Downstream effects of word frequency.

Item Type	open;article;thesis
Authors	Slattery, Timothy J.
DOI	10.7275/7765833
Download date	2025-03-26 05:51:13
Link to Item	https://hdl.handle.net/20.500.14394/46003

DOWNSTREAM EFFECTS OF WORD FREQUENCY

A Thesis Presented

by

TIMOTHY J. SLATTERY

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTERS OF SCIENCE

May 2005

Psychology

DOWNSTREAM EFFECTS OF WORD FREQUENCY

A Thesis Presented

by

TIMOTHY J. SLATTERY

Approved as to style and content by:

Alexander Pollatsek

Alexander Pollatsek, Chair

Keith R.

Keith Rayner, Member

Don Fisher

~~Don Fisher~~, Member

Melinda Novak

Melinda Novak, Department Chairperson
Department of Psychology

TABLE OF CONTENTS

Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION

As you read silently to yourself, it is likely that you still hear an inner voice reading the words. It is also likely that you subjectively notice pauses in this inner voice stream that correspond to the ends of sentences. As it turns out, it is not only your inner voice that pauses at the ends of sentences, but your eyes as well. Rayner et al. (1989) used eye tracking to show that a word at the end of a sentence is looked at longer than if the same word had appeared earlier in the sentence. This effect was dubbed the *sentence wrap-up effect*. Rayner et al. (1989) postulated that this effect was due to the need to complete any unfinished processing before moving on to the next sentence.

As I will show in this thesis, there is quite a bit of evidence that suggests the sentence and the clause are both important psychological units in language processing. After reviewing this evidence, I will elaborate my position on what the sentence warp-up effect may be in terms of language processing. Then, in the experiment that follows, I will explore the sentence wrap-up effect in attempt to see if the amount of processing early in a sentence can affect the magnitude of sentence wrap-up. That is, if the sentence contains more difficult words, do the eyes pause longer at the sentence end before moving on to the next sentence? Additionally, does sentence processing difficulty carry over to following text, and if so, what would it mean? Finally, the method of manipulating sentence difficulty in this experiment is a novel one and future application of this manipulation in research will be discussed. But before moving on to discuss sentence wrap-up, it will be helpful to give a quick review of some of the theories and models of working memory in sentence processing.

CHAPTER 2

WORKING MEMORY AND SENTENCE PROCESSING

Working memory has been theorized to play a role in many of the tasks that cognitive psychologists study today. In fact, it may be impossible to find a task in which working memory, in some way, hasn't been postulated to have an effect. This is true in the realm of sentence comprehension, where many models of reading comprehension have been postulated that heavily rely on working memory (Just and Carpenter 1992, Waters and Caplan 1996, MacDonald and Christiansen 2002).

Just and Carpenter's 1992 capacity theory of comprehension predicts that individual differences in working memory, as assessed by the Daneman and Carpenter reading span task, can predict individual differences in reading comprehension. The Daneman and Carpenter reading span task requires people to read sets of unrelated sentences and then recall the last word from each sentence in the set. The sets differ in the number of sentences they contain and an individuals reading span score is calculated based on the set sizes they are able to remember completely. This measure has been shown to correlate with reading comprehension measures. For instance, correlations between reading span and verbal SAT are between .5 and .6 in some studies (Daneman & Carpenter, 1980; Masson & Miller, 1983). Additionally, Just and Carpenter claim that the reading span measure was highly correlated (.7-.9) with the ability to correctly answer a question about a passage in various studies.

In the Just and Carpenter capacity theory, people with a higher reading span measure are assumed to have a larger capacity (or more available activation) than participants with a lower reading span measure. The capacity is used to maintain elements in working memory and create the linguistic connections between elements that are necessary to understand the meaning of a sentence. When a person runs out of capacity, they suffer memory decay and lengthened reading times. Therefore, high span individuals would be expected to have improved sentence comprehension relative to low span individuals, just as the correlational evidence suggests.

Most of the research used to support the capacity theory used the "self paced reading" technique. In this technique, a sentence is presented on a computer screen to the reader piece by piece and the reader must push a button to proceed to the next piece—problems with this technique will be discussed later in this paper. Just and Carpenter present self paced reading data claiming to show that high span, but not low span, participants are able to use apparently semantic information (animacy of the head noun) to help disambiguate a "garden path" sentence (i.e., a sentence that initially leads the reader to construct one syntactic structure, but the sentence ends up having a different syntactic structure). The example they used came from Ferreira and Clifton (1986). Ferreira and Clifton used sentences like the following:

1. The evidence examined by the lawyer shocked the jury.
2. The defendant examined by the lawyer shocked the jury.

Sentence one is not a garden path sentence, because the head noun is inanimate. But in sentence two, there is a temporary syntactic ambiguity. A reader is faced with two interpretations of the phrase "The defendant examined"—either the defendant had examined something, or the defendant was examined. According to Just and Carpenter, high span participants are able to use the inanimate head noun to help with their syntactic parse because they have enough capacity to do so, while low span participants do not.

Just and Carpenter also predicted that when an ambiguity is resolved in favor of the more frequent interpretation, high span participants will take longer to read the sentence than low span participants. Their reasoning is that, in these cases, high span participants have the capacity to maintain multiple interpretations in working memory while low span participants can only maintain one (the correct one in these sentences). The additional load of two interpretations on the high span participants causes them to move slower than the low span participants. McDonald, Carpenter and Just (1992) used sentences such as 3-6 below to test this hypothesis.

3. The experienced soldiers warned about the dangers before the midnight raid.
4. The experienced soldiers spoke about the dangers before the midnight raid.
5. The experienced soldiers warned about the dangers conducted the midnight raid.
6. The experienced soldiers who were told about the dangers conducted the midnight raid.

Sentences 3 and 5 above are temporarily ambiguous at the verb "warned". Readers could interpret this as the main verb as in sentence 3 or as a participle as in sentence 5 which is a garden path sentence. McDonald et al. claim that interpreting "warned" as the main verb of the sentence is the preferred interpretation and the one that low span participants will chose. However, according to their theory, high span participants will create and maintain both interpretations until the ambiguity is resolved. This should cause high span participants to read sentence 3 slower than low span participants. They claim that this is the case but offer only a graph showing slower reaction time for high span participants relative to low and medium span participants as proof—without any statistical tests. According to their theory high span participants should have had an advantage over low span participants when reading sentences like 5. In these sentences high span but not low span participants should have the correct interpretation available when they reach the disambiguating information. Again, Just and Carpenter state that high span readers are better at answering comprehension questions about these sentences but again, they failed to provide any statistical tests to back this claim up. Another prediction of the Just and Carpenter model is that an extrinsic memory load will affect low span participants more than high span participants. Again, they gave self paced reading evidence showing just this type of interaction, but failed to give any statistical evidence.

The Just and Carpenter model sparked an argument that has persisted in the literature for a decade now. In 1996, Waters and Caplan wrote a very negative critique of Just and Carpenter's capacity theory. They criticized Just and Carpenter's use of statistical evidence (or lack thereof) as well as the reliability of the Daneman and Carpenter reading span task. However, it appears that the main objection of Waters and Caplan is that the capacity theory assumes that readers use the same working memory resources to parse a sentence as they use "…for conscious, controlled verbally mediated processes." According to Waters and Caplan, the working memory involved in sentence interpretation is completely separate from the working memory involved in holding a digit or word load in some phonological loop for later recall. They site evidence showing that stroke victims and Alzheimer's patients, who have limited short-term and working memory, are still able to compute meaning from many different syntactic structures despite having very low scores on the Daneman and Carpenter reading span task (Martin, 1993; Waters, Caplan, & Hildebrandt, 1991; Rochon, Waters, & Caplan, 1994).

Just and Carpenter's (1992) data suggested that high span readers were capable of using the animacy of a noun to disambiguate a syntactic ambiguity. According to Waters and Caplan this should not be possible due to the modular nature of language comprehension. In their re-analysis of the Just and Carpenter data, they showed that while high span participants were able to benefit from the inanimate noun condition, the animacy effect did not interact with syntactic complexity. When they compared the reading time data for inanimate nouns in the ambiguous vs. the unambiguous sentence types, it is clear that high span readers were still led down the garden path to the same extent. It would seem from this that language comprehension still has a modular nature.

Waters and Caplan also attacked the data from the McDonald et al. (1992) study of syntactic ambiguity (sentences 3-6 above). They stated that they had been unable to replicate the McDonald et al. finding that high span participants read sentences like 3 slower than low span participants. They also pointed out that the McDonald et al. results for the garden path sentences such as 5, showing that high span participants read slower but were more accurate than low span participants, were not statistically significant. Additionally, had the results been significant, they would have been just as easily explained by a speed accuracy trade off. Clearly, the evidence in favor of the Just and Carpenter model was not airtight.

In the Waters and Caplan view of language comprehension, it is necessary to differentiate the processes that are interpretive (and therefore under the automatic and modular language working memory) from what they term post-interpretive processes, which can interact with other working memory resources. However, their attempt to make this distinction was less than successful. It should also be pointed out that while Waters and Caplan have a theory of working memory for sentence comprehension; it is not a computational model. This fact, coupled with the lack of a distinct boundary between what is interpretive (and therefore modular) and what is post interpretive, make it very difficult to test the claims made by Waters and Caplan.

While these two camps continued to debate these issues in the literature (Just and Carpenter 1996, Waters and Caplan 1999), MacDonald and Christiansen added another model to the fray. MacDonald and Christiansen (2002) felt that the language processing differences that exist between individuals and arise through an interaction of biological and experiential factors could be better modeled using a connectionist network. They argued that there is no distinction between working memory and knowledge in their connectionist network model. This is probably the most profound difference between their view and the views of the other camps. In both the Just and Carpenter model and the Waters and Caplan theory, working memory is a space for elements to be placed and for processing to be conducted on these elements. However, the elements themselves are considered 'knowledge' as are the productions that need to be carried out to complete processing. This means that it should be possible to damage either the knowledge or the working memory while leaving the other component intact. In the MacDonald and Christiansen model this is not the case. The network is both the working memory and the knowledge. In order to simulate the high span and low span readers, they implement their network with different levels of training—the more training the network has the higher its "span". With these networks, they simulated the human data from King and Just (1991). In King and Just (1991), subjects performed a self-paced reading task. The stimuli used were subject relative (7) or object relative (8) sentences.

