3
2004

This is to certify that the
dissertation entitled

DISFLUENT SPEECH AND THE VISUAL WORLD: AN APPLICATION OF THE VISUAL WORLD PARADIGM TO THE STUDY OF SPOKEN LANGUAGE COMPREHENSION

presented by

KARL GREGORY DAVID BAILEY

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Psychology

Major Professor's Signature

July 8, 2004

Date

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

DISFLUE
VISCA

DISFLUENT SPEECH AND THE VISUAL WORLD: AN APPLICATION OF THE VISUAL WORLD PARADIGM TO THE STUDY OF SPOKEN LANGUAGE COMPREHENSION

By

Karl Gregory David Bailey

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

2004

ABSTRACT

DISFLUENT SPEECH AND THE VISUAL WORLD: AN APPLICATION OF THE VISUAL WORLD PARADIGM TO THE STUDY OF SPOKEN LANGUAGE COMPREHENSION

By

Karl Gregory David Bailey

Input to the language comprehension system frequently deviates from ideal delivery. One major type of deviation is speech disfluency: the pauses, ums, uhs, repeated words and false starts that occur throughout spontaneous speech. While a few studies have examined the effects of disfluent speech on the language comprehension system, there is relatively little understanding of the online effects of disfluent speech. In this dissertation, I use a newly rediscovered methodology known as the visual world paradigm to examine some effects of disfluent speech on syntactic parsing and reanalysis which were initially described in studies using offline tasks. This methodology takes advantage of a link between eye movements within a constrained visual world and the processing of spoken language. An initial study intended to demonstrate the value of this paradigm for the study of the online effects of disfluency on syntactic parsing found mixed results. Some measures showed evidence of the effects of disfluencies on parsing, but other measures showed no such effects. In addition, previously reported effects from the particular visual world used in this initial study failed to replicate. As a result, several experiments were used to examine the visual world task in greater depth with the goal of better adapting the task to the study of disfluency. A modified version of the original disfluency study

indicated that this paradigm is more suited to the study of lexical processing and referent resolution, rather than syntactic processing. A final experiment using a new visual world task, however, provided online information about the relative difficulty of various kinds of disfluency. In addition to these experiments, I also describe plans to examine current procedures for analyzing data in the visual world paradigm and develop a suite of analyses for use in visual world experiments.

Copyright by
KARL GREGORY DAVID BAILEY
2004

For Rosemary Joy.

ACKNOWLEDGEMENTS

It would have been very difficult for me to have completed this dissertation without the help and support of a great number of people. My advisor, Fernanda Ferreira, was instrumental in helping me grow as a scientist throughout my graduate school career. I appreciated her ability to give me guidance and advice when necessary, but to allow me to learn independently at times as well. I would also like to thank the members of my dissertation committee for their support and assistance, John Henderson, Zach Hambrick, and Joyce Chai, as well as Rose Zacks, Tom Carr, Allan Munn, and Eric Altmann, who all served on various committees while I was a student at Michigan State.

I would like to thank the members of the Eyelab for the many hours spent discussing disfluencies and head mounted eye tracking. Liz Davis spent countless hours coding data for this project, and without her help, I would still be counting fixations and saccades.

Nathan and Aaron Christiansen, Rick and Caryn Jordan, and Lloyd Caesar made sure that I remained sane while in graduate school, and I appreciate their friendship and support, not to mention their continuing interest in my research. I would also like to thank the members of Jackson, East Lansing (University), and Lansing Seventh-day Adventist churches for their kindness and fellowship.

An army of macros, pivot tables, and perl scripts were pressed into service in the course of analyzing, formatting, and presenting the data in this dissertation. Despite their lack of sentience (for what do they know of dissertations or graduate students?), I would like to acknowledge their existence, if only to encourage the use of such tools by other researchers.

La
course of
My paren
Richard a
discuss m
I would l
for supp
this proj

Last, but certainly not least, my family supported me over the entire course of my life, were wonderful role models, and deserve a great deal of thanks. My parents, Rudi and Arlene Bailey, my sister, Kieren Bailey, and my in-laws, Richard and Arlene Bauer, encouraged me to pursue this degree, listened to me discuss my work, and made this whole process a great deal less stressful. Finally, I would like to thank my wife, Rosemary, for her love, for her patience with me, for supporting our household, and for donating many hours of volunteer labor to this project by listening, proofreading, typing, and reading.

LIST OF
LIST OF
KEY TO
INTROD
Disflu
The
Pre
Study
Ass
Struc
THE VI
Ch
Form
Desc
Gr
Cl
Stat
Le
Al
Fi
St
Sum

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xii

KEY TO ABBREVIATIONS xvii

INTRODUCTION 1

Disfluencies and Spontaneous Speech 1

The Disfluency Schema 5

Previous Research on Disfluent Speech 9

Studying Online Speech Processing: The Visual World Paradigm 20

Assumptions of the Visual World Paradigm 30

Structure of this Dissertation 33

THE VISUAL WORLD PARADIGM: DATA ANALYSIS 37

Characteristics of the Constructed Data Set 37

Format of Data 39

Descriptive Analyses 42

Graphical Representation 43

Clustering 49

Statistical Analyses 57

Length of Analysis 57

Alignment and Misalignment 60

Fixations and Saccades 61

Statistical Tests 63

Summary 67

COMPREHENSION OF DISFLUENT SPEECH: AN INITIAL STUDY 70
Experiment I 70
Material and Methods 75
Data Analysis 79
Results and Discussion 82
THE VISUAL WORLD PARADIGM: TASK EFFECTS 94
Experiment II 94
Material and Methods 98
Results and Discussion 101
THE VISUAL WORLD PARADIGM: DISPLAY AMBIGUITY 117
Experiment IIIA 117
Material and Methods 119
Results and Discussion 121
Experiment IIIB 129
Material and Methods 131
Results and Discussion 133
Experiment IIIC 141
Material and Methods 142
Results and Discussion 144
COMPREHENSION OF DISFLUENT SPEECH: FILLED PAUSES 151
Experiment IV 151
Material and Methods 154
Results and Discussion 155

COMPR
Exper
Mater
Result
GENER
Summ
Eye M
Disflu
Futur
REFER

COMPREHENSION OF DISFLUENT SPEECH: REPEATS AND REPAIRS.....164
Experiment V.....164
Material and Methods.....170
Results and Discussion.....173
GENERAL DISCUSSION.....186
Summary of Results.....187
Eye Movements in the Visual World Paradigm.....195
Disfluencies and Language Comprehension.....197
Future Directions.....201
REFERENCES.....204

Table 1. E
repre
regio
singl

Table 2. C

Table 3. C

Table 4. C

Table 5. C

Table 6.
sub

Table 7.
sub

Table 8.
sub

Table 9
sub

Table 1
by

Table 1
su

Table 1
su

LIST OF TABLES

Table 1. Eye movement pattern data generated for the sample data set. Solid black areas represent fixations in region A; solid gray areas, region B; lightly textured areas, region C; and heavily textured areas, region D. White spaces represent saccades. A single block (▮) represents $1/30^{th}$ of a second. 41

Table 2. Grouping of sequences in two cluster solution for Y Utterance Type..... 56

Table 3. Grouping of sequences in five cluster solution for Z Utterance Type..... 56

Table 4. Contingency table used in Multiway Frequency Analysis of sample data..... 66

Table 5. Objects used in Experiments I-IV..... 75

Table 6. Utterance types used in Experiment I. Segments for analysis are indicated by subscripts in the example utterances..... 76

Table 7. Utterance types used in Experiment II. Segments for analysis are indicated by subscripts in the example utterances..... 99

Table 8. Utterance types used in Experiment IIIA. Segments for analysis are indicated by subscripts in the example utterances..... 120

Table 9. Utterance types used in Experiment IIIB. Segments for analysis are indicated by subscripts in the example utterances..... 132

Table 10. Utterance types used in Experiment IIIC. Segments for analysis are indicated by subscripts in the example utterances..... 144

Table 11. Utterance types used in Experiment IV. Segments for analysis are indicated by subscripts in the example utterances..... 154

Table 12. Utterance types used in Experiment V. Segments for analysis are indicated by subscripts in the example utterances..... 171



Figure 1.

Figure 2.
(199

Figure 3.

Figure 4.
for

Figure 5.
are
seg

Figure 6.
Utt
and

Figure
san
.....

Figure 8
vis
Bo
dis
ob
sor

Figure 9
dis
for

Figure
dis

Figure
dis
the

Figure
ob

LIST OF FIGURES

Figure 1. The disfluency schema (Clark, 1996). 6

Figure 2. One referent and two referent conditions of the type used in Tanenhaus et. al (1995). 25

Figure 3. Displays assumed in the generation of the sample data set. 38

Figure 4. Probability of fixating each region in the sample data set at each sampling point for both conditions in sample data set. 44

Figure 5. Alignment of words in utterances generated for the sample data set. Utterances are aligned to utterance onset. Each row represents a single utterance, and each segment a single word. 45

Figure 6. Alignment of words in utterances used in one condition of an actual experiment. Utterances are aligned to utterance onset. Each row represents a single utterance, and each segment a single word. 45

Figure 7. Probability of fixating regions of interest during each time segment in the sample data set. Bin sizes for each segment were calculated on a trial by trial basis. 47

Figure 8. Displays used in Experiment I. Shades are used to differentiate regions in the visual world, and do not reflect the actual colors of objects used in the experiment. Boxes on the left and right indicate differences between display conditions in distractor/alternative objects and incorrect/early goals respectively. The two middle objects in gray are irrelevant distractor objects and locations that were referenced on some filler trials. 74

Figure 9. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor or alternative object during the NP1 segment of the utterance. Proportions for the fluent control utterance are included for comparison. 84

Figure 10. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor or alternative object for each segment of the utterance. 86

Figure 11. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor or alternative object during PP1 segment of the utterance. Proportions for the fluent control utterance are included for comparison. 87

Figure 12. Proportion of trials with a fixation in (A) and with a saccade to (B) the target object for each segment of the utterance. 89

Figure 13
early
utter

Figure 14
those

Figure 15
utter
II. S
pres

Figure 16
utter
II. S
pres

Figure 17
the
Exp
repr

Figure 18
the
Exp
repr

Figure
utt
tw
pr

Figure
utt
II.
pr

Figure
go
"u

Figure
di

Figure
th

Figure 13. Proportion of trials with a fixation in (A) and with a saccade to (B) incorrect or early goal during PP1 segment of the utterance. Proportions for the fluent control utterance are included for comparison. 90

Figure 14. One referent and two referent displays used in Experiment II were similar to those described in Trueswell, et al. (1999). 97

Figure 15. Proportion of trials with a fixation on the target object for each segment of the utterance in one referent (A) and two referent (B) display conditions In Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions 103

Figure 16. Proportion of trials with a saccade to the target object for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions 103

Figure 17. Proportion of trials with a fixation on the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions. 104

Figure 18. Proportion of trials with a saccade to the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions. 104

Figure 19. Proportion of trials with a fixation on the incorrect for each segment of the utterance and 300 millisecond windows after utterance offset in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions. 108

Figure 20. Proportion of trials with a saccade to the incorrect goal for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions 108

Figure 21. Proportion of trials with a fixation in (A) and with a saccade to (B) the correct goal during the PP1 segment of the utterance in Experiment II. "Ambiguous" and "unambiguous" refer to utterance types 114

Figure 22. Early and late disambiguation versions of the one referent and two referent displays for Experiment IIIA. 120

Figure 23. Proportion of trials with a fixation on the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions In



Experiment IIIA. Solid lines represent early disambiguation conditions, and dashed lines represent late disambiguation conditions. 122

Figure 24. Proportion of trials with a saccade to the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions In Experiment IIIA. Solid lines represent early disambiguation conditions, and dashed lines represent late disambiguation conditions 123

Figure 25. Proportion of trials with a fixation in (A) and with a saccade to (B) the incorrect goal during the PP1 segment of the utterance in Experiment IIIA. "Ambiguous" and "unambiguous" refer to utterance types..... 125

Figure 26. Proportion of trials with a fixation in (A) and with a saccade to (B) the correct goal during the PP1 segment of the utterance in Experiment IIIA. "Ambiguous" and "unambiguous" refer to utterance types..... 127

Figure 27. Displays created for Experiment IIIB. All four possible displays were seen an equal number of times by participants. 130

Figure 28. Percent trials in which a target object is moved to the early/incorrect goal. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types. 133

Figure 29. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor object during the PP1 segment of the utterance in Experiment IIIB. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types. 135

Figure 30. Proportion of trials with a fixation in (A) and with a saccade to (B) the early incorrect goal during the PP1 segment of the utterance in Experiment IIIB. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types. 137

Figure 31. Proportion of trials with a fixation on the early/incorrect goal during and after the utterance in one referent (A) and two referent (B) display conditions In Experiment IIIB. Solid lines represent fully ambiguous display conditions, and dashed lines represent temporarily ambiguous display conditions..... 138

Figure 32. Proportion of trials with a fixation in (A) and with a saccade to (B) the late/correct goal during the PP1 segment of the utterance in Experiment IIIB. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types. 139

Figure 33. Displays created for Experiment IIIC. All four possible displays were seen an equal number of times by participants. 143

Figure 34. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor object during the PP1 segment of the utterance in Experiment IIIC. "Ambiguous" and "unambiguous" refer to utterance types and "ambiguous display" and "unambiguous display" to display types. 146

Figure 35. Proportion of trials with a fixation in (A) and with a saccade to (B) the incorrect goal during the PP1 segment of the utterance in Experiment IIIC. "Ambiguous" and "unambiguous" refer to utterance types and "ambiguous display" and "unambiguous display" to display types. 147

Figure 36. Proportion of trials with a fixation in (A) and with a saccade to (B) the correct goal during the PP1 segment of the utterance in Experiment IIIC. "Ambiguous" and "unambiguous" refer to utterance types and "ambiguous display" and "unambiguous display" to display types. 148

Figure 37. Fully ambiguous displays used in Experiment IV 152

Figure 38. Percent trials in which a target object is moved to the early and late goals in one referent (A) and two referent (B) displays. "TO" and "DO" refer to target object and distractor object respectively 155

Figure 39. Proportion of trials with a saccade to the target object (A) and distractor object (B) during the NP1 segment of the utterance in Experiment IV 158

Figure 40. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor object during the PP1 segment of the utterance in Experiment IV 159

Figure 41. Proportion of trials with a fixation in (A) and with a saccade to (B) the early goal during the PP1 segment of the utterance in Experiment IV 161

Figure 42. Proportion of trials with a fixation in (A) and with a saccade to (B) the late goal during the PP1 segment of the utterance in Experiment IV 162

Figure 43. Visual world for Experiment V. The grid used in the actual experiment utilized dark background colors in order to produce better video images. 168

Figure 44. Proportion of trials with a fixation on (A) or saccades to (B) the target object for each segment of the utterance in Experiment V 173

Figure 45. Proportion of trials with a fixation on (A) or saccades to (B) the type match object for each segment of the utterance in Experiment V. 174

Figure 46. Proportion of trials with a fixation on (A) or saccades to (B) the color match object for each segment of the utterance in Experiment V. 175

Figure 47. Proportion of trials with a fixation on (A) or saccades to (B) the unrelated object for each segment of the utterance in Experiment V. 176

Figure 48. General format of visual world used in fluent utterance experiments..... 188

ANOVA
MFA: M
NP: Nou
OM: Or
PM: Pre
PP: Pre

KEY TO ABBREVIATIONS

ANOVA: Analysis of Variance
MFA: Multiway Frequency Analysis
NP: Noun Phrase
OM: Optimal Matching Algorithm
PM: Precedence Matching
PP: Prepositional Phrase

Disfluen

T

itself. So

articles

the spea

Court o

that rep

individu

along w

regular

way. A

course

psycho

uttera

instan

psych

ideal.

(Clark

would

referr

psych

INTRODUCTION

Disfluencies and Spontaneous Speech

The way we think about spoken language is very different from the product itself. Some of the best evidence for this is in local newspapers. Newspaper articles regularly quote individuals – but they never quote the exact speech that the speaker produces. Rather, there is an understanding, upheld by the Supreme Court of the United States (Masson v. New Yorker Magazine, Inc. et al., 1991), that reporters may "clean up" the actual speech stream that is produced by the individual being quoted. Thus, repeated words, ums and uhs, and false starts, along with ungrammaticality and instances of dialect specific language are regularly repaired by reporters.

Psycholinguists have approached the study of language in much the same way. Although a significant proportion of utterances produced by speakers in the course of spontaneous speech are anything but clear and fluent, most psycholinguistic experiments use utterances that are just that, if they are utterances at all (sentence stimuli in reading studies are presented visually, for instance).

This simplification of reality has led to the tacit, if not overt, assumption in psycholinguistic research that language is produced in spontaneous speech is ideal. Ideal speech, or ideal delivery, is a flawless performance in a given situation (Clark & Clark, 1977). In a world filled with ideal speakers, theoretically, language would be produced in a manner that would correspond to what Chomsky (1965) referred to as *competence*, rather than the *performance* that is observed. That is, psycholinguists have been concerned with the comprehension of language that

reflects s

compreh

that is s

fact, mo

syntacti

of what

actually

of psych

intervie

were tr

repeate

In fact,

of edit

compr

recogn

1996).

Thus,

accou

numb

Ravin

conve

other

reflects speakers' internalized knowledge of language, as opposed to comprehension of the sort of language that speakers actually produce, language that is sometimes a poor reflection of speakers' underlying competence. And, in fact, models of language comprehension, especially those that focus on the syntactic parser, are built to operate on ideal sentences that are more a reflection of what the speaker knows about their language than what language a speaker actually produces.

