller & Fellbaum (1991), which were given earlier. For instance problem has the following entry:
3 senses of problem

Sense 1

problem, job — (a state of difficulty that needs to be resolved; "she and her husband are having problems"; "it is always a job to contact him"; "urban problems such as traffic congestion and smog")
=> difficulty -- (a situation or condition almost beyond one's ability to deal with and requiring great effort to bear or overcome: "grappling with financial difficulties")
=> condition, status -- (a condition or state at a particular time: "a condition (or state) of disrepair"; "the current status of the arms negotiations")
=> state -- (the way something is with respect to its main attributes; "the current state of knowledge"; "his state of health"; "in a weak financial state")

Sense 2

problem -- (a question raised for consideration or solution; "our homework consisted of ten problems to solve")
=> question, head -- (the subject matter at issue; "the question of disease merits serious discussion"; "under the head of minor Roman poets")
=> subject, topic, theme
=> message, content, subject matter, substance -- (what a communication that is about something is about)
=> communication -- (something that is communicated between people or groups)
=> social relation -- (a relation between living organisms; esp between people)
=> relation -- (an abstraction belonging to or characteristic of two entities or parts together)
=> abstraction -- (a concept formed by extracting common features from examples)
The one classification that matches Jackendoff's for problem is [State]. But the category abstraction does not correspond to any of Jackendoff's. It is a vague classification, perhaps corresponding in some way to an "abstract" [Property]. So individual words might fit into several categories, some perhaps too vague to be useful for the current purposes. Similarly, dream might be a [State/Event] (or Psychological feature according to Wordnet), theory a [State/Property], and story an [Act] (or Abstraction). It seems that none of these items can be categorized by a single constituent in a useful way. Some of them seem to fit equally into two different categories. At worst, it might be impossible to categorize these words according to a single constituent. Perhaps they refer to concepts that have a rich set of interconnected beliefs at their core (see e.g. Keil, 1989), and Jackendoff's theory does not do well with such concepts. One reason for this might be
that the formalism is based almost exclusively on the analysis of the behavior of verbs, and
the principles he arrives at might not fully capture the representation of nominals.

The construction of items for experiment 6 revealed another potentially interesting
aspect in the study of anomaly. That is most of the anomalies in the experiment were
constructed by changing a [Thing] constituent into some other constituents. Very rarely,
the change was between two of the other constituents, and almost never was one of these
constituents changed into a [Thing]. The practical reason for this is that it was easier to
construct items in this way. But this seems contrary to the hypothesis that merely
substituting one constituent for another should result in an anomaly. That is, if anomaly is
a result of inserting a lexical item of the wrong category, then inserting a [Thing] into a
slot meant for a [Property] should be as bad as inserting a [Property] into a slot for a
[Thing]. To see what might be going on, consider the following examples.

(40) a. Dogs walk.
    b. Sincerity walks.
    c. Sincerity is a virtue.
    d. Dogs are a virtue.

It is clear that the a and b sentences impose a much stronger restriction than c. and d.
Thus, changing a [Thing] into a [Property] results in a violation while the opposite
substitution does not. But the example gives us a clue as to why this might hold generally.
The demands on the constituents are determined primarily by the verb in the construction,
and so its demands are the most relevant. In the above case, the verb *walk* which selects for [Things], is very specific in its demand for a conceptual type. However, *be* in c. and d. is much less specific: it can accommodate a [Property], but also a [Thing]. We cannot look at the constituents without also looking at the context in which they appear.

There are other possible explanations for why substituting a [Thing] with something else is the best way to reach an anomaly. The least interesting one is that the category [Thing] is very general, and can appear with a larger variety of verbs than any of the other constituents, making it almost impossible to form an anomalous sentence by replacing a word with a [Thing]. Another possibility is that words which canonically denote [Thing]s can, in certain contexts, be extended to some other category by some inference rule. For instance, in sentence (d) above, *dogs* might be construed as referring to some sort of [State] or [Property] to do with the abstract concept of “doghood”. The most interesting possibility, however, stems from some ideas in Jackendoff (1983), in which he considers how verbs referring to spatial concepts can be used in other semantic fields. Since verbs tend to refer to actions, they predominantly describe the actions of [Things]. However, this is not always the case. For instance, we could say “*The meeting went from ten to twelve*”, which does not involve [Things]. Jackendoff (1990) suggests that the use of verbs in “abstract” domains is accomplished by changing the designated constituent of the function (in this example, *went* is the function GO which can be seen with several constituents, as GOspatial or GOTemp, as it appears here). The important idea is that the spatial domain, which involves “things” as its major conceptual role players is primary, and that other semantic fields are derivative on this spatial one. But the direction
of derivation (from spatial to "abstract") suggests that the spatial domain is a superset of all others. That is, while certain spatial verbs might not be applicable in some semantic domains, there should be no verbs in the other domains that cannot be used in the spatial domain. Thus, verbs which have a prevalent use with [Property], [State], and so on, should also be useable with [Thing], but verbs whose main use is with [Thing] might not be appropriate with many words belonging to some other category. But if this were true, then we should not be able to construct sentences where the replacement of some conceptual element with a [Thing] was involved in the anomaly. We therefore need to look at examples where the anomaly did seem to be created, in part, by such a change.

Consider the following anomalous/plausible sentence pairs from the experiment:

(41)  

a. The advice went home after the end of the cucumber.

b. The dancers went home after the end of the ball.

(42)  

a. Theories often move the time of the meeting from red to oil.

b. Managers often move the time of the meeting from morning to night.

(43)  

a. Solitude knows that it is not easy to convert discretion into white rice.

b. Authorities know that it is not easy to convert violence into serenity.

Perhaps the most important thing to note is that in none of these examples is the relevant substitution critical in determining the anomaly. Either the major violation comes in that part of the sentence where a [Thing] is replaced, or, the substitution creates a violation only in that idiosyncratic combination. Examples in (43) above illustrate this point. In
sentence a., *white rice* is not ruled out per. se. What is ruled out is the whole process of the conversion of a [Property] into *white rice*, in which the anomaly arises from the mixing of conceptual types. Thus, converting *brown rice* into *white rice* would be perfectly reasonable, but *discretion* into *white rice* is not. In (41), we can see that the sentence is in fact open to multiple interpretations, and *cucumber* is anomalous under only one of these. Thus, if *end of the cucumber*... is interpreted as referring to an event in which the *cucumber* somehow came to its “end” (perhaps it was eaten), then the phrase is perfectly acceptable (if a little odd). That is, the word *cucumber* is taken as a “short hand” for the entire event, the inference being licensed by demands of the sentence as a whole. The anomaly only arises if this inference is not made, and *cucumber* is interpreted strictly as a [Thing]. Finally in (42), the sentence strictly specifies that the expected constituent has to be a [Time]. Somehow this specific imposition blocks the sort of re-interpretation that we saw in sentence a. However, this is an unusual sentence since the strict and overt specification of the required element is unusual, and would seem heavy handed in ordinary speech (e.g. *The thing I want is a hammer*. Or *The emotion I feel is anger.*)

These examples suggest an interesting possibility that has received little discussion, that anomalous sentences are predominantly of the type where a [Thing] constituent is replaced by some other in an inappropriate fashion. A cursory investigation of the examples suggested by Lyons (1995), Katz (1972), Chierchia and McConnell-Ginet (1993), support this observation. The examples also show that the restrictions on conceptual constituents are very subtle and possibly multiply determined, such that some
aspects of the word can license an insertion which another aspect prohibits (thereby leading to multiple interpretations at times). For the purposes of this experiment, then, we made use of the clearest possible examples of conceptual constituents of an inappropriate type being inserted into a sentence, which is the reason for the predominance of the cases where a [Thing] was replaced by some other conceptual constituent.

One shortcoming of the experiments reported in this thesis is the somewhat chaotic state of some of the early experiments. It was argued that this was due to the fact that sentences were all of the AP/NP type, and that this construction was not well suited for constructing anomalous sentences. However, the theoretical machinery which could explain why this was the case, was not presented. Using semantic anomaly in the construction of semantic theory could prove important, just as the use of ungrammatical sentences in syntactic theory has proven important. The problem with anomaly, though, has been that our intuitions in this domain are limited (though see Cruse, 1986), who gives many interesting tests for eliciting semantic intuitions). These experiments suggest a new way in which anomaly can be defined, and therefore used in the construction of semantic theory. As an example, we will look at a modern, and highly involved theory of lexical semantics, and show how the theoretical constructs interact with the experimental evidence reported here. The theory we adopt here is Pustejovsky's (1995) theory of the generative lexicon. However, since the theory itself is quite complicated, a detailed explanation is given in Appendix A, which can be consulted by interested readers unfamiliar with the theory. In the next few pages, I attempt to give a summary of the key
features. Perhaps the easiest way to do this is to look at an example of a lexical representation in Pustejovsky’s formalism.

This is a part of the representation of a sentence in which someone “bakes a cake”. The first thing to note is the Event Structure, which specifies how the events contained in the representation are related to one another. Here, the main event $e_1$ is a process, which results in $e_2$, a state. The process refers to the baking act, and the state to the conclusion of the process. The Argument structure describes how the syntactically realized arguments

---

**EVENTSTR**

\[\begin{align*}
E_1 &= e_1; \text{process} \\
E_2 &= e_2; \text{state} \\
\text{REST} &< \\
\text{HEAD} &= e_1
\end{align*}\]

**ARGSTR**

\[\begin{align*}
\text{ARG1} &= 1; \text{animate\_ind} \\
\text{FORMAL} &= \text{physobj} \\
\text{ARG2} &= 2; \text{artifact} \\
\text{CONST} &= 5 \\
\text{FORMAL} &= \text{physobj} \\
\text{D-ARG1} &= 3; \text{material} \\
\text{FORMAL} &= \text{mass}
\end{align*}\]

**QUALIA**

\[\begin{align*}
\text{create\_lcp} \\
\text{FORMAL} &= \text{exist}(e_2, 3) \\
\text{AGENTIVE} &= \text{bake\_act}(e_1, 4, 5)
\end{align*}\]
of the sentence relate to each other in semantic structure. The first argument is an
animate_individual, which is the subject of the sentence; in this case, the baker of the
cake. The second argument is the object; in this case the cake, which is an artifact.
Finally, the Default argument refers to the fact that the cake is made from some material.
Note also that the second argument has a slot labeled CONST which represents the
constitutive components of the object in question. In this case this is linked with the
default argument in the representation. This also shows some of the complex internal
structures and cross-references possible with this representational formalism. This
mechanism is also similar to the one we saw in Jackendoff’s theory, where default
interpretations are assigned for verbs like drink and butter. The Qualia structure defines
the type of the concept, which basically situates it in a hierarchically ordered structure of
representations. Basically, the typing system allows lexical entries to inherit features from
other entries in the hierarchy. The nature of the taxonomy is an empirical issue, but a small
set of non-overlapping hierarchies of the type suggested by Miller & Fellbaum (1991)
seems to be involved. The qualia structure is also important because it defines aspects of
the referent of the entry. The AGENTIVE quale defines factors which were involved in
the creation of the object. This could refer to a creator if the object was an artifact, or it
could specify that the object is a natural kind. In this example, the factor was a bake_act
involving the process, which was performed by whoever fills the first Argument e₁, on the
material specified by e₂. The FORMAL quale describes the relations between elements in
the representation. In this case it simply states that e₂ relates to the state at the conclusion.
of the act, such that \( e_2 \) exists. In addition, many lexical items specify a TELIC quale, which describes the purpose, or function of the object. For instance, the TELIC role of \textit{book} would be specified as

\[
\text{TELIC} = \text{read}(e, w, x, y)
\]

where \( e \) is bound to some event in event structure, \( w \) to whoever is reading the book, and \( x \) and \( y \) to type information specified in the argument structure of \textit{book}, where \( x \) is bound to \texttt{information/content} and \( y \) to \texttt{physical/object}. The whole specification, then, specifies that the purpose of a \textit{book} can be described as a function which involves a relation between an event, a person, and a physical object that contains information.

The most important thing to note is that each lexical entry contains a great deal of information that has to do with how the word fits into a causal structure. Thus, the function and origin of an object are included. In order to do this successfully, the representation involves defining the relations between different relevant aspects of the representation; e.g., the FORMAL quale of has an element bound to the second element in \texttt{argument structure}. This results in a representation which has a large number of interrelated constituents, but each constituent is independent in the sense that they can be modified by transformational rules without changing the overall relational structure.