7. The clerk that insulted the shopper fled the store.

8. The clerk that the shopper insulted fled the store.

In King and Just's analysis of the reading times for the main verb, they found shorter reading times for high span participants than low span participants, shorter reading times within subject than within object relatives, and an interaction of reading span with sentence type. MacDonald and Christiansen's model was able to capture all three of these effects. As MacDonald and Christiansen point out, their model is an improvement over the capacity theory of Just and Carpenter which was unable to simulate the interaction between reading span and sentence type.

When looking to explain individual differences in language comprehension, all these models differ in theoretically important ways. However, for the sake of this study, their similarities are even more important—all of these theories assume that working memory has, in some sense, a limited capacity. And this assumption appears to be the norm in the working memory literature (Baddeley & Hitch, 1974; Kintsch, 1998; Kaakinen, Hyöna, & Keenan 2003). Language comprehension should also face constraints due to this limited capacity. These constraints should force a comprehender to clear or compact the elements in working memory from time to time or face mistakes in comprehension as the elements are forgotten.

All of the models of working memory discussed so far deal with fairly short sentences. While, none of the models explicitly dealt with the issue of clearing working memory or making a more compact representation of its contents to allow for incoming information, they do make certain assumptions about the clearing of working memory. In these models, there is zero memory for a sentence once the simulation is done—with the exception of the connectionist network during training. Therefore, the models assume that working memory is cleared after every sentence. However, these models were all dealing with short sentences. In written language, it is theoretically possible to string a single sentence out indefinitely by embedding clauses within each other (e.g. The daughter of the General who the soldier driving the jeep with the flat…). Sentences like this are rare (thank goodness!) as people have a great deal of difficulty correctly interpreting them. One very plausible reason for this difficulty – and one that is assumed in some way by all the models discussed – is that working memory gets overloaded. Let's take the example sentence from above but give it an ending: "The daughter of the general who the soldier driving the jeep with the flat saluted ate breakfast." In order to parse this sentence correctly, it is necessary to hold all of the sentence's elements in some sort of working memory until the very end. Only then can you correctly assign the general's daughter as the one who ate breakfast. As already mentioned, people have a hard time correctly grasping sentences like these, which suggests that it is possible to overwhelm working memory. This is not surprising nor would it be difficult to model. For instance, the capacity theory of comprehension would predict that as a sentence grew in length, a person would become increasingly likely to forget some of the elements held in working memory thus causing the comprehension of the sentence to be poor. What is surprising is that many very long sentences, such as the previous one, can be understood with only a modest amount of effort. It would appear that a reader does not necessarily have to wait until the end of a sentence in order to clear or compact the elements in working memory.

The notion of clearing or compacting the elements in working memory to make room for new information is not novel. In fact, Miller suggested it as far back as 1957. Since then, numerous studies have attempted to shed light on the topic. Many have provided valuable insights into when working memory is compacted and what information if any is lost during compaction.

CHAPTER 3

SPOKEN LANGUAGE COMPREHENSION AND THE CLAUSE

Let me begin by discussing studies of spoken language comprehension before moving on to discuss language comprehension in reading. Early studies by Bever and Fodor (1965) showed the importance of sentence structure. Participants listened to recordings of spoken sentences containing a non-speech sound (such as a beep) and were then asked to recall where the sound had occurred in the sentence. Interestingly, the participant's perception of the non-speech sound tended to migrate to clause breaks. Subsequent studies showed that, while listening to speech, the end of a clause or sentence can affect a number of attention-mediated tasks such as lengthening the reaction time to "clicks" (Abrams and Bever 1969), slowing the orienting response to shock (Bever et al. 1969), making discrimination of tones (Holmes and Forster 1972), and "click" detection (Bever, Hurtig, and Handel 1975) more difficult, and was also shown to suppress evoked potentials (Seitz 1972). It was hypothesized that the loss of attention at the ends of clauses and sentences was due to recoding the information into a more compact representation.

Carroll and Tanenhaus (1978) argued against the simplified notion that the clause was necessarily the unit of segmentation, and they differentiated between functionally complete clauses and functionally incomplete clauses.

They defined a functionally complete clause as one that contains a participant, verb, and object such as:

Mark punched Bob when Bob called him stupid.

Here the words in italics represent a functional clause. However, the following would be considered functionally incomplete:

After falling Jane went to the emergency room.

Participants listened to the sentences which contained a tone, much like Bever and Fodor (1965) and were asked to recall where they had heard the tone. Their perception of the tone tended migrated to the clause boundaries like it had in the Bever and Fodor study and this effect was greater for functionally complete clauses. So it would seem that not all clauses are created equal, and that functionally complete clauses are more apt to draw attention than incomplete clauses. This suggests that recoding may be more likely to occur once all the elements (participant, verb, and object) are present.

In sum, the above data suggest that recoding may take place at specific times during language comprehension, most notably at the ends of functionally complete clauses and sentences. One important question that remains is: if such a recoding takes place, how is the representation changed and what aspects of the information (if any) are lost? Jarvella (1971) showed that recall from two adjacent clauses was worse when the clauses came from different sentences. He had participants listen to two 1500 word passages. Three adjacent target clauses were imbedded within the passages. These clauses were termed the <u>context</u>, *previous,* and **immediate** clauses and were heard in that order. The critical difference between items was whether the sentence boundary occurred between the previous and immediate clause (termed the short condition) or whether the sentence boundary was between the context and the previous clause (termed the long condition).

- Long:"...The tone of the document was threatening. *Having failed to disprove the charges,* **Taylor was later fired by the President**."
- Short:"...The document had also blamed him for *having failed to disprove the charges.* **Taylor was later fired by the President**."

As participants listened to the passages there was a test pause at different places in the target region and the participants were to write down as much of the previous material word for word as they could remember. The important result for this discussion is that when the test pause occurred in the immediate clause (the last of the three clauses), memory for the previous clause was better in the long condition – where it occurred in the same sentence as the immediate clause. Also, when the test pause occurred in the previous clause (middle target clause) memory for the context clause (first target clause) was better in the short condition where the context clause and previous clause were in the same sentence. So memory for two adjacent clauses is dependent on whether the two clauses are part of the same sentence or different sentences.

What can be taken from these results is that following the end of the sentence, verbatim memory for a clause gets worse. Thus it appears that information is lost during sentence recoding in exchange for freeing up of working memory resources. But what kind of information is lost? There is evidence that one aspect of the sentence that is lost is its syntax (Sachs 1967). Sachs had participants listen to passages in which one of four versions of a target sentence was imbedded. These sentences had the same general meaning though they had different syntax. The passages terminated at different intervals from the target sentence and a test sentence was then presented to participants. The participant's task was to judge whether or not the sentence had appeared in the original passage. The test sentence could be either identical to the original target sentence, or different from it in one of three ways: it could have a different meaning, it could be changed from an active to a passive construction while maintaining the same meaning, or it could have an altered form which maintained meaning and deep structure. When the test sentence was presented immediately after the target sentence, participants were quite accurate at rejecting all three of the altered versions of the test sentence. However, at longer intervals they became much more likely to accept all but the meaning altered versions of the incorrect test sentences. This finding indicates that their memory for the syntax of the sentence was impaired much more than their memory for the sentences meaning.

CHAPTER 4

EYE-TRACKING METHODOLOGY AND IMPORTANT FINDINGS

All of the studies discussed to this point have dealt with spoken language comprehension. A critical question for this paper is how these results transfer to the study of reading. There are many parallels between reading comprehension and spoken language comprehension but do clauses and sentences have the same special status for recoding during reading as they do during listening tasks?

The formal cognitive study of reading is over a century old and has shed light on many important aspects of reading. As with any research field, the advances in technology have provided powerful tools for investigation. In the field of reading research, advances in eye-tracking technology have been paramount. The coupling of computers with eye-trackers in the 1970's started a revolution in reading research and remains today as the leading methodology for the study of psycholinguistics.

There are a number of reasons for eye-tracking's prevalence in the study of reading. First and foremost in my opinion is that it allows the study of natural silent reading of text. Prior to eye-tracking, reading research consisted of a number of artificial laboratory tasks such as self paced reading, lexical decision, and rapid serial visual presentation (RSVP). The problem with all of these tasks is that they are different from normal reading. Some would argue that the lexical decision task has little if any relevance to real reading as the task is to decide if a single string of letters represents a word or not. While this can tell us something about word identification, word identification is only part of reading. Self paced reading studies resemble natural reading more than lexical decision tasks do because participants have to integrate the text. In these studies, the participants see only part of the text at a time and must press a button in order to see the next part of text. Sometimes this is done word by word with each button press eliciting the next new word. One major problem with this technique is that participants tend to adopt a strategy of pressing the button at a fairly consistent rate which has little to do with the moment to moment language comprehension processes. Another major problem with this technique is that, as eye-tracking has shown (Rayner 1975) words in the parafovea (the area around the center of fixation) are also at least partially processed during normal reading. This is impossible in word by word self paced reading. The self paced reading technique also fails to allow for regressive eye movements back to previous words. However, eye-tracking has shown that during normal reading between 10 and 15% of fixations are regressions to previous words (Rayner 1989). Sometimes self paced reading studies are conducted where more than one word is presented at a time and each button press brings in a new set of words. While this technique does allow for **some** parafoveal preview and regressions, it still suffers from the fact that the groups of words chosen to be presented together by the investigator would not necessarily be viewed as a single unit during normal reading. In this way, the investigator may force structure where it normally would not occur. It has also been noted that in general, self paced reading tasks cause participants to adopt a strategy of pressing a button at a standard rate. This is not to say that self paced reading is incapable of offering insight as a methodology, but it not as fine grained a measure as eye-tracking.

Clearly eye-tracking is a powerful tool but, even before accurate eye-trackers, many important aspects of eye movements were known. For example, it was well known that the eyes moved in saccades or jumps from word to word rather than moving smoothly across the page. The advances in eye-trackers have made it possible to measure these movements and the length of the fixations between them with a high degree of accuracy. These measures can be used to infer many of the cognitive processes in tasks such as reading. For instance, the amount of time a word in a sentence is fixated prior to moving on to another word (either forward or backward) is called *gaze duration* and can be taken as a measure of the amount of initial processing (mainly lexical access) required for that word at that point in the sentence. There are many other measures that are commonly used in eye-tracking research such as the *first fixation duration* on a word, or the *second fixation duration, regression data* (eye movements back to previous text), *regression path duration* (first pass time plus all the time spent regressing from that word or region until the eyes move to the right of the region), as well as data on word skipping to name just a few (for a more complete discussion of the measurements used in eye-tracking research and a more thorough review of the field see Rayner, 1998).