How is a given native speaker's performance at odds with the typical object of psycholinguistic study? Recall the modifications that reporters make to their interviewees' speech before publication. While the messages that the speakers were trying to convey remain intact, the starts and stops in speech, ums and uhs, repeated words, and corrections are all removed before quotations are published. In fact, to do otherwise results in an ungrammatical string of text. While this sort of editing seems relatively easy upon introspection, as the language comprehension system seems to be relatively insensitive to the conscious recognition of deviations from ideal delivery (Lickley, 1995; Lickley & Bard, 1996), it is nevertheless true that deviations must be processed at some level. Thus, a complete model of the language comprehension system must take into account the fact that these deviations are present, and occur in greater or lesser numbers depending on the topic of conversation at hand (Schacter, Christenfeld, Ravina, & Bilous, 1991; Bortfeld, et al., 2001; Oviatt, 1995), and the conversational role of the speaker (Branigan, Lickley, & McKelvie, 1999), among other things.

While some of these departures from the ideal are slips of the tongue (leading to nonwords) or actual ungrammaticalities, the vast majority are silent or overt interruptions of the utterance, repeated words, and corrections or modifications of parts of the utterance. All of these can be referred to as *disfluencies*, a term which highlights the disruption to the flow of fluent speech. While some disorders can result in severe disfluent speech, some amount of disfluency is present in the spontaneous speech of any speaker. In fact, disfluencies are a relatively common feature of spontaneous speech; these disfluencies are differentiated from those due to a language disorder by referring to the former as *disfluency* and the latter as *dysfluency*. A conservative estimate suggests that a disfluency occurs once every 16 or so words, on average (Fox Tree, 1995; Bortfeld, et al., 2001). This measure only takes into account disfluencies where there is an overt repetition, repair, or filler, and thus does not account for silent hesitation disfluencies. Nevertheless, it is clear that disfluencies must regularly be dealt with in some way by the language comprehension system. Whether they are filtered out at an early level, used pervasively, or something in between, the input to the language comprehension system contains disfluencies. Thus, we should expect to find that processes for dealing with disfluencies are part of the language comprehension system.

The majority of research examining disfluencies has been and continues to be conducted in two fields, language production and computational speech analysis. Interest in disfluencies in these two fields is, in some sense, pragmatic: comparing performance with competence is a time tested method of determining the structure of the language production system (e.g. Levelt, 1983; Hartsuiker &

Kolk, 2001; Oomen & Postma, 2001), and parsing spontaneous speech (with its associated disfluencies) is an important computational problem (e.g. Charniak & Johnson, 2001; Stolcke et al., 1998). In language production, disfluent speech is viewed as a deviation from fluent speech; the input that a listener must deal with is thus assumed to be suboptimal. In fact, Levelt (1989) refers to this as the listener's continuation problem, suggesting that the listener must somehow edit or process the input in order to determine how it properly continues. This framing of disfluencies as a continuation problem for the listener has led to the hypothesis that disfluencies should always be damaging to comprehension. Likewise, researchers who apply computational techniques to speech parsing consider disfluencies to be damaging to comprehension (at least to the degree that computers are performing some sort of comprehension task). This view has driven a search for an editing signal (Hindle, 1983) in order to better recognize and filter (or at least mark) disfluencies, and researchers have suggested that this process may occur in humans as well (e.g. Shriberg, 1994).

However, the characterization of disfluencies as harmful to comprehension does not necessarily need to follow from the observation that some disfluencies are clearly the result of error correction. That is, it is reasonable to expect that the language comprehension system is built to deal with noise in the input. Moreover, it may not necessarily be true that all disfluencies are the result of some sort of error in production. For instance, disfluencies may arise from delay due to the difficulty in word form retrieval (Beattie & Butterworth, 1979; Maclay & Osgood, 1959; Levelt, 1983) or may be produced as a result of conversational behavior (Bortfeld, et al., 2001; Branigan, Lickley, &



McKelvi
compre
disfluen
betwee
are able
signal t
the sev
if it is u
compr
itself.

can be
effects
speech
suppo
of the
to brie
speec
thus it

instar
previ
used

McKelvie, 1999). It is, in fact, possible to make the claim that instead of harming comprehension, disfluencies can have exactly the opposite effect. That is, disfluencies may be used as overt signals that help to coordinate joint actions between people involved in a conversation (Clark, 1996). In this view, speakers are able to control the type of disfluency that they produce such that they can signal to the listener not only a problem with word retrieval, for example, but also the severity of that problem (Clark & Fox Tree, 2002). In this case, the disfluency, if it is used by the listener, might be helpful in guiding the language comprehension system, as well as helping to direct the conversational interaction itself.

The characterizations of disfluencies as always harmful or always helpful can be considered the extreme versions of possible hypotheses concerning the effects of disfluencies on comprehension. It seems unlikely, though, that disfluent speech should always be helpful or harmful, regardless of the situation; this is supported by what empirical evidence exists. Before proceeding to a discussion of the effects of disfluencies on language comprehension, however, it is important to briefly discuss a general schema for the structure of disfluencies. Variations in speech disfluency might have distinct effects on language comprehension, and thus it is helpful to be able to separate disfluencies into types.

The Disfluency Schema

A disfluency schema was introduced by Levelt (1983) in order to identify instances of a particular type of disfluency involving the replacement of previously uttered fragments with new fragments. The schema has since been used in work on disfluencies ranging from studies of self monitoring in

product

recogni

coordin

these ca

various

Clark's

ideal h

1977).

deliver

disflue

the ide

resume

of disf



Figure

can us

non-r

production (Levelt, 1983; Blackmer & Mitton, 1991) to automatic parsing and recognition (Nakatani & Hirschberg, 1994; Shriberg, 1994) to joint action and coordination in dialogue (Clark, 1996). While the schema structure is the same in these cases, each has used a completely different set of terms to describe the various parts of the schema. I will, for the most part, follow the terminology in Clark's (1996) version, as it is most applicable to comprehension.

Each disfluency (Figure 1) is bracketed by bits of ideal delivery (again, ideal here refers to flawless presentation in a given situation; Clark & Clark, 1977). The speech that immediately precedes the disfluency is the *original delivery* and the speech that immediately follows is the *resumed delivery*. The disfluency itself can be decomposed into three parts, which correspond to how the ideal delivery is suspended, what happens while it is suspended, and how it resumes. Variations in these three components serve to describe the entire range of disfluent and fluent speech.

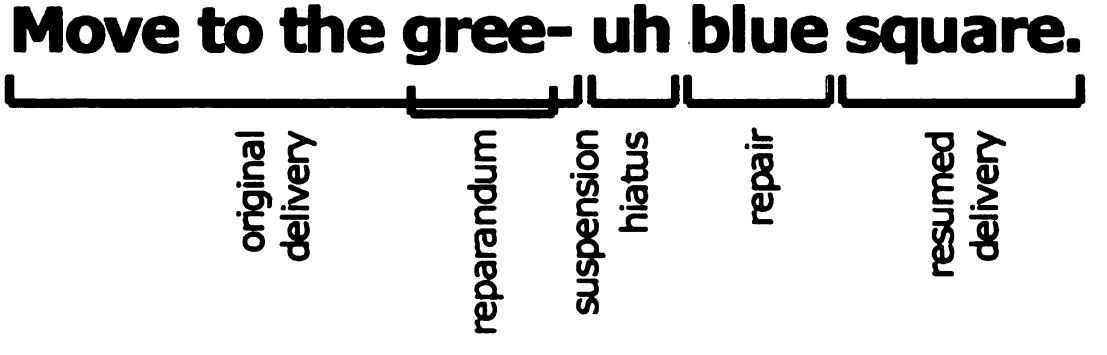


Figure 1. The disfluency schema (Clark, 1996).

Ideal delivery is interrupted at the *suspension point*. The suspension point can usually be identified by some phonological feature, such as elongation, or non-reduction of vowels. In certain cases, a word is actually interrupted. When



no clear marking of the suspension point is evident, the content of the next component of the disfluency schema can be considered to mark the suspension.

After the ideal delivery has been suspended, there is often some amount of time that elapses before the resumption of speech. This pause between the suspension of the original delivery and the resumption of ideal delivery is the *hiatus*. The hiatus is not necessarily a complete suspension of speech, however. It may include a filler (e.g. "uh" or "um") or an editing expression (e.g. "I mean" or "no"). As with the possible absence of a clear suspension point, this component may not be realized in practice; a disfluency may lack a hiatus.

In order to begin ideal delivery once again, the speaker must resume in some way. At the *resumption point* the speaker resumes delivery in a manner that may modify a portion of the preceding delivery. Resumptions may modify or change the message from the original delivery, or may involve no change to the original delivery at all. Modifications and changes occur by the addition, substitution, or deletion of words relative to the original delivery (as shown in (1)-(3); italics indicate original and resumed delivery, curly brackets indicate the hiatus, and square brackets indicate resumption material). These three types of disfluency are all considered self repairs, and the material that is added, substituted, or deleted is referred to as the repair. The material that the repair replaces, "green" in the cases of (1) and (2) and "greenish blue" in the case of (3), is referred to as the reparandum (the underlined portion of the original delivery in (1)-(3) is the reparandum; it is replaced by the repair material in square brackets).

(1) *Move to the <u>green</u>* {– uh –} [light green] *square.* (addition)

there is

where

clear su

disflue

when

resum

deleti

contin

in the

one ty

can, in

Howe

disflu

Pause

imme

disflu

involv

repair

suspe

(2) *Move to the <u>green</u>* {– uh –} [blue] *square.* (substitution)

(3) *Move to the <u>greenish blue</u>* {– uh –} [blue] *square.* (deletion)

Some resumptions involve continuation and repetition, but not self repair; there is no identifiable repair or reparandum. Continuation picks up immediately where the original delivery left off prior to suspension. Thus, when there is no clear suspension, no hiatus and immediate continuation upon resumption, the disfluency schema describes ideal delivery. Repetition, on the other hand, occurs when part of the original delivery is repeated prior to continuation. Both of these resumption types can occur in concert with the addition, substitution, and deletion types discussed above; that is, in almost all cases, the resumption continues the original delivery at some point, and may repeat some of the words in the original delivery (this is true of (1) and (3) above).

This disfluency schema can also be applied recursively; the resumption of one type of disfluency can be the original delivery of the next. Thus, disfluencies can, in some cases, form complex clusters if a speaker has significant difficulties. However, in the vast majority of experiments discussed here, singleton disfluencies, rather than disfluency clusters, will be the object of study.

In summary, disfluencies can be grouped into three major categories. Pause disfluencies involve only a suspension and hiatus of some sort and immediate continuation, without reference to the original delivery. Of those disfluencies which involve reference to the original delivery in some way, repeats involve repetition of some part of the original delivery but no self repair, while repairs always involve self repair. Either of these can involve any type of suspension or hiatus.

perspe

conver

disflue

betwe

these h

have s

compr

most h

studie

nonre

mean

studie

comp

One s

pause

comp

gener

comp

are m

The disfluency schema can be added to the previously discussed perspectives from language production, computational analysis, and conversational interaction as a possible source of hypotheses about the effects of disfluency on language comprehension. We might, for example, find differences between types of resumption material. In the next section, I will discuss which of these hypotheses have been examined, and present evidence that disfluencies can have significant effects on language comprehension.

Previous Research on Disfluent Speech

The majority of studies that have examined the effects of disfluencies on comprehension have considered the effects of filled pause disfluencies. Of these, most have focused on the fillers "uh" and "um", although there have been a few studies examining other features associated with pause disfluencies, such as nonreduced vowels (Fox Tree & Clark, 1997) and editing expressions (e.g. "oh", "I mean", "you know"; Fox Tree & Schrock, 1999). There have also been a few studies that have looked at the processing of repeat and repair disfluencies.

Filled pause disfluencies have been shown to affect language comprehension at every level from word recognition to metalinguistic judgments. One successful strategy has been to attempt to demonstrate that patterns of filled pause disfluency in production corpora can be used by the language comprehension system at various levels of interest. And indeed, the evidence is generally consistent with the hypothesis that disfluencies are useful to comprehension.

There is evidence, for instance, from production studies that filled pauses are more likely to occur before open class than closed class words (e.g. Maclay &

Osgood

closed

more d

result.

adjecti

compr

word f

when

pressi

difficu

pictur

effect

studie

disflu

in an

press.

prece

This a

have

and f

simil

there

exper

Osgood, 1959). Because open class words are generally lower frequency than closed class words, this might suggest that the retrieval of an open class word is more difficult, and thus delayed, and a disruption of fluent speech occurs as a result. This pattern has also been found within a single class of words (color adjectives; Levelt, 1983). One might predict, then, that the language comprehension system could use this information to select a lower frequency word from a set of possibilities, or at least to reject a very high frequency word when a disfluency occurs. There is some suggestion from a forced choice button pressing experiment (Corley & Hartsuiker, 2003) that this is the case when the difficulty of retrieving the word is reflected in the display by blurring one of the pictures; however, the effect of frequency was overshadowed by a very strong effect of the presence of disfluency.

The effect of the presence of disfluency has been reported in several studies, and is seen in comparisons of matched utterances with and without disfluencies. Stated as a general case, it appears that when a filled pause appears in an utterance prior to a target word that must be responded to by a button press, that button press occurs more quickly than when the target word is not preceded by a filled pause. I will refer to this effect as the filled pause advantage. This advantage has occurred in several paradigms. Word monitoring paradigms have demonstrated this effect for both "uh" and "um" pauses (Fox Tree, 2001), and for the editing expression "oh" (Fox Tree & Schrock, 1999). In addition, a similar finding was reported for repetition disfluencies (Fox Tree, 1995), although there was some evidence that the advantage may have been an artifact of the experimental paradigm in that case. It has been argued that the word monitoring

paradigm measures the amount of processing required to comprehend an utterance at a given point by introducing a dual task situation (that is, monitoring for a specific word) and measuring the amount of time required to respond in the secondary task. However, the dual task nature of this paradigm makes it difficult to determine the presence of a disfluency advantage.

In a task requiring a button press response to the instruction in the utterance, Brennan and Schober (2001) found the same filled pause advantage, suggesting that the effect cannot be attributed solely to the dual task situation that is present in word monitoring. In the Brennan and Schober (2001) experiments, participants received instructions such as those in (4)- (6) while viewing a display that contained either two or three color squares.

(4) Move to the yellow- purple square.

(5) Move to the yel- purple square.

(6) Move to the yel- uh purple square.

Participants were able to respond to utterances like the example in (6) by pressing a button corresponding to the correct square more rapidly than when responding to a corresponding utterance that had the disfluency replaced by a pause of the same length. The same filled pause advantage was found in the Corley and Hartsuiker (2003) study of word accessibility using a similar paradigm, although the disfluencies in their study were only filled pauses, and not the repair disfluencies (both with and without fillers in the hiatus) used in Brennan and Schober's (2001) study.

The button pressing paradigm is similar to word monitoring in that the participants did not necessarily need to comprehend the entire utterance; they

merely

the wo

situati

an arb

task di

respon

seems

preser

due to

of the

& Fox

result

If a sp

or ph

word

evide

curre

know

unsu

Liste

less s

as be

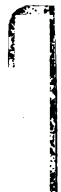
know

merely needed to monitor for the appropriate color adjective. Of course, unlike the word monitoring paradigm, there was no need to introduce a dual task situation, as participants were responding to a display instead of monitoring for an arbitrary word. Nevertheless, this paradigm, just like the word monitoring task discussed earlier, can be viewed as a task that requires a button press in response to a single word.

The advantage conveyed by filled pause disfluencies in these experiments seems to be general; that is, participants seem to be better off if a disfluency is present than if it is absent. Several researchers have suggested that this may be due to the fact that the presence of a disfluency focuses the listener on the portion of the utterance that follows (Fox Tree, 1995, 2001; Fox Tree & Clark, 1997'; Clark & Fox Tree, 2002; Fox Tree & Schrock, 1999; Brennan & Schober, 2001). This result is predicted by the conversational interaction perspective discussed earlier. If a speaker indicates that there is some difficulty in producing a particular word or phrase, it may be to the advantage of the listener to pay close attention to that word or phrase, as it may be correspondingly difficult to understand. There is evidence that at the metalinguistic level, listeners use disfluencies to estimate the current state of the speaker to some extent. When speakers answer general knowledge questions, they are more likely to produce a disfluency when they are unsure of their answer (Smith & Clark, 1993; Brennan & Williams, 1995). Listeners seem to be sensitive to this tendency, and thus rate speakers as being less sure when they produce an answer containing a disfluency and rate speakers as being more likely to actually know the answer when speakers claim not to know the answer, but are disfluent (Brennan & Williams, 1995).

However, the filled pause advantage is not the only effect on comprehension that we might expect to find. Recall, for instance, that disfluencies tend to precede open class words that are lower in frequency. If we take the frequency of a word as a measure of its accessibility, all things being equal, we might expect listeners to take the presence of a disfluency as indicative of a speaker intention to refer to a less accessible concept. One clear manipulation of accessibility is whether a concept is new or given with respect to a particular discourse. A few studies suggest that listeners may use the presence of a filled pause as an indication of the speakers intention to refer to a new concept. Barr (2001) reports results consistent with this in a task where participants had to point to a shape that was being discussed. The abstract shapes in the display were either familiar to the participants or were new. When a description of a new object was preceded by an "um", participants were faster to point to that object. Likewise, in eyetracking studies (Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Altmann, et al., 2003), participants were more likely to fixate a previously unmentioned (discourse new) object when the utterance contained a disfluency ("uh") prior to the referring noun.

As with the button pressing experiments above, these experiments involve forced choices from very small sets of possible referents. In addition, the participants in experiments conducted by Arnold and colleagues (Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Altmann, et al., 2003) could have completed their task by monitoring for a single noun. This may be a cause of some concern, as the results may be due to participants treating the task as a visual search task (within a very small set) as opposed to a language



comprehension task. Thus, these effects may not be generalizable to the real world comprehension of spontaneous speech (even though several of the studies reviewed here actually used spontaneous, rather than experimenter produced, utterances).