Pustejovsky (1995) specifies a number of transformational rules by which elements of a lexical entry can be modified. Basically, these processes involve the modification of constituent elements in a lexical entry, by overwriting existing information with a few
specification obtained through inheritance over the type hierarchy, or from other words in the sentence. For instance, the verb *drive* might require a type of *vehicle* as an object. And, a *car* might be specified as a sort of *vehicle*, which would license its interpretation in the sentence *I drive a car*. However, *Honda* might not directly be typed as a *vehicle*. Instead, it would be specified as a type of *car*. However, through inheritance the feature *vehicle* would be inserted into the entry for *Honda*, licensing its interpretation in *I drive a Honda*. This is called *Type-Coercion*. There are many more subtle facts that can be derived from these rules, as well as the type hierarchy. Details of these are found in the appendix. The important point is that the power and flexibility of the representation allows us to explain a number of our previous observations.

Recall example (41), repeated below. The point was that the use of *cucumber* in the first sentence is licensed because there is a sense in which it is taken to be an [Event]. Thus, it is not really an anomalous phrase.

a. The advice went home after the end of the *cucumber*.
b. The dancers went home after the end of the *ball*.

But this is a straightforward application of *type coercion*, licensed by the TELIC role of *cucumber*. This would specify that a *cucumber* is a kind of *foodstuff*, one of whose purposes is *to be eaten*. But *eating* is an *event* which has a beginning and an end. So, through inheritance, the *end of the* is taken to refer to the end of the *event* which the *cucumber* participates in.
The second claim we can now evaluate is that the AP/NP constructions are potentially more interpretable than sentences where the subject or object of a sentence is substituted. When an adjective modifies an NP, there must be some relation found between them, if the meaning is well formed. The way Pustejovsky's theory works is that the modifier attempts to alter some relevant aspect of the lexical entry of the NP it is modifying, through the mechanism of selective binding. This mechanism allows adjectives to affect meaning in the appropriate way, by modifying just one sub structure within a lexical item, rather than the whole item itself. Consider the adjective *fast*, and how it modifies the noun *car*. The qualia structure of *car* includes the TELIC role $\text{drive}(e,y,x)$, where $e$ is an event, $x$ is bound to the car, and $y$ to the person driving the car. It is this TELIC role that *fast* modifies, such that the events involved with driving occur *fast*. But what about *fast game*? This is somewhat more difficult, since there is a real ambiguity, between a *game that is played fast*, and *a game that is over very quickly*. The first of these is a straightforwardly similar to the *fast car* example, with the TELIC role this time being $\text{play}(e,y,x)$. The second, however, requires some steps in the derivation. Intuitively, the phrase refers to a time period, whose passage is *fast*. What is modified in this example is not the TELIC role, but the FORMAL role, if we assume that this encodes something like $\text{play_event}(e',y,x)$, and it is the event as a whole which is *fast*. What these examples are meant to illustrate is that modification can occur at more than one point in the lexical entry, and that these points can themselves change their identity through inheritance from other entries. Thus, the degrees of freedom for finding some suitable modification is vast.
Consider how such a process might be able to explain the relative difficulty in generating anomaly (and hence stable results in the matching task) with AP/NP sentences.

Let us look at the example of *timber shirt*, and *architectural shirt*, from before. The first example is straightforward. The entry for *shirt* would specify that it was an artifact, made by someone, from some material. In turn, *timber* is a kind of material, which can straightforwardly satisfy the requirement. *Architectural shirt* is more difficult. Let us first try and determine what sorts of roles *architectural* can modify, by looking at some of the constructions it typically appears in.

(44)  
   a. architectural restoration  
   b. architectural knowledge  
   c. architectural history  
   d. architectural survey  
   e. architectural museum

The first thing to note is that the NPs that *architectural* modifies tend not to refer directly to physical objects. This might seem a little odd, given that the word *architecture* is defined as a “style and method of design and construction; orderly arrangement of parts” (American Heritage dictionary). Thus, it is intimately about “things”, but the NP which is typically modified is not a [Thing], as shown in examples (44a) to (d), which are all “abstract” concepts. In order to see how the modification works, we need to determine what sort of taxonomic hierarchy each example might fit into. In order to do this, it was
useful to consult "Wordnet", which is George Miller's electronic catalogue of semantic relations. "Wordnet" can be used to establishing what an inheritance hierarchy for a particular word might look like, since it makes the hierarchy (as proposed by Miller) explicit. Below we see the taxonomy of some of the words given in the examples. (Note: Wordnet often gives many different senses of a word, but, due to space limitations, only the most relevant entries are given here).

5 senses of restoration

Sense 2

renovation, restoration -- (the state of being restored to its former good condition; "the inn was a renovation of a Colonial house")

=> improvement -- (a condition superior to an earlier condition: "the new school represents a great improvement")

=> condition, status -- (a condition or state at a particular time: "a condition (or state) of disrepair"; "the current status of the arms negotiations")

=> state -- (the way something is with respect to its main attributes; "the current state of knowledge"; "his state of health"; "in a weak financial state")
Sense 4
restoration, reclamation, renewal, revival, rehabilitation -- (the
conversion of waste land into land suitable for
use of habitation or cultivation)
=> repair, fix, fixing, mending, reparation -- (the act of
putting something in working order again)
=> improvement -- (the act of improving something: "Their
improvements increased the value of the
property")
=> change of state -- (the act of changing something
into something different in
essential characteristics)
=> change -- (the act of changing something; "the
change of government had no impact
on the economy"; "his change on
abortion cost him the election")
=> action -- (something done (usually as
opposed to something said); "there
were stories of murders and other
unnatural actions")
=> act, human action, human activity --
(something that people do or
cause to happen)

1 sense of knowledge
Sense 1
cognition, knowledge -- (the psychological result of perception and
learning and reasoning)
=> psychological feature -- (a feature of the mental life of a
living organism)
history sensel — (the aggregate of past events: "a critical time in
the school's history")

=> past, past times, yesteryear, yore -- (the time that has elapsed; "forget the past")

=> time -- (the continuum of experience in which events pass from the future through the present to the past)

=> abstraction -- (a concept formed by extracting common features from examples)

Sense 2
history, account, chronicle, story -- (a record or narrative description of past events: "a history of France"; "the story of exposure to lead")

=> record -- (anything (such as a document or a phonograph record or a photograph) providing permanent evidence of or information about past events; "the film provided a valuable record of stage techniques")

=> evidence -- (information that makes something evident)

=> information, info -- (a message received and understood that reduces the recipient's uncertainty)

=> message, content, subject matter, substance -- (what a communication that is about something is about)

=> communication -- (something that is communicated between people or groups)

=> social relation -- (a relation between living organisms; esp between people)

=> relation -- (an abstraction belonging to or characteristic of two entities or parts together)

=> abstraction -- (a concept formed by extracting common features from examples)
1 sense of museum

Sense 1
museum — (a depository for collecting and displaying objects having scientific or historical or artistic value)
 => depository, deposit, repository — (a place where things can be deposited for safekeeping)
 => facility, installation — (something created to provide a particular service; "the assembly plant is an enormous facility")
 => artifact, artefact — (a man-made object)
   => object, inanimate object, physical object — (a nonliving entity)
   => entity — (something having concrete existence; living or nonliving)

There are two important points to note from these taxonomies. First, none of the words end in a “physical object” or “entity” (except for museum, to which we return shortly) as we suggested earlier. Second, to obtain the full range of possible modifications we need to consider every node in the hierarchy, not just the maximal one. In addition, it is possible that these nodes themselves participate in related but independent hierarchies which might open the way to other unexpected modifications. For instance, one of the entries in sense 2 of history is record. While a record in this sense is an abstract notion referring to information, that information is typically implemented in some sort of physical object. Thus, history might be able to acquire some features relevant to physical objects, through this link. For instance, if one university student complains to another that he was "...forced
to read way too many pages of history last night.", then we could easily ascertain that for some reason it was necessary for him to read many pages from his history book. Thus, history inherits some features from the physical object book that happens to contain information about it. Of course, the details of this kind of "indirect" inheritance would differ from example to example. However, Pustejovsky's theory is unique in that it provides a direct way of implementing this process through his use of dot objects (see appendix A for details). Essentially, a word like history would be a dot object, meaning that it would explicitly be typed as both a physical object and a information_type. To see why this is important, consider again example (44e). We see from the Wordnet output that museum is, ultimately, a type of entity. However, the immediate superordinate is repository for things of historical or scientific value, and the knowledge which is contained in those things. This could give access to concepts in a different taxonomy. In Pustejovsky's theory the connection would be made explicit without having to derive it from the taxonomy, by specifying museum as a dot object something like physical_structure.knowledge_thing.repository. This is important because it could allow us to place constraints on the kinds of links we can make between taxonomies, while still allowing certain words to participate in orthogonal hierarchies, in a principled way. Returning to the example, we can see how a dot object works with modification. When the complex typed object is modified by architectural, it is only the second part of the type specification which is affected by the modification. That is, architectural modifies the kind of "stuff" that the museum contains, and not the physical structure itself.
This last example is important for two reasons. First, it shows that, in spite of appearances to the contrary, example (44e) is consistent with a-d, in that they all show that architectural does not modify a physical entity. Secondly, it shows the flexibility available when we allow inheritance of features along a complex taxonomic web, and when these features can selectively affect the internal constituents of a lexical representation. Thus, in a system like Pustejovsky's, the constraints on modification, though principled, are very flexible.

There is also a general point about adjectival modification that these examples clarify. Namely, what is the semantic relationship between the AP and the NP in these constructions? Jackendoff (1990) has an overly simplified treatment, where the AP is simply construed as a “property” of the NP. But we already saw that this approach is not sufficient to describe the properties of the sentences used in these experiments. In particular, the approach is at a loss to describe non-predicative adjectives whose interpretation cannot be construed as the intersection of two sets. Thus, while red car could be taken as denoting the set of entities which is the intersection of the set of “red things” and the set of “cars”, fast car cannot be interpreted as the intersection of “fast things” and “cars”, since the set of things that are to be counted as “fast” is defined in relation to the NP it modifies (see Chierchia and McConnell-Ginet, 1993), for an introduction to formal semantic treatments of adjectival modification). Pustejovsky's formalism offers a new way of describing the effects of adjectival modification, by specifying the internal composition of lexical items in great detail. Thus, the way in which
the adjective relates to the NP can be specified in exact detail, and subtle differences in this relation between different NP's can be captured. In the example above, architectural modifies restoration and knowledge in different ways, and these ways are captured in exact detail by the analysis of the internal structure of the lexical items.

Consider now what happens in the putatively uninterpretable examples. Returning to the example of architectural shirt, we first note that shirt is that it is a sort of physical object, and does not seem to be involved in independent type hierarchies. As such, it should not be modifiable by architectural, if the previous discussion is correct. However, before concluding that there is no possible way the modification could be construed, we need to look to see if some internal constituent of the lexical entry for shirt might not license some sort of modification. To this end we note that the TELIC role of shirt specifies that its purpose is that it is meant to be worn. The definition might be as follows (compare to the book example above):

$$\text{TELIC} = \text{wear}(e,w,x)$$

Thus, the purpose of shirt is to be involved in a wearing event, with a person $w$. The event is, in addition, specified as an action in the event structure. But, a human action or activity is modifiable by architectural, as we see from the analysis of restoration. Thus, architectural shirt could be taken as a shirt which is worn during activities related to architecture. If this line of reasoning is true, then it could explain why it was so difficult to obtain strong item effects in the AP/NP sentences. That is, many of the items which
were putatively anomalous, might in fact have an interpretation. And to complicate matters even more, it is possible that only some of these kinds of sentences were successfully interpreted, given the time constraints in the matching task. However, without a detailed analysis of each word combination, it is not possible to say when a modification is possible and when it is not, or at least, when it ought to be possible. This lack of homogeneity in the sentence set might be why the AP/NP construction gave such variable results in the experiments.

The experimental results, while supporting the notion of inheritance of features in a lexical taxonomy, also suggest some constraints on the process. That is, if the hypothesis explaining the results is correct, then there should be a constraint on the nodes that given other nodes can inherit from. As a concrete example, in the sentence Sincerity went for a walk, it should not be possible to construe sincerity as a kind of animate entity or, as we put it before, a [Thing]. If it were possible, then this sentence should not have taken longer to match than a plausible control. This implies that there must be orthogonal inheritance hierarchies for lexical items belonging to different ontological classes. A hierarchy that contains sincerity cannot also contain dog, nor can there be an indirect path leading from one to the other. This conflicts to some extent with our previous suggestion, that inheritance paths can “jump” hierarchies if there is a node which links two separate taxonomies. The proposed interpretation of the experimental results suggests that this is either wrong, or at least that the number and type of these links must be carefully constrained. In particular, while Vossen & Copestake (1993) might be correct in
concluding that taxonomies in lexical structure "... are more like tangled hierarchies than neat tree structures ...", it seems that the details of the "tangling" must respect the boundaries of ontological kinds.

Let us now consider a number of problems with the proposed theory. There are a number of results reported in Murray (1982), that are not easily explained by the current hypothesis. In a number of experiments which studied the effects of plausibility on sentence matching, Murray obtained strong effects with items that are highly implausible, but certainly not anomalous in the sense suggested. For instance, consider the sentences below. (For each sentence $x$, we show a sentence $x'$ that is the corresponding plausible sentence in the experiment).