As mentioned in the introduction, Rayner et al. (1989) explored sentence wrap-up in an experiment similar to the listening studies of Jarvella and Sachs. They created passages of text that were identical except for the placing of punctuation so that a word was either at the end of a sentence or at the end of a sentence internal clause. They reported that when a word ends a sentence it is fixated longer than when the same word does not end a sentence. They also explored clause wrap-up and found that it affects more than just the duration of the fixation on a word. Rayner, Kambe, and Duffy (2000) provided evidence that in addition to increasing reading times, clause wrap-up can also effect the length of the subsequent saccade. Specifically, when a fixation was in a group of words that ended a clause, the saccade out of that region was longer than when the same group of words was not the end of a clause. They postulated that when wrap-up was completed, the eye-movement system was less hindered by higher order comprehension processes and thus was able to move further forward. Clearly eye-tracking methodology is sensitive enough to detect the same types of clause effects that were found in the early listening studies.

Traxler, Bybee, and Pickering (1997) used eye-tracking in a study of clausal integration designed to refute aspects of the Millis and Just study (1994) delayed integration hypothesis. The delayed integration hypothesis asserted that when connectives such as ‘because’ are used between two clauses, the content of the first clause will be held in working memory until the second clause has been completely processed – and then the two clauses will be integrated. Millis and Just used self paced reading methodology to investigate this hypothesis. However, this was a poor methodology to choose for such a study. In addition, the evidence that Millis and Just provided to support their hypothesis was “null” evidence: they interpreted a failure to show that two experimental groups differed significantly as support for their theory. The reason “null” evidence is suspect is that failing to find a difference between groups could have been a result of low statistical power, or more importantly in this case, using a measure that lacks the sensitivity needed to find a difference. This is the same basic argument used by Traxler et al. to justify reinvestigating the effects of connectives on clause integration using the more sensitive and accurate measures provided by eye-tracking. They set out to provide evidence that clausal integration can begin prior to the reader having processed two complete clauses. To do this they created pairs of sentences that each had two clauses bridged by a connective. In each of the pairs of sentences, the second and final clause was the same—only the initial clause differed. The difference between the initial clauses was that the whole sentence would either be (9) causal or (10) diagnostic in structure (Traxler et al., 1997 pg. 485).

9. Heidi felt very proud and happy because she won first prize at the art show.

10. Heidi could imagine and create things because she won first prize at the art show.

In an earlier study, Traxler, Sanford, Aked, and Moxey (1997) had shown that causal sentences took less time to comprehend than their diagnostic counterparts. By comparing the fixation data for the easier causal and more difficult diagnostic sentences, they hoped to find differences prior to the end of the sentence. If participants took more time processing the earlier words in the second clause when the clause integration was difficult then they would have convincing evidence against the Millis and Just delayed integration hypothesis. This is in fact what they found. First pass reading times were longer for the more difficult diagnostic clause integration in the region before the last few words of the sentence. This shows that the processing of relations between clauses does not have to wait until the end of the second of two clauses to begin. However, they still found a large effect of difficulty in first pass reading time at the very end of the sentence. So although the integration of two clauses can begin before both of the clauses have been read, the added processing for this integration may continue past the reading of the second clause.

The hypothesis I am posing is that processing difficulty encountered early in a sentence can affect the time needed to wrap-up the sentence when compared to a sentence with less processing difficulty but having the same meaning. At first glance it may appear that the Traxler et al. results just mentioned have already shown this effect. However, while these results clearly show that clause integration can begin before, and continue to the end of the sentence, caution should be used in interpreting this as an effect on the recoding of the sentence (note that the Traxler et al. were not attempting to make this conclusion). It is important to keep in mind when interpreting these results in terms of sentence wrap-up that these sentences differed in their meaning as well as their structure. As mentioned earlier, meaning is the essential aspect of the sentence that is kept in the recoding process. While it is worth noting that sentences with different meanings would be recoded differently, this is far from surprising. One would expect that a sentence with a more abstract meaning or with many elements might be more difficult to recode than one that is very straightforward or has only a few elements. These types of differences amount to wrapping up different packages. Recall that my hypothesis is that the time spent wrapping up a sentence can be affected by the amount of processing occurring earlier in the sentence when meaning is controlled. Additionally, the Traxler et al. items consisted of single sentences. It would be interesting to know whether the processing of difficult materials stops after the sentence boundary, or whether it would continue into a subsequent sentence.

In this study, participants read sentences that have the same basic meaning—the relationships between the words in the sentence remain the same—but differ in processing difficulty early in the sentence. The key question is whether this difficulty transfers to the recoding process. In order to keep sentence structure and the basic sentence meaning (gist) relatively constant while manipulating processing difficulty, I used a novel word frequency manipulation. However, before going into the manipulation used in this study, let me review some of the findings on word frequency.

One of the most robust findings in the eye-tracking literature on reading is the effect of word frequency (Rayner, 1977; Just and Carpenter, 1980; Inhoff and Rayner, 1986; Rayner and Duffy, 1986; Henderson and Ferreira, 1990; S. C. Sereno, 1992; Vitu, 1991; Raney and Rayner, 1995). Word frequency is a measure of the number of times a word appears in a large corpus of text. For instance, the Francis and Kucera frequency database used just over a million words of text to determine word frequency. The finding that has been replicated time and again is that words of low frequency are fixated for a longer time than words of the same length with a higher frequency. Most believe that this effect is due to *lexical access*, the process of identifying the word. The more common it is to see a particular word in print, the easier it is to identify. In fact, one of the most successful models of eye movements during reading, the E-Z Reader model (Reichle, Rayner, Pollatsek, 2003) uses word frequency as one of the main predictors of fixation durations.

One important aspect of the word frequency effect is its apparently short lived nature. The effect is believed by most to be confined to lexical access, and once the word is identified, it's frequency no longer plays a prominent role. However, it has been shown to have a small effect on the word immediately following (Rayner and Duffy, 1986). That is, a word following a low frequency word tends to be fixated longer than the same word following a high frequency word. This effect has been referred to as the *spillover effect.* However, there is some confusion on the nature of this spillover effect. The confusion is due to the effects of parafoveal preview—the processing of words to the right of fixation. In order to explain the effect of parafoveal preview it will be necessary to describe two very important techniques in eye-tracking research: the moving window technique and the boundary technique.

In the moving window technique (McConkie and Rayner, 1975) the experimenter controls how much of the text a participant can view by altering the size of a "moving window". This is made possible thanks to the marriage of computer and eyetracker. The eyetracker informs the computer of where the participant's eyes are during a fixation and the computer displays the text clearly in the window around that fixation. However, the computer alters the text outside of the window by changing the normal text into meaningless strings of letters. While the eye is moving, the tracker communicates with the computer, and prior to the eyes stopping on a word, the computer readjusts the display so as to keep the window of clear text around the fixation the same size. Using this moving window technique, it has been possible to ascertain the perceptual span, the area from which information about the text can be gained during reading. On average, this span extends 3 or 4 letter spaces to the left of fixation and 14 or 15 letter spaces to right of fixation.

The boundary technique (Rayner, 1975) is similar to the moving window in that some part of the text has been altered. However, unlike the moving window technique, in a boundary change, the text changes only once—when an invisible "boundary" is crossed. The boundary is usually placed just before the changed word so that the experimenter can control the types of information presented before the boundary is crossed and investigate what kinds of information are extracted from the parafovea. You might think that this manipulation would be easily noticed by the participants. But since the change in the display takes place during a saccade, it is not. During a saccade, the information coming into the brain from the retina is suspended. This is important because if we were aware of everything that the retina "saw" during an eye movement the world would appear as a blur every time we moved our eyes.

Henderson and Ferreira (1990) used the boundary change technique to investigate how attention is allocated between the fovea and parafovea. In their first experiment they manipulated foveal load (the difficulty of foveal processing) on word n to see how this affected the ability to process word 'n+1' in the parafovea. The foveal difficulty was manipulated by having word 'n' be either low or high frequency. In the control condition, word 'n+1' was identical before and after the boundary was crossed. There were two experimental conditions, one in which the preview word was similar to the target word and one in which it was different. If attention to word 'n+1' was the same when word n was high or low frequency, then one would expect the dissimilar preview to cause the same inflation on fixation times compared to the control condition. However, if foveal load affects the distribution of attention then one would expect the high frequency condition to suffer more from a dissimilar preview then the low frequency condition. This latter effect is in fact what Henderson and Ferreira found. In their second experiment they replicated this finding using predictability to influence foveal load instead of frequency showing that the effect can be extended to higher level processes. It should be noted however, that Rayner, Kambe, and Duffy in 2000 failed to find an effect of foveal load on clause final words and argued that wrap-up processes differ from those used to investigate foveal load by Henderson and Ferreira. As a crude comparison, wrap-up could be viewed as downloading information while lexical access, predictability and parafoveal preview could all be viewed as uploading information.

The finding that the reading time associated with the low lexical frequency of word 'n' can spillover to word 'n+1' is robust and has been replicated. However, there is still some confusion over how this effect should be interpreted. It is possible that the increased reading times on 'n+1' occur because the eyes have moved on from word 'n' too soon—before lexical access completed. This would mean that while a reader is now viewing word 'n+1'they are actually still uploading word 'n'. But it is more likely, that the low frequency of word 'n' causes a decrease in parafoveal preview of word 'n+1' while the eyes are still at 'n'. This is the mechanism posited by the E-Z Reader model (Reichle et al. 2003). This decrease in parafoveal processing would mean that once the eyes get to word 'n+1' more needs to be done to process it. For the purpose of this paper it is important to realize that manipulating the frequency of a word will affect the reading times of the word immediately following it. If one uses a word frequency manipulation to look at the effects of processing difficulty on sentence wrap-up, it will be crucial to separate the effects of spillover from those of wrap-up with some sort of buffer region.