A final level at which filled pause disfluencies appear to have effects is at the level of syntactic parsing. In addition to occurring prior to less accessible words, filled pause disfluencies also tend to occur prior to complex constituents (Clark & Wasow, 1998; Ford, 1982; Hawkins, 1971). That is, filled pauses are more likely to occur prior to or at the initiation of, for example, a clause (or after the first, usually closed class, word) than within the clause. If listeners were sensitive to this distribution, they might take the presence of a disfluency as indicative of the initiation of a new complex constituent. The tendency of disfluencies to be reported at clause boundaries in detection tasks (Lickley, 1995; Lickley & Bard, 1996) suggests that listeners might expect this to occur (or that more processing resources are free at the beginnings of complex constituents). In a study that measured participants' judgments of the grammaticality of garden path sentences (where the initiation of a new complex constituent is ambiguous), Bailey and Ferreira (2003) found that the tendency for participants to be garden pathed could be reduced if the disfluency was placed such that it predicted the less preferred (but ultimately correct) structure. Moreover, the form of the disfluency did not seem to be critical, as it still occurred when environmental noises (cats meowing, telephones ringing, etc.) were used. This finding is in opposition to other studies (Fox Tree, 2001; Clark & Fox Tree, 2002; Smith & Clark, 1993; Barr, 2001; but cf. Brennan & Williams, 1995) that have found

effects of the form of the filled pause disfluency. However, the grammaticality judgment task used by Bailey and Ferreira (2003) may be different from the word recognition, discourse accessibility, and metalinguistic tasks described earlier. In addition, a confound between the phonological form and the length of disfluencies like "uh" and "um" in naturally produced stimuli may make studies that use such stimuli more difficult to interpret. The Bailey and Ferreira (2003) study did not suffer from this confound.

Up to this point, the focus of this review has been on filled pause disfluencies. Research concerning this type of disfluency suggests that there are several situations where filled pauses may aid comprehension. This finding is consistent with the conversational interaction perspective, which suggests that disfluencies can be used as cues by the listener. Note that the term "cue" here refers only to the listener's use of disfluencies and not the speaker's intentions. While, some researchers have suggested that disfluencies may be intended as overt signals by the speaker (e.g. Clark & Fox Tree, 2002), it is possible to be agnostic to the speaker's intent and still predict that listeners may make use of any cue that has predictive validity. The language production and computational analysis perspective analyses, in addition to the disfluency schema, do not make any strong predictions concerning filled pause disfluencies. At most, these perspectives suggest that some sort of filtering of filled pauses should take place.

However, what these perspectives do make predictions about are the effects of repeat and repair disfluencies. It is clear that such disfluencies can be identified even before the first word in the resumption can be recognized (Lickley & Bard, 1998). Thus, if the language comprehension system wanted to filter

repetit

uttera

possib

would

dynan

canno

modif

that n

Take

1999)

repar

in re

these

proc

diffic

Blac

diffe

delet

com

that

proc

invo

are

than

repetition disfluencies (which do not add or modify any new information to the utterance) in some way (as has been suggested by Shriberg, 1994), it would be possible at quite an early stage. However, even if such filtering did take place, it would require some amount of resources and would affect temporal processing dynamics. Moreover, there is reason to believe that the filtering hypothesis cannot be correct concerning repair disfluencies, as they not only require the modification of the original delivery, but also contain material in the reparandum that may be referred to in the resumed delivery. For example, in the utterance *Take the oranges to Elmira um I mean take them to Corning* (Core & Schubert, 1999), the pronoun "them" refers to "the oranges", which is part of the reparandum – the text that is replaced by the repair. Storage of the information in repairs might lead to storage costs, but, in theory, repeats would not incur these costs because the reparandum and the repair are identical. Moreover, production accounts might suggest that repetition is indicative of much less difficulty than repair (or at least repair caught earlier by the production system; Blackmer & Mitton, 1991). In addition, the disfluency schema suggests that different types of resumptions (i.e. repetition, addition, substitution, and deletion, compared to continuation alone) might have differing effects on comprehension, depending on the amount of modification of the original delivery that is necessary. All of these perspectives, then, lead to the prediction that some processing difficulty should be found in disfluencies that have a resumption involving more than just continuation. Predictions about repetition disfluencies are less clear, but, at the very least, repetitions should be less difficult to deal with than should repairs.

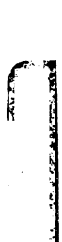
Unfortunately, the consistency of results across studies that is present in the literature on the effects of filled pauses on comprehension is not present in the few studies of repeat and repair disfluencies that exist. Fox Tree (1995) used a word monitoring task to examine the processing of repeat and repair disfluencies, and found that while repetition disfluencies had no effect on monitoring times, repair disfluencies slowed monitoring times, suggesting that greater processing was required. The Brennan and Schober (2001) study discussed above, on the other hand, found that certain repair disfluencies (specifically, those involving mid-word interruptions and filled pauses (as in (5) and (6)) were easier to process as compared to fluent and edited controls. Moreover, under time pressure, all repair disfluency times were found to be easier to process than controls. Brennan and Schober were unable to make any statements about repetition disfluencies due to their low frequency in the source corpus for their stimuli; however, those repeats that were included were also processed more quickly than fluent controls. This result is in opposition to the Fox Tree (1995) finding that repair disfluencies slow processing.

To make matters more complex, Ferreira, Lau, & Bailey (in press) report grammaticality judgment studies that indicate that the syntactic frame selected by the original delivery may continue to affect the parse of the sentence even after the repair disfluency. In one of their experiments, participants heard utterances such as those in (7) - (10).

(7) Simon says that you should drop the frog.

(8) Simon says that you should put- uh drop the frog.

(9) Simon says that you should put the frog.



(10) Simon says that you should drop- uh put the frog.

The utterance in (7) should be judged grammatical because the verb "drop" does not require a goal, while the utterance in (9) should be judged ungrammatical because the verb "put" does require a goal. The disfluent utterances in (8) and (10) should be judged the same way as (7) and (9) respectively if the repair process runs to completion and the syntactic frames associated with the reparandum verb are fully replaced by the frames associated with the repair verb . However, (8) is judged less grammatical than (7), while (10) is judged more grammatical than (9), suggesting that the syntactic frames (or theta roles) selected by the original verb are influencing the judgment of grammaticality, and by inference, the parse of the sentence. Thus, it seems that the processing of repair disfluencies may not always involve complete replacement of the original delivery with the material in the resumption, even if the resumption involves substitution.

The idea that repair disfluencies do not result in a complete replacement of the syntactic and semantic structure previously built by the parser has been formally described in model of parsing based on the Tree-Adjoining Grammar (TAG; Joshi & Schabes, 1997) formalism. In this model (Ferreira, Lau, & Bailey, in press; Ferreira & Bailey, in press), an operation called Overlay is proposed for repairs. Overlay involves building a three dimensional tree structure at points in parsing where normal parsing operations fail, reanalysis is impossible and repair is necessary. This three dimensional structure is constructed by matching root nodes in the originally parsed structure and the repair fragment and 'overlaying' the repair fragment on the original structure. As a result, the original structure is



not obliterated, and can thus be accessed (although not as easily), and in cases where argument structures differ, can even continue to be visible to future parsing operations.

In summary, then, the answer to the question "Do disfluencies help or harm comprehension?" is not a straightforward yes or no. Instead, the question must be answered with respect to the type of disfluencies, the part of the comprehension process involved, and the methodology used. Filled pause disfluencies seem to benefit comprehension, although there have been questions raised about whether this benefit is an artifact (e.g. Corley & Hartsuiker, 2003) of the types of task used to study filled pause disfluencies. The benefit derived also appears to depend on the location of the disfluency (Bailey & Ferreira, 2003). The evidence pertaining to repeat and repair, on the other hand, does not lend itself to a clear conclusion.

As has been discussed throughout this section, there may be significant limitations in the methodologies that are currently being used to study comprehension of disfluent speech, leading to effects that may not be generalizable, or disagreement about how disfluencies are processed. With the exception of the Arnold, Fagnano, & Tanenhaus (2003) and Arnold, Altmann, et al. (2003) studies, all of these studies have been conducted using either offline or dual task paradigms which cannot clearly identify the locus of any effect. In addition, studies in which participants may have been relying on word recognition strategies (i.e. the vast majority of studies concerning filled pause disfluencies) may not have required syntactic parsing at all. Thus, they may have very little to say about how language comprehension *in vivo* deals with

disfluencies. In the next section, I will describe in some detail a paradigm that may allow the effects of disfluencies on comprehension to be studied moment by moment in a situation where participants will need to fully parse disfluent utterances.

Studying Online Speech Processing: The Visual World Paradigm

It is clear from the methodologies discussed in the previous section that online methodologies for studying speech that do not introduce additional tasks are scarce. One major reason for this is that the modal online methodology for studying language comprehension, eyetracking, is much more easily applied to reading. Because reading requires that participants fixate a word (or very close to a word) in order to process it (Rayner, 1998), it is possible to not only infer where additional processing is required in a sentence, but also to infer how reanalysis occurs when a sentences is misanalyzed. Unfortunately, disfluencies do not naturally occur in writing, so eyetracking cannot be used in this manner to study the processing of disfluencies.

A related task, the moving window paradigm (Just, Carpenter, & Wooley, 1982), where sentences are segmented into sections and a button must be pressed in order to read the next section, gives patterns of results similar to those from eyetracking studies (Rayner, 1998). An auditory version of this task has been developed (Ferreira, Anes, & Horine, 1996; Ferreira, Henderson, Anes, Weeks, & McFarlane, 1996; Pynte, 1978), and elicits effects that are comparable to eyetracking of written sentences, but it may not be suitable for studying disfluencies because not all prosodic information in utterances is preserved after segmentation. Disfluencies may have specific prosodic characteristics (Fox Tree,

1995; Brennan & Schober, 2001; Shriberg, 1994), and so it is not clear that the auditory moving window is the best task for studying the online effects of disfluencies. Moreover, this task cannot control the amount of time that disfluencies take up in an utterance.

Recently, however, a task that involves reference to objects that are visually present has been rediscovered, and shows some promise as a task for studying the online effects of disfluencies on comprehension. This methodology is known as the visual world paradigm, and relies on a link between the language comprehension system and the movement of the eyes through a set of related objects.

This link was first described by Cooper (1974) in a study of the relationship between text processing and eye movements. In his study, participants listened to stories of between one and three minutes in length while viewing a three by three array of line drawings of objects. For half of the participants, all of the objects were related in some way to words used throughout the stories. For the other half, none of the objects were. In the cases where the objects were related, the relation was either direct (an object existed in the display that was an exact representation of a word in the text) or indirect (an object existed that was semantically associated to a word in the text), and either required the interpretation of a word with respect to the current context or did not. Cooper (1974) found that participants viewing relevant displays were more likely to fixate a particular (semantically relevant) object while listening to a particular word than were participants viewing non relevant displays (for the non relevant displays, the object in the corresponding location on the grid was used as a



control). This effect was stronger the more directly related the object was to the particular word in the utterance, and was weaker when more inferences were required to interpret the word. Moreover, eye movements to a related object were initiated quite rapidly, so that the vast majority of fixations had occurred within one second of word onset, and a sizeable majority within 200 milliseconds of the end of a word. Although appropriate eye movements were not made in the majority of cases, enough were made to suggest that there can be a link between spoken language comprehension and eye movements to related objects in the visual world.

This paradigm has since been adapted for studies of the time course of word recognition (Allopena, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Dahan, Tanenhaus, & Chambers, 2002; McMurray, Tanenhaus, & Aslin, 2002; Tanenhaus, Magnuson, Dahan, & Chambers, 2000). In general, the results from these studies have been used to argue for the cohort model of lexical access. In studies examining the cohort model (e.g. Allopena, Magnuson, & Tanenhaus, 1998), participants hear an utterance like "Pick up the candle" while viewing a display which includes an object referred to by that noun (the target; e.g. "candle"), and object with a name that begins with the same phoneme (a onset cohort competitor; e.g. "candy"), along with some set of related and unrelated distractors, depending on the experiment. The probabilities of fixating the target and the onset cohort competitor are identical immediately after word onset, and are higher than unrelated or rime cohort competitors. However, immediately after the listener

hears a phoneme which differentiates between the target and the onset cohort competitor, the probability of fixation rises to one, while the probability of fixating the cohort competitor falls to zero. This has been taken as evidence for the incremental processing of phonological information.

These visual world studies used to support the cohort model all involve instruction that involve some sort of motor response ('pick up', 'move', 'put, etc.). Recall that in the majority of cases in the Cooper (1974) study, participants did not fixate a related object upon hearing a word that was semantically related to something in the display (probabilities of fixation ranged between 20% and 40% depending on condition; probabilities in control conditions were between 10% and 12%, consistent with the probability of randomly fixating a single object in a nine object display). Cooper (1974) argued that this was because participants could be in one of three visual modes. In interaction mode, fixation of objects was related to concurrent language input. On the other hand, in free scanning mode or point fixation mode, participants responded in a manner unrelated to concurrent language input. In free scanning mode, participants move their eyes through the set of objects, but do not look at objects that are related to the concurrent utterance. Point fixation mode occurs when participants fixate a single point for an extended period of time. Cooper (1974) also noted that participants tended to move between these three modes in the course of an experiment, resulting in the low, but still significant, probability of fixating the associated object in his study. In cohort effect experiments, however, participants may be forced to remain in interaction mode (or spend more time proportionally in interaction mode) because it is necessary for them to



manipulate the display in some way. The introduction of an active component of the task, then, seems to greatly reduce the amount of noise present in the eye movement data, possibly by forcing participants to remain in interaction mode.

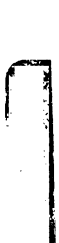
While the spoken word recognition studies are a good demonstration of the utility of using the visual world paradigm to study online spoken language processing, effects can be found beyond the word level, as should be evident given the original Cooper (1974) study. In fact, the visual world paradigm has been applied to questions of reference resolution (Dahan, Tanenhaus, & Chambers, 2002), language development (Trueswell, Sekerina, Hill, & Logrip, 1999), and syntactic processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) This is good news, as the many interesting questions (and contradictory answers) concerning the processing of disfluent speech lie in the areas of syntactic and semantic analysis and reanalysis, especially in the case of repair disfluencies. Two general applications of the visual world paradigm may be useful for studying these questions.

Several visual world studies have suggested that referential ambiguities may be parsed differently depending on the objects in the display. One situation where this has been studied (Tanenhaus, et al., 1995; Trueswell et al., 1999; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Chambers, Tanenhaus, & Magnuson, in press) involves the prepositional phrase ambiguity in (11) which can have a meaning identical to either (12) or (13).

(11) Put the apple on the towel in the box.

(12) Put the apple that's on the towel into the box.

(13) Put the apple onto the towel that's in the box.



Figu

(11)

tha

inc

eith

irre

con

to t

am

the

the

pre

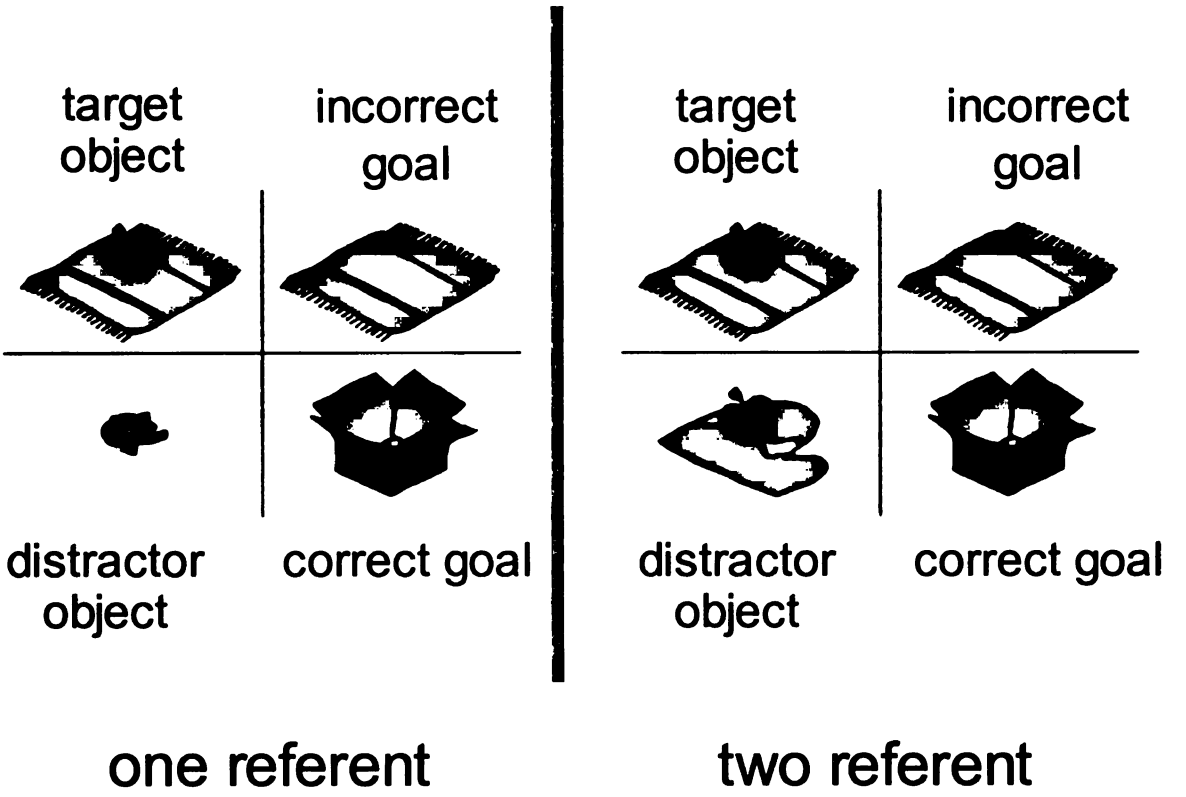


Figure 2. One referent and two referent conditions of the type used in Tanenhaus et. al (1995).