(45)  

a. The book had clean coils.  

a'. The road had sharp bends.

b. The brothers returned the station.

b'. The engineer designed the machine.

c. The knights bronzed the rocks.

c'. The sailors scrubbed the deck.

d. The hydrogen had been seized.

d'. The soldier had been wounded.

One possible difference between these items and the ones reported in this thesis is the choice of baseline. Recall Ratcliff (1983) results which suggested that sentences describing
stereotyped events are processed faster than those which describe non-stereotyped, but plausible events. It seems possible that Murray’s plausible sentences were more highly stereotyped than the ones used in the current thesis, and that this was the source of the discrepancy between the results.

Another possibility arises from the fact that each of Murray’s experiments also included ungrammatical strings that were constructed by jumbling words in a sentence, as illustrated below.

(46) a. Olga snow washed the over.
    b. Was walking servant the daily.
    c. The recovered the cream railway.

None of the experiments reported in the present thesis included such items. It is possible that this difference in the design of the experiments can explain why Murray obtained effects with implausible sentences which were very similar to the ones which showed no effect here. Consider how a subject might approach the task. Initially they focus on the physical form of the stimulus, as the instructions suggest. However, they soon notice (perhaps unconsciously) that the syntactic and semantic structure of the sentences allows them to perform the task more efficiently, if they make use of such information. (Consider, for instance, what subjects would do if they were to match a set of two sentences, or an entire paragraph, at once. Surely they would be forced to rely on a semantic representation to some extent). As a consequence they no longer rely on a purely physical comparison.
By making use of the meaning of the sentence, subjects show the sorts of effects we have been discussing. However, consider what would happen to this strategy if one third of the items were structurally ill formed. Clearly, it is not advantageous to rely on the interpretation for these sentences, and perhaps subjects would become especially sensitive to any violation to the form of the sentences they were reading. In such a case, subjects might adopt a strategy where sentences without a clear and straightforward meaning might be treated as ill formed, and matched at a lower and less efficient level, essentially ignoring output from the semantic level. The implausible sentences used by Murray (1982) might therefore have been taken to be ill formed by the subjects. In fact, one aspect of the results directly supports this possibility. In approximately one half of these experiments, Murray (1982) found no difference in matching between the implausible and ungrammatical sentences. In the remainder, there was a small effect by subjects, but not by items. This is consistent with the claim that subjects treat the implausible sentences as though they were "ungrammatical". Fortunately this hypothesis can be tested very easily, by replicating Murray's experiment, and leaving out the ungrammatical sentences.

Finally, these two alternatives might both be at work. If indeed subjects are tempted to treat highly implausible sentences as if they were ill formed, then this temptation would surely be greatest when the plausible sentences were highly stereotyped. That is, the factor of stereotypy might interact with whether or not ungrammatical sentences are included in the experiment. But the most important aspect of these possibilities is the implication that subjects can alter their strategy on the basis of the
overall composition of the items, and thereby change the pattern of results in the sentence matching task.

Up until now, we have been working under the assumption that sentence matching is a task that is able to tap into various possible levels of mental representation, without subjects awareness that these levels influence the response. As Murray (1982) argues, other tasks like grammaticality judgment would not be as suitable, because the response criterion might become confounded by the experimental manipulation. That is, if subjects are told to decide if a sentence was "grammatical" or not, they might become confused about the grammatical status of grossly implausible sentences, and delay their response. However, we are now faced with the possibility that the composition of the item set does influence the way that the task is performed. This is clearly an important issue to investigate if we wish to use the sentence matching task to study the effects of a single level of representation without the influence of conscious strategies. Consider, for instance, the experiments studying the effects of various sorts of ungrammaticality on matching times. We have already considered a number of possible reasons why subjacency constraint violations do not have an effect on matching. But now, we have a further possibility. Consider the following examples from Stevenson (1984). In each case the (a) sentence is the grammatical control from which the (b) sentence is constructed.
(47) a. Which yacht has Al heard that Ed sailed?
    b. Which yacht has Al heard the rumor that Ed sailed?

(48) a. Which horse did Helen deny that she startled?
    b. Which horse did Helen deny the claim that she startled?

(49) a. What did the cook like to curry?
    b. What did the cook like the plan to curry?

In earlier discussion we claimed that the reason why there was no grammaticality effect is that in both cases, a meaning is derivable from the string. But now the opposite possibility is also suggested, if we accept the possibility that the difference in interpretability between the two kinds of sentences is not great. That is, subjects find it too difficult to interpret the constraint violation sentences and hence the lower level representations control matching times for these items (since semantic information is essentially ignored). However, the respective controls, in many cases, also sound “odd”, and are treated the same way that the constraint violation sentences are.

These considerations clearly pose a challenge to any theory of the sentence matching task. In fact, Forster (1987) considered similar challenges and raised the possibility that the controlling level in sentence matching can vary between individual subjects, and even within subjects, as a function of the particular set of experimental conditions. However, as Forster points out, this does not contradict the assumptions of the “race model” interpretation of the matching task. Instead, it cautions us to be especially
careful in interpreting the results of any single experiment, since a number of different controlling levels are potentially involved. This contradicts claims that sentence matching can provide unambiguous insight into a single and very precise level of syntactic representation (Forster & Stevenson, 1987). Instead, the task can be used to probe several levels of representation, depending on the nature of the items involved. If there are no ungrammatical sentences, then the subjects tend to focus on meaning. If there are ungrammatical strings of various sorts involved, then the subjects might not be able to use a high level semantic interpretation, and therefore make use whatever lower level structure can be induced from a given string. The nature of this representation, of course, can be influenced by the exact ungrammaticality in any given string.

It is important to decide between the alternative accounts of how ungrammatical strings are processed in the matching task. The claim, that some types of ungrammatical sentence can be assigned a semantic representation that is indistinguishable from one that is constructed for a grammatical sentence, has interesting consequences for theories of semantic interpretation. The implication for the matching task would be that a simple principle can explain the results obtained with the entire set of possible sentences. On the other hand, the possibility that the controlling level changes in response to the item set leads to more interesting theories of the relationship between the matching task and syntactic processing, but leads to no interesting claims about semantic interpretation under these circumstances.
To study the effect that ungrammatical strings may have on the matching task, a set of plausible and implausible sentences would need to be tested in two separate experiments, one of which would also include ungrammatical strings in the item set. If implausible sentences were matched slower than plausible ones only in the presence of ungrammatical strings, then this would suggest that the controlling level for the matching task does change in response to the item set. If this is true, then much of the past research into the effect of various types of ungrammaticality on the matching task could be seen in a different light. That is, the properties of intermediate levels of syntactic representation would control the matching task, but only when subjects are encouraged to ignore higher level representations. On the other hand if the manipulation failed, and Murray's implausible items were slower to match than his plausible ones in both experiments, then the question of whether or not there are levels of plausibility that interpretive processes are sensitive to, would become critical. This would imply that the plausibility manipulation in the present thesis was not strong enough. This would not be a favorable outcome for the present thesis for the additional reason that it contradicts the claim that pragmatic plausibility considerations have no consequences at all for the tasks that were investigated. The third and most favorable outcome for the ideas presented in this thesis, however, would be if the plausibility effects obtained by Murray (1982) were due to his plausible sentences being stereotyped in some sense. To test this, we could obtain ratings for Murray's sentences in a cloze task, where subjects are asked to predict the final word in sentences. If Murray's plausible sentences had a high cloze frequency, they could be
compared with sentences where the final word is replaced with equally plausible words of lower frequency. If there was a difference, then we could conclude that plausibility was not the relevant factor, and we could maintain the hypothesis that the matching task was controlled at the semantic level. This would entail, of course, more experiments to investigate the various possibilities that were suggested for how the different kinds of ungrammatical sentences could nevertheless have well formed semantic representations.

Clearly, then, the question of how ungrammatical strings in the item set affect the matching task is critical. However, we should not let this question obscure the results that were obtained in this thesis with grammatical sentences. The results support the hypothesis that the matching task is, under some circumstances at least, sensitive to high level interpretive processes. In addition, they suggest some interesting implications for the properties associated with these processes. This thesis demonstrates how experimental psychology can reveal important constraints that formal theories of representation need to address, and it also shows how these formal theories can help to clarify otherwise contradictory empirical observations.
Pustejovsky attempts to explain how a word can contribute in slightly different, but related ways to a variety of linguistic environments in which it can appear. His theory accounts for the polysemous behavior of words, without the need for stipulating a separate lexical entry for each particular meaning of that word in different contexts. There are two major aspects of the theory that allow this. First, lexical entries are represented as templates which encode various kinds of information about each word that many other theories do not, as well as the relations between various aspects of this information. Secondly, various pieces of this information to be changed independently according to the linguistic environment, by a number of well defined operations. While Pustejovsky's theory has many unique features, in many ways it is similar to theories in the computational linguistics literature where detailed lexical structures are also needed, in implementations of lexically driven theories of grammar (e.g. Saint-Dizier & Viegas, 1995). As we will soon see, one of the predominant representational devices used in these, and Pustejovsky's theory, is inheritance. Inheritance is a key process in knowledge representation systems, since it allows a way of storing information that reduces redundancy. By structuring information in a network of linked nodes and allowing nodes to inherit information from other nodes they are connected to, the more specialized
instances of a given concept do not need to store information that they have in common with a more general concept. For instance, a node labeled *robin* can be linked to one standing for *animal* via a *is-a* link. Then, the lexical representation of *robin* would not need to explicitly state that robins have blood, or internal organs, and so on, since this would be a part of our knowledge about *animal*. Inheritance mechanisms would allow the inference that these facts are true of *robin* as well. Inheritance networks have enjoyed extensive use in both psychology and artificial intelligence (Collins & Quillian, 1969; Brachman, 1985; Tourzetzky, 1986). While intuitively the process seems simple, there are a large number of technical difficulties and issues that must be faced when actually implementing an inheritance network. Discussions of many of these issues can be found in Brachman (1985), Brachman (1983), Touretzky (1986), and, concerning specifically lexical representations, Briscoe, Copestake, & De Pavia (1993).

In Pustejovsky's formalism, then, each lexical entry has information represented at four distinct levels:

a) Argument Structure specifies the number, and type of logical arguments associated with a word.
b) Event structure definition of event types which are involved with the word. These include STATE, PROCESS, TRANSITION.
c) Qualia structure specifies different "modes of explanation". This level includes information about how an object (for nouns) came into being, how things are used, relations between arguments, or events, as specified above, etc. Qualia
structure is itself divided into a number of levels, as we will see.

d) Lexical Inheritance
Structure
Specifies how a word is related to other words in the lexicon, where relationships can license the application of certain rules.

There are three rules which operate on these representations, and which are responsible for the generative nature of the lexicon:

a) Type Coercion
lexical item inherits a semantic interpretation from a governing phrase, changing its semantic, but not syntactic type.

b) Selective binding
lexical item operates specifically on some sub-structure of a phrase, without changing the overall type.

c) Co-composition
several words within a phrase behave as functors, generating new, non-lexicalized senses. Several words can simultaneously modify sub-structures of each other.

The first level of representation, the argument structure, captures a minimal but essential component of lexical semantics, and the mapping between the semantics and syntax. It describes the number and type of logical arguments participating in the semantics of a given word. Pustejovsky distinguishes between four types of arguments that lexical items can have:
a) TRUE ARGUMENTS: Syntactically realized parameters of the lexical item.

   e.g. "John arrived late."

b) DEFAULT ARGUMENTS: Arguments which are necessarily defined internally to the lexical item, but do not necessarily appear in the syntax.

   e.g. "John built the house (out of bricks)."

c) SHADOW ARGUMENTS: Arguments which appear in the semantics of the lexical item, but can only be expressed by special operations on the semantics, or discourse specification.

   e.g. "John buttered his toast with an expensive butter."

   Vs. *"John buttered his toast with butter."

d) TRUE ADJUNCTS: Parameters which modify the expression as a whole, but are not tied to the semantics of a particular lexical item.

   e.g. "John drove down to New York on Tuesday."

The majority of the theoretical claims center around the first three of these. The main point to note for our purposes is that the argument structure can specify default assignments, and therefore selectional restrictions, in a word's lexical entry. In the example we used before, for instance, the verb drink would have liquid specified as a shadow argument.