In order to investigate the effects of early processing difficulty on later recoding processes, I used a novel manipulation of word frequency—manipulating the frequency of three words in a row. If the sentence wrap-up effect is due to readers forming a more compact representation of the sentence relationships, then manipulating the frequency of more than one word will increase the chances of finding such an effect by increasing the number of possible relationships affected. Most studies that have looked at word frequency, manipulated the frequency of a single word. I know of only one study in which the frequency of two consecutive words was manipulated (Rayner et al., 1989). In the Rayner et al. 1989 study, sentences included adjective noun pairs that were either both high or both low in frequency, and as controls, other sentences were presented that only had the high or low frequency noun (unmodified). They found that the frequency of the adjective affected both the first fixation duration and the gaze duration on the adjective. This same word frequency main effect was found on the first fixation and gaze duration of the noun. There was also an effect of semantic integration evidenced by longer first fixations and gaze durations on the nouns if they were preceded by and adjective than if they were unmodified. However, there was no interaction between word frequency and the presence of the adjective on these measures. Unfortunately, the data on the regions following the noun, which might shed light on the current question, were not provided. It should also be mentioned that this study did not control for the meaning of the frequency manipulated words.

CHAPTER 5

EYE MOVEMENT MODELS

An additional benefit of using three high frequency words or three low frequency words in a row is that it offers the possibility of looking more closely at how lexical access affects eye movements. Word frequency has been shown to affect how long the eyes stay fixated on a word, but less is known about how it affects where the eyes go next. It is known, for example, that low frequency words tend to be refixated more often than high frequency words, and that frequency has a small effect on the probability that a word will be skipped (Rayner et al., 1996). However, it is unknown whether the effect of frequency on fixation duration will cumulate over successive words. That is, these results could prove informative in understanding the extent to which "unfinished business" from a word influences processing of the next word. Such data may prove a challenge for current models of eye movement control during reading, and a discussion of two of the current models may help to clarify why these data would be of interest: the E-Z Reader model (Reichle et al. 2003), and the SWIFT model (Engbert, Longtin, Kliegl 2002). These two models differ in a number of ways. However, in many circumstances they make similar predictions.

The EZ Reader (7) Model

EZ reader is considered a "sequential attention shift" or SAS model. In EZ reader, attention that enables word identification moves from word to word in a sequential manner. However, just because attention moves in a sequential manner does not mean that the eyes have to follow such a strict pattern. EZ reader is capable of modeling the skipping and regressive eye movements that occur during reading by decoupling the decision of when to move the eyes from the decision of where to move the eyes. There have been numerous versions of the EZ reader model to date, the most recent version being EZ reader 7. EZ reader 7 is comprised of three systems, visual, oculomotor, and word identification. The visual system is responsible for early processing and attention selection. The oculomotor system controls eye movements in two stages, labile or M_1 and non-labile or M_2. During M_1 of oculomotor control, it is possible to cancel a saccade. Once M_2 has begun a saccade can no longer be cancelled and will be executed at the end of the stage. The word identification system also consists of two stages referred to as L_1 and L_2. In earlier versions of the model, the first stage (L_1) was referred to as a familiarity check of the word. In EZ reader 7 the authors liken L_1 to a check of the words orthography without access of its phonology or meaning which are left up to L_2. The time required for both stage $L_1$and stage L_2 depend on the frequency and predictability of the word being attended to. The lower the frequency and predictability of the word the longer these stages will take. When a reader fixates word 'n', the word identification system begins L_1. When L_1 completes, a signal is sent to the oculomotor system to move the eyes to word 'n+1' and M_1 begins. Once L_2 of the word identification system finishes, a signal is sent to the visual system to move attention to word 'n+1'. This shift of attention to word 'n+1' causes the word identification system to begin L_1 on word 'n+1'. Meanwhile, back in the oculomotor system, the completion of M_1 will cause the start of M_2, and at the end of this stage a saccade will be executed. EZ reader can account for skips because the timing of the transitions within and between systems can cause a number of different eye movement patterns. For instance, if M_1 for the saccade from 'n' to 'n+1' completes prior to the completion of L_1 for word 'n+1', a saccade will be executed to word 'n+1'. However if L_1 for word 'n+1' occurs first, a new saccade program will be initiated to word 'n+2' and word 'n+1' will be skipped.

In EZ reader 7, a refixation saccade is planned on a word with a probability of 0.07 times the length of the word being fixated as long as this value is less than one—otherwise the probability of planning a refixation is equal to one. Since the information needed to determine if a refixation should be planned is low-spatial frequency information (word length), the information from the early processing of the visual system provides the refixation cue to the oculomotor system. This means that a refixation saccade can begin M_1 as soon as the word is fixated rather than having to wait until L_1 has completed. Of course this refixation saccade can be cancelled by a subsequent saccade plan as described above, and this is more likely to happen if the word currently fixated is of high frequency than if it is of low frequency.

So what would the EZ reader model predict for the current experiment? At first glance it seems like the model would predict an increasing effect of frequency over the three word frequency region. This would be because the second and third low frequency words would be processed more slowly than their high frequency counterparts not only because of frequency differences between these words but because they also would benefit from less parafoveal processing than their high frequency counterparts. When a reader first fixates the initial low frequency word 'n', the signal to move the eyes to 'n+1' will occur later for low frequency words than for high frequency words. However, the shift of attention to word 'n+1' that occurs after the signal to move the eyes has been sent to the oculomotor system is further delayed in low frequency words relative to high frequency words. This is the cause of the parafoveal preview benefit for high frequency words. It would appear then that in our experiment when the eyes reach the second low frequency word 'n+1', they will have had less of a preview of this word relative to the second high frequency word thus they would have additional processing to due beyond the normal increased processing for the lower frequency. However, this neglects to consider the model's predictions for refixations.

A refixation is planned on a word with a probability that is a function of the word's length. Given two words of equal length, one high and one low frequency, on average an equal number of refixation saccades will be planned to both. However, more of these refixation saccades will be cancelled for the high frequency word. Only if stage M_1 of the oculomotor system finishes before stage L_1 of the word identification system will a planned refixation saccade be executed. When this occurs a large amount of processing will have occurred prior to the second fixation. How does this extra processing affect the parafoveal preview of the next word? According to the EZ reader model, it doesn't. Let's take two cases of the same low frequency word. If L_1 for word identification finishes before M_1 for the refixation no refixation will occur. In this case the parafoveal preview of word 'n+1' will be $t(M_1)+t(M_2)-t(L_2)$. But, if M_1 for the refixation is faster than L_1 for word identification the word will be refixated. When this happens, the signal to plan a saccade to 'n+1' and the beginning of stage L_2 still occur at the same time—when L_1 finishes. Since the constant interval of $t(M_1)+t(M_2)$ begin at the same time as the interval for L_2 in both situations, the parafoveal preview will be $t(M_1)+t(M_2)-t(L_2)$ whether there is a refixation or not.

It is clear that the EZ reader 7 model would predict that the difference in processing difficulty between the low and high frequency words will be larger for the second word than for the first of the three. However since this effect is due completely to reduced parafoveal preview, and since this reduction in parafoveal preview will be the same for the third word as the second, there should be no additional increase in the processing difference on the third word in the frequency manipulated region. Therefore, given that the size of the frequency manipulation stays constant over the three word region and given that predictability over the three word region is also constant, we would expect to see larger effects on first fixation time, first pass time, and refixation probability for the second and third word positions than for the first word position.

One of the current shortcomings of the EZ reader model is that it does not yet predict interword regressions. The EZ reader model is meant to be a model of eye movements during reading when higher level comprehension processes are not hindered. Regressions then, according to EZ reader, are more likely to occur when processing gets difficult. Therefore we may expect to find more regressions on the second or third word of the low frequency stimuli.

The Swift Model

SWIFT by contrast, is a "guidance by attentional gradient" or GAG model. This model differs from EZ reader 7 in a number of ways, but the most prominent difference is that in SWIFT lexical processing for multiple words is occurring in parallel. The first instantiation of the SWIFT model had three basic assumptions or principles. The first is that lexical processing of words occurs in parallel over an attentional window and, within this window, processing is fastest at the fovea and slows down the further the word is from the fovea. The second principle is that the timing of a saccade is separated from the selection of the target of a saccade. The third principle is that the word that is currently fixated can inhibit saccade generation.

The first principle assumes that there is an attentional window in which lexical processing takes place. According to the SWIFT model this window encompasses four words: the word currently fixated ‘n’, one word to the left of fixation ‘n-1’, and two words to the right of fixation ‘n+1 and n+2’. All of the words within this window receive lexical processing. However, the rate of processing depends on the words eccentricity or distance from fixation. Like the EZ reader model, lexical processing in SWIFT takes place in two stages. The first stage is the lexical preprocessing stage. During this stage lexical activity for a word is increasing and continues to increase until it reaches its maximum activation. A word’s maximum activation (a_n) is a function of the word’s frequency (f_n) and predictability (p_n)-- $a_n = (1 - p_n)\{\alpha - \beta \log(f_n)\}$. This means that the lower the predictability or frequency of the word, the higher it’s maximum activation. To simplify things, I am assuming for the moment that predictability is constant across the three word positions for the current experiment. Once activation for a word reaches its maximum, the word enters the lexical completion stage. During this stage, the word’s activation is decreasing and continues to decrease until it reaches zero. Additionally, the first stage of lexical processing is assumed to be a faster than the second stage. Therefore, on average, the slope up to a word’s maximum activation will be steeper than the slope back down to zero.

The second principle of SWIFT is that the timing of a saccade is separated from the selection of the saccade target. Saccade preparation in SWIFT has both a labile stage and a non-labile stage much like EZ reader. However, in the SWIFT model, saccades can not only be cancelled during the labile stage but can also be modified during the target selection period that runs concurrently with the labile stage and ends 16.5 ms sooner. At the end of this target selection stage, a target word is chosen with a probability proportional to the words activation at that time. In this way, SWIFT is able to predict skips, refixations, and regressions all with the same underlying mechanism. Let's look at each of these situations in turn.