In these experiments, participants are asked to respond to sentences like (11) by carrying out an appropriate action. They must do this with a set of objects that has an apple sitting on a towel (the target object), a towel by itself (the incorrect goal), and a box (the correct goal), and a distractor object which could either be a possible referent (i.e. another apple; two referent condition) or an irrelevant object (e.g. an apple; one referent condition). The two display conditions (Figure 2) are intended to make the experiment conceptually similar to that of the spoken word recognition experiments described earlier: syntactic ambiguities would lead to an equal probability of fixating competing objects, and the probabilities would diverge once the ambiguity was resolved. As noted above, the sentence in (11) has two locations of ambiguity in the absence of context. The prepositional phrase "on the towel" can either be a possible destination for the

apple to be placed, or it can be a relative clause modifying the noun "apple", in which case it denotes a current location at which a apple may be found. Given that syntactic parsing is incremental, both of these interpretations are possible immediately after hearing the word "towel" in either the one or two referent display. The same sort of ambiguity occurs with the prepositional phrase "in the box", which can either be the location at which a towel may be found, or a destination to which the apple should be moved. Both of the displays described above allow only the destination interpretation of "in the box", as there is no towel inside a box in either display.

A competition can be set up, then, in the one referent condition, between the towel that the apple (i.e. the target object) is sitting on, and the one which is by itself (i.e. the incorrect goal). The two referent display, however, should not show evidence of the same competition according to some theoretical models (e.g. the referential model; Crain & Steedman, 1985; Altmann & Steedman, 1988). These model make the following argument: Because there are two apples in the display, participants require modifiers to determine which apple in the display is the referent of the word "apple". Therefore, listeners may interpret the prepositional phrase "on the towel" as a modifier. As this parse is obligatory in the sense that it is required to disambiguate the referent of "apple", there should be no tendency to interpret "on the towel" as a possible reference to a goal. Thus, we should expect a greater probability of fixating the distractor object in the two referent condition when compared to the one referent condition (where the distractor object is not mentioned in the utterance), and a greater probability of fixating the incorrect goal in the one referent condition compared to the two

referent co

both in the

al., 1999; S

(Chambers

goal. This s

of spoken l

these diffe

Whi

restrict the

considerin

utterances

fixation in

examinatio

involve the

fill themat

instance, t

things, and

interpreta

in the disp

case, the c

displays w

utterances

best exemp

1999), par

referent condition. This is, in fact, the pattern of results that has been reported, both in the proportion of trials with looks to (Tanenhaus et al., 1995; Trueswell et al., 1999; Spivey et al., 2002) and in the percentage of time spent looking at (Chambers, Tanenhaus, & Magnuson, in press) the distractor object and incorrect goal. This suggests not only that the visual world may affect the online processing of spoken language, but again highlights the fact that this paradigm can capture these differences.

While the preceding application of the visual world paradigm appears to restrict the set of possible syntactic or semantic interpretations that the parser is considering at particular points, there have also been demonstrations that spoken utterances can reduce the set of objects that are considered as possible targets for fixation in a visual world. These studies are similar in some ways to the examination of indirect reference in Cooper's (1974) study, but specifically involve the interpretation of verbs. All verbs have associated entities that typically fill thematic roles (MacDonald, Pearlmutter, & Seidenberg, 1994). Eating, for instance, typically involves edible things, throwing typically involves throwable things, and driving typically involves drivable things. Thus, it is possible that the interpretation of a verb might allow participants to pick out (i.e. fixate) an object in the display that can be eaten prior to that object being explicitly named. In this case, the competition set up in the experiment is not between two different displays while holding the utterance constant. Rather, two slightly different utterances are compared while holding the display constant. In the study that best exemplifies this version of the visual world paradigm (Altmann & Kamide, 1999), participants' eye movements were tracked while they completed a version

of the pictu

heard eithe

typical eve

much larg

(14

(15

The

(14), speci

hearing th

after heari

increment

paradigm

verb type

Kamide, S

It i

behavior i

different v

where con

to differen

studies ha

fixating a

condition

broken int

two secon

of the picture sentence verification task. While viewing a display, participants heard either an utterance like (14), which has a verb with a particular set of typical event participants, or an utterance like (15), which has a verb that has a much larger set of typical event participants.

(14) The boy will eat the cake.

(15) The boy will move the cake.

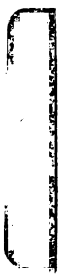
The display contained only a single typical event participant related to (14), specifically, a cake. Participants initiated eye movements to the cake prior to hearing the word "cake" in response to (14), but only initiated those movements after hearing "cake" in response to (15). Once again, this is support for incremental interpretation of sentences and evidence that the visual world paradigm is sensitive to these incremental processes. Other studies manipulating verb type have also found similar results (Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003).

It is worthwhile to pause at this point and note that the eye movement behavior in the visual world studies described here has been measured in different ways depending on the type of study. In some studies, usually those where competition was expected to occur between two different displays, leading to differences in the probability of fixating an object (e.g. Spivey et al., 2002), studies have used either a coarse grained comparison of the probability of fixating a region (or the time spent fixating a region) on a given trial across conditions, or a similar, but slightly more fine grained, version where trials were broken into time windows of anywhere from a few hundred milliseconds to over two seconds. However, other studies, usually those where the display was held

constant (e
saccade wa
particular t
the field as
In Chapter
In s
between sp
situations
in pictoria
processing
that requi
the amou
movemen
Although
appear to
that can
applicati
of syntac
objects f
demonst
particula
propose
course o
viewed a

constant (e.g. Altmann & Kamide, 1999), used the average time at which a saccade was launched, or the probability of launching a saccade within a particular time window. Thus, it appears that there is no good agreement within the field as to how the results of these experiments should be properly analyzed. In Chapter 2 of this dissertation (Data Analysis), we will return to this issue.

In summary, then, there is a fair amount of evidence that there is a link between spoken language comprehension and eye movement behavior in situations where relevant objects are present, either in physical form or depicted in pictorial form. Moreover, this link can be employed in order to study the online processing of spoken language, especially with the addition of a task component that requires participants to interact with the display in some way, thus reducing the amount of noise in fixation data, most likely by forcing participants' eye movement behavior into an interaction mode of eye movement behavior. Although a standard method of data analysis has not been agreed upon, there appear to be useful applications of the visual world paradigm in the literature that can be exploited in order to study the processing of disfluencies. One application involves the use of competing objects in a display to show the course of syntactic ambiguity resolution. The other involves the selection of particular objects for use in displays that are related to verbs in corresponding utterances to demonstrate the presence of incremental interpretation with respect to a particular display. The former can be used to test hypotheses (such as those proposed by Bailey & Ferreira, 2003) about the effects of disfluencies on the time course of parsing ambiguous sentences. In addition, if any particular disfluency is viewed as a temporary ambiguity in need of resolution, the visual world paradigm



may be able to shed light on the processing of repeat and repair disfluencies. The latter application can be used to examine how repair disfluencies, especially those involving verbs, proceed and whether complete revision of the associated mental representations (Ferreira, Lau, & Bailey, in press) occurs.

Assumptions of the Visual World Paradigm

Use of the visual world paradigm requires several assumptions about both linguistic and visual processing, and how the two are linked. The vast majority of these are well supported by research from other paradigms; however, there has been relatively little explicit recognition of exactly how these assumptions might affect the data that is collected in visual world experiments. Moreover, there seems to be a tacit, if not explicit, assumption that there is a direct connection between linguistic processing and eye movements. This assumption, that language and eye movements are linked in some manner, is necessary as fixations on objects are the dependent measure used to make inferences about linguistic processing. As was discussed in the previous section, this seems to be a valid assumption. What is not clear is how direct the connection is. For instance, behaviors that show no sign of linking do occur (Cooper, 1974). Moreover, it is clearly necessary that the link between eye movements and linguistic processing must operate through some mediating cognitive mechanisms such as memory and attention.

To see that mediating mechanisms must exist, let us begin with the assumptions about linguistic processing necessary for the visual world paradigm. We will begin with a general model of language comprehension (e.g. Frazier & Rayner, 1982; Frazier & Clifton, 1996; MacDonald, Pearlmutter, & Seidenberg,



1994; Tanenhaus & Trueswell, 1995) that involves the retrieval of information from a lexicon, the construction of a syntactic representation (by a parser), the interpretation of the meaning of the utterance, incorporation of the utterance into some internal representation of the discourse and the formation of metalinguistic information about the utterance (such as hypotheses about the confidence of the speaker), if necessary. Most importantly, it is also necessary to assume that analysis of an utterance proceeds incrementally (Altmann & Kamide, 1999; Marslen-Wilson & Welsh, 1978; Sedivy, Tanenhaus, Chambers & Carlson, 1999; Traxler, Bybee & Pickering, 1997), and that ambiguities are dealt with by making an immediate commitment (whether by heuristic selection, weighted activation, or some other process) to an analysis. Thus, fixations should be driven by the immediately preceding linguistic material. This general model is consistent with the vast majority of sentence processing theories. A further assumption must be made, however, that as part of the processing of linguistic material, attention is directed to possible referents of that material. The possible referents may be internal – that is, be part of the current discourse model in memory – or they may be external and part of the visual world. Finally, in order to make fixations in this paradigm interpretable, it is also necessary to assume that ambiguities are dealt with in a serial manner; that is, if two ambiguities (lexical, syntactic, or semantic) are present in an utterance, the first is resolved before the second, and thus shifts of attention related to the second ambiguity will not be initiated until the first ambiguity is completely resolved. This assumption is currently untested, but seems reasonable.



Once a link is assumed between linguistic processing and attention, then, we must further assume a model of eye movement control that links eye movements to attention in order to draw conclusions about language from eye movement patterns. As the visual would paradigm involves both comprehension of language and perception of a visual display, it is important that the model be general enough to deal with both linguistic and visual processing. One such model (Reichle et al., 1998; Reichle, Rayner, & Pollatsek, 1999; Henderson & Ferreira, 1990; Henderson & Ferreira, 1993; Henderson, 1992) is based on research in both reading and scene perception. In this model of eye movement control, a fixation begins with attention directed to the center of fixation. When attention is moved (in the visual world paradigm, most likely due to the incremental processing of the current linguistic material), programming of an eye movement begins. Once the program reaches a certain threshold, a saccade is launched. If the locus of attention is moved to a new location prior to the threshold being reached, a second program is initiated, which may result in a very short fixation at the first location, cancellation of the first fixation, or an intermediate fixation. It is important to note that the degree to which attention is being moved around the scene is determined by low level properties of the scene (e.g. saliency), the amount of preview time prior to utterance onset, and interference with previous scenes, in addition to the concurrent linguistic input.

In addition to the assumption that language and eye movements are linked through attention, however, it is also necessary to assume that participants in this task are using the visual world in place of short term memory (Ballard, Hayhoe, & Pelz, 1995; Gajewski & Henderson, in press). That is, we must assume that

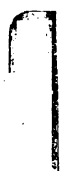


participants are not simply searching their memory for a possible referent in the cases when they do not launch a saccade. Because the objects in the visual world are often identified prior to the onset of an utterance, or at least prior to the onset of the critical word or phrase, listeners could simply search their short term memory for the referents in the utterance and only make eye movements necessary for a motor response. This would be similar to Cooper's (1974) point fixation mode. In order to encourage participants to use eye movements within the visual world as a stand in for short term memory search, the relative cost of eye movements should be kept low (Ballard, Hayhoe, & Pelz, 1995).

The visual world paradigm, then, relies on a series of assumptions that connect language to attention, and the latter to eye movement control. Thus, the connection between language and eye movement patterns is not direct; other factors besides language may be responsible for eye movements. In addition, the possibility of memory search and low level visual factors may also affect eye movement control. It is clear, then, that the visual world paradigm involves the interaction of a large number of psychological mechanisms. Because of this complexity, researchers must take care when drawing inferences about language comprehension from eye movement patterns.

Structure of this Dissertation

In this dissertation, I will endeavor to demonstrate that the visual world paradigm can be used for studying the processing of disfluent speech. I will begin by describing an initial experiment, using the visual world paradigm, that attempted to determine whether filled pauses could be used online as cues to help parse an utterance when they occurred in particular locations relative to sentence



structure (as suggested by Bailey & Ferreira, 2003). This experiment was intended as a proof of concept, in which an established finding from psycholinguistics using ideal utterances was affected by the presence of disfluencies. However, in the course of analyzing the data from this study, it became clear that the visual world paradigm was not necessarily as robust as might be expected, as established findings did not replicate. In order to identify possible reasons for this failed replication, I will then describe several further visual world experiments involving utterances containing no disfluencies that attempt to explore the boundaries of the visual world paradigm itself (and thus do not directly concern disfluent speech). In addition, I will suggest a plan for developing additional analyses than may shed light on how participants are behaving in response to the visual world paradigm. I will then return to the issue of disfluent speech processing and report a revised version of the initial experiment and a new visual world experiment considering repeat and repair disfluencies.

The research described in the proposal has implications not only for the visual world paradigm, but also for theories of language and especially sentence comprehension. Several authors (Arnold, Fagnano, & Tanenhaus, 2003; Brennan & Schober, 2001; Bailey & Ferreira, 2003; Ferreira, Lau, & Bailey, in press) have noted that models of sentence comprehension should take speech disfluency into account; they should become models of spontaneous utterance comprehension. However, in order for this to occur, studies which examine the parsing of utterances online must be conducted and the results must be compared to and combined with the research, generally using reading paradigms, that has lead to



the development of current models of parsing. The research proposed here will aid this by testing two hypotheses concerning the effects of disfluencies on parsing (Bailey & Ferreira, 2003; Ferreira, Lau, & Bailey, in press). Thus, a major aim is to move from the suggestion that models of sentence comprehension should take disfluencies into account to suggestions of how models might take disfluencies into account. In addition, this research builds on the proposal (e.g. Ferreira, Lau, & Bailey, in press) that the parsing of disfluent speech might utilize mechanisms and processes that are similar (or perhaps identical) to those used in the reanalysis of temporarily ambiguous garden path sentences.

The experiments here also clearly have implications for questions of what information is used by the language comprehension system and when that information is used. The visual world paradigm has been used (e.g. Tanenhaus, et al., 1995) to argue against an architecture of language where the parser only considers certain sources of information initially (e..g. garden path theory; Frazier, 1987), in favor of an architecture that considers any or all sources of information immediately (e.g. constraint-based approaches; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). If, however, eye movements are driven by characteristics of the display, by task demands, or by shallow, good enough parsing (Ferreria, Bailey, & Ferraro, 2002), these arguments may not be valid. In addition to these fundamental questions about architecture, the experiments described in this proposal may also shed light on where and when disfluencies inform processing.

Finally, this research should shed light on basic questions about how visual cognition and language comprehension are related. There is renewed



interest in the comprehension of spontaneous speech, as speaking and hearing seem to be the modal forms of human language use (Trueswell & Tanenhaus, in press). In part because of this interest, there are a number of researchers using eye movements to understand speech comprehension and production (see, for example, many of the papers in Henderson & Ferreira, 2004). In order to properly interpret the results of these studies and to develop an understand of speech comprehension, it is necessary to build a theory of how language comprehension and visual cognition interact. Each experiment described here is a possible building block for such a theory.



THE VISUAL WORLD PARADIGM: DATA ANALYSIS

The visual world paradigm is based on the assumption that there is a cognitive link between spoken language and eye movements. However, as was discussed in Chapter 1, that link is by no means direct. Likewise drawing conclusions about the language comprehension system from analysis of eye movement data is not necessarily straightforward.

In this chapter, I will briefly describe the type of data collected in visual world experiments, and the descriptive and statistical methods used to analyze such data. In order to more clearly illustrate the format of the data and the analysis methods, I will make use of a constructed sample data set.

Characteristics of the Constructed Data Set

The sample data set is of the type that would be collected in a simple experiment with one independent variable (utterance type) that has two levels (labeled in the set as Y and Z). In constructing the data set, it was necessary to generate the utterances that would be used in the experiment. Each utterance was made up of four words (labeled Word1, Word2, Word3, and Word4). These were data generated based on a hypothetical experiment in which participants would view one of eight different displays while listening to a corresponding utterance from either utterance type Y or Z, and then make a behavioral response. Each participant would view each display only once and would hear only one of the two utterance types for each display. An equal number of each utterance type would be heard by each participant, and each utterance type-display pairing would occur an equal number of times in the course of the experiment. As a result, each participant would encounter four trials in each

condition total. This design, of course, is simpler than what is found in most visual world studies (typically two or more independent variables are investigated simultaneously), but the number of observations per cell for each participant and the characteristics of the generated utterances are representative.

The displays assumed in the generation of this data set would be similar to the four object array depicted in Figure 3. This type of display is typical of the displays used in visual world experiments, whether the experiment involves real world objects or computer displays. In the former, the visual world generally consists of as many as four clusters or groups of objects, usually with between four and eight objects involved. In the latter, the visual world generally consists of several clip art objects, often mismatched in size and scale. For the purposes of this data set, the display was segregated into four regions, each corresponding to an object, and labeled A, B, C, and D, starting in the upper left hand corner of the display and proceeding in a clockwise manner.



Figure 3. Displays assumed in the generation of the sample data set.



The constructed data set, then, consists of a total of 32 trials composed of fixations in the regions of interest, fixations outside the regions of interest, and saccades between fixations. These trials were generated from a point corresponding to the onset of an utterance and ended at a point corresponding to a behavioral response in order to be directly comparable to trials generated in the course of a visual world experiment. Each trial was randomly generated from one of a set of basic patterns similar to those produced in visual world expeirments. Parameters for the corresponding utterances were also randomly generated from a basic pattern. Each generated trial was assigned to a participant, display, and utterance according to the experimental design described above. These data were then placed in a data file and analyzed using the data analysis tools typically used for visual world experiments.