The next level to consider is the Event Structure. This is where the details of the event type of a lexical item, including any participating sub-events, are spelled out. The
three major event types considered are PROCESS, STATE, and TRANSITION. The event type is broken down into participating subevents, and the relation between them is specified. The subevents are bound to the parameters specified in the argument structure, where the specifics of the binding relation is made explicit in the Qualia structure (which will be considered soon). Pustejovsky defines the following vocabulary for designating the relations expressible in the event structure:

\[
\begin{align*}
\leq & \quad \text{means "partial order of", or, "part of" } \\
& \text{e.g. } a_1 \leq a_2 \text{ designates } a_1 \text{ is a part-of } a_2 \\
< & \quad \text{means "strict partial order of", or, "precedes" } \\
& \text{e.g. } a_1 < a_2 \text{ designates } a_1 \text{ precedes } a_2 \\
\circ & \quad \text{means "overlap"} \\
& \text{e.g. } a_1 \circ a_2 \text{ designates } a_1 \text{ overlaps with } a_2 \\
\subseteq & \quad \text{means "inclusion"} \\
& \text{e.g. } a_1 \subseteq a_2 \text{ designates } a_1 \text{ is included in } a_2 \\
* & \quad \text{means "heads"} \\
& \text{e.g. } \{a_1, a_2^*\} \text{ designates the set of events } a_1, a_2, \text{ which is headed by } a_2
\end{align*}
\]

Using this vocabulary it is possible to give a detailed account of the constituent structure of the main event types. Typically, an event type $e$ will be composed of two
subevents $e_1$ and $e_2$, which can be written with a bracket notation or drawn as a tree diagram:

$$[e_1 < e_2]$$

is equivalent to

```
  e2
_/ \
|  |
e1 e2
```

This system can be used to define a large variety of events. For instance the example above refers to an event $e$ composed of two subevents occurring in strict temporal order. Causative verbs would have event structures of this sort. For instance the verb *build* would be described as $e_1$, a *process*, strictly preceding $e_2$, a *state*. Further specification in the Argument and the Qualia structures would bind the *process* as being initiated by some animate actor, and the *state* as being the resultant end product, namely the thing that was built.

Another event structure might look like $[e_1 \circ e_2]$, which would be the event structure for verbs like *accompany*, where the two subevents overlap. The two participants in this event are bounded together for the duration of the *accompany* act.

Finally, the "*" allows us to demarcate one or both of the subevents in the structure as the *head* of the event. One should interpret this notion of *headedness* as a mechanism for focusing "interpretive importance" on members of the event structure, so
either one or both members might carry the brunt of the interpretive load. This mechanism adds greatly to the range of word meanings that can be captured effectively by the formalism. Consider the three event structures below:

(1) a. \([e_{1}\star < e_2]\) e.g. build
    b. \([e_{1} < e_2\star]\) e.g. arrive
    c. \([e_{1}\star < e_2\star]\) e.g. give

In a., the focus is on \(e_1\), since the verb is primarily describing the process that is taking place. Sentence b., on the other hand focuses on the outcome of the event, and in c., both participants have equal weight since the final state is as important to the meaning of the verb as the process which lead up to it.

The next level of representation, at which the argument and event structures are bound together and relations between them are defined, is the Qualia Structure. The qualia structure is itself composed of four main parts; CONSTITUTIVE, FORMAL, TELIC, and AGENTIVE. A brief description of these is given below. Note that not all words will have each of the qualia roles specified, since that particular role might have no contribution to the lexical item’s semantic content. Second, even though the descriptions below are given in terms appropriate to the way in which the roles might be filled for objects, the roles can also be filled in similar ways for words that do not refer to concrete entities. So, for instance, the agentine role for kill is kill_act since the act of killing was brought about by a kill_act. The qualia roles, then, are:
(a) **CONSTITUTIVE** - the relation between an object and its constituents, or proper parts. This might include the material from which it is made, its weight, or any other parts and component elements.

(b) **FORMAL** - this makes reference to the overall semantic type that the word can be categorized under. For complex types (to be defined shortly), the FORMAL qualia specifies how the elements of the complex type relate to each other.

(c) **TELIC** - the purpose and function of an object.

(d) **AGENTIVE** - factors which were involved in the creation of an object. This could refer to a creator if the object was an artifact; it could specify that the object was a natural kind; or it could specify a causal chain involved in the creation of the thing.

For an example of how qualia roles are filled, consider the word *novel*. To begin with, we can specify the **CONSTITUTIVE** role by noting that the primary component of a *novel* is its narrative content. A *novel* is distinguished by its being a kind of a *book*, which becomes its **FORMAL** quale. The purpose is to be *read*, so the **TELIC** quale becomes *reading*. Finally, it came to be through an act of *writing*, which fills the **AGENTIVE** role. To establish relations between meaning elements within a lexical item, as well as elements of lexical items in combinations, these roles are best thought of as functions which define the relations between variables. The variables themselves can then be filled by other words in an expression. So, the (partial) qualia structure for *novel* would look something like the
following where, for instance, the TELIC role defines the function *read* over the variables \( x \) and \( y \).

\[
\text{FORMAL} = \text{book}(x) \\
\text{TELIC} = \text{read}(y,x)
\]

Here, the variable \( x \) refers to the novel itself, \( y \) to someone whose identity could be filled by a person who appears in the expression. Not only would this provide a mechanism for assigning roles in simple sentences like *Mary read the novel*, but would also make it possible to interpret sentences like *Mary enjoyed the novel*. The way in which the novel was *enjoyed* by Mary is given by the TELIC role, which specifies the purpose or function of an object. In this case, a novel is *to be read*, and thus the sentence can be interpreted as *Mary enjoyed reading the novel*. Note also that the ambiguity in the sentence *Mary began the novel* can also be accounted for, since the verb *begin* can refer to either the TELIC role as before, or the AGENTIVE role, which specifies that a *book* is created through an act of *writing*. Thus, Mary could begin *reading* the novel, or she could begin *writing* it.

Below, we see an example of a well specified lexical entry, which demonstrates many important aspects of the formalism discussed thus far.
The **AGENTIVE** quale tells us that there is a *build_act* relation between an event $e$, (which here represents a process, as specified by the event structure) the syntactic subject, $arg_1$, and the syntactic object $arg_2$. The **FORMAL** quale tells us that the object ($arg_2$), an artifact, comes to be in a state, $e_2$, of existence at the end of the *build_act*. (Note that the little square notation allows us to specify an entire argument structure slot as a variable in the relations established in the **qualia** structure). Finally, $arg_2$ specifies a **CONSTITUTIVE** role which states that the object that $arg_2$ refers to is made of whatever material the default argument refers to. If no specific material is specified then it is assumed that some material, whose identity is not specified, is involved. The event structure, which should be recognizable from example (1a) indicates that there is a strict ordering between the
process, which brings about the state, and that the interpretive force is to be carried by the process that is being described.

Perhaps this is a good place to pause a while and address an apparent problem with the formalism. That is, there seems to be a circularity in definitions, where the word *build* is defined in terms of the concept *build act*. But it must be noted that the lexical entry is not meant to give an exhaustive definition of the word, as a conceptual "definition" would. In Pustejovsky's view, the role of the lexicon is to provide a set of mechanisms which allow a single template to assume a number of different but related guises in a number of linguistic environments. Pustejovsky is not obliged to posit conceptual structures which can exhaustively define meanings of words. The role of a lexical item is to establish the specific place of that item with respect to other items in lexical space, to describe mechanisms which allow the word to move in creative yet constrained ways in lexical space, and to provide "hooks" to other kinds of conceptual representations which can further specify the meaning of the word. The relationship between lexical and conceptual structures, if they are taken to be separate, is of course of great importance. But a discussion of this issue is beyond the scope of this thesis.

Another important thing to note is that a large part of the definition involves specifying its place in a *type hierarchy*. The example shows that a given lexical item can be involved with several different types. For instance, the Qualia structure specifies the overall type of the lexical item, which in this case is a verb of *creation*. The argument
structure specifies that the ARG1 is of type animate, ARG2 an artifact, and so on. Each of these concepts is a part of a hierarchy. An artifact, for instance, is a kind of object.

It is in part the typing of lexical items that gives the theory the power to explain the highly productive yet constrained nature of word use, since it allows the implementation of well behaved methods of inheritance as mentioned earlier. Inheritance allows a given lexical item to assume slightly different characteristics depending on the different contexts it can appear in. But before describing how this is achieved, it is important to see why a typing system is needed to constrain inheritance.

One of the processes available in inheritance networks is the arbitrary cancellation of features in a given feature structure. This is necessary because many categories we wish to represent include exceptions, which share most but not all of their features with the other members of the class. We want these exceptions to correctly inherit some of the features from a higher node, but not the conflicting ones. The classic example involves the penguin, which is a bird, but can't fly. The property that birds fly would be stored under the entry for bird, since typically, birds can fly, and it is convenient to store this fact in just one place and not under the entry for every different type of bird. Then, a link from the distinct bird types to the bird node would allow this specific feature to be inherited for each distinct bird from the bird node. The property can fly is a default feature that is inherited by nodes that are more specific than the one which contains the information. But penguins can't fly, and the lexical entry penguin explicitly specifies this fact. This information then blocks the inheritance of the default information from the more general
entry, for this specific lexical entry. All other defaults, however, are inherited as normal. While this seems like an attractive way to represent facts about both exceptional instances and the general case, it is in fact very problematic. Brachman (1985), for instance, points out that, in systems which allow arbitrary cancellation of features, it becomes impossible to meaningfully interpret the knowledge stored in a network. The reason is that there is nothing to disallow the possibility of representing a “fact” that a penguin is a bird which isn’t an animal and isn’t a thing. If we can cancel the first feature, then why not the second two? For related reasons, we lose all sense of constituency between concepts. For instance, we would like to be able to infer that a three legged elephant is, in actual point of fact, an elephant, since the latter concept is a constituent part of the former. But we cannot do this in a system which allows arbitrary cancellation of features. If we can arbitrarily cancel the default feature has four legs, then why can’t we also cancel is a elephant?

The point is that there must be some restriction on the kinds of information that can be canceled, but these restrictions have to be imposed on the inheritance mechanisms by other means. Unfortunately, a suitable set of constraints for a large enough piece of conceptual space has not been found. This goal would be similar to finding a distinction between necessary and non-necessary features, which has been a large enterprise in philosophy and psychology for many years, with very limited success (e.g. Fodor, 1983). In fact, one of the most successful Knowledge Representation formalisms in AI, which was spawned from such concerns, does not allow cancellation of features at all. In KL-
ONE (Brachman & Schmolze, 1985), there is a "conceptual core" which consists exclusively of definitional knowledge, and a separate sub-system is used to deal with assertions about exceptional cases.

These problems are important for implementations of lexical knowledge bases which allow inheritance, such as the one we are considering. If the system was allowed to arbitrarily cancel features, then there could not be a notion of selectional restrictions, for instance, since we could just cancel offending features if the situation demanded it.

One solution that is adopted in many theories of lexical structure, including Pustejovsky's, is the introduction of semantic types to constrain inheritance. Semantic types form the "backbone" of the lexicon by providing a hierarchically structured set of nodes, around which lexical items are specified. The hierarchy of types is established in terms of an order of specialization. For instance a brick might be typed as a sort of construction material which in turn can, ultimately, be regarded a special case of the type physical object. The type hierarchy determines where a given concept can inherit information from. Importantly, the knowledge stored in the type hierarchy is non-defeasable (Copestake, Sanfilippo, Briscoe, & De Pavia, 1993). Thus, it would not be possible to construe of a brick that was not also a physical object. But the type system only captures certain key aspects of our knowledge. The distinctions made possible by the type system are not sufficient to capture the richness of all the words in the lexicon. Another mechanism is needed to allow the inheritance of non-typed information. Of course this would also involved many complex processes fraught with problems. However
we will not discuss these here, since the two important points can already be made. First, there is a distinction made between *defeasible* and *non-defeasible* information in a given lexical item. Importantly, the kind of knowledge involved in a typing system involves a restricted set of concepts, much like Jackendoff's (1983) notion of conceptual constituents. These can be seen as placing absolute requirements on the semantics of the word. Secondly, the typing system determines whether or not a lexical item can inherit from some given other node. For instance, a *bird* can inherit from both *animal* and *physical_object*, but not from *animal* and *artifact*, since these are incompatible types. Once again, this places absolute constraints on the ways in which words can be used.

There is also a very important innovation in the way that Pustejovsky uses type hierarchies that differentiates his work from many others in the computational linguistics literature. The methods of typed inheritance we have seen thus far, allow a given node to inherit from any set of compatible nodes in just a single type hierarchy. However, Pustejovsky notes that, many times, lexical items seem to have distinct (but related) senses which allows them to participate in one or other of a number of different type structures. For instance, in a previous example we were content type a *novel* as a kind of *book*. But now consider how we extend the type *book*? Is it a *thing*? Is it an *object*? Is it both of these? While these possibilities all seem to be true, any one of them fails to capture the full set of linguistic extensions available to the lexical item. Consider the following sentences:

(2)  
  a. The book was very thick.  
b. The cake was very thick.
c. The book was about the war.

d. ?The cake was about the war.