Skips will occur in the SWIFT model when word 'n+1' is completely processed in the parafovea as does the EZ reader model. However, SWIFT also predicts a second type of word skipping, one in which word 'n+1' isn't completely processed but is skipped due to word 'n+2' winning the target selection process. Since targets are chosen with a probability proportional to their activation, this second type of skipping will be more likely to occur when word 'n+2' has a higher activation. All other things being equal the model would predict that a word preceding a low frequency word would be skipped more often than if it had preceded a higher frequency word.

Refixations will occur when the currently fixated word wins the target selection process. This will be more likely when the word has high activation. Therefore, predictability being constant, low frequency words will be refixated more often than high frequency. However, as in the second case of skipping described above, the SWIFT model makes the interesting prediction that a word preceding a low frequency word will be refixated less often, and skipped more often than if the word had preceded a higher frequency word.

Regressions will of course occur in SWIFT when a word to the right of the fixated word wins the targeting selection. Since lexical activity for a word does not change once the word is outside of the attentional window, words that are not fully processed the first time through the window can still be the target of a regression. In fact the model predicts that most regressions will occur once the end of the sentence has been reached due to the fact that competition for target selection will be very low at this point. It also makes the prediction that the fixation prior to a regression will be just as long as a fixation prior to a forward eye movement.

It isn't quite as easy to see what the SWIFT model would predict for our three word frequency manipulation. The first thing to note is that the SWIFT model does not predict a decrease in parafoveal preview when fixating more difficult words the way that EZ reader does. This means that SWIFT will not predict increased reading times for the second and third word positions that EZ reader predicts. It does however predict an increase in refixations (and therefore gaze durations), and regressions out of the third word position. Let's assume that the eyes are on the first of the three words. If this word is a low frequency word, it will have a higher maximum activation than its high frequency counterpart. This should make the low frequency word more likely to be refixated. However, as I mentioned earlier, words that precede a low frequency word should be less likely to be refixated given the competition for the saccade to go to the upcoming word. This means that the first low frequency word will be refixated less than it would have been had it been followed by a high frequency word. Therefore, for the first word position, it is a toss up as to whether or not there will be more refixations on the low than the high frequency version of the stimulus and the values of the free parameters may have a large influence on how this pans out. First fixation durations should still be longer for low than high frequency words but the effect on gaze durations might not be as pronounced. The situation is much the same for the second of the three low frequency words, as it also has a low frequency word to its immediate right. However, at the third word position, there is no longer a lower frequency word to the right and therefore the model would predict that there should be an increase in refixations at this position relative to the high frequency version as well as more regressions back to the earlier low frequency words that may have been under processed the first time through.

I will look briefly at how the predictions from these two models fit with the observed eye movement measures in the results section, and leave a more detailed look at the predictions for appendix A.

CHAPTER 6

METHOD

Participants. Thirty-four adults from the UMASS community participated in the eye-tracking portion of this experiment. All had normal vision or corrected to normal with contact lenses. Additionally, all participants were native speakers of American English and were naive to the purpose of the experiment. They were either provided with extra credit for psychology classes or were paid eight dollars for their participation.

Apparatus. A Fourward Technologies Dual Purkinje Eyetracker (Generation VI) was used to record participants' eye movements. Eye Movements were recorded from the right eye but viewing was binocular. Participants were seated 61 centimeters from the computer screen. Head movements were minimized by use of a bite bar and head rests.

Procedure. On arriving for the experiment, participants were presented with an informed consent sheet that also gave a brief description of the instructions for the study. Then, a bite bar was prepared for them. Next, they were familiarized with the eyetracker and given detailed verbal instructions on the experiments procedures. Once a participant was settled onto the bite bar, the eyetracker was calibrated for the participant. The accuracy of the calibration was checked after each sentence. Participants read the sentences at their own rate and signaled that they were finished reading a sentence by pressing a button. Multiple-choice questions with two alternative answers were asked after each sentence in order to check for comprehension. Participants answered with a button press. The first five trials were practice trials to get participants comfortable with the task. Experimental items were presented randomly along with 108 filler items in one of two counterbalancing conditions.

Stimuli. There were two conditions in this design--high frequency and low frequency, manipulated within a sentence pairs. There were 24 pairs of sentences. Each had a low frequency and a high frequency version, and the pair of sentences only differ in the three critical frequency words. The following are the LF and HF versions (respectively) of a stimulus used in the experiment (**frequency** region, *buffer* region, wrap-up region):

- The **rival warriors ambushed** *the vulnerable* guard patrol. During the battle, the commander fled out of fear.
- The **enemy soldiers attacked** *the vulnerable* guard patrol. During the battle, the commander fled out of fear.

Note that the first sentence of each pair is broken into three regions. The reason for this will be described in detail shortly. The frequency manipulation used in this study was different from those in other studies. Most other studies used only a single word frequency manipulation. In the present study, the frequency of three consecutive content words were manipulated--either all high or all low frequency. The three manipulated words were usually consecutive. However, for 10 of the 24 items, short words such as "of" or "the" intervened between two of the frequency manipulated words (see stimuli appendix). In the earlier discussion of word frequency, it was noted that the effect of word frequency can carry over to the next word. What I am hypothesizing is that there is an additional downstream effect of word frequency. Therefore, it was important to include some sort of *buffer region* immediately following the frequency manipulated region. If the wrap-up region of the first sentence had come immediately after the frequency manipulation the results would be obviously confounded with a spillover effect. The buffer region consisted of two to three words with an average length of 12.58 characters and a range of 9 to 18 characters. A region of this size should be long enough to remove any spillover effect from contaminating the wrap-up region. The wrap-up region of the sentence was, on average, 16.04 characters long. Given these values, the wrap-up target region will occur on average between the 38th and 54th character positions on the display, which places the region just to the right of center screen. The buffer region, wrap-up region, and the second sentence were identical in both the LF and HF versions of the stimuli. Additionally, the entire first sentence of each stimulus was on the first line as was the beginning of the second sentence. This was done to prevent return sweeps (large eye movements that bring the eye to the beginning of the next line) from affecting the wrap-up procedure at the end of the first sentence.

In most word frequency studies, there is a strict demarcation for deciding if a word is high or low frequency. Although, studies differ on what they consider high and low frequency, most experimenters would consider words with 100 or more occurrences per million to be high frequency and words with 10 or fewer occurrences per million to be low frequency. In the present study, the difference between high and low frequency words wasn’t as strict. The reason for this is that, for each stimulus pair, the frequencies of three words were being manipulated while the approximate meaning and length were controlled. These additional constraints made creation of the stimuli quite difficult. However, as the frequencies of the three words were being manipulated to create differences in the processing difficulty in the sentence, it is enough for our purposes that the words in the low frequency version of a sentence were substantially lower than their counterparts in the high frequency version. The LF versions of the word triples in this experiment had an average word frequency of 11.45 occurrences per million and the HF versions had an average word frequency of 162.48 occurrences per million (Francis and Kucera 1982). For some of the items, the difference between the LF and HF versions was only 20 occurrences per million. However, as already mentioned, the effect of word frequency on processing time is roughly logarithmic. What this means is that the difference in the time spent looking at a word with a frequency of 1 and a word with a frequency of 20 is much larger than that between a word of frequency 51 and a word with frequency 70. A log of the base 10 transformation was therefore used on the raw frequency scores for the words. The average difference was 1.24 for the first word position, 1.15 for the second and 1.26 for the third. This is an important control when examining the eye tracking record for accumulating effects of frequency. If the log transformed difference scores varied considerably between the first and third word positions, it would be difficult to interpret accumulating effects.

As mentioned above, the three words in the high and low frequency version of an item were also controlled for length. The primary goal was that the length of the region encompassing the three words was close to the same in both conditions. The average length of the LF regions was 22.08 characters and for the HF regions it was 22.20 characters.

Whenever possible, the lengths of the individual words within the region were controlled for as well.but this was not always possible. The average length of the words in characters by position were: LF1=7.3, HF1=7.4, LF2=7.6, HF2=7.2, LF3=6.6, HF2=6.9.

The major question was whether there would be downstream effects of word frequency on reading times (i.e., effects that occur after any spillover effects). When analyzing the data for the effects of word frequency, I was primarily concerned with four regions. These were the *frequency* region, the **buffer** region, the wrap-up region, and the first word of the second sentence. Since most of these regions consist of multiple words, first pass time on a region was the primary reading time measure. Earlier, I defined the measure of gaze duration as the cumulative time spent fixating a word from first fixating it until leaving it given that the word was not skipped. When looking at regions with multiple words, gaze duration is called first pass time. Thus first pass time is the cumulative time spent fixating a region from first entering it until leaving it given that the region was not skipped. Additionally, I looked at the number of first pass fixations in each region and the number of regressions from each region in order to provide a more comprehensive look at the sentence processing in this experiment. I was not concerned with skipping percentages for these regions as they are long regions and very rarely skipped (less than 2% of the time for the multiple word regions). Following the analysis of downstream effects, I present an analysis of the individual words in the frequency region to see how the multiple low vs. high frequency words effect fixation durations, saccade lengths, and number of fixations.

CHAPTER 7

RESULTS

This experiment was a 1 factor (low vs. high frequency) within-participant design. Items were removed from the data analysis due to blinks or track losses. In all, 6.25% of the data was removed for these reasons. Additionally, fixations shorter than 90 ms or longer than 1000 ms were truncated from the data set.

The following table contains the measures used in the downstream analysis for the regions in question. They were analyzed region by region.

Table 1: Sentence Regions

Region	Frequency		Buffer		Wrap up		Begin	
Frequency	Low	High	Low	High	Low	High	Low	High
First pass ms	932.0	754.4	422.5	399.1	504.4	508.7	323.4	306.0
# fixation	3.68	3.14	1.77	1.68	1.90	1.94	1.18	1.09
% regressions out	5.53	5.32	9.74	5.50	13.56	14.94	3.59	3.06

Frequency region: When looking for downstream effects of word frequency it was first important to verify that there was an effect of frequency at the source. If our manipulation of word frequency in the early parts of the sentence doesn't effect reading times there, then a failure to find a downstream effect might be due to the failure of the frequency manipulation to actually influence processing difficulty early in the sentence. In fact, there was a 177 ms frequency effect in first pass reading time for the frequency region, $F_1(1,33) = 74.26$ $p < 0.001$, $F_2(1,23) = 42.71$ $p < 0.001$. Paralleling this result is the finding that there were more first pass fixations in the low frequency version of the stimuli (mean = 3.68) than in the high frequency version (mean = 3.14), $F_1(1,33) = 32.57$, $p < 0.001$, and $F_2(1,23) = 28.87$, $p < 0.001$. However, although the percentage of regressions out of this region was slightly higher for the low frequency versions, the difference was not close to significant, F's < 1. It should be noted that the frequency region was very near the beginning of the sentence and therefore there was not much to regress to from this region.