Format of Data

Two different general types of eye tracker can be used to conduct visual world experiments. With computer displays, a table mounted eye tracker is often used. This type of eye tracker requires participants' heads to be stabilized in a chin rest, but allows computer scoring of the eye position data. When participants need to manipulate objects, on the other hand, a table mounted eye tracker cannot be used, as the participants' heads would move as they reached for objects. In experiments where object manipulation is required, a head mounted eye tracker is used. This type of eye tracker stabilizes the eye image on the camera by attaching the camera to the participant's head via a visor. However, the head mounted tracker is not anchored to the world. Thus, automatic methods do not currently exist for computer scoring of the eye position data output by the head

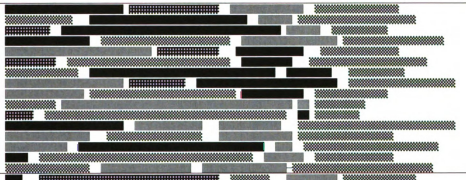
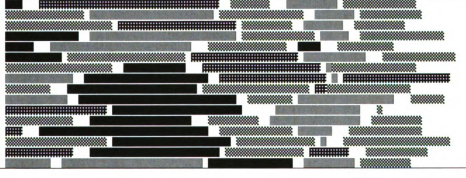


mounted eye tracker; hand scoring must be used instead. Of course, regardless of the type of eye tracker used, the data set that ultimately results will have the same format.

Specifically, the data from one trial for one participant in a visual world experiment can be represented as a string. This string is composed of tokens selected from an alphabet of tokens that match all possible regions in the visual world that could be fixated. This includes a catchall "other" region for fixations away from the objects. Take, for example, the visual world assumed while constructing the sample data set. There are four possible regions of interest in the world shown in Figure 3 that could be fixated. These regions form the component alphabet (consisting in this case of the tokens A, B, C, D, and X; region X is the catch all region) that makes up strings corresponding to patterns of eye movements. In a visual world experiment, the string is initialized simultaneous with the onset of the utterance, and the termination point is a predetermined behavioral response – a button press or a hand movement, for instance – or after the passage of a certain amount of time. If the sampling rate of the eyetracker used in the experiment is 30 Hz (that is, if the position of the eye is sample 30 times every second), the string would be incremented by one character every $1/30^{th}$ of a second until the point of termination. Incrementing the string simply involves matching the current eye position to a fixation in a region of the display and recording the appropriate token (or recording a non-fixation state such as a blink, track loss, or a saccade). Thus, every trial, even if it lasts only a few seconds, can consist of well over one hundred individual observations, which make up a unique string. For example, the string corresponding to the response

made by Participant #1 in response to Utterance #1 in the sample data set was "AAAAAAAAAAAAAAAAAAAA.DDDDDDDDDDDDDDDD.BBBBBBBBBBBB. .CCCCCCCCCCCCCC", where each letter represents the region being fixated during a single frame in the eye tracking video being analyzed, and each dot represents a frame during which an eye movement (saccade) was taking place. Table 1 graphically depicts all of the strings generated for the sample data set.

Table 1. Eye movement pattern data generated for the sample data set. Solid black areas represent fixations in region A; solid gray areas, region B; lighter textured areas, region C; and darker textured areas, region D. White spaces represent saccades. A single block (▮) represents 1/30th of a second.

Utterance Type	Eye Movement Patterns (Strings)
Y	
Z	

It is worthwhile to note that any pattern of eye movements in a visual world experiment can only be interpreted with respect to the concurrent utterance and display. That is, the string representing the pattern of eye movements for a particular trial was generated in response to a specific linguistic and visual stimulus presented during that trial. Because the words in each

linguistic stimulus occur at a particular point in time relative to the onset of the utterance and the other words in that utterance, and because this point in time differs for each utterance, strings cannot be directly compared without correcting for differences in word onset timing. Differences in the individual components of the display may change the interpretation of the utterance as well. However, it is difficult to compare responses to a particular utterance across different displays. This is because it is not possible to correct for these differences through string manipulation as can be done to correct for differences across utterances. That is, utterances differ in the length of their individual segments, which can be matched on a trial by trial basis, while displays differ in their component parts, which are present throughout the trial and thus cannot be adjusted in the same way.

As can be seen in the constructed data set, visual world experiments typically have a within participants, repeated measures design. Thus, a number of data strings are produced for each cell in the design matrix for each participant (each participant is identified by a number at the beginning of each line in Table 1). The data contained in eye movement pattern strings are typically averaged in some way in order to descriptively and statistically analyze the resulting patterns of eye movements. However, because of the time required to set up displays in some visual world experiments, the number of observations in each cell can be quite limited.

Descriptive Analyses

The general purpose of descriptive analysis is to determine what the typical patterns of response are, and whether those patterns differ as a result of the independent variables in the experiment. Most graphical presentations of eye



movement data used in the literature are purely descriptive because they are not associated with any inferential statistical analysis. In addition to these graphical analyses which have become more or less standard in the literature, an alternative technique based on the observation that eye movement patterns can be represented as strings is presented as a method of identifying possible subgroups in eye movement patterns.

Graphical Representation

The simplest method of describing data from visual world experiments involves simply aligning the strings corresponding to each trial in the experiment to some point in the utterance. After aligning the strings, fixations are summed for each region in the display at each sampling point, and then the total for each region is divided by the total number of trials. This produces a graph of the proportion of trials on which each region of interest was fixated for each point in time (Figure 4). Other graphs may be produced with bins that are larger than a single frame. In these cases, the proportion of trials with a fixation in the region of interest within the duration of the bin is calculated for each region. Thus, if each bin in a given analysis contains 20 samples, then a fixation would be counted for a given trial as long as that fixation occurred sometime during those 20 samples. This would happen even if a fixation in another region also occurred during that time window.

Several issues arise with this sort of graphical representation of eye movement patterns. The first has to do with the process of alignment. In the graphs in Figure 4, the strings were aligned to the onset of the utterance (the same alignment as in Table 1). While this is probably the most conceptually

simple alignment, other alignments are possible and perhaps preferable. The strings can be aligned at the onset of any word or phrase in the utterance, or at the offset of any word or phrase. Usually, the alignment point is the onset of the word or phrase (hereafter, segment) of interest. However, because utterance stimuli differ in the lengths of their segments, certainty that a given peak is in response to a certain segment of an utterance generally decreases as the amount of time between the segment of interest and the point of alignment increases (Altmann & Kamide, 2004). In addition, eye movements generated in response to a given segment that occurs much later or earlier than the point of alignment can be spread out over regions of the graph that can range from several hundred milliseconds to well over a second. In other words, when strings are aligned to the onset or offset of one segment of an utterance, they are as a result almost certainly misaligned to some degree with respect to all other words and phrases in that utterance.

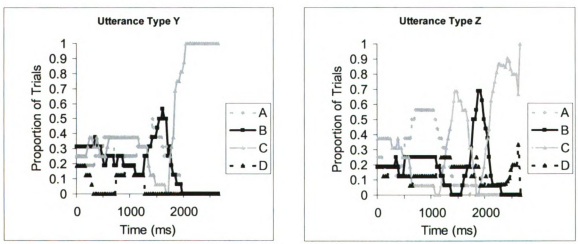


Figure 4. Probability of fixating each region in the sample data set at each sampling point for both conditions in sample data set.

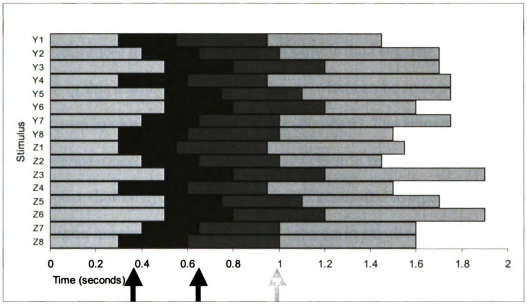


Figure 5. Alignment of words in utterances generated for the sample data set. Utterances are aligned to utterance onset. Each row represents a single utterance, and each segment a single word.

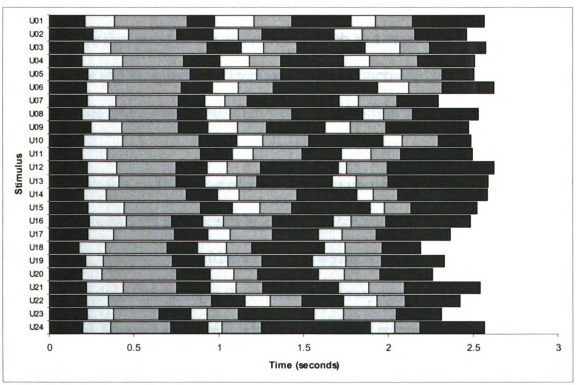


Figure 6. Alignment of words in utterances used in one condition of an actual experiment. Utterances are aligned to utterance onset. Each row represents a single utterance, and each segment a single word.

These misalignments are depicted in Figure 5, which shows the randomly generated utterances used in generating the sample data set. The variation in word length in this sample data set is similar to what can be seen in the actual utterances used in visual world experiments (Figure 6). The arrows in Figure 5 show the average onsets of Word2, Word3, and Word4 respectively.

Suppose that an eye movement is always generated to a particular region of interest upon hearing a certain phrase of interest in an utterance. Because the onset of any particular downstream or upstream word precedes the average onset of corresponding words in some utterances and is later than the average in others, eye movements from various trials that are launched in response to the phrase of interest will not occur at the same time on the graph, unless the graph is aligned to that point in the utterance for each and every trial. Instead, eye movements will be spread out over an artificially increased time span. We will return to the discussion of misalignment in the section on statistical analysis below.

One solution to downstream and upstream misalignment is to generate a series of graphs with alignments at each possible point. However, this is a time consuming, complex, and ultimately unsatisfying solution, as it still cannot account for misalignments in the offsets of the very segments to which the graphs are aligned. Another solution might be a technique referred to as time warping, in which sections of the eye movement record corresponding to the various segments of the appropriate utterance are either stretched or compressed in time such that corresponding sections for each trial take up the same amount of time. This, however, distorts the eye movement record. A different procedure that

provides much the same result (i.e. tying each eye movement record to the corresponding utterance for that trial) is a trial by trial analysis of the probability of fixation in each utterance segment (or whatever other analysis might be of interest; Altmann & Kamide, 2004). In such an analysis, the bin sizes used to construct the graph vary by the length of each segment of the particular utterance encountered on each trial. The trial by trial analysis corresponding to the moment by moment analysis in Figure 4 is depicted in Figure 7. Note that one result of this type of analysis is a decrease in the grain size of the graphical depiction, as would happen with any increase in bin size. Although there is a decrease in grain size, the decrease is directly motivated by tying the eye movement record to the concurrent utterance throughout the utterance. In a moment by moment graph of the type shown in Figure 4, the eye movement record can only be tied to the concurrent utterance at a single point, the point of alignment. The benefit achieved in tying the entire eye movement record to the corresponding utterances seems greater than any loss in grain size.

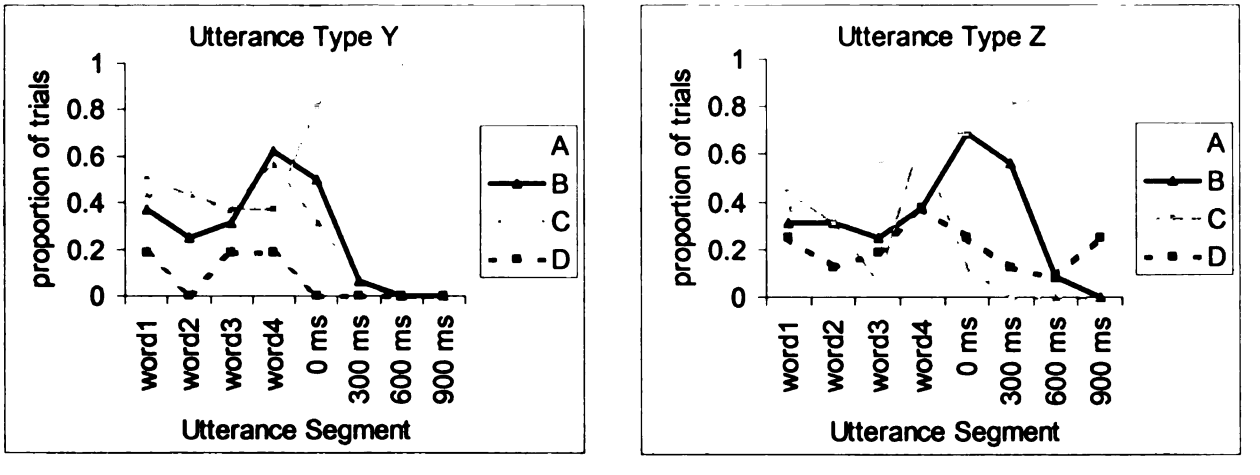
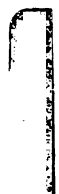


Figure 7. Probability of fixating regions of interest during each time segment in the sample data set. Bin sizes for each segment were calculated on a trial by trial basis.



The graphical analyses discussed so far have used the proportion of trials on which a particular region of interest is fixated at a particular point in time as an estimate of the probability of fixating that region of interest in response to a particular utterance in a particular visual world. The probability of fixating a region of interest is then used to infer how language comprehension processes are operating. Again, this inference also must assume that attention mediates the connection between language comprehension and eye movements, and that the visual world is being used in place of a short term memory representation of the world (Ballard, Hayhoe, & Pelz, 1995).

However, there are other possible estimates of the probability of fixating a region of interest. For instance, the frequency of fixation in a region of interest during a certain time period not only measures the tendency to fixate that region, but also the tendency to refixate. Refixations may be of interest because they can indicate situations when decisions are still being made about which of a set of possible referents is being referred to by a component of an utterance.

A third estimate calculates the proportion of time spent fixating a given region during a selected time period. This measure is useful for identifying regions that draw either very long fixations, or a great number of shorter fixations. While this analysis does conflate a frequency measure with a duration of fixation measure, it is useful because it corrects for inequalities in bin size that result from component probability analysis. Measures like the probability of fixation or the frequency of fixation show decreases for shorter words because there is less chance of a fixation landing in a region of interest during a shorter

time period. The proportion of time spent fixating corrects for this by dividing by the total number of frames in the bin for that trial.

Finally, the probability of saccade during a segment can directly capture whether a new fixation in a region was launched in response to a particular segment of an utterance. This measure may underestimate the tendency of a particular component of an utterance to cause an eye movement to a particular region in cases where that region is already being fixated because attention will remain in the same region of interest. On the other hand, this measure can correct for overestimations of the probability of fixation in cases where the region of interest is already being fixated and the participant is in a point fixation mode.

While there are some circumstances where, for instance, proportion of time spent fixating or frequency of fixation are better analyses than a probability of fixation or saccade (with large components or very slow speech, for instance), in general, a variety of measure should be used in order to converge on the correct description of the eye movement patterns in the data. Viewing the data in several ways will give a researcher a more clear picture of the relationship between the eye movement patterns and the independent variables in the experiment.

Clustering

When viewing the graphs described in the previous section, there is a temptation for researchers to draw conclusions about the "typical" pattern of fixation based on peaks and troughs in lines corresponding to various regions of interest. As mentioned earlier, however, the data displayed in the graphs are the result of averaging and thus two or more subgroups of "typical" behavior may be

present in the data. Clearly, then, a method for identifying groups of similar eye movement patterns within the data set is necessary.

In many areas of the social and biological sciences, one method used to group data consisting of individual strings is clustering. Clustering is an exploratory method used to isolate homogenous groups within a set off possible cases. Homogenous groups of cases are clustered by first calculating a distance metric based on some measurements in the data set. This distance metric is calculated for each possible pairing of cases in the data set, thus forming a distance matrix. This distance matrix is used to cluster cases through some iterative process. Depending on the iterative alogirthm, the number of clusters must be prespecified based on the researcher's intuitions, or is a result of the distance matrix and parameters selected. The makeup of the clusters will also depend on the particular iterative clustering algorithm used to group similar cases.

Because different clustering methods can produce different solutions and because the appropriate number of clusters cannot be to determined before running the analysis, some caution should be taken in interpreting the results of clustering analyses (Aldenderfer & Blashfield, 1984). There are, however, some ways of dealing with these issues. The difficulty in predetermining the correct number of clusters to use can be mitigated by running a hierarchical analysis and examining the solutions for a variety of numbers of clusters because cluster analysis is being presented here as an exploratory descriptive analysis. The hierarchical clustering algorithm (as opposed to other clustering algorithms) which identifies the least similar case on each iteration seems like a reasonable

algorithm to use because it requires no initial assumptions about the makeup of the clusters or the number of clusters. It also results in a dendrogram (a relationship diagram) showing the relationship between all the cases in the data set. Ideally, clustering should be applied to each condition separately because all strings to which clustering will be applied will then have been generated in response to a single display type and utterance type and should logically comprise a superset. This also keeps the data set to a manageable size.

Different solutions can arise, however, for different reasons than merely the iterative clustering algorithm selected. The distance metric, (i.e. the estimate of how similar each string is to every other string) is a major determinant of the final solution. The most common distance metric used with string data is the Levenshtein distance, which was developed to compare DNA sequences (Sankoff & Kruskall, 1983; Abbott & Tsay, 2000, Gusfield, 1997). This metric is computed by calculating the minimum number of deletions, insertions, and substitutions required to transform one string into another. The total number of deletions, insertions, and substitutions is standardized across strings of different lengths by dividing by the number of characters in the longest string in the pair, resulting in a distance that ranges from 0 to 1. The process of computing the Levenshtein distance is referred to as Optimal Matching (OM; Sankoff & Kruskall, 1983; Abbott & Tsay, 2000) because there are multiple possible combinations of operations that can transform one string into another. Thus, an algorithm must be used to determine the minimum number of operations required for transformation - that is, the optimal solution. OM has been used not only for analyses of DNA sequences, but also life histories (Abbott & Tsay, 2000; Abbott,

1995; Abbott & Hrycak, 1990) and eye movement patterns (Brandt & Stark, 1997).