Sentences a and b have similar interpretations because they both make reference to physical objects, which can have thickness predicated of them as a property. However, while sentence c is perfectly acceptable, d is meaningless. Thus, while in one context book and cake have similar contributions to meaning, in other contexts they don't. The reason for this seems to be that book can accrue properties by being a physical object just like cake, but it is not limited to being an object. More precisely, it refers to a physical object which is also "something that holds information". Thus, while a cake can't be about anything, a book, which is also a holder of information can be "about" whatever its information content describes. What is needed is a notation that can represent things which have multiple types associated with them.

In order to capture this observation, Pustejovsky introduces the notion of dotted types. A dotted type refers to a way of combining two types. However, a dotted type is very different from a more standard combination of feature structures obtained through conjunction or unification (Briscoe et al., 1993), where the features of two types are combined to produce a third, more specific type. Dotted types, in contrast, are not more specific than their constituents. They are new structures which retain fully the types of both constituents, but add to this a combination of them. They can thus participate in the inheritance hierarchy of either of the parent nodes, depending on the circumstance in which they are used. Let us look at a few examples.
(3)  
a) Mary broke the window.
b) John crawled through the window.

In the first sentence window refers to the actual physical object that can be broken, the second to an aperture which can be climbed through. The two distinct types for window, then, are phys_obj and aperture. In order to capture both possibilities in a single entry we form the dotted type: phys_obj.aperture. The complete lcp (Lexical Conceptual Paradigm = Lexical Item) would then have the complex type \{phys_obj.aperture, phys_obj, aperture\}. The lexical item is identified with the type that happens to be appropriate in a given context, enabling it to inherit information from, and unify with, other entries in the same type hierarchy. In addition dotted types can express a third meaning that is literally the combination the other two interpretations. For instance, consider the complex dotted type construction. This can refer either to the process, to the end result of the process, or to the dotted combination of the two which would be interpreted as something like "the entire event of building, from beginning to end". The three sentences below illustrate each of these possible interpretations:

(4)  
a) The construction was arduous and tedious.
b) The construction is standing on the next street.
c) The house’s construction was finished in two months.
These, then, are the main representational elements of the generative lexicon. Let us now see how the generative rules work on them to determine the interpretation of sentences. Since lexical items have a rich internal structure, the generative rules can operate independently on the various constituent structures. For instance, a given lexical item is not restricted to inheritance of properties along a single pathway, but can simultaneously participate in a number of hierarchies, which allows for inheritance along a number of orthogonal dimensions at the same time. We have seen a similar property entailed by dotted types. Let us consider the generative rules which the rich internal structure of lexical items, and the inheritance structures between them. There are three such rules:

a) Type Coercion - lexical item inherits a semantic interpretation from a governing phrase

b) Co-composition - several elements within a phrase behave as functors, generating new, non-lexicalized senses. That is, lexical items can enrich the semantic specifications of other words.

c) Selective Binding - lexical item operates on a specific part of the sub-structure of a phrase, without changing the overall type of the phrase

Type coercion is the mechanism whereby lexical items can be used in a semantic combination for which they are not specifically typed. Consider first the special case of Subtype coercion. This is a process which allows a word to be placed in a position that is normally reserved for its superordinate. For instance, consider the sentence "Mary drives a
Honda to work.” Now, one might imagine that the verb *drive* selects for a *vehicle* in its argument structure, as shown here:

\[
\begin{align*}
\text{drive} & \quad \text{EVENTSTR} = \begin{cases} 
E_1 = e_1: \text{process} \\
E_2 = e_2: \text{process} \\
\text{RESTR} = < 
\end{cases} \\
\text{ARGSTR} & = \begin{cases} 
\text{ARG1} = x: \text{human} \\
\text{ARG2} = y: \text{vehicle} 
\end{cases} \\
\text{QUALIA} & = \begin{cases} 
\text{FORMAL} = \text{move} (e_1, y) \\
\text{AGENTIVE} = \text{drive} \_ \text{act} (e_1, x, y) 
\end{cases}
\end{align*}
\]

In order to say “Mary drives a car to work”, *car* would need to be typed as a *vehicle*, since that is what is required by the argument structure of *drive*. Thus the lexical entry of *car* would include something like the following information:

\[
\begin{align*}
\text{car} & \quad \text{ARGSTR} = \begin{cases} 
\text{ARG1} = x: \text{vehicle} 
\end{cases} \\
\text{QUALIA} & = \begin{cases} 
\text{FORMAL} = \text{move} (e_1, y) \\
\text{AGENTIVE} = \text{drive} \_ \text{act} (e_1, x, y) 
\end{cases}
\end{align*}
\]

To accommodate the original sentence, *Honda* would be typed as a *car*, which as we have seen is typed, in turn, as a *vehicle*. This would license the insertion as before, since we have established the inheritance path Honda \(\supset\) car \(\supset\) vehicle.
Now we are ready to consider the full fledged *True Complement Coercion*, where a lexical item can actually change its semantic type because of its environment. Consider the following sentences.

(5)    a) John began *reading a book*.

b) John began *a book*.

Suppose that the lexical entry for *begin* is something like the following, and that the required complement of the verb is an *event*:

```
begin
EVENTSTR = [E1 = transition
             E2 = transition
             RESTR = <]

ARGSTR = [ARG1 = x: human
           ARG2 = e2]

QUALIA = [FORMAL = P(e2, x)
           AGENTIVE = bake_act(e1, b, e2)]
```
In sentence (5a), the type is satisfied, and the phrase can be constructed. But what of (b)?

The lexical entry for book is:

```
book
ARGSTR =
ARG1 = x: info
ARG2 = y: physobj
QUALIA =
info.physobj.lcp
FORMAL = book (x)
TELIC = read (y,x)
```

In order to bind this entry into a slot that expects an event, there must be able to coerce the lexical item to assume an event interpretation. Thus, the fact that begin requires a particular semantic type will result in book changing its semantic type, in this particular composition. In this particular case it is easy to see how the coercion takes place. The qualia structure of book makes two event readings available. First, the TELIC role gives us read, and the AGENT role gives us write, each of which are associated with event readings. Either one of these can be used as the event associated with book, and will therefore satisfy the typing requirement, giving the two possible interpretations.

It is important to note that this type of coercion can only occur if there is an appropriate type available in the qualia of a given lexical item. In this example there were two events described in the qualia structure. If there were no such events, then coercion would not be possible. However, there is a very large space of possible compositions in
which coercion is made possible, because of the inheritance mechanisms which allow items
to assume the types of other items in a given hierarchy. For instance, it is possible that
*Complement Coercion* could be applied after first applying *Subtype Coercion*. For
example, we could say “John began *the duck*”, and easily interpret it to mean that he
began *eating* it. *Duck* is a sub-type of *animal* which is a subtype of *food*, for which the
TELIC role would involve the act of *eating*. But for other compositions this particular
type path might be irrelevant, and some other superordinate might become relevant. For
instance, consider the dialogue: “*John went to see the new Donald Duck movie. He
decided that he hates that stupid Duck*”. Here, the meaning in the composition would be
obtained through some path involving *cartoon characters*. The point is, once again, that
lexical items are not limited to being construed as belonging to only a single *type*. This
vast and rich pattern of possible interrelations is partly responsible for the infinitely rich
productivity of lexical meaning.

As a final example of the subtleties of complement coercion, and one which shows
how other cognitive abilities connect to lexical processes, consider the sentence “*Mary
believes John.*” Now, *believe* takes an object of type *proposition*, and *John* does not have
such a role in its qualia, nor is there an obvious subtyping path to a word that does. The
process that needs to be undertaken is an interesting interplay between inferential
processes which have “real world knowledge” as their domain, and lexical knowledge.

First, since *believe* requires a *proposition* as complement, we must try and coerce
the existing complement into such a reading. However, purely lexical processes fail to
achieve this goal for the reasons outlined above. But the lexical entry for the word believe needs a proposition as a complement, and this proposition needs to be in some relation to John. However, the lexical entry cannot tell what that relation is, or where to find it in all conceivable sentences. It just tells us that there must be one if the composition is to be valid, and possible syntactic positions in which it might normally be found. But in non-canonical expressions, it is up to “higher” cognitive processes to tell us how to find it. In the present example, given what we know about humans (and John being human), we know that it is possible to have such a proposition related to John through some form of communication. Thus it is inferred that Mary believed some proposition that John communicated to her. In this example we see that the complete meaning cannot always be given by the lexicon. However, the conceptual work undertaken to form an interpretation was driven by the expectancy for a proposition complement by the verb. That is, the lexicon had a critical role in guiding inferential processes.

These examples show the power and subtlety that is possible with this system. But, it is important to keep in mind that the mechanistic processes responsible for sentence interpretations are precisely defined, and have precise logical limitations. Perhaps most importantly, it must be appreciated that interpretation does not proceed in ad hoc and random ways. There are not infinite and unpredictable ways to interpret every new lexical combination (contrary to, say, Fodor, 1983). Instead, interpretation is guided by the very same information and processes that words require for their uses in canonical environments.
The next mechanism to consider is co-composition. Co-composition refers to an operation where details of the role specifications of a lexical item are significantly altered by other words in a phrase. Unlike with coercion operations, the semantic type of the complement is not forced to change; it is merely complemented. As an illustration consider the following:

(6)  
   a) John baked the potato.  
   b) John baked the cake.

The first sentence captures a change of state in which John's actions merely alter the state of the entity in (somewhat) superficial ways. On the other hand, sentence b. describes a creative act, where the cake is actually created from scratch. The complement is able to change the meaning of the verb, in this example by changing the event type that is specified in the event structure. Let us see how this is achieved by looking at the structure of the verb bake:
This reflects the “state change” sense of the verb, which is taken to be the canonical definition. The AGENTIVE role of *potato* specifies that it is a natural kind, which can undergo a state change as required. On the other hand, *cake* is an artifact which is created through some sort of human act. This is encoded in the qualia structure of *cake*, which has the following entry:

\[
\text{AGENTIVE} = \text{bake}\_\text{act}(e_1, w, y),
\]

which can unify with the AGENTIVE role of *bake*. As a result of this unification, the feature structures of the verb are modified through a process referred to as *function application with qualia unification* (Pustejovsky, 1995, p. 124). The end result of this process is the augmented representation shown below:
Note that the event structure has changed into one that specifies a new *state* as the end result of the process, which is bound by the FORMAL Quale to the second argument, an *artifact*, in the argument structure.

The final generative mechanism in the lexicon is *Selective Binding*. This is perhaps the most straightforward of the three operations. Its primary role is to assign a sense in which adjectives modify their compliments. Thus, phrases like *red car* would involve selective binding. But this mechanism is especially useful for adjectives like *fast*, which can add unique contributions to each complement they modify. These cannot be taken as simple predicative adjectives, unlike adjectives like *red*, whose contribution to the interpretation of a phrase in which it appears is more straightforward. If we have the
phrase *red cars*, for instance, then its interpretation can be stated in terms of set intersections: \( \{ \text{red things} \} \cap \{ \text{cars} \} \). However, the interpretation of the phrase *fast cars* cannot be so simple, because what *fast things are* cannot be independently established. *Fast* is a relative term whose interpretation depends on whatever it is modifying. Its interpretation **depends** on the identity of the syntactic head it is in construction with.

The mechanism of *selective binding* allows adjectives to affect meaning in the required way, by modifying just one substructure within a lexical item, rather than the whole item itself. Consider in somewhat more detail the adjective *fast*, and how it modifies the noun *car*. Recall that the qualia structure of *car* includes the TELIC role \( \text{drive}(e,y,x) \), where \( e \) is an event, \( x \) is bound to the car, and \( y \) to the person driving the car. It is this TELIC role that *fast* modifies, such that the events involved with driving occur *fast*. But what about *fast game*? This is somewhat more difficult, since there is a real ambiguity, between a *game that is played fast*, and a *game that is over very quickly*. The first of these is a straightforwardly similar to the *fast car* example, with the TELIC role this time being \( \text{play}(e',y,x) \). The second, however, requires some steps in the derivation. Intuitively, the phrase refers to a time period, whose passage is *fast*. What is modified in this example is not the TELIC role, but the FORMAL role, if we assume that this encodes something like \( \text{play_event}(e',y,x) \), and it is the event as a whole which is *fast*.

The previous example shows, in addition, that adjectives can modify several of the qualia representations, not just TELIC. Other adjectives like *expensive*, or *opaque*, which refer to objects rather than events or activities that the objects are associated with, also
modify the \textsc{formal} quale. Thus the adjective must specify what semantic type it modifies, and this type must be available in the noun that is being modified. As we saw, \textit{fast} modifies events, which are represented as described in the various nouns. If there are two events being described, then the modification might be ambiguous. If no events are specified, the modification fails.