Buffer region*:* Previous studies found that effects of word frequency can "spillover" onto the next word. For this reason I added a buffer region to the stimuli to remove any spillover effect from our downstream analysis. Note that the buffer region consisted of the same words for both the LF and HF versions of the stimuli so that any effect in that region had to come from what was read before. There was a 23 ms effect of frequency on the first pass reading times for the buffer region, which was only marginally significant $F_1(1,33) = 2.97, p = 0.094$, and $F_2(1,23) = 3.01, p = 0.096$. The number of first pass fixations in this region differed but this difference was also not significant by participants $F_1(1,33) = 2.46, p = 0.127$, but was significant by items $F_2(1,23) = 4.37, p = 0.048$. The proportion of regressions from the buffer region was almost twice as large when it had been preceded by the LF words than when preceded by the HF words. This effect was significant by participants $F_1(1,33) = 4.77, p = 0.036$, but only marginal by items $F_2(1,23) = 3.30, p = 0.082$.

Wrap-up region*:* This region ended the first sentence and consisted of the sentence's last phrasal constituent. Rayner et al. (1989) had found that a word was fixated longer if it ended a sentence than if it was sentence internal. It has been hypothesized that this effect is due to the recoding that occurs at the end of the sentence. For this reason, I had predicted that the first pass times for this region would be longer in the LF versions of the stimuli due to a greater amount of processing needed to recode the more difficult LF words. However, this was not the case. The mean first pass reading time was actually four ms longer for the HF versions of the stimuli. The effect was far from statistically significant either by participants or items (Fs <1). The results for number of fixations and regressions mirror those for first pass times. All F values were less than 1 for these analyses both by participants and by items. The relative difficulty participants had, reading the earlier parts of the sentence, seems to have disappeared.

First word of the second sentence: The results from the wrap-up region suggest that the manipulation of frequency used in this study did not effect the time needed to complete later sentence recoding. However, as Waters and Caplan point out, there is a difference between sentence interpretation effects and post-interpretive effects. It is possible that the frequency manipulation could cause such a post-interpretive effect. Therefore, I also analyzed the same first pass measures for first word of the second sentence. There was a 17 ms effect in first pass reading times on the first word of the second sentence. This effect was only marginal $F_1(1,33) = 3.48$ $p = 0.071$, and $F_2(1,23) = 3.92$ $p = 0.060$. There were also more fixations in this region for the LF stimuli. This effect was significant, $F_1(1,33) = 8.66$ $p = 0.006$, $F_2(1,23) = 7.50$ $p = 0.012$. Although there were more regressions from this region for the LF sentences, the effect was not significant (Fs < 1).

An alternative measure to first pass reading time is eyegaze time. The eyegaze measure is identical to first pass reading except in cases where the word or region was skipped. With eyegaze, when the word is skipped, the region is extended one character to the left, and if a fixation is found there, it is included into the measure. In this case, the character to the left of the word region is the period at the end of the first sentence. It may be the case that the eyes occasionally land on the period of the first sentence but the reader is really processing the next word—allowing it to be skipped. The times for these cases should be included in our reading measure as they pertain to the reading of the first word of the second sentence. When using this eyegaze measure, the effect only increased from 17 ms to 20 ms, but was much more reliable, $F_1(1,33) = 4.814$ $p = 0.035$, $F_2(1,23) = 8.031$ $p = 0.009$. One final analysis was done considering only those fixations that were launched from ten character spaces or less to the left of the beginning of the second sentence, eliminates the small number of trials where a reader skipped the last ten characters (about two words). Eliminating these trials also eliminates trials where the reader regressed from the end of the first sentence and then makes a large eye movement to the beginning of the next sentence. More plainly, this measure should ensure that, on average, the reader's eyes have been away from the frequency region for more time. When first pass time is calculated this way, the effect increases from 17 ms to 24 ms and is significant $F_1(1,33) = 4.79$ $p = 0.036$, $F_2(1,23) = 7.12$ $p = 0.014$.

More on the frequency region*:* The frequency region of these stimuli also provide an opportunity to test out a few predictions of the EZ reader and SWIFT models. To do this it will be necessary to dissect the frequency region into its individual words. Table 2 below presents the data for the individual words in the frequency region. Keep in mind that these data will not necessarily add up to the first pass times presented earlier for the whole frequency region. For example, the number of regressions out of the combined frequency region will not include inter-word regressions from within the combined frequency region.

Table 2: Frequency Region Words

Position	First word			Second word			Third word		
Frequency	Low	High	Diff1	Low	High	Diff2	Low	High	Diff3
First pass ms	302	247	55	324	278	46	313	273	41
First pass ms/char	41.2	33.7	7	43.2	38.9	4	48.6	40.0	9
# of fixations	1.09	0.96	0.13	1.06	0.93	0.13	0.96	0.94	0.02
% regressions out	4.2	5.9	-1.7	15.9	11.5	4.4	14.4	10.0	4.4
Length in char.	7.3	7.4	-0.2	7.6	7.2	0.3	6.6	6.9	-0.3
Log10Freq F&K	0.89	2.14	-1.25	0.91	2.07	-1.15	0.85	2.11	-1.26

For the raw gaze duration measure, given constant predictability over word position, EZ reader predicts that the times for the low frequency words should be inflated on the second and third word positions relative to the first while SWIFT predicts that only the third word position should be inflated. However, the effect of frequency is not larger for the second and third word positions. If anything the effect size appears to be shrinking over consecutive words. For the first word the effect is ~55ms, for the second it is ~46ms, and for the third it is ~40ms. This would seem to go against both of the models. A repeated measures ANOVA with position and frequency as independent variables was conducted on the raw gaze duration measure. There was a main effect of position $F_1(2,66) = 6.09, p = 0.004$, $F_2(2,46) = 8.53, p = 0.001$, indicating that the second and third words had longer gaze durations than the first word. There was also a main effect of frequency $F_1(1,33) = 54.31, p < 0.001$, $F_2(1,23) = 24.47, p < 0.001$, reiterating the fact that the manipulation had in fact caused processing difficulty. Individual t-tests show that the low frequency words had longer gaze durations for each position. For the first position, $t_1(33) = 5.44, p < 0.001$, $t_2(23) = 4.73, p < 0.001$; for the second position, $t_1(33) = 5.38, p < 0.001$, $t_2(23) = 3.24, p = 0.004$; for the third position, $t_1(33) = 3.83, p = 0.001$, $t_2(23) = 2.33, p = 0.029$. However, the interaction of position with frequency was not significant for gaze duration (both Fs < 1), thus the apparent trend of a shrinking frequency effect over positions is not reliable.

In controlling for the frequency manipulated words more importance was given to the length of the entire three word region. As can be seen in table 2, there are slight differences in the average length and frequency of the individual words. To try and better compensate for the differences in word length, an analysis on the millisecond per character gaze duration was done. The results for both analyses were very similar. However, the trend of a decreasing frequency effect over the positions is not seen in the millisecond per character times.

The average number of fixations on a word is directly related to the probability that the word is refixated. Given our assumptions about the predictability of the words, the EZ reader model, predicts an increase in refixations for the second and third word positions while SWIFT again predicts that an increase in refixations will most likely occur for the third word position.

Each high frequency word was fixated, on average, .09 less than each low frequency word, $F_1(1,33) = 20.05, p < 0.001$, $F_2(1,23) = 43.04, p < 0.001$, indicating that the low frequency words were refixated more often than the high frequency words. There was also a marginal effect by participant for position $F_1(2,66) = 3.01$, $p = 0.056$, $F_2(2,46) = 1.28$, $p = 0.29$ suggesting that the first two words in the region were more likely to be refixated than the third word, however this may have been due to the average third word being relatively shorter than the first two. The interaction between position and frequency was not significant $F_1(2,66) = 1.84, p = 0.17$, $F_2(2,46) = 1.19, p = 0.31$.

again suggesting that the effects of frequency are not getting larger. Again, neither model would predict this.

The SWIFT model predicts an increase in regressions that should be largest on the last word of the region. While regressions and the fixation durations prior to regressions are outside the current scope of the EZ reader model, the model does assume that regressions are more likely to occur when higher level processing fails. This assumption would suggest that regressions should be higher for the second and third word positions as the reader must then integrate more difficult words into the sentence interpretation. In contrast, the SWIFT model predicts that an increase in regressions shouldn't occur until the third word position. For the percent of regressions from a word, there was a main effect of position $F_1(2,66) = 10.04, p < 0.001$, $F_2(2,46) = 5.44$, p = 0.008, indicating that readers were more likely to regress from the second and third words than from the first. The main effect of frequency was marginal by items only $F_1(1,33) = 2.29, p = 0.140$, $F_2(1,23) = 3.69, p = 0.067$, however, the interaction of position with frequency was marginal by participants and significant by items $F_1(2,66) = 2.48, p = 0.091$, $F_2(2,46) = 3.67, p = 0.033$. Looking at the simple effect of frequency at each of the regions, we see that the percentage of regressions from the first frequency manipulated word are slightly, and not significantly, larger in the high frequency condition $t_1(33) = 0.85, p = 0.404$, $t_2(23) = 1.50, p = 0.146$. For the second position, opposite trend is seen but is only significant by items $t_1(33) = 1.59, p = 0.121$, $t_2(23) = 2.11, p = 0.046$. For the third position, the effect is almost as large as it had been in the second position but it is marginal by subjects and items $t_1(33) = 1.85, p = 0.073$, $t_2(23) = 1.85, p = 0.078$. It would appear that even though the EZ reader 7 model is not designed to predict regressive eye movements, the model's assumptions about what causes regressions may be more on track then the assumptions made by SWIFT.

CHAPTER 8

DISCUSSION

In the current study, a novel manipulation of word frequency was used to investigate the effects of early sentence processing on final sentence wrap-up. This involved manipulating the frequency of three consecutive words[1] while controlling for word meaning and length. It was hypothesized that this design would provide evidence of increased reading times on the sentence final constituents of the low frequency versions of the stimuli due to lengthened recoding or "wrap-up" processes. It was also hoped that the novel frequency manipulation would provide insight for eye movement modeling endeavors.