On the surface, DNA sequences and eye movement patterns seem very similar. DNA sequences are strings generated from a four token alphabet; likewise, eye movement patterns can also be represented as strings, and have alphabets of more or less comparable sizes. However, the two types of strings differ greatly in the methods by which they are generated. These differences may have major ramifications for similarity measures. New DNA sequences arise because of the very operations used to calculate Levenshtein distance, namely insertion, deletion, and substitution of nucleotides (which correspond directly to tokens in the alphabet). Thus, the similarity of two DNA sequences can be directly related to how many of these operations have taken place on the strings.

The operations of insertion, deletion, and substitution do not map onto the generation of separate eye movement patterns nearly as well. Because each eye movement pattern is generated anew, and has not been transformed from some previously existing eye movement pattern, it is difficult to say exactly what cognitive processes involved in eye movement programming are modeled by insertion, deletion, and substitution. Moreover, eye movement patterns, especially those in the visual world paradigm, are strictly ordered with respect to time; that is, the eyes fixate objects in a particular order for cognitive reasons. While DNA sequences are also strictly ordered for reasons related to protein folding, it is also nevertheless true that substitution, insertion, or deletion of any character can occur without affecting the operations underlying the evolution of DNA sequences. In the visual world paradigm, however, differences in the

characters making up the strings and the positions of those characters within the string are the basis for drawing conclusions about the processes used to generate the eye movement patterns (in this case, language comprehension).

Elzinga (2003) has described just this issue in another field that uses strings that are strictly ordered with respect to time: sociological life histories. He notes the following problem with OM as a measure of similarity: Let us suppose that we want to compare two strings (representing life histories, as described by Elzinga, 2003, or eye movement patterns, as will be described here), ABCAB and ABCDAB, to a third string, ABABAB, and determine which is more similar. To determine the optimal distance between these strings, we need to first set the relative costs of substitutions and insertions or deletion (insertions and deletions are considered to be the same operation in OM and are referred to collectively as indels). Because there is no clear theoretical reason for setting the cost of one of these operations higher than the other in analyses of eye movements, the relative cost is set to 1 for both. This results in a Levenshtein distance of 0.667 between ABABAB and both ABCAB and ABCDAB. However, it is not clear that this is a result that makes sense when working with eye movement patterns. Specifically, the pattern ABCDAB contains fixations in two regions that were not fixated in ABABAB, while ABCAB contains fixations in only one noncommon region. Likewise, a comparison of ABCAB and ABCDAB to BACBA results in ABCDAB being more similar (a Levenshtein distance of 0.333) than ABCAB (0.200), even though ABCDAB contains a fixation in a noncommon region, while the same regions are fixated the same number of times in BACBA and ABCAB. In addition,

the distances are relatively low compared to the earlier comparisons to the string ABABAB, even though there are fewer noncommon regions over all.

An alternative distance metric based on which tokens precede which others in the string was proposed by Elzinga (2003). This metric is calculated by a Precedence Matching (PM) algorithm that compares all possible ordered k-tuples in a string to determine similarity. A k-tuple is an ordered string *k* tokens in length that is made up of a subset of the tokens in a string. The length of any k-tuple cannot be greater than the length of the string, and the tokens need not be adjacent, although their order in the original string with respect to each other may not be changed. The number of matching k-tuples in the two strings being compared is standardized by the number of within string matches and the lengths of the strings to arrive at a distance metric that ranges from 0 to 1.

The process of counting possible k-tuples begins with a comparison of individual tokens (i.e. $k = 1$) in the two strings. For ABCAB and ABABAB, these tokens would be A, B, C, A, and B for the former and A, B, A, B, A, and B for the latter. These two lists of tokens are then compared. In this case, there are 10 matches for a *k* of 1. The comparison procedure is repeated for pairs ($k = 2$; ABCAB: AB, AC, AA, AB, BC, BA, BB, CA, CB, AB; ABABAB: AB, AA, AB, AA, AB, BA, BB, BA, BB, AB, AA, AB, BA, BB, AB), triples ($k = 3$), and so on, up to the maximum length of the individual strings. While PM can be accomplished with paper and pencil, it becomes quite time consuming as the number of computations required to calculate the distance metric is approximately exponential with the length of the strings (Elzinga, 2003). As a result, an algorithm (Elzinga, 2003) has been developed that calculates matches in

polynomial time. Using PM, ABCAB (0.464) is identified as more similar to ABABAB than ABCDAB (0.381), reflecting the presence of two noncommon tokens in ABCDAB compared to just one in ABCAB. Likewise, BACBA is more similar under this measure to ABCAB (0.470) than to ABCDAB (0.338), again because of a non common token. Importantly, the similarity values in the comparisons to ABABAB and BACBA are of approximately the same magnitude in both comparisons because the majority of tokens are the same and are in relatively the same location in the string. Precedence, then, seems to be a better index of similarity than Levenshtein distance when considering eye movement patterns.

Applying the clustering analysis (calculating the distance metric using PM and using a hierarchical clustering algorithm) to the sample data demonstrates the value of identifying homogenous subgroups. In the Y Utterance Type condition, a two cluster solution includes one cluster where the two final fixations are in regions B and C respectively in each case and a second cluster in which the final fixations are in regions A and C respectively in each case (Table 2). Solutions with other numbers of clusters do not result in readily identifiable groups. In the Z Utterance Type condition, on the other hand, it is the five cluster solution that yields identifiable homogenous groups. These five clusters are defined by the set of regions fixated in the cases in each cluster, but not necessarily the particular order of fixation (Table 3).

Table 2.

particip
1
1
1
1
3
4
4
3
4
4
2
2
2
2
3
3

Table 3.

particip
1
2
2
3
4
1
1
3
4
3
4
1
2
3
4
2

and re

import

experi

Table 2. Grouping of sequences in two cluster solution for Y Utterance Type.

participant	utterance	sequence	cluster
1	1	ADBC	1
1	2	CABC	1
1	3	DABC	1
1	4	ACBC	1
3	8	ABBC	1
4	6	BABC	1
4	7	ACBC	1
3	6	CBBC	1
4	5	BCBC	1
4	8	CBBC	1
2	1	BDAC	2
2	4	BDAC	2
2	2	DCAC	2
2	3	CAAC	2
3	5	BCAC	2
3	7	DCAC	2

Table 3. Grouping of sequences in five cluster solution for Z Utterance Type.

participant	utterance	sequence	cluster
1	5	ADCBC	1
2	6	ACDBC	1
2	7	CADBC	1
3	3	DABC	1
4	1	DACBC	1
1	6	CBCBC	2
1	7	CDCBC	2
3	4	CACBC	2
4	2	CACBC	2
3	1	CACDC	3
4	3	DACDC	3
1	8	ABCBC	4
2	5	ABCAC	4
3	2	BACBC	4
4	4	BBABC	4
2	8	BADBD	5

Again, it is important to note that these analyses are descriptive in nature, and require judgment on the part of the experimenter. Nevertheless, they provide important information about what eye movement patterns are typical in a given experiment, or whether typical patterns exist at all. This information can then be

used t

partici

results

Statist

summ

drawn

chang

compa

windo

there a

statisti

made e

in the

experi

sampl

Howe

differ

infere

especi

analy

used to conduct post hoc analyses, or to determine whether particular participants, subgroups of utterances, or trials were responsible for the pattern of results seen in graphical depictions of the data.

Statistical Analyses

While descriptive analyses are very helpful in discovering and summarizing patterns in eye movement data, statistical inferences should also be drawn about whether manipulations of utterance type, task, or visual display changed the pattern of eye movements. These tests generally take the form of comparisons between the probability of fixation or saccade during a selected time window in each of the conditions. As with the selection of graphical displays, there are several factors that must be taken into consideration when selecting a statistical test. In fact, many of the points made here follow directly from those made earlier. Specifically, care must be taken in selecting the data to be included in the test and the type of statistical test itself.

Length of Analysis

The time window over which statistical analyses in a visual world experiment can be conducted ranges, in theory, from the length of a single sample to the length of the trial from onset of utterance to behavioral response. However, the former is, in practice, too small a range to detect significant differences between conditions, and the latter is too large to properly allow inferences about online processing.

Surprisingly, however, longer ranges have been the norm in many studies, especially those concerned with syntactic effects. Time windows used in these analyses have often ranged from 800 milliseconds (Trueswell, et al., 1999) to well

over 2 seconds (Tanenhaus, et al., 1995; Spivey, et al., 2002; Chambers, Tanenhaus, & Magnuson, in press). Typical measures include the proportion of trials with fixation in a region of interest at some point during the time window (e.g. Spivey, et al., 2002), and the proportion of the time window spent fixating the region of interest (e.g. Chambers, Tanenhaus, & Magnuson, in press).

The practice of using relatively long time windows is a cause for concern for several reasons. First, the longer the time window, the less informative it is about the locus of an effect of linguistic input on eye movement behavior. Thus, if a time window is long enough such that multiple words, not to mention multiple phrases, are encountered and if the measure reported is the probability that on a given trial there is a fixation somewhere within that time window, it is impossible to say exactly what word or phrase is driving the effect. In other words, the claim that the task is "online" in any sense is seriously compromised. Moreover, if the analyses of long time windows is combined with descriptive graphs that suffer from the upstream or downstream misalignments described earlier in this chapter, it will be difficult to locate effects due to the a particular linguistic segment (i.e. a critical word or phrase) in the concurrent utterance.

The use of relatively long time windows to select data for statistical analysis is also difficult to support based on the assumptions underlying the visual world paradigm. That is, there is no basis in the assumptions made by the vast majority of models of language comprehension to consistently tie syntactic or semantic parsing, never mind word recognition, to the eye movement pattern two or more seconds after that input. This is especially true given the assumption

of incremental processing, which is one of the motivations behind the use of an online measure of language processing.

Another common practice (Trueswell, et al., 1999; Chambers, Tanenhaus, & Magnuson, in press) in selecting an appropriate time window is to delay the start of the window by 200 milliseconds from the onset of the segment of interest in the utterance, in order to account for the time needed to plan an eye movement. The logic of this delay is that any eye movements that occur during the first 200 milliseconds of the segment cannot have been planned during that segment and must therefore be due to processes that were initiated prior to segment onset. However, the empirical basis for this 200 millisecond estimate lies in experiments (Rayner, et al, 1983; Saslow, 1967) that were very different in both the cue to respond and the types of target presented in visual world experiments (Altmann & Kamide, 2004). That is, the 200 millisecond estimate is based on conditions unlike the conditions faced by participants in visual world experiments. When a related experiment was conducted using the sorts of cues and utterances typically found in the visual world paradigm (Altmann & Kamide, 2004), researchers found that 60% of saccades to regions of interest were launched within the first 200 milliseconds of the cueing expression. Thus, the use of a 200 millisecond delay may miss some eye movements of interest.

In addition to the use of a 200 millisecond delay, some experiments (e.g. Chambers, Tanenhaus, & Magnuson, in press) eliminated any fixation from the analysis if it began before the segment of interest in the utterance. Again, this may underestimate effects, as participants' responses to a segment in an utterance that references the region that they are already fixating may be to

continue to focus their attention on that region. The practice of eliminating fixations that begin prior the analysis time window, when combined with a 200 millisecond delay, would, of course, further underestimate effects.

Finally, in a visual display where eye movements are cheap (i.e. about 15° of visual angle between objects; Ballard, Hayhoe, & Pelz, 1995), the probability that a fixation in the region of interest will occur during a long time window often tends towards one. As a result, ceiling effects can occur, and no difference between conditions may be found. This is of great concern, because it is in theory only when eye movements are cheap that the eye movement record is a good measure of attention and, by inference, of language comprehension processes.

Researchers, then, often ignore many issues that are directly related to the initial choice of the visual world paradigm as a methodology when they select a time window over which a statistical analysis will be conducted. If the visual world paradigm is to be used to determine the locus of particular effects in the processing of language, a time window on the order of the length of the segment of interest in the utterance is most likely the best choice. However, simply selecting one time window length and applying it to each trial can lead to other problems.

Alignment and Misalignment

As was mentioned in the earlier discussion of the descriptive graphs, the alignment of strings at one point in the graph can lead to both upstream and downstream misalignments. This is also a concern when a single window length is selected and applied to all trials because a window length based on the average length for some segment of an utterance will include additional segments on

some trials and will fail to include all of the segment of interest on others. That is, the eye movement record will be based on more input than intended on some trials and on less on other trials. Thus, some fixations will occur because of later linguistic input, while other fixations will be missed because the entire segment does not fall within the time window.

As in the descriptive graphs, a simple solution to this problem is to calculate the probability of a fixation or saccade (or whatever measure is appropriate) by setting the time window size on a trial by trial basis. This also allows statistical analyses to be tied directly to the descriptive graphs; the values for any segment on the x axis of a descriptive graph will be those compared in the statistical test.

Fixations and Saccades

Given that setting the size of the time window on a trial by trial basis is the most accurate method for selecting data for statistical analysis, it is still necessary to decide what feature of the eye movement record should be the source of values for the statistical test. At the broadest level, this is a decision about whether to count trials with a saccade launched to a region of interest or trials with a fixation in a region of interest. Of course, these two eye movement behaviors are linked (Altmann & Kamide, 2004); a saccade is always followed by a fixation. However, because the statistical analysis will be performed on data within a particular time window, counts of fixations and saccades can produce different patterns of data (Altmann & Kamide, 2004). This is because a saccade can only be launched at one point in time, while a fixation can last over a long time span. Thus, saccades are an index of an immediate shift of attention to another region of the display,

while fixations are an index of the tendency for attention to remain at a location. Thus, if the hypotheses being evaluated predict immediate shifts of attention at some point during the utterance, the probability of a saccade would be a better measure, while if the hypotheses suggest a tendency for attention to remain on a certain object, the probability of fixation over a number of segments, or fixation duration would be better suited.

As noted earlier, frequency of fixation and frequency of saccade launch can also be useful measures because they index the tendency to refixate a region of interest. When a referring expression in the utterance is ambiguous, eye movement patterns show a tendency for participants to move their eyes rapidly between all of the candidate referents. This behavior is reflected in a frequency measure, but not in a measure involving the proportion of trials during which a fixation or saccade launch event occurs, because the latter is only concerned with whether at least one event of interest occurs during the time window on a given trial. By definition, then, refixations are not counted.

The proportion of time spent fixating within a time window is also a useful measure because it can differentiate between fixations that are very short in duration (and thus presumably do not require a great deal of processing before attention is directed to another region) and fixations that are longer in duration (and thus presumably require more processing before a shift in attention occurs). However, this measure conflates refixations and long fixations, and thus needs to be used along with a frequency of fixation measure in order to differentiate between cases where a high proportion of time spent is due to a single long

fixation that might be indicative of especially difficult processing, or is due to several shorter fixations that might be indicative of rapid shifts in attention.

Statistical Tests

The final consideration in drawing statistical inferences from head mounted eye tracking data is the type of statistical test to be used. It has been common practice in visual world experiments to calculate the proportion of trials in which a fixation (or a saccade, depending on the analysis), occurs for each participant and then to submit these proportions to a an Analysis of Variance (ANOVA). This process is permissible, because a proportion can be conceptualized as the average of a number of Bernoulli trials (that is, trials where the outcome is either success (1) or failure (0)). As a result, the Central Limit Theorem can then be invoked and the normality assumption of ANOVA will be satisfied. However, the Central Limit Theorem can only be invoked if the sample size n for each participant and condition is sufficiently large. Given that many visual world experiments only have two (Chambers, Tanenhaus, & Magnuson, in press) or three (Spivey, et al., 2002) observations in each cell and only six or eight participants in total, this criterion is likely not met. A sufficiently large n is necessary because the proportion cannot otherwise be a good estimate of the underlying probability and because the distribution of proportions can become highly skewed. This is especially true in cases where the underlying probability is very large or small.

A statistical test that retains the logic of ANOVA, but that does not require the assumptions of ANOVA and is better suited for dealing with categorical data (i.e. frequencies) is multiway frequency analysis (Vokey, 2003; Vokey, 1997;

Wicke
squar
both
deper
Hiera
partit
via m
Simp
over
error
assur
theor

table
frequ
unifo
whet
exper
indiv
done
marg
and t
is con
table

Wickens, 1989). Using hierarchical tests that calculate the likelihood-ratio chi-square statistic G^2, multiway frequency analysis makes it possible to examine both unique main effects and interactions of independent variables on a single dependent variable while using frequencies, rather than proportions. Hierarchical analyses are possible because G^2 is additive; thus G^2_{total} can be partitioned in the same way that variance is partitioned in ANOVAs conducted via multiple regression. Moreover, the use of multiway frequency analysis avoids Simpson's paradox (Vokey, 2003; Vokey, 1997), the situation where collapsing over heterogeneity in individual participant frequencies can lead to a Type I error. G^2 is distributed as χ^2 for sufficiently large N and does not require assumptions about either random sampling or population parameters in a set theoretic approach (Rouanet, Bernard, & Lecoutre, 1986).

A multiway frequency analysis is conducted by first building a multiway table of observed frequencies and comparing it to a multiway table of expected frequencies. For G^2_{total}, the expected frequency for each cell in the table is a uniform distribution of frequencies. However, G^2_{total} is not informative about whether an effect is present that is specifically due to the manipulation(s) in the experiment. Thus, the structural effects of the experiment design and the individual effects due to differences in participants must be removed. This is done by computing $G^2_{structural}$ using a table of expected frequencies based on the marginal distributions of the variables and the participants in the experiments, and the interactions of the participants and the independent variables. This table is constructed by way of an iterative algorithm (Vokey, 1997). Importantly, any table of expected frequencies constructed for a higher order effect (the interaction

between an independent variable and participants, for instance) necessarily takes into account all of the lower order effects subsumed by it (e.g. the marginal distributions of the independent variable and participants). The difference between G^2_{total} and $G^2_{structural}$ can then be calculated in order to determine whether differences from expected frequencies are due to the structure of the experiment and differences in participants' responses, or to the particular manipulations in the experiment.