We can now summarize the main points that this brief overview of the theory leaves us with. We have seen some of the reasons why this theoretical approach turns out to be so powerful. First, by dividing lexical items into a number of independent but connected levels, and those levels into their constituent parts, we are able to see how words can change aspects of their meaning, while still maintaining their semantic integrity; only parts of the representation change, with the bulk remaining constant. The result of this is that while two uses of a given word might have somewhat different contributions to meaning in two different phrases, these meanings are related in intuitively powerful ways. The theory can capture this fact by showing how the separate meanings derive from the very same lexical "template". Another consequence of the decomposition is that it allows words to partake in several orthogonal lattice structures which describe independent relationships between words. Any given word can inherit features from different sets of these hierarchies, depending on the context. And finally, the rules themselves are made powerful enough to bring about the necessary changes within a word that can explain their expressive flexibility. Yet, as noted earlier, while the system allows for tremendous generativity, it is one that is based on strict rules, and obeys well defined constraints. The
approach begins to bring order to the otherwise chaotic state of affairs that philosophers like Fodor (1983) have despaired over for many years
Appendix B

Instructions for the plausibility rating task.

You are about to see a number of sentences. Some of the sentences describe very ordinary events in the world. Some of them, however, describe strange and unusual circumstances. I want you to read each of the sentences, and write a number between 1 and 7, describing how "implausible" or "strange" the event that the sentence describes, is. Please assign the number according to the following scale:

1 2 3 4 5 6 7
Completely normal Strange Very strange

Here are a couple of examples:

"The dog goes for a walk every day." would be a 1.

"John adores Cathy so much, he buys her diamond rings every morning." would be about 2.7, say.

"Colorless green ideas sleep furiously." would be a 7.

Make sure you use the whole range of numbers. Good luck!
Appendix C

Items from experiment 1. Each experimental file is shown separately for all experiments, to show the counterbalancing. The conditions are anomalous, implausible, and plausible.

File 1

Anomalous

The woman liked Peter's senile shirt.
Ships entered the Socratic harbor at a slow speed.
Crops could not be planted in the shallow climate.
The yacht rolled continuously on the cozy ocean.
Last night Helen wore her genetic dress to dinner.
Last night Karen wore her secretive necklace from Spain.
His wife was annoyed by the wealthy typewriter.
In Chicago Alice bought a desperate camera.
After she stood on the slow nail she swore.
Factories with grammatical machines have increased profits.
The barber disliked cutting hair with early scissors.
The temples in Greece are a sticky spectacle.
Jane slept under the energetic blankets until noon.
Sue went to buy some spectral cushions.
He helped the tinted student before the examination.
Her painfully shady glands are a sign of infection.

Implausible

Clive wanted to read the furry contract before signing.
John invited the electronic girl to the party.
Hikers believe that walking snakes are a hazard.
Simon was unable to eat the elastic cheese.
Without any rain they had mirrored tanks all summer.
Andrew struggled through the tiny jungle for days.
To reach the summit they scaled a glass cliff.
In the mountains the explorers discovered simmering tribes.
Yesterday Ruth ate a frosty apple for lunch.
The patients were upset about the synthetic music.
Jane's two sour horses require careful treatment.
Alice was excited by the portable opera.
The driver had difficulty lifting Tom's juicy luggage.
The woman prepared some sandwiches for her moldy children.
He was allergic to the concrete flowers.
People crossing thorny rivers must take great care.
Plausible

Those old sentimental songs can make Jan cry.
The agent showed Joan the vacant house.
Charities appreciate wealthy people who donate money.
Sonia baked a moist cake with raisins and nuts.
Ann's old rubber gloves had many holes.
Pam usually stayed in a comfortable hotel after traveling.
Michael often carried his black umbrella to work.
Phillip swept the broken plate into a bin.
The messenger delivered the urgent telegram to Joanne.
To ride stubborn donkeys requires skill and patience.
The plow was pulled by the powerful tractor.
That old nylon carpet will burn easily.
We were surrounded by the large hostile crowd.
He gave the money to a grateful beggar.
The most violent prisoners are difficult to control.
The teachers strict rules frightened the children.

Anomalous

Those old tertiary songs can make Jan cry.
The agent showed Joan the prompt house.
Charities appreciate eastward people who donate money.
Sonia baked a calisthenic cake with raisins and nuts.
Ann's old corrupt gloves had many holes.
Pam usually stayed in a mutual hotel after traveling.
Michael often carried his constant umbrella to work.
Phillip swept the fluent plate into a bin.
The messenger delivered the hospitable telegram to Joanne.
The plough was pulled by the hourly tractor.
To ride architectural donkeys requires skill and patience.
That old collapsed carpet will burn easily.
We were surrounded by the large indirect crowd.
He gave the money to a populous beggar.
The most transmitted prisoners are difficult to control.
The teacher's greasy rules frightened the children.
Implausible

The woman liked Peter's timber shirt.
Ships entered the emerald harbor at a slow speed.
Crops could not be planted in the hectic climate.
The yacht rolled continuously on the pink ocean.
Last night Helen wore her ceramic dress to dinner.
Last night Karen wore her woolen necklace from Spain.
His wife was annoyed by the concrete typewriter.
In Chicago Alice bought a salty camera.
After she stood on the hairy nail she swore.
Factories with soft machines have increased profits.
The barber disliked cutting hair with glass scissors.
The temples in Greece are a raunchy spectacle.
Jane slept under the leafy blankets until noon.
Sue went to buy some gravel cushions.
He helped the melting student before the examination.
Her painfully exploding glands are a sign of infection.

Plausible

Clive wanted to read the legal contract before signing.
John invited the attractive girl to the party.
Hikers believe that poisonous snakes are a hazard.
Simon was unable to eat the smelly cheese.
Without any rain they had empty tanks all summer.
Andrew struggled through the tropical jungle for days.
To reach the summit they scaled a vertical cliff.
In the mountains the explorers discovered primitive tribes.
Yesterday Ruth ate a sweet apple for lunch.
The patients were upset about the loud music.
Jane's two lame horses require careful treatment.
Alice was excited by the dramatic opera.
The driver had difficulty lifting Tom's heavy luggage.
The woman prepared some sandwiches for her hungry children.
He was allergic to the fragrant flowers.
People crossing swift rivers must take great care.
**Anomalous**

Clive wanted to read the handicapped contract before signing.
John invited the transplanted girl to the party.
Hikers believe that automatic snakes are a hazard.
Simon was unable to eat the absurd cheese.
Without any rain they had quick tanks all summer.
Andrew struggled through the surgical jungle for days.
To reach the summit they scaled a faithful cliff.
In the mountains the explorers discovered spurious tribes.
Yesterday Ruth ate a foggy apple for lunch.
The patients were upset about the oily music.
Jane's two cryptic horses require careful treatment.
Alice was excited by the spilled opera.
The driver had difficulty lifting Tom's diluted luggage.
The woman prepared some sandwiches for her constant children.
He was allergic to the patron flowers.
People crossing libelous rivers must take great care.

**Implausible**

Those old transparent songs can make Jan cry.
The agent showed Joan the singing house.
Charities appreciate ceramic people who donate money.
Sonia baked a silent cake with raisins and nuts.
Ann's old liquid gloves had many holes.
Pam usually stayed in a rusty hotel after traveling.
Michael often carried his spicy umbrella to work.
Phillip swept the sliced plate into a bin.
The messenger delivered the bronze telegram to Joanne.
To ride luminous donkeys requires skill and patience.
The plough was pulled by the feather tractor.
That old glass carpet will burn easily.
We were surrounded by the large plastic crowd.
He gave the money to a magnetic beggar.
The most compact prisoners are difficult to control.
The teacher's growing rules frightened the children.
Plausible

The woman liked Peter's cotton shirt.
Ships entered the shallow harbor at a slow speed.
Crops could not be planted in the arid climate.
The yacht rolled continuously on the stormy ocean.
Last night Helen wore her elegant dress to dinner.
Last night Karen wore her intricate necklace from Spain.
His wife was annoyed by the noisy typewriter.
In Chicago Alice bought an expensive camera.
After she stood on the rusty nail she swore.
Factories with efficient machines have increased profits.
The barber disliked cutting hair with blunt scissors.
The temples in Greece are a unique spectacle.
Jane slept under the warm blankets until noon.
Sue went to buy some velvet cushions.
He helped the nervous student before the examination.
Her painfully swollen glands are a sign of infection.
Experiment 2

File 1

Anomalous

The contract that Clive signed was calisthenic and binding.
The girl who was transplanted liked to sing.
Snakes which are written are a serious hazard.
The cheese which Simon refused was very syncopated.
The tanks for rain water were quick all summer.
The jungle in Brazil is surgical and dense.
The cliff which the climbers scaled was almost transmitted.
The tribes which the explorers discovered were spurious.
The apple Ruth ate was foggy and crisp.
The patients enjoyed music unless it was oily.
Horses that are cryptic need careful treatment.
Opera excited Alice when it was spilled.
The luggage which Tom carried was quite mutual.
She prepared sandwiches for her children who were erased.
Those flowers are both beautiful and patrons.
Rivers that are libelous are dangerous to cross.

Implausible

Songs which are transparent can make Jan cry.
The house which Joan liked was singing.
People who are ceramic often donate money.
The cake which Sonia baked was silent and sweet.
Gloves made from liquid can tear easily.
The hotel Pam stayed in was rusty and cheap.
The umbrella Michael carried was spicy and large.
The plate which was sliced belonged to his mother.
The telegram Joanne received was bronze and brief.
Donkeys which are lucinous are difficult to ride.
The tractor which pulled the plough was feathered.
Carpet made from glass will burn easily.
The crowd which marched towards us was plastic.
The beggar who received the money was magnetic.
Prisoners who are compact are difficult to control.
Rules which are growing can frighten children.
Plausible

Peter's shirt was made of cotton.
The harbor was too shallow for the large ships.
The climate was unfavorable because it was so arid.
The ocean frightened Thomas when it was stormy.
The dress which was the most elegant was Helen's.
The necklace which Karen wore was intricate and expensive.
He disliked the typewriter because it was noisy.
The camera which Alice bought was expensive.
The nail she stood on was rusty and bent.
Machines which are efficient can greatly increase profits.
The scissors which the barber disliked were blunt.
The spectacle of the Greek temples is unique.
The blankets Jane used were warm and soft.
Sue bought new cushions covered with velvet.
The student who was nervous missed the examination.
Glands which are swollen are a sign of infection.

File 2
Anomalous

Peter's shirt was hyperbolic.
The harbor was too Socratic for the large ships.
The climate was unfavorable because it was so expensive.
The ocean frightened Thomas when it was subtracted.
The dress which was the most genetic was Helen's.
The necklace which Karen wore was secluded and expensive.
He disliked the typewriter because it was clustered.
The camera which Alice bought was habitual.
The nail she stood on was slow and bent.
Machines which are grammatical can greatly increase profits.
The scissors which the barber disliked were early.
The spectacle of the Greek temples is sticky.
The blankets Jane used were catastrophic and soft.
She bought new cushions which were spectral.
The student who was tinted missed the examination.
Glands which are shady are a sign of infection.
Implausible

The contract that Clive signed was furry and binding.
The girl who was electronic liked to sing.
Snakes which are walking are a serious hazard.
The cheese which Simon refused was very elastic.
The tanks for rain water were mirrored all summer.
The jungle in Brazil is tiny and dense.
The cliff which the climbers scaled was almost glass.
The tribes which the explorers discovered were simmering.
The apple Ruth ate was frosty and crisp.
The patients enjoyed music unless it was synthetic.
Horses that are sour need careful treatment.
Opera excited Alice when it was portable.
The luggage which Tom carried was extremely juicy.
She prepared sandwiches for her children who were moldy.
Those flowers are both beautiful and concrete.
Rivers that are thorny are dangerous to cross.

Plausible

Songs which are sentimental can make Jan cry.
The house which Joan liked was vacant.
People who are wealthy often donate money.
The cake which Sonia baked was moist and sweet.
Gloves made from rubber can tear easily.
The hotel Pam stayed in was comfortable and cheap.
The umbrella Michael carried was black and large.
The plate which was broken belonged to his mother.
The telegram Joanne received was urgent and brief.
Donkeys which are stubborn are difficult to ride.
The tractor which pulled the plough was powerful.
Carpet made from nylon will burn easily.
The crowd which marched towards us was hostile.
The beggar who received the money was grateful.
Prisoners who are violent are difficult to control.
Rules which are strict can frighten children.
File 3

Anomalous

Songs which are tertiary can make Jan cry.
The house which Joan liked was syncopated.
People who are eastward often donate money.
The cake which Sonia baked was calisthenic and sweet.
Gloves which are ambient can tear easily.
The hotel Pam stayed in was mutual and cheap.
The umbrella Michael carried was written and large.
The plate which was spilled belonged to his mother.
The telegram Joanne received was rainy and brief.
Donkeys which are architectural are difficult to ride.
The tractor which pulled the plough was hourly.
Carpet which collapsed will burn easily.
The crowd which marched towards us was indirect.
The beggar who received the money was populous.
Prisoners who are transmitted are difficult to control.
Rules which are greasy can frighten children.