As I have shown, the novel word frequency manipulation used in this study produced a large difference in processing time (177ms) in the frequency region. Additionally, the effects on the individual frequency manipulated words were significant for all three word positions. Therefore, the requirement that the sentences in a pair differ in early processing difficulty was fulfilled.

The manipulation also caused a spillover effect on first pass reading that was marginal by participants and items. However, the wrap-up region (the remainder of the first sentence) did not show an effect of the earlier frequency manipulation as originally hypothesized. This result may appear to contradict the finding of a sentence wrap-up effect reported in previous studies. However, the present manipulation was different from those showing a sentence wrap-up effect. The sentence wrap-up effect is the finding that a word is fixated longer when it is sentence final then when it is sentence internal. Here, the sentence final words for each sentence pair are always the same. I was hypothesizing a difference in processing time on the sentence final words due to the word frequency differences early in the sentence. From these results, it appears that frequency of the earlier words is not affecting the speed at which a person is able to complete sentence wrap-up. This could have been due to the fact that both sentences in a pair had not only the same meaning but also the same syntactic structure. Since syntax has been shown to be one of the aspects about the sentence that is lost with subsequent sentences (Sachs 1967), it is possible that recoding is largely a syntactic process. An experiment designed to look at this possibility will be discussed in the section on future research.

Although the effect of frequency in the buffer region was only marginal in first pass time, and non significant in the wrap-up region, there was an effect of frequency on the very next region—the first word of the second sentence. Readers were more likely to refixate the first word of the second sentence when the first sentence contained the low frequency words. Gaze durations for this word were also larger if the previous sentence contained the low frequency words.

One possible reason that the reading times for the first word of the second sentence are longer in the low frequency version of the stimuli is that it is a late effect of recoding the sentence (wrap-up). This would assume that frequency did affect sentence wrap-up but that for an unknown reason, the effect occurred in the region after the end of the first sentence. However, this would contradict the previous findings that clause and sentence wrap-up happen on the clause or sentence final word (Rayner et al. 1989, Rayner et al. 2000). Alternatively, the memory representation for the first sentence might not be as strong in the LF version of the stimuli causing readers to spend more time on the first word of the second sentence as they try to integrate the meaning of the first sentence with the new information. This latter explanation seems more plausible. For a large number of the stimuli, the second sentence usually began with a connecting word such as 'However' (see stimuli appendix). Words such as these inform the reader as to how the upcoming text will fit together with the previous text. For example, consider the following sentence pair: "Tim was rather short. However, Tim's brothers were all____." The word 'however' in conjunction with the information from the first sentence allows the reader to fill in the blank at the end. It is possible that the memory representation for the LF first sentence is harder to use in this type of across-sentence comparison. If the effect were due in some way to memory or forgetting, then one would expect the effect to get larger when more time passed between when the eyes left the frequency region and when they landed on the first word of the second sentence. However, if the effect were due to sentence wrap-up, then the opposite prediction would seem to apply—the more time they have to create sentence relationships prior to the end of the sentence the less they would have to do once they get there. The analysis considering only those fixations that were launched from ten or fewer character spaces to the left of the beginning of the second sentence provides evidence for the interpretation that this effect is due in some way to memory. By considering only these instances, we remove from the data set times when the reader either skipped the end of the first sentence or regressed from the end of the first sentence and then made a long eye-movement to the beginning of the second sentence. Removing these instances increases the likelihood that the eyes have been away from the frequency region for a longer amount of time. Recall that in the initial analysis, the effect was 17ms and only marginal while in this alternate version it was 24 ms and significant by both subjects and items.

CHAPTER 9

MODEL PREDICTIONS AND FUTURE DIRECTIONS

The analyses of the pattern of eye movement measures observed in the frequency region largely fails to support either of the leading eye movement models of reading discussed herein. The EZ reader model clearly predicts that the frequency effect for the second and third word positions should be larger due to a lack of parafoveal preview and this result was not found in any of the analyses. The SWIFT model would seem to predict that the third LF word would have the largest difference in reading time and the most regressions to previous words. This is due to the fact that at the third word position, there is no longer a lower frequency word to the right of fixation drawing the eyes forward, making it more likely for the third word to be refixated and more likely to regress to previous words. Contrary to the prediction, the refixation rate was lower for the third LF word position and the difference between LF and HF regression rates was no larger than it had been in the second word position. However, judging these models on their lack of fit to the current data should be done with caution for a few reasons. First, both of the models discussed use both word frequency and word predictability in determining fixation and gaze durations. In the current study we have assumed that word predictability remains constant over the three word positions for high and low frequency stimuli. While I believe that this is a reasonable assumption, it is one that will need to be tested by having a separate group of participants complete offline predictability ratings. Given that our assumptions about predictability hold true, it must be said that our predictions about the SWIFT model performance could be less than accurate as this model is not nearly as transparent as the EZ reader model. It would be best to have the SWIFT model simulate the reading times for these sentences to confirm our predictions.

When looking for such higher level effects in the eye movement record, it is important that sentence meaning (gist) be controlled for. It would be difficult or impossible to interpret an effect such as this if meaning was not well controlled. This being said, I believe that experimental designs such as the one used here, where multiple words are manipulated while keeping meaning constant, may have additional research applications. One interesting application would be to use a multiple word frequency manipulation in subject vs. object relative sentences such as:

- The guards that arrived with the politician were heavily armed. (LFE)
- The guards that the politician arrived with were heavily armed. (LFH)
- The police that arrived with the president were heavily armed. (HFE)
- The police that the president arrived with were heavily armed. (HFH)

This would allow investigation of whether word frequency can interact with syntax and if syntax can affect sentence wrap-up. The specific prediction being that the difference between the LF hard and easy items would be greater than the difference between the HF hard and easy items.

Another use for a, meaning controlled, word frequency study would be to test model predictions. For example, one of the predictions of the SWIFT model that I feel confident in making is that a words preceding a low frequency word should be refixated less often and skipped more often than if the word had preceded a higher frequency word. This prediction is particularly interesting as it is rather unintuitive. While the current study could not shed any light on the validity of this prediction, it could be easily tested using a similar manipulation of two word pairs again controlling for word meaning so as to keep higher level processes constant.

In closing, while sentence wrap-up appears to be a relatively stable event (i.e. unaffected by earlier word frequency), word frequency can affect eye movement measures on later text—after spillover effects have disappeared. This downstream effect of word frequency appears to be due, at least in part, to memory. Additionally, sentence initial words and or connectives may be particularly sensitive to higher level memory effects during reading.

APPENDIX A

MORE ON THE FREQUENCY REGION

To keep from overlooking a possible effect in the three word frequency region I have done a number of additional analyses on subsets of the stimuli from the complete study. The first of these analyses concerns the short words that intervened between frequency manipulated words in some items. Recall that 10 of the 24 items contained a short word such as "of" or "the" between two of the frequency manipulated words. In all but one of these cases, the intervening word occurred between the second and third frequency manipulated word. It is possible that this short word could have attenuated any cumulating effect of frequency. Table 3 presents the data for this subset of stimuli.

Table 3: Analysis Without Intervening Words

	First word		Second word		Third word	
Frequency	Low	High	Low	High	Low	High
first pass ms	302.98	258.98	327.00	278.84	320.03	276.21
first pass ms/char	41.28	36.64	44.71	39.58	49.86	38.76
# of fixations	1.10	0.98	1.07	0.89	0.95	0.94
regress out	3.08	4.90	14.94	8.47	11.82	9.31
mean length	7.35	7.26	7.46	7.08	6.50	7.23
mean frequency	9.13	156.84	12.35	120.44	8.59	187.18

In this subset of the data, for gaze duration, there was a main effect of position $F_1(2,66) = 3.36, p = 0.041, F_2(2,26) = 3.89, p = 0.033$, as well as a main effect of frequency $F_1(1,33) = 27.41, p < 0.001, F_2(1,13) = 9.27, p < 0.009$. The interaction of position with frequency was not significant for the raw gaze duration measure $F_1(2,66) = 0.77, p = 0.468, F_2(2,26) = 0.55$ $p = 0.588$. However, as table 3 shows, there were some differences in the mean number of characters between the LF and HF versions. The results for the analysis in milliseconds per character show some differences. The main effect of position was significant by subjects only $F_1(2,66) = 7.49, p = 0.001$, $F_2(2,26) = 1.82, p = 0.182$. The main effect of frequency was still significant $F_1(1,33) = 35.88, p < 0.001$, $F_2(1,13) = 4.84, p < 0.047$. The interaction of position with frequency was significant by subjects only $F_1(2,66) = 3.66, p = 0.031$, $F_2(2,26) = 0.50$ $p = 0.610$. The millisecond per character analysis is much the same as the analysis of the raw times except for the hint of an interaction between position and frequency. It may be that the additional short words ease the effect of processing multiple low frequency words but this conclusion would require additional experiments to confirm.

For the number of first pass fixations, there was a main effect of frequency $F_1(1,33) = 15.68, p < 0.001$, $F_2(1,13) = 17.61, p = 0.001$. There was also a significant effect of position by participants only $F_1(2,66) = 3.60, p = 0.033$, $F_2(2,26) = 1.03, p = 0.37$. The interaction between position and frequency did not approach significance $F_1(2,66) = 2.44, p = 0.095$, $F_2(2,26) = 1.59, p = 0.224$. These results mirror those from the complete data set.

For the percent of regressions from a word, there was a significant main effect of position by participants which approached significance in the items analysis $F_1(2,66) = 7.98, p = 0.001$, $F_2(2,26) = 3.31, p = 0.052$. There was no main effect of frequency $F_1(1,33) = 2.17, p = 0.151$, $F_2(1,13) = 1.39, p = 0.260$, nor was there an interaction of position with frequency $F_1(2,66) = 2.27, p = 0.111$, $F_2(2,26) = 2.57, p = 0.096$. While some of the statistical tests that were significant in the complete data set no longer are in this subset, the trend in the data is very similar.