The logic used in calculating main effects and interactions in multiway frequency analysis is the same as is used in hierarchical multiple regression (Vokey, 2003). After the removal of the structural effects of the experiment and any possibly correlated effects of the same or lower order, whatever portion of G^2 remains (ΔG^2) can be attributed to a particular manipulation or interaction of manipulations. This is quite useful, as it allows the researcher to partial out, for instance, the effects of display in cases where more than one display is used in order to look at the effects of linguistic input, or the effects of differences in participants' individual eye movement patterns. Degrees of freedom in this analysis are calculated as in a between subjects ANOVA.

Let us suppose that for the constructed data set, we would predict that Utterance Type Y would elicit a greater number of fixations on region C during Word3 than would Utterance Type Z. An examination of the graphs in Figure 7 indicates that this pattern of results is present numerically. A multiway frequency test conducted to see if the two proportions are significantly different from each other would begin by first building a multiway contingency table (Table 4)

show
(regi
Table
Parti
1
2
3
4
each
mode
parti
inter
abse
parti
avera
parti
parti
signi
fixati
the sa

showing the number of trials with and without a fixation in the region of interest (region C) during the segment of interest (Word3) for both utterance types.

Table 4. Contingency table used in Multiway Frequency Analysis of sample data.

	Utterance Type Y		Utterance Type Z	
Participant	Fixation Present	Fixation Absent	Fixation Present	Fixation Absent
1	1	3	0	4
2	1	3	1	3
3	2	2	0	4
4	2	2	0	4

A unique likelihood-ratio chi-square statistic (ΔG^2) is then calculated for each effect or interaction by calculating G^2 for the full model and for a nested model that contains all of the structural and possibly correlated effects. In the particular analysis described by the contingency table above, we would be most interested in the effect of utterance type on the dependent variable (presence or absence of fixation). The main effect of participant and the interaction of participant and utterance type are also tested in order to determine whether the averaged proportion displayed in the graph is due to the behavior of only a few participants, or due to a pattern of response across the entire group of participants in the experiment. The test of the main effect of utterance type is significant ($\Delta G^2 = 5.06$, $df = 1$, $p<0.05$), thus supporting the hypothesis that more fixations on region C should occur during Word3 of Utterance Type Y than during the same segment of Utterance Type Z. The main effect of participant ($\Delta G^2 =$

4.07, df = 15, p>0.05) and interaction of participant and utterance type (ΔG^2 = 3.38, df = 15, p>0.05) were not significant, suggesting that any differences in response patterns were not due to the responses of a small subset of participants.

A corresponding ANOVA conducted on the arcsin transformed proportions for each participant indicated a marginal effect of utterance type ($F_{1,3}$ = 8.00, p=.066). Although it is impossible to draw conclusions based on this single example, it is encouraging to see that both the typically used parametric test and the more appropriate nonparametric test are in agreement. Indeed, one strategy (Altmann, 2004) for dealing with the tension between using a nonparametric test and the pressure to use a statistical test that is well accepted in the peer review system is to use a nonparametric test to confirm the results of the parametric test.

Summary

Information about the position of the eye is collected at intervals of a few tenths of a second or less during every trial of a visual world experiment. After calibrating known eye position to fixations on particular regions of the visual world, the eye position record can be transformed into a record of the regions of the visual world fixated throughout each trial. The resulting eye movement record is most easily considered as a string. When considering possible data analysis techniques, the string format of the data must be considered.

Based on the review of data analysis techniques presented in this chapter, we can draw some conclusions about the strategies for the analysis of eye movement data in visual world experiments. First, care must be taken to avoid downstream and upstream misalignments in both descriptive graphs and

statistical analyses. This can be done by calculating time windows for analysis on a trial by trial basis (Altmann & Kamide, 2004). Although some amount of resolution is lost by this process, the descriptive graphs and statistical analyses can be directly tied to each other and to the individual segments of utterances presented on each trial. Such an analysis is similar to the types of analyses used in eye tracking studies of reading (Rayner, 1998). Moreover, this method provides a principled method for determining the time windows used to select data for statistical analysis and allows the locus of any effects to be easily determined.

In addition to descriptive techniques (e.g. graphs) that average over participants and trials, researchers should also examine the data for homogenous subgroups of eye movement patterns in each condition. Doing so prevents researchers from falling prey to the temptation to treat the averaged data as a typical pattern of response when in actuality several patterns may exist. One method for identifying subgroups that shows promise is cluster analysis. In cluster analysis, eye movement patterns are grouped based on their similarity with each other. While most researchers in the past have used the optimal matching algorithm developed for clustering of DNA sequences to calculate similarity, this measure may not be the best for eye movement data. A different algorithm based on the order of fixations in the eye movement record appears to have properties better suited to studies of eye movement data. Caution should be exercised when interpreting the results of cluster analyses, however, as clustering has not yet developed a strong statistical basis and thus different techniques can yield widely different solutions in some cases.

When considering statistical analyses, researchers must select measures based on the hypotheses that they are testing. If a hypothesis predicts an immediate shift of attention upon hearing a certain word, then the researcher should use a measure (i.e. saccade launch) that reflects this prediction. Likewise, if a hypothesis predicts a great deal of time will be spent fixating a particular region of interest, a probability of fixation or proportion of time spent fixating measure may be more appropriate.

Lastly, because the design of most visual world experiments limits the number of observations in each cell in the design, a multiway frequency analysis is more suitable for analyzing the eye movement data than the more commonly used ANOVA. Hierarchical multiway frequency analysis partials out variance due to structural factors, as well as possible correlated effects, and allows for the asymmetrical designs (multiple independent variables, one dependent variable) typical of psycholinguistic research.

COMPREHENSION OF DISFLUENT SPEECH: AN INITIAL STUDY

Experiment I

Few studies of the effects of disfluencies on comprehension to date have made clear predictions about their effects on parsing. One exception is the Bailey & Ferreira (2003) study examining the role of filled pause disfluencies (e.g. "uh") in guiding the syntactic parse of an utterance. In that study, the authors proposed that differences in the proportion of sentences judged as grammatical were driven by the locations of filled pause disfluencies. In their experiments, participants judged the grammaticality of sentences that contained a noun phrase that was temporarily ambiguous as to whether it was the object of the current clause or the subject of a new clause. Disfluencies could appear before an ambiguous head noun (in bold) as in (16) or after the ambiguous head noun as in (17).

(16) Sandra bumped into the busboy and the uh uh **waiter** told her to be careful.

(17) Sandra bumped into the busboy and the **waiter** uh uh told her to be careful.

Bailey and Ferreira (2003) found that sentences with disfluencies before the head noun of the ambiguous phrase were judged grammatical more often than sentences with disfluencies after the head noun. This suggests that the process of reanalysis was more difficult in (17) than in (16), although these utterances differed only in the position of the disfluency.

Two reasons were suggested for this effect of disfluency location on parsing. The first is that disfluencies take up time (consistent with Ferreira &

Henderson, 1999), allowing an incorrect partial analysis to remain active longer and making reanalysis more difficult. Thus, (17) is more difficult to reanalyze because "waiter" is initially assigned as an object of "bumped into". The presence of a disfluency prior to the disambiguating word in (17) means that this assignment is maintained for longer in (17) than in (16) prior to disambiguation. The other explanation is that the parser may be sensitive to the higher probability of a disfluency at the beginning of a complex constituent such as a clause. Thus, when a disfluency occurs in the immediate vicinity of a possible clause boundary (as in (16)), the parser may take this as evidence that there is, in fact, a boundary, and would thus be less likely to be garden pathed.

Because these two explanations are confounded in (16) and (17), Bailey and Ferreira (2003) compared sentences where the disfluency always occurred prior to the ambiguous head noun and thus could not affect the amount of time that it was incorrectly assigned as an object. Utterances like (16), where the disfluency occurs at a place that is consistent with the eventual structure of the utterance were compared with utterances like (18), where the disfluency was inconsistent.

(16) Sandra bumped into the busboy and the uh uh **waiter** told her to be careful.

(18) Sandra bumped into the uh uh busboy and the **waiter** told her to be careful.

Again, sentences with a consistent disfluency were judged grammatical more often than those with an inconsistent disfluency. Moreover, although the sentences with disfluencies patterned in the same manner as sentences with

modifying words (e.g. the tall and handsome waiter; the waiter that was handsome; see also Ferreira & Henderson, 1991, 1999), in experiments comparing utterances like (16) and (17), there was no such effect of modifying words in the experiment where (16) and (18) were compared. Disfluencies, then, may affect syntactic parsing in ways that words cannot because they have a distribution with respect to syntax that differs from the distributions that words and word classes have. However, as noted in the introduction, the grammaticality judgment task is generally viewed as an offline task. Thus, there is no way of knowing when the parser makes use of the disfluency information.

In order to determine whether the disfluency immediately affects parsing, an ideal test would be to present participants with completely ambiguous sentences that have a disfluency in a location that is consistent with only one possible structure, while monitoring their eye movement behavior in a an online task. Recall utterance (11), repeated here for ease of discussion, which can have the meanings in (12) and (13).

(11) Put the apple on the towel in the box.

(12) Put the apple that's on the towel into the box.

(13) Put the apple onto the towel that's in the box.

The online technique best suited to studying spoken language comprehension is the visual world task. In a two referent visual world (Figure 8) containing a apple on a towel (target object), an apple by itself (alternate object), a towel in a box (early goal; early here refers to the point in the utterance at which the goal would be identified), and a box by itself (late goal), both of the interpretations in (12) and (13) are possible (although (12) is more felicitous). I

will refer to such a visual world as an ambiguous display. If the hypothesis that disfluency location can cue a particular structure is correct, then an utterance such as (19) should elicit a different pattern of eye movements than an utterance such as (20). Brackets in (19) and (20) indicate the hypothesized structure that may be constructed if the parser uses the presence of a disfluency to indicate the initiation of a complex constituent. I will refer to the disfluency in (19) as a theme biased disfluency, as it is hypothesized to bias the parser towards a heavy theme interpretation. Likewise, I will refer to the disfluency in (20) as a goal biased disfluency.

(19) Put the uh uh [apple on the towel] in the box.

(20) Put the apple on the uh uh [towel in the box].

If this information is used immediately, we should see more looks to the target object (i.e. the apple on the towel) upon hearing "apple" after a theme biased disfluency as in (19) than in an utterance such as (20) where the disfluency has not yet occurred at the word "apple". We should also see fewer looks to the distractor or alternative object in the theme biased disfluency condition. Later in the utterance, after "on the towel", we should see fewer looks to the early goal (i.e. the towel in the box) after a theme biased disfluency as in (19) than in (20), both overall and immediately after hearing "towel".

The same sorts of effects should also be seen in unambiguous one referent and two referent displays (Figure 8). In the two referent display, the theme biased disfluency in (19) should aid in the resolution of the visual ambiguity between the two possible referents of "apple" relative to the as yet unheard disfluency in (20) or the fluent utterance in (11). In the one referent display, on

the other hand, the goal biased disfluency in (20) may lead to more garden pathing, and thus more looks to the incorrect goal, as participants may expect that the complex constituent "the apple on the napkin" will not be produced.

This experiment compared the eye movement behavior of participants who viewed the three types of display (ambiguous, one referent, and two referent; Figure 8) described above, and listened to either of the types of disfluent utterances shown in (19) and (20) or the type of fluent utterance shown in (11).

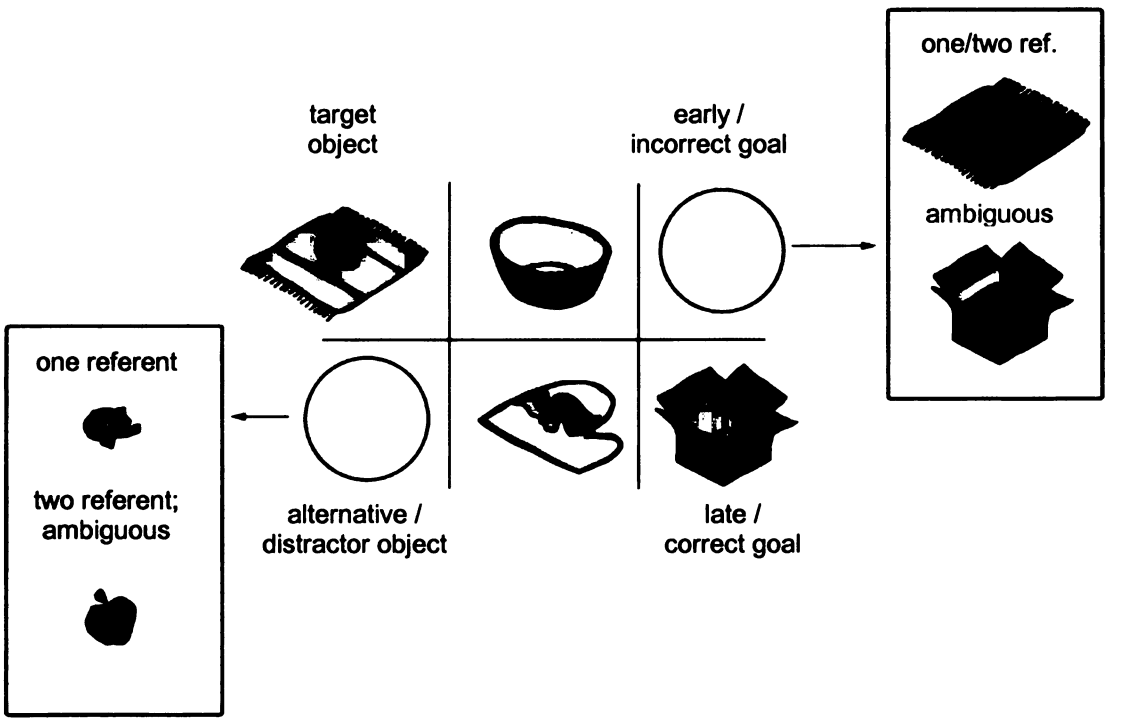


Figure 8. Displays used in Experiment I. Shades are used to differentiate regions in the visual world, and do not reflect the actual colors of objects used in the experiment. Boxes on the left and right indicate differences between display conditions in distractor/alternative objects and incorrect/early goals respectively. The two middle objects in gray are irrelevant distractor objects and locations that were referenced on some filler trials.

Material and Methods

Participants. Eight participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($10.50). All participants were native speakers of English and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported here.

Materials. Seventy two prepositional phrase ambiguity utterances (the structure exemplified by (16)) were constructed for this experiment using the nouns in Table 5. Utterances were recorded and digitized using the Computerized Speech Laboratory (Kay Elemetrics) at 10 kHz, and then converted to wav format. Each utterance was recorded in two ways: once as a fluent utterance, and once as an utterance with two disfluencies, as in (21).

(21) Put the uh uh apple on the uh uh towel in the box.

Table 5. Objects used in Experiments I-IV.

possible target objects	possible "in" goals	possible "on" goals
apple	basket	felt
banana	bowl	glove
bat	box	napkin
brush	flowerpot	sock
candle	mug	sponge
car	pan	towel
dinosaur		
football		
frog		
jeep		
lizard		
plane		

Utterances like (19) and (20) were then created by excising the appropriate disfluency. Participants are relatively insensitive to the removal of disfluencies from utterances (Fox Tree, 1995; Fox Tree & Schrock, 1999; Brennan & Schober, 2001; Fox Tree, 2001) and thus this procedure was used to control the prosody of the disfluent utterances. The removal of a single disfluency from an utterance did not result in utterances that participants found odd or strange. A disfluency-removed version of the utterances was also created by removing both disfluencies with the intention of comparing the disfluency removed utterance to the fluent control; however, this proved to be very disruptive to the prosody of the utterance, leading to confusion on the part of the participants and thus this condition will not be discussed further. Each participant heard only one version of any given utterance. For this initial study, a single order of trials was created. Utterance types used in this experiment are shown in Table 6.

Table 6. Utterance types used in Experiment I. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance (segments for analysis indicated by subscripts)
Fluent	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
Theme Bias Disfluency	/$_{VERB}$ put /$_{NP1}$ the uh apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
Goal Bias Disfluency	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the uh towel /$_{PP2}$ in the box./

In addition to the critical items, 288 filler utterances were created. These involved a variety of syntactic structures. Participants heard four filler items and one critical item for each of 72 displays. The critical item appeared an equal number of times in the second, third, and fifth instruction positions.

Displays consisted of a two by three grid (Figure 8). Each object or goal type appeared equally often in each position. The one referent and two referent displays were conceptually equivalent to those used in previous experiments (Tanenhaus, et al., 1995; Spivey, et al., 2002), while the ambiguous display was added because it did not force the syntactic interpretation of the utterance.

Apparatus. Participants wore an ASL model 501 head mounted eyetracker (Applied Science Laboratories). This eyetracker consists of a pair of cameras that are securely fastened to a participant's head. One camera records the scene that the participant is currently viewing. The other, an infra-red camera, records the movement of the participant's left eye. An infra-red emitter is housed within this camera, and illuminates the eye. Because this emitter is attached to the participant's head, the corneal reflection stays relatively stable across eye (and head) movements while the pupil moves. By measuring the offset between these two components of the eye image as the eye is moved to foveate different parts of the scene, it is possible to identify where in the scene the participant is looking. This is realized as a small cross on the scene image that corresponds to the center of fixation and moves as the eye moves. The location of this cross is accurate to within a degree of visual angle under optimal conditions.

The merged scene and fixation position video data were recorded at 30 Hz. The video was then stepped through frame by frame, starting with the onset of each critical utterance and ending with the movement of an object to a new location. Each change of state in fixation position was recorded and converted to a frame by frame string format for further analysis.

Procedure. After participants were introduced to the objects and apparatus and had provided informed consent, the eyetracker was adjusted and placed on their head. Depending on the height of the participant, and the relative structure of their eyes, eye sockets, and eyelashes, participants either stood or were seated at a table with a central partition and a surface that could be rotated. Participants' eye positions were then calibrated to the scene by recording their eye positions while they looked at nine predetermined positions.