Implausible

Peter's shirt was made of timber.
The harbor was too emerald for the large ships.
The climate was unfavorable because it was so hectic.
The ocean frightened Thomas when it was pink.
The dress which was the most ceramic was Helen's.
The necklace which Karen wore was woolen and expensive.
He disliked the typewriter because it was concrete.
The camera which Alice bought was salty.
The nail she stood on was hairy and bent.
Machines which are soft can greatly increase profits.
The scissors which the barber disliked were glassy.
The spectacle of the Greek temples is raunchy.
The blankets Jane used were leafy and soft.
She bought new cushions covered with gravel.
The student who was melting missed the examination.
Glands which are exploding are a sign of infection.
Plausible

The contract that Clive signed was legal and binding.
The girl who was attractive liked to sing.
Snakes which are poisonous are a serious hazard.
The cheese which Simon refused was very smelly.
The tanks for rain water were empty all summer.
The jungle in Brazil is tropical and dense.
The cliff which the climbers scaled was almost vertical.
The tribes which the explorers discovered were primitive.
The apple Ruth ate was sweet and crisp.
The patients enjoyed music unless it was loud.
Horses that are lame need careful treatment.
Opera excited Alice when it was dramatic.
The luggage which Tom carried was extremely heavy.
She prepared sandwiches for her children who were hungry.
Those flowers are both beautiful and fragrant.
Rivers that are swift are dangerous to cross.
Experiment 3

File 1

*Plausible, adjectival modifier*

Finally John was satisfied with his new comfortable pillow.
Some old cabinets were built to last forever.
There was a beautiful clean counter in his kitchen.
Peter was very jealous of Tim's new luxury house.
Thick strong pillars were needed to support the tall building.
It is good to have dark lenses in sunglasses.
Jane always made sure she carried a soft handkerchief.
Heavy bottles break easily when they are dropped.
The government banned dangerous toys for young children.
The woman bought a beautiful plate at the store.
Yesterday John ate a tasty salad for lunch.
He asked the woman to look after his large suitcase.

*Plausible, NP modifier*

The concrete pipe burst from the bitter cold.
That is the most valuable gold watch ever made.
John's plastic telephone broke at the party.
There was a terrible shortage of leather shoes.
Helen brought her plastic mug to work.
The new silk jackets are machine washable.
Everyone loves a delicious chocolate cake for dessert.
Jeff's wife disliked his worn out canvas sneakers.
Rich people buy diamond rings for their spouses.
The wooden mast on the deck was a real hazard.
Hot coffee in paper cups can be hard to carry.
Steve helped build a wooden cross for the church.
Implausible material

Passengers never like riding on the old cement train carriages. Plastic sandwiches are always nice to have for lunch. People buying woolen necklaces must take great care. In winter it is always nice to have some steel socks. Timber shirts feel sensual and decadent to wear. Classic books with carbon bindings are a pleasure to own. People with babies are happy for titanium diapers. Cement tortillas are a very versatile snack food. The horse with the timber saddle won the race. Grocery stores now offer the choice of stone bags. Silk cables are needed to support heavy loads. The man with the flour gloves cleaned the bathroom.

Selectional restriction violation

Complacent rugs are excellent for home decoration. Richard dropped the feverish bowling ball onto his foot. The painting was mounted onto the anxious frame. Some trusting knives are very difficult to sharpen. The mother put a frightened mat in front of the door. Last night they filled up the large clever tank. The hungry trophy was placed right on the pedestal. Mary's old obsessed bracelet needed some restoration. Some people don't approve of talking pizza. John's admiring dollar is worth a lot more these days. There was a useful mourning table in the back garden. Stacey was always rather fond of stubborn jeans.
Turkish rugs are excellent for home decoration.
Richard dropped the heavy bowling ball onto his foot.
The painting was mounted onto the antique frame.
Some kitchen knives are very difficult to sharpen.
The mother put a large mat in front of the door.
Last night they filled up the large gas tank.
The victory trophy was placed right on the pedestal.
Mary's old dirty bracelet needed some restoration.
Some people don't approve of vegetarian pizza.
John's historic dollar is worth a lot more these days.
There was a useful picnic table in the back garden.
Stacey was always rather fond of tight jeans.

Finally John was satisfied with his new feather pillow.
Some wooden cabinets were built to last forever.
There was a beautiful marble counter in his kitchen.
Peter was very jealous of Tim's new brick house.
Thick concrete pillars were needed to support the tall building.
It is good to have glass lenses in sunglasses.
Jane always made sure she carried a lace handkerchief.
Glass bottles break easily when they are dropped.
The government banned metallic toys for young children.
The woman bought a ceramic plate at the store.
Yesterday John ate a tuna salad for lunch.
He asked the woman to look after his leather suitcase.
Implausible material

The feather pipe burst from the bitter cold.
That is the most valuable cotton watch ever made.
John's glass telephone broke at the party.
There was a terrible shortage of brass shoes.
Helen brought her nylon mug to work.
The new ivory jackets are machine washable.
Everyone loves a delicious wooden cake for dessert.
Jeff's wife disliked his worn out emerald sneakers.
Rich people buy cardboard rings for their spouses.
The elastic mast on the deck was a real hazard.
Hot coffee in brick cups can be hard to carry.
Steve helped build a uranium cross for the church.

Selectional restriction violation

Passengers never like riding on the old vomiting train carriages.
Unhappy sandwiches are always nice to have for lunch.
People buying playful necklaces must take great care.
In winter it is always nice to have some angry socks.
Trusting shirts feel sensual and decadent to wear.
Classic books with bored bindings are a pleasure to own.
People with babies are happy for amused diapers.
Pregnant tortillas are a very versatile snack food.
The horse with the surprised saddle won the race.
Grocery stores now offer the choice of envious bags.
Frightened cables are needed to support heavy loads.
The man with the excited gloves cleaned the bathroom.
Files

Plausible, adjectival modifier

Passengers never like riding on the old noisy train carriages. Tasty sandwiches are always nice to have for lunch. People buying expensive necklaces must take great care. In winter it is always nice to have some warm socks. Soft shirts feel sensual and decadent to wear. Classic books with opulent bindings are a pleasure to own. People with babies are happy for disposable diapers. Soft tortillas are a very versatile snack food. The horse with the western saddle won the race. Grocery stores now offer the choice of recycled bags. Strong cables are needed to support heavy loads. The man with the waterproof gloves cleaned the bathroom.
**Plausible, NP modifier**
Woolen rugs are excellent for home decoration.
Richard dropped the marble bowling ball onto his foot.
The painting was mounted onto the wooden frame.
Some steel knives are very difficult to sharpen.
The mother put a rubber mat in front of the door.
Last night they filled up the large steel tank.
The gold trophy was placed right on the pedestal.
Mary's old silver bracelet needed some restoration.
Some people don't approve of chicken pizza.
John's silver dollar is worth a lot more these days.
There was a useful plastic table in the back garden.
Stacey was always rather fond of denim jeans.

**Implausible material**
Finally John was satisfied with his new concrete pillow.
Some paper cabinets were built to last forever.
There was a beautiful woolen counter in his kitchen.
Peter was very jealous of Tim's new rayon house.
Thick foam pillars were needed to support the tall building.
It is good to have paper lenses in sunglasses.
Jane always made sure she carried a brass handkerchief.
Vinyl bottles break easily when they are dropped.
The government banned diamond toys for young children.
The woman bought a furry plate at the store.
Yesterday John ate a silver salad for lunch.
He asked the woman to look after his liquid suitcase.

**Selectional restriction violation**
The pregnant pipe burst from the bitter cold.
That is the most valuable thinking watch ever made.
John's mourning telephone broke at the party.
There was a terrible shortage of angry shoes.
Helen brought her envious mug to work.
The new happy jackets are machine washable.
Everyone loves a delicious happy cake for dessert.
Jeff's wife disliked his worn out grumpy sneakers.
Rich people buy sarcastic rings for their spouses.
The loving mast on the deck was a real hazard.
Hot coffee in sad cups can be hard to carry.
Steve helped build a thirsty cross for the church.
File 3

*Plausible, adjectival modifier*

The old pipe burst from the bitter cold.
That is the most valuable antique watch ever made.
John's black telephone broke at the party.
There was a terrible shortage of good shoes.
Helen brought her souveneir mug to work.
The new designer jackets are machine washable.
Everyone loves a delicious sweet cake for dessert.
Jeff's wife disliked his worn out smelly sneakers.
Rich people buy expensive rings for their spouses.
The rotted mast on the deck was a real hazard.
Hot coffee in disposable cups can be hard to carry.
Steve helped build a new cross for the church.

*Plausible, NP modifier*

Passengers never like riding on the old wooden train carriages.
Turkey sandwiches are always nice to have for lunch.
People buying pearl necklaces must take great care.
In winter it is always nice to have some wool socks.
Silk shirts feel sensual and decadent to wear.
Classic books with leather bindings are a pleasure to own.
People wath babies are happy for cotton diapers.
Flour tortillas are a very versatile snack food.
The horse with the fiberglass saddle won the race.
Grocery stores now offer the choice of paper bags.
Steel cables are needed to support heavy loads.
The man with the rubber gloves cleaned the bathroom.

*Implausible material*

Icy rugs are excellent for home decoration.
Richard dropped the platinum bowling ball onto his foot.
The painting was mounted onto the polyester frame.
Some cement knives are very difficult to sharpen.
The mother put a feather mat in front of the door.
Last night they filled up the large rayon tank.
The bitumen trophy was placed right on the pedestal.
Mary's old tar bracelet needed some restoration.
Some people don't approve of lead pizza.
John's leather dollar is worth a lot more these days.
There was a useful oat bran table in the back garden.
Stacey was always rather fond of aluminum jeans.

Selectional restriction violation

Finally John was satisfied with his new thoughtful pillow.
Some angry cabinets were built to last forever.
There was a beautiful pregnant counter in his kitchen.
Peter was very jealous of Tim's new talkative house.
Thick friendly pillars were needed to support the tall building.
It is good to have happy lenses in sunglasses.
Jane always made sure she carried a sleeping handkerchief.
Hungry bottles break easily when they are dropped.
The government banned envious toys for young children.
The woman bought a thoughtful plate at the store.
Yesterday John ate a singing salad for lunch.
He asked the woman to look after his laughing suitcase.
Experiment 4

Anomalous

He was given the hyperbolic handkerchief for his wedding anniversary.
Last night at Albertson's, subtracted hinges were on sale.
Yesterday afternoon Tom made the genetic fence.
I believe no one who admired the grammatical meringue.
John noticed that the spectral surgeon was overweight.
Within a matter of minutes the clustered cat was hungry.
She watched the hyphenated leopard feed the ravenous cubs.
At the theater the audience applauded the syllogistic dancer.
The misaligned fuel stained the concrete floor in the garage.
The existential milk was removed by the health department.

Implausible

The happy retaliation for the dreadful lie was unexpected.
Sally finally admitted that the rubber bowl was disgraceful.
The pleasant wound from the knife battle was bathed.
The note said that the heavy gas required attention.
He lived in a porous apartment on the East side.
The advertisement said that periodic health should be maintained.
The fragrant charcoal in the cold fireplace was annoying.
After he got very old he always avoided quiet exertion.
According to the rules the steel cards were discarded.
Jane always hated the way the wet odor filled the room.

Plausible

Yesterday afternoon the fruity syrup splashed onto the table.
The government was forced to fund the architectural restoration.
After the rescue, the resuscitated diver was hungry.
People worked on the terraced slopes every hour of the day.
The protruding handle on the ugly green door was red.
There was a gradual cessation of applause after the long speech.
The narrated incident in today's newspaper was humorous.
He dropped a ceramic dish while emptying the dishwasher.
The engraved emeralds were cracked during the daring robbery.
Her scholastic proficiency was renowned through the entire village.
Anomalous

Yesterday afternoon the architectural syrup splashed onto the table. The government was forced to fund the fruity restoration. After the rescue, the terraced diver was hungry. People worked on the resuscitated slopes every hour of the day. The gradual handle on the ugly green door was red. There was a protruding cessation of applause after the long speech. The ceramic incident in today's newspaper was humorous. He dropped a narrated dish while emptying the dishwasher. The scholastic emeralds were cracked during the daring robbery. Her engraved proficiency was renowned through the entire village.

Implausible

He was given the brass handkerchief for his wedding anniversary. Last night at Albertson's, lace hinges were on sale. Yesterday afternoon Tom made the strawberry fence. I believe no one who admired the aluminum meringue. John noticed that the moulting surgeon was overweight. Within a matter of minutes the swearing cat was hungry. She watched the skating leopard feed the ravenous cubs. At the theater the audience applauded the purring dancer. The curdled fuel stained the concrete floor in the garage. The flaming milk was removed by the health department.