Clean First Pass*:* It is possible that the larger percentage of regressions from the second and third high frequency words is masking an interaction of frequency with word order. To examine this possibility I removed all the trials in which a participant made an inter-word regression within the frequency manipulated region. This amounted to the removal of 24.45% of the data used in the complete analysis. Table 4 below shows the data over the three words.

Table 4: Clean First Pass

	First word		Second word		Third word	
Frequency	Low	High	Low	High	Low	High
# of fixations	1.18	0.98	1.06	0.95	0.97	0.93
first pass ms	313.45	252.30	337.41	286.14	320.11	279.89
first pass ms/char	42.61	34.99	44.89	39.65	49.44	40.53

Results for this subset of the data are again strikingly similar to the complete set of data. In fact only the main effect of position on the number of fixations differs in significance between the two data sets. In the complete data set, the effect for position was marginal and only by participants. However, in this subset, the main effect of position on number of fixations is significant $F_1(2,66) = 6.93, p = 0.002, F_2(2,46) = 3.74, p = 0.031$.

It is also possible that the effect of word frequency in this study is affecting how far the eyes move forward each time they leave a word. Table 5 below presents the mean length of the first pass forward saccade from each word in characters as well as the average launch position from the word. The launch position is again represented in characters but is counting backward from the end of the word therefore a larger number of characters means an earlier launch position.

Table 5: Saccade Analysis

	First word		Second word		Third word	
Frequency	Low	High	Low	High	Low	High
Length	8.03	8.34	8.48	8.82	8.36	9.17
Launch site	4.71	4.78	4.60	4.99	4.54	5.05

For saccade length, there was a significant main effect of frequency $F_1(1,33) = 19.02, p < 0.001$, $F_2(1,23) = 10.07, p = 0.004$. The effect of position was significant by participants only $F_1(2,66) = 7.36, p = 0.001$, $F_2(2,46) = 2.19, p = 0.124$. There was no significant interaction between position and frequency $F_1(2,66) = 1.96, p = 0.149$, $F_2(2,46) = 2.22, p = 0.120$.

For launch site, there was a main effect for frequency $F_1(1,33) = 7.48, p = 0.010$, $F_2(1,23) = 4.58, p = 0.043$. There was no effect for position $F_1(2,66) = 0.05, p = 0.948$, $F_2(2,46) = 0.006, p = 0.994$, nor was there an interaction between position and frequency $F_1(2,66) = 1.26, p = 0.290$, $F_2(2,46) = 0.88, p = 0.420$.

These results show that the saccades leaving the high frequency words are longer than those leaving the low frequency counterpart. However, they are launched from a further distance away from the next word and this difference in launch site is on the same magnitude as the difference in saccade length. The difference in launch site is largely due to the greater number of fixations on low frequency words which implies more refixations.

As mentioned before, when controlling for word length in this study, priority was given to the size of the three word region and not to the individual words. As a result, in a number of the stimuli, words differ by more than one character. Also, the whole region of one of the stimuli differs by 2 characters between the HF and LF version.

This was the most that any of the stimuli differed with respect to length. It has been shown that one of the most important aspects of the text for predicting the length of a saccade is the length of the word the eye is moving to (Rayner, 1998). Therefore to be confident in our analysis of saccade lengths and launch positions, stimuli that contained words or regions that differed by more than a single character were removed and the data was reanalyzed. This left a pool of 16 items. The three word frequency region for these 16 items has a mean difference in length of zero and the mean difference for individual words was less than +_0.13 characters for all three words. This analysis showed the same trends and significance as the analysis on the complete data set.

APPENDIX B

LIST OF EXPERIMENTAL STIMULI

The following is a list of the experimental items. The words inside the parenthesis and separated by the slash are the low and high frequency version words respectively. When the items were displayed experimentally, the line return would be located so that the first few words of the second sentence were still on the first line.

1. The election (analyst tallied the ballots/official counted the votes) for the Senate race twice. Preventing errors was very important.

2. The (guards shielded the politician/police protected the president) during his controversial speech. Afterward, they escorted him out of the building.

3. The (laborers hoisted the crates/employees raised the boxes) onto the delivery truck. Within minutes, it was loaded and ready to go.

4. The (producer notified the actors/director informed the writer) of the last minute scene change. Originally, there was to be a blizzard but now there will be an avalanche.

5. The old (buddies yelled at the cabby/friends shouted at the driver) who wildly cut them off. Fortunately, they avoided an accident.

6. The (teen ignited the blaze/child started the fire) in the garage by accident. However, he was too afraid to tell anyone.

7. The (suspect dumped the razor/killer dropped the knife) near the bloody corpse. Investigators hope it will lead to a conviction.

8. The (ranger steered the tank/officer drove the truck) toward the front lines. Simultaneously, the enemy began a mortar attack.

9. The (rival warriors ambushed/enemy soldiers attacked) the vulnerable guard patrol. During the battle, the commander fled out of fear.

10. The (rookie athlete crushed/young player destroyed) the long standing world record. Consequently, he gained many new fans.

11. The (adult learners swiftly/older students quickly) answered the teacher's question. Unfortunately, most of them were wrong.

12. The (jazz guitarist stunned/blues musician pleased) the crowd with his new song. Luckily, a record producer was there to hear the performance as well.

13. The (diligent janitor mopped/careful employee washed) the dirty stockroom floor. Cleanliness is important to the store manager.

14. The (stout laborers hauled/strong workers carried) the rocks from the quarry. Pulleys were used to help with the largest boulders.

15. The (venomous rattler slew/dangerous snake killed) the youngster while she slept. Panic spread quickly through the small village.

16. The (albino mare ambled/white horse walked) across the flourishing meadow. Eventually, it stopped to drink from the stream.

17. He ordered a (mug of frothy ale/glass of red wine) before looking at the menu. Thankfully, they had his favorite dish.

18. The (acute molar ache/sharp tooth pain) was caused by a severe cavity. Unfortunately, it couldn't be filled and had to be pulled.

19. The (chef's soiled apron/cook's dirty clothes) got cited in the inspectors report. Ironically, the inspectors pants were covered in mud.

20. They saw a (giant flock of gulls/large group of birds) fly overhead just before sunset. Earlier that day, they had also seen a bald eagle.

21. They stayed in a (rugged brick cabin/strong stone house) during the loud thunderstorm. Lightning struck a nearby tree which landed only ten feet from the door.

22. She took the (advanced physics test/difficult science test) yesterday before her math class. Grades will be posted outside the professors office in a week.

23. His new car had a (damaged rear axle/broken front door) after the minor crash. Insurance will cover everything over the deductible.

24. He bought (attractive purple mums/beautiful red flowers) for his wife's birthday party. Sadly, her allergies kept her from enjoying the gift.

BIBLIOGRAPHY

Abrams, K., Bever, T.G. (1969). Syntactic Structure Modifies Attention During Speech Perception and Recognition. Quarterly Journal of Experimental Psychology, 21, 280-290.

Bever, T.G., Hurtig, R. (1975). Detection of a Nonlinguistic Stimulus is Poorest at the end of a Clause. Journal of Psycholinguistic Research, 4, 1-7.

Birch, S., Rayner, K. (1997). Linguistic Focus Affects Eye Movements During Reading. Memory and Cognition, Vol. 25(5) 653-660.

Engbert, R., Longtin, A., Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. Vision Research, 42, 621-636.

Gordon, P., Hendrick, R., Johnson, M. (2001). Memory Interference During Language Processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol. 27(6) 1411-1423.

Henderson, J.M., Ferreira, F. (1990). Effects of Foveal Processing Difficulty on the Perceptual Span in Reading: Implications for Attention and Eye Movement Control. Journal of Experimental Psychology: Learning, Memory, and Cognition. Vol. 16(3), 417-429.

Jarvella, R. (1971). Syntactic Processing of Connected Speech. Journal of Verbal Learning and Verbal Behavior, Vol. 10 409-416.

Just, M.A., Carpenter, P.A. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. Psychological Review, Vol. 99(1) 122-149.

Just, M.A., Carpenter, P.A., Keller, T.A. (1996). The Capacity Theory of Comprehension: New Frontiers of Evidence and Arguments. Psychological Review, Vol. 103(4) 773-780.

Kaakinen, J.K., Hyönä, J, Keenan J.M. (2003). How Prior Knowledge, WMC, and Relevance of Information Affect Eye Fixations in Expository Text. Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol. 29(3), 447-457.

MacDonald, M.C., Christiansen, M.H. (2002) Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). Psychological Review, Vol.109(1) 35-54.

Meyer, B., Talbot, A., Florencio, D. (1999). Reading Rate and Prose Retrieval. Scientific Studies of Reading, 3(4) 303-329.

Millis, K.K., Just, M.A. (1994). The Influence of Connectives on Sentence Comprehension. Journal of Memory and Language, 33, 128-147.

Rayner, K. (1975a). The Perceptual Span and Peripheral Cues in Reading. Cognitive Psychology, 7, 65-81.

Rayner, K. (1975b). Parafoveal Identification During a Fixation in Reading. Acta Psychologica, 39, 271-282.

Rayner et al. (1989). Eye Movements and Online Language Comprehension Processes. Language and Cognitive Processes, 4 (special issue), 21-49.

Rayner, K., Kambe G., Duffy S. (2000). The Effect of Clause Wrap-up on Eye Movements During Reading. The Quarterly Journal of Experimental Psychology, 53A (4), 1061 1080.

Reichle, E.D., Pollatsek, A., Fisher, D.L., Rayner, K. (1998). Toward a Model of Eye Movement Control in Reading. Psychological Review, Vol. 105(1) 125-157.

Reichle, E.D., Rayner, K., Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. Behavioral and Brain Sciences, 26, 445-526.

Sachs, J. (1967). Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse. Perception and Psychophysics, Vol. 2(9).

Traxler, M.J., Bybee, M.D., Pickering, M.J. (1997). Influence of Connectives on Language Comprehension: Eye-tracking Evidence for Incremental Interpretation. The Quarterly Journal of Experimental Psychology, 50A (3), 481-497.

Traxler, M.J., Sanford, A.J., Aked, J.P., et al. (1997). Processing Causal and Diagnostic Statements in Discourse. Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol.23(1) 88-101.

Waters, G.S., Caplan, D. (1996). The Capacity Theory of Sentence Comprehension: Critique of Just and Carpenter (1992). Psychological Review, Vol. 103(4) 761-772.