Plausible

The sudden retaliation for the dreadful lie was unexpected. Sally finally admitted that the chipped bowl was disgraceful. The jagged wound from the knife battle was bathed. The note said that the flammable gas required attention. He lived in a roomy apartment on the East side. The advertisement said that bodily health should be maintained. The smudged charcoal in the cold fireplace was annoying. After he got very old he always avoided extreme exertion. According to the rules the bent cards were discarded. Jane always hated the way the gassy odor filled the room.
File 3
Anomalous

The chipped retaliation was for the dreadful lie unexpected.  
Sally finally admitted that the ambiguous bowl was disgraceful.  
The inverted wound from the knife battle was bathed.  
The note said that the jagged gas required attention.  
He lived in a bodily apartment on the East side.  
The advertisement said that roomy health should be maintained.  
The extreme charcoal in the cold fireplace was annoying.  
After he got very old he always avoided smudged exertion.  
According to the rules the gassy cards were discarded.  
Jane always hated the way the bent odor filled the room.

Implausible

Yesterday afternoon the salty syrup splashed onto the table.  
The government was forced to fund the animal restoration.  
After the rescue, the crippled diver was hungry.  
People worked on the slimy slopes every hour of the day.  
The tiny handle on the ugly green door was red.  
There was a very cute cessation of applause after the long speech.  
The skinny incident in today's newspaper was humorous.  
He dropped a leather dish while emptying the dishwasher.  
The softened emeralds were cracked during the daring robbery.  
Her forced proficiency was renowned through the entire village.

Plausible

He was given the lace handkerchief for his wedding anniversary.  
Last night at Albertson's, brass hinges were on sale.  
Yesterday afternoon Tom made the aluminum fence.  
I believe no one who admired the strawberry meringue.  
John noticed that the clever surgeon was overweight.  
Within a matter of minutes the moulting cat was hungry.  
She watched the purring leopard feed the ravenous cubs.  
At the theater the audience applauded the skating dancer.  
The flaming fuel stained the concrete floor in the garage.  
The curdled milk was removed by the health department.
Experiment 6

File 1

*Anomalous*

Happiness climbed the meeting very quickly and gracefully.
Solitude knows that it is not easy to convert ability into white rice.
The paint transformed an ordinary yawn into a beautiful harvest.
Last night sincerity entered the color and then left in a multiplication.
Despite the weather, sarcasm kept the meeting at subservience.
As a party trick Harry sometimes drank sincerity from a crystal imagination.
The advice went home after the end of the cucumber.
The problem ran around under the accident all night long.
Yesterday we saw the predicament jump over the dream and land with a thump.
It is annoying when sorrow falls onto the vandalism in the kitchen.

*Implausible*

As often happens, cranberries thoughtlessly put butter on top of the mouse.
The disobedient crayon clawed the expensive finger.
It is always a pleasure to eat charcoal at a nice mortuary.
Toilets often move the time of the meeting from dusk to dawn.
In spite of the recent challenge, condiments still own turtles in America.
Sausages sold the movie set to rabbits for fifty dollars.
Sue carefully placed mercury inside the book on the kitten.
The plant decided to feed the kettle some carrots.
He received admiration after demanding a refund for the hippopotamus.
The river is where oranges saw cockpits float downstream towards the needle.

*Plausible*

The explorers found that students hiked over the rocks very easily.
The dog suddenly changed from a sweetheart into a monster.
Stockholders decided to trade the stocks for some bonds.
Finally the gorilla walked into the trap and screamed.
The salesman said that the computers ranged from affordable to expensive.
The customer decided to pay a lot of money for the car.
The un-ethical physician was sued for habitual malpractice.
Mike read that athletes swam the charmel in just three hours.
Every afternoon the man exits the elevator with a big frown on his face.
Never loan a piece of jewelry to a dishonest woman if you want to get it back.
File 2

Anomalous

As often happens, eternity thoughtlessly put color on top of the prayer.
The disobedient revival clawed the expensive enigma.
It is always a pleasure to eat anomaly at a nice allegation.
Theories often move the time of the meeting from red to oil.
In spite of the recent challenge, six o'clock still owns reciprocity in America.
Courtesy sold the square root to happiness for five dollars.
Sue carefully placed persuasion inside the drive on the miracle.
The boredom decided to feed the thought some volume.
He received shallowness after demanding a refund for the serenity.
The river is where disaster saw subtraction float downstream towards the story.

Implausible

The explorers found that suitcases hiked over the pond very easily.
The dog suddenly changed from a genius into a fanatic.
Hunchbacks decided to trade the cathedrals for some gypsies.
Finally the television walked into the salad and smiled.
The coconut said that the computers ranged from devilish to hilarious.
The cottage decided to pay a lot of money for the feather.
The un-ethical fountain was sued for habitual barking.
Mike read that bicycles swam the channel in just three weeks.
Every afternoon the chair exits the cup with a big hairball on its legs.
Never loan a piece of a cat to a very large flea if you want to get it back

Plausible

Mountaineers climbed the cliff very quickly and gracefully.
Authorities know that it is not easy to convert criminals into good citizens.
The paint transformed an ordinary house into a beautiful home.
Last night politicians entered the room and then left in a great hurry.
Despite the weather, everyone kept the meeting at six o'clock.
As a party trick Harry sometimes drank ketchup from a crystal goblet.
The dancers went home after the end of the ball.
The mouse ran around under the table all night long.
Yesterday we saw the kangaroo jump over fence, and land with a thump.
It is annoying when sugar falls onto the floor in the kitchen.
Anomalous

The explorers found that proficiency hiked over the angle very easily. The dog suddenly changed from a slogan into a meeting. Disobedience decided to trade the distance for some apologies. Finally the experience walked into the departure and wept. The celebration said that the computers ranged from misfortune to communism. The destruction decided to pay a lot of boredom for the conflict. The un-ethical legibility was sued for habitual caricatures.
Mike read that uniformity swam the frustration in just three cats.
Every afternoon disaster exits the secret with a big product on its face. Never loan a piece of confusion to an educated guess if you want to get it back.

Implausible

Matches climbed the building very quickly and gracefully.
Lawnmowers know that it is not easy to convert water into gasoline. The paint transformed an ordinary sausage into a beautiful sofa. Last night cucumbers entered the room and then left in a hovercraft. Despite the weather, pencils kept the meeting at lunch-time. As a party trick Harry sometimes drank vomit from a crystal thimble. The worms went home after the end of the explosion.
The computer ran around under the plate all night long. Yesterday we saw the automobile jump over the moon, and land with a thump. It is annoying when sludge falls onto the couch in the kitchen.

Plausible

As often happens, mother thoughtlessly put oil on top of the salad. The disobedient kitten clawed the expensive couch. It is always a pleasure to eat lobster at a nice restaurant. Managers often move the time of the meeting from morning to night. In spite of the recent challenge, government still owns timberlands in America. Marianne sold the old books to students for fifty dollars. Sue carefully placed tin foil inside the cooker on the cabinet. The woman decided to feed the dog some red meat. He received compensation after demanding a refund for the television. The river is where tourists saw crocodiles float downstream towards the village.
Experiment 7. For this experiment, the pairs of files are collapsed, since the two versions differ only by one word. Both words are given, separated by a slash.

File 1 and 2

*Anomalous*

Happiness climbed the banquet/dinner very quickly and gracefully.
Solitude knows that it is not easy to convert discretion/capability into white rice.
One word transformed an ordinary yawn/cough into a harvest.
Sincerity entered the comfort/convenience and then left in a multiplication finally.
Despite the weather, sarcasm/anger kept the meeting at subservience.
As a trick at parties, Harry sometimes drank innocence/sincerity from a crystal imagination.
The guidance/advice went home after the end of the cucumber last night.
The problem ran around under the catastrophe/accident all night long.
Yesterday we saw the predicament/difficulty jump over the dream and land with a thump.
Ed is annoyed when sorrow falls onto the devastation/vandalism in the kitchen.

*Implausible*

As often happens, cranberries thoughtlessly put comfort/convenience ahead of gratification.
The sarcastic/angry crayon clawed the expensive finger.
It is always tedious/boring to eat charcoal at a nice mortuary.
Toilets often move the site of the occasion/event from dusk to dawn.
Condiments still possess reservoirs/computers in America.
Sausages sold the canal/channel to rabbits for fifty dollars.
Sue carefully placed colors/hues inside the book on the kitten.
The plant decided to feed the kettle some expensive/precious carrots.
He got happiness/despair after demanding a refund for the hippopotamus.
Oranges saw barriers/fences float downstream towards the needle.
Plausible

The explorers found that students hiked the distance/stretch very easily.
The dog suddenly changed from coughing/yawning to biting.
Stockholders decided to trade the product/stocks for some bonds.
Finally the gorilla walked into the confusion/commotion and groaned.
The painter said that the hues/colors ranged from red to green.
The customer decided to pay a lot of money for the great comfort/convenience after all.
The unethical physician was sued for medical stupidity/disasters last week.
Mike read that athletes swam the channel/canal in just three hours.
Every afternoon the man exits the elevator with much despair/happiness on his face.
Never reveal a secret/mystery to a dishonest woman if you want to keep it.

File 3 and 4

Anomalous

As often happens, eternity thoughtlessly put color/hue ahead of the prayer.
The disobedient revival clawed the expensive/precious enigma.
It is always a pleasure to eat questions/anomaly at a nice allegation.
Theories/ideas often move the site of the meeting from red to oil.
Six o'clock still possesses compensation/reciprocity in America.
Courtesy sold the square root to despair/happiness for five dollars.
Sue carefully placed the canal/channel inside the drive on the miracle.
The boredom/tedium decided to feed the thought some volume.
He received shallowness after demanding a refund for the moderation/serenity.
Stupidity/incompetence saw subtraction float downstream towards the story.

Implausible

The explorers found that suitcases hiked over the pond with great difficulty/predicament.
The dog suddenly changed from being stupid/incompetent into a fanatic.
Hunchbacks decided to trade the ideas/theories for some gypsies.
Finally the television walked into the beer/wine and smiled.
The coconut said that the entertainment ranged from vandalism/devastation to torture.
The kitten/dog decided to pay a lot of money for the feather.
The unethical fountain was sued for occasional excitement/happiness last week.
Mike read that bicycles swam the tub/pool in just three weeks.
Every afternoon the chair exits the cup with a big secret/mystery on its mind.
Never reveal an question/anomaly to a very large flea if you want to keep it.
Plausible

Mountaineers climbed the cliff very excitedly/blissfully and gracefully. Authorities know that it is not easy to convert violence into moderation/serenity. One word transformed an ordinary sentence into a tough anomaly/question. Politicians entered the room and then left in great predicament/difficulty after the conference. Despite the tedium/boredom, everyone kept the meeting at six o'clock. As a trick at parties Harry sometimes drank warm wine/beer from a crystal goblet. The dancers went home after the end of the occasion/event last night. The mouse ran around under the pool/tub all night long. Yesterday we saw the kangaroo jump over the barrier/fence and land with a thump. Ed acts angry/sarcastic when sugar falls onto the floor in the kitchen.

File 5 and 6

Anomalous

The explorers found that proficiency hiked over the mystery/secret very easily. The dog/kitten suddenly changed from being a slogan into a meeting. Disobedience decided to trade the stretch/distance for some apologies. Finally the event/occasion walked into the departure and wept. The celebration said that the beer/wine ranged from misfortune to communism. The destruction decided to pay a lot of excitement/happiness for the conflict. The unethical legibility was sued for habitual fences/barriers last week. Mike read that uniformity swam the frustration in just three tubs/pools yesterday. Every afternoon disaster exits the secret with a big product/stock on its face. Never reveal a confusion/commotion to an educated guess if you want to keep it.

Implausible

Matches climbed the building very capably/discretely and gracefully. Lawnmowers know that it is not easy to convert sincerity/innocence into pleasure. One wish transformed an ordinary yawn/cough into a roar. Cucumbers entered the room and then left in confusion/commotion after the conference. Despite the stretch/distance the pencils kept the meeting at lunch time. As a trick at dinners/banquets Harry sometimes drank vomit from a crystal thimble. The worms went home after the end of the useless guidance/advice last night. The computer ran around under the products/stocks all night long. Yesterday we saw the automobile jump over the moon, and land with moderation/serenity. It is an catastrophe/accident when sludge falls onto the worm in the kitchen.
As often happens mother thoughtlessly put innocence/sincerity ahead of prosperity.
The disobedient kitten/dog clawed the expensive couch.
It is always a pleasure to eat lobster at a nice banquet/dinner.
Nasty people often move the site of the defestation/vandalism from city to city.
Government still gives advice/guidance in America.
The scientist sold his old ideas/theories to students for fifty dollars.
Sue capably/discretely placed tin foil inside the cooker on the cabinet.
The woman decided to feed the dog some precious/expensive red meat.
He obtained compensation/reciprocity after demanding a refund for the television.
Tourists saw wreckage float downstream from the accident/catastrophe.
REFERENCES