The experiment made use of a table that had an opaque divider in the center. The table top could be rotated on a central pivot. This allowed the experimenter to set up objects for the next trial while the participants was completing a trial, thus minimizing the amount of time participants had to preview the display. At the beginning of each trial, the table top was rotated to reveal a new set of objects. Participants were instructed to look at a central red square prior to receiving the instructions for each trial in order to indicate to the experimenter that they were prepared to begin. The participants then heard the five instructions (four filler and one critical) for that set of objects. After completing the final instruction for that set of objects, the table top was rotated, revealing a new set of objects and beginning a new trial. Calibration was monitored throughout the experimental session, and participants were recalibrated if necessary. Each session took approximately 90 minutes to complete.

Design. The four utterance conditions (Table 6) were paired with the three display conditions (Figure 8) to create 12 different conditions for this experiment. As noted earlier, the disfluency removed utterances were not considered because

the removal of two disfluencies in a single utterance was too disruptive. Neither fluent condition was included in the statistical analyses. Six trials in each condition were presented to each participant, for a total of 72 trials.

Data Analysis

As discussed in Chapter 2, previous experiments using the type of visual world used here have reported differences in the amount of attention directed to the distractor object (e.g. the apple not on the towel, or the unrelated distractor) and the incorrect goal. Two specific measures of attention (as evidenced by eye movements) have been used. The first is the proportion of trials on which a fixation in the region of interest occurred (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002). This measure indicates the probability of attention being directed to that particular region, but does not allow inferences about the amount of processing that occurred when applied to long segments of time, as it is not sensitive to the total number of fixations and refixations, nor the amount of time spent fixating the region of interest. A second measure, the proportion of time spent fixating the region of interest, has been used for the purpose of measuring the amount of processing during a segment of time (Chambers, Tanenhaus, & Magnuson, in press). While researchers have claimed that this measure is reflective of both the number and duration of fixations in a given region during the processing of an utterance (Chambers, Tanenhaus, & Magnuson, in press), in actuality these two sources of information are conflated in this measure. Thus, while proportion of time spent fixating is a measure of the amount of processing required, it is still a rather coarse measure when applied over the duration of an entire trial.

In this study, both the probability of fixation and the probability of saccade were measured during short time segments. The probability of saccade is a measure that has been used quite frequently in the literature of anticipatory eye movements in the visual world (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Altmann & Kamide, 2004; Altmann, 2004). The probability of a saccade to a region of interest during a time segment is a good measure of the tendency for attention to shift to that region during that particular segment of time. Moreover, any saccade can only be launched at one point in time. The probability of a saccade is a conservative measure of attention; if a participant fixates a region of interest prior to the start of a trial and does not move his or her eyes during the trial, a probability of saccade measure will not be artificially inflated in the same way that a probability of fixation measure would be. Of course, the probability of saccade is not a perfect measure for the same reasons that it is conservative: The region of interest may already be being fixated, and thus no saccade can be launched. Accordingly, I will use both the probability of fixation and the probability of saccade as convergent measures to describe how attention is being deployed at points of interest during the processing of utterances in this experiment.

The time segments used for calculating the probabilities of fixation and saccade will be generated based on the phrases in each utterance on a trial by trial basis (Altmann & Kamide, 2004) as described in Chapter 2. Each utterance will be divided into four segments as in (22) and will be referred to using the corresponding labels in (23). Segments for all utterance types can be seen in Table 6.

(22) / Put / the apple / on the towel / in the box. /

(23) / verb / NP1 / PP1 / PP1 /

A phrase by phrase analysis was selected in this case instead of a word by word analysis for several reasons. First, the prepositions in these utterances ("on" and "in") are very short, as are all of the determiners (recall that these are spoken utterances). Thus, there are very few fixations or saccades in total that occur during these words. As result, the probability of any eye movement behavior during a short word is deceptively low when compared to the longer nouns and verbs. In addition, it is not immediately clear whether we should expect saccades to be launched to possible goals at the preposition (e.g. "on the towel) or at the noun (e.g. "on the towel") in a PP. Most previous studies have assumed the latter (Trueswell, et al., 1999; Chambers, Tanenhaus, & Magnuson, in press), but research on anticipatory eye movements in the visual world paradigm (e.g. Altmann & Kamide, 2004) might suggest the former, especially as the preposition the point at which it would be possible for the parser to postulate a phrase referring to a goal. At any rate, it is likely that some eye movements will be generated during each word in the PP, especially if the parser acts upon each constituent in the PP immediately (under the assumption of incrementality), and thus the entire phrase will be treated as a segment here.

The counts of trials with and without a fixation in (or saccade to) the region of interest during the segment of interest for each participant were submitted to a multiway frequency analysis (Vokey, 2003). This analysis indicates whether the frequencies in each condition differ from what might be expected if the counts were evenly distributed, after accounting for any

covariation with the structure of the experiment and any possibly correlated effects or interactions. Covariance is accounted for by conducting nested calculations of the likelihood-ratio chi-square statistic G^2 using the same logic as hierarchical multiple regression. A significant ΔG^2 (that is, a ΔG^2 value that is higher than the appropriate critical value from the χ^2 distribution) suggests that there is a relationship between the tested effect or interaction and the dependent variable.

Finally, all analyses in this experiment were conducted separately for display type, except when it was necessary to compare results to previously reported findings. Because it is impossible to control for differences in the visual properties of the different display types when different display types are made up of different, if overlapping, sets of objects, it is likewise impossible to say whether a difference in eye movements between display types is due to differences in the interpretation of the utterance in the context provided by the display, or differences in the ambiguity and saliency of the display itself. Comparisons of the effects of utterance variables are thus more easily made within each display type, and predictions can be more easily generated for each display type separately.

Results and Discussion

Fixations and saccades during each segment of the utterance (Table 6) were counted and were submitted to a one way (disfluency location) multiway frequency analysis for each display or a one way (number of referents in the display) multiway frequency analysis of responses in the fluent utterance conditions in those cases where it was necessary to examine the effects of

differences in display. Segments were calculated for each trial independently to account for differences in word length.

The hypotheses generated by Bailey and Ferreira (2003) suggest that the location of disfluencies in the two referent and ambiguous display conditions should affect the proportion of looks to the distractor object. Recall that in these displays (Figure 8), the distractor object is identical to the target object; that is, the distractor object may be temporarily considered as a possible referent of the first NP. Thus, if the presence of a disfluency is used by the parser to predict an upcoming complex constituent, we should expect to see fewer immediate looks to the distractor object in the theme biased disfluency condition. In the one referent condition, on the other hand, the distractor object is never a possible referent of any noun in the utterance; thus, the location of a disfluency should not make the distractor object any more or less likely to be fixated.

This hypothesis, however, is not fully supported by the data collected in Experiment I. While a main effect of disfluency location is present in the two referent display ($\Delta G^2(1, \underline{N} = 96) = 5.56$, $p < 0.05$), it is not in the predicted direction. Rather, the probability of a saccade to the distractor object during the NP1 segment of the utterance was higher in the theme biased disfluency condition (0.25) than in the goal biased condition (0.083). This pattern can be seen in Figure 9B. The same pattern is present numerically, although it is not significant for either the one referent ($\Delta G^2(1, \underline{N} = 96) < 1$, $p > 0.1$) or the ambiguous ($\Delta G^2(1, \underline{N} = 96) < 1$, $p > 0.1$) display. Analyses of the probability of fixation were all nonsignificant ($p > 0.1$) for each display, except for the ambiguous display, which showed a marginal effect of disfluency location ($\Delta G^2(1,$

N = 96) = 3.75, $p < 0.1$). This marginal effect was also in the opposite direction of that hypothesized by Bailey and Ferreira (2003): More fixations on the distractor object occurred in the theme biased disfluency condition than in the goal biased disfluency condition.

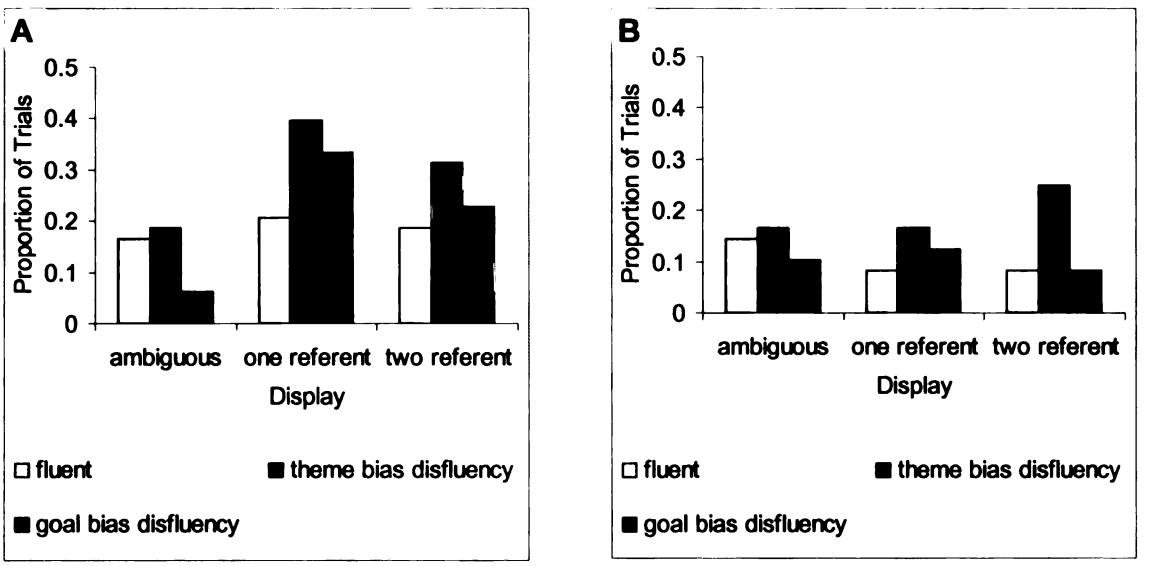


Figure 9. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor or alternative object during the NP1 segment of the utterance. Proportions for the fluent control utterance are included for comparison.

Previous studies that examined the processing of prepositional phrase ambiguities (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002) have found that there are more fixations on the distractor object in the two referent condition than the one referent condition in response to the same utterance. In the current study, however, no such relationship exists. When the proportion of trials with a fixation on the distractor object was compared in the fluent utterance condition for both the one referent and two referent display, no effect of display was found ($\Delta G^2(1, N = 96) < 1$, $p > 0.1$). The same was true for the proportion of trials with a saccade to the distractor object ($\Delta G^2(1, N = 96) < 1$, $p > 0.1$).

One possible reason for this somewhat surprising finding is that there were additional distractors in the displays used in this study. The presence of additional objects that were not referred to by the concurrent utterance may have been responsible for the failure to replicate previously reported effects because they might have drawn additional fixations early in the utterance and thus reduced the overall probability of fixating the distractor or alternative object. Furthermore, the number of objects present in the display (six) may have exceeded the working memory capacity available to participants. Thus, in addition to eye movements made because of activation of a concept in memory (Altmann, 2004), there may have been eye movements made for the purposes of exploring the display or for refreshing memory of what objects exist in the word. This would result in an increased tendency for any distractor object to be fixated, whether or not it was referred to by the concurrent utterance. If true, however, the presence of exploratory eye movements in more complex displays would suggest that participants' eye movements can be decoupled from linguistic processing simply by increasing the complexity of the display, and that the visual world paradigm must thus be limited to very simple displays. In fact, this suggestion is consistent with the relatively high level of eye movements unrelated to concurrent linguistic input found in the original Cooper (1974) study, which involved nine objects (line drawings) in an array.

Graphs showing the probability of fixation in or saccade to the alternative object in the ambiguous display for each of the four utterance segments are shown in Figure 10. (The bars for the ambiguous display in Figure 9 correspond to the points for the NP1 segment in Figure 10.) These graphs indicate that while

the expected relationship between disfluency location and probability of fixation or saccade did not occur at NP1, it did occur in the very next segment, PP1. The effects of disfluency location at PP1 on both the probability of fixation ($\Delta G^2(1, \underline{N} = 96) = 4.68$, $p < 0.05$) and saccade ($\Delta G^2(1, \underline{N} = 96) = 6.55$, $p < 0.02$) were significant and were in the direction predicted by the disfluency cueing hypothesis. The same numerical pattern was seen for both the probability of fixation and saccade in the one and two referent display conditions (Figure 11), but the differences were not significant ($p > 0.1$ for all analyses). Thus, it is possible that, in the ambiguous display condition at least, the effect of disfluency was simply delayed, and did not emerge until the following segment.

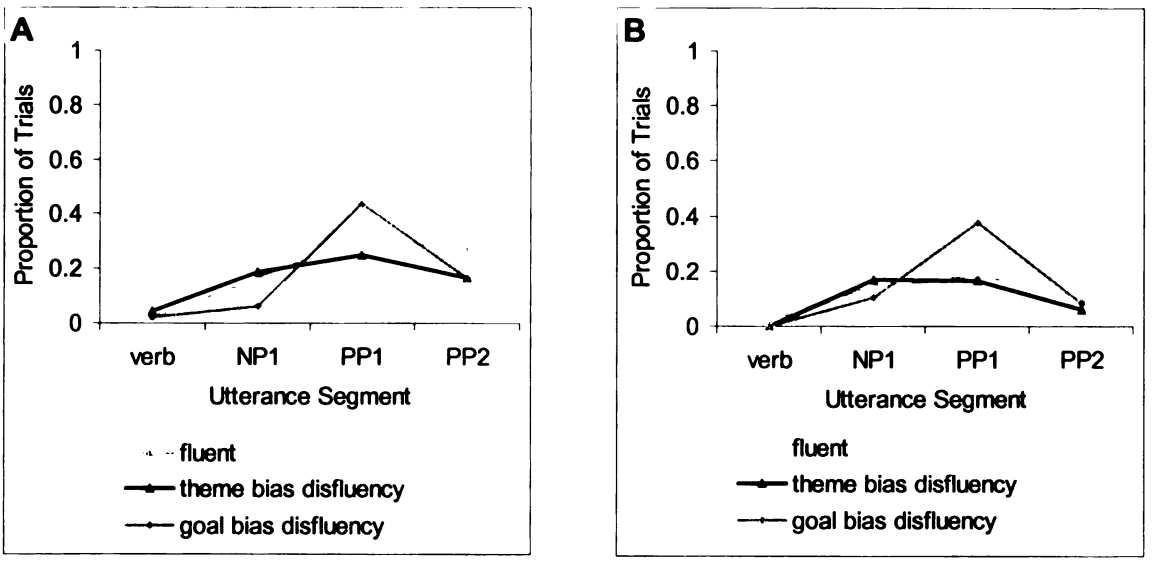


Figure 10. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor or alternative object for each segment of the utterance.

There is another possible explanation for the reversal of direction between the NP1 and PP1 segments in the relationships found between disfluency location and eye movements. In both cases, the condition that actually contained the disfluency had a higher probability of fixation or saccade. One reason for this

might be that the presence of a disfluency simply adds time to the utterance and this extra time allows for more eye movements, thus increasing the probability of fixating a particular region. However, this is not the most likely explanation, as it would predict that the same pattern should hold true during the disfluency for any region in the display, a prediction which is easily falsified by examining the irrelevant distractors and the possible goal object and noting that there is no effect of disfluency location on fixations on or saccades to these regions (note, for instance, the lack of effect in the one referent display where the distractor is irrelevant).

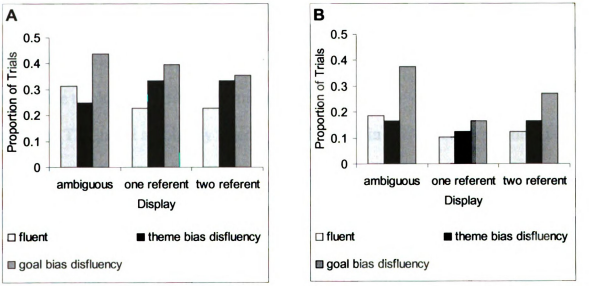


Figure 11. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor or alternative object during PP1 segment of the utterance. Proportions for the fluent control utterance are included for comparison.

A modified version of this hypothesis might suggest that the increase in eye movements due to additional time should only affect the objects that are currently being considered as possible referents; that is, the target object and the distractor or alternative object (except in the one referent condition, where the

distractor object is not a possible referent). This explanation accounts for the eye movement patterns in the ambiguous display, but cannot account for the lack of a significant difference in the two referent display, nor can it account for the large tendency to fixate an irrelevant distractor in the one referent display (Figure 11).

An alternative explanation might suggest that the disfluencies in the experiment are being interpreted as cues, not to the syntactic structure of the utterance, but rather to a difficulty in establishing the appropriate referring expression. To see why this might be the case, we must examine the task from the perspective of participants in this type of visual world experiment. In addition to being faced with an utterance that is fully ambiguous, participants are also faced with one or more ambiguities in the copresent visual world. In addition, the task that they are required to perform involves connecting referring expressions in the utterance to objects in the world and then acting on those objects in a timely manner. The initial ambiguity faced by participants in this experiment, then, is not syntactic. Instead, it is the referential ambiguity faced upon encountering NP1. Thus, when a disfluency is encountered, participants may attribute the speaker's difficulty to a problem in establishing reference and, thus, may make a series of rapid eye movements in order to predict or anticipate which of the possible referents might be the correct one. Anticipatory eye movements might also be made simply for the purpose of guessing what might be the source of the speaker's difficulty.

Although this account does not explain the lack of an effect of disfluency location during PP1 in the two referent condition (this may simply be due to the low number of participants in this experiment, as the expected pattern of

fixations and saccades is present), it does explain the increase in fixations on the irrelevant distractor during the one referent condition. Because in the one referent condition participants are able to match NP1 to a single object in the display immediately, they may be using the disfluency in the goal bias condition during PP1 to scan the display in order to account for why the speaker is being disfluent.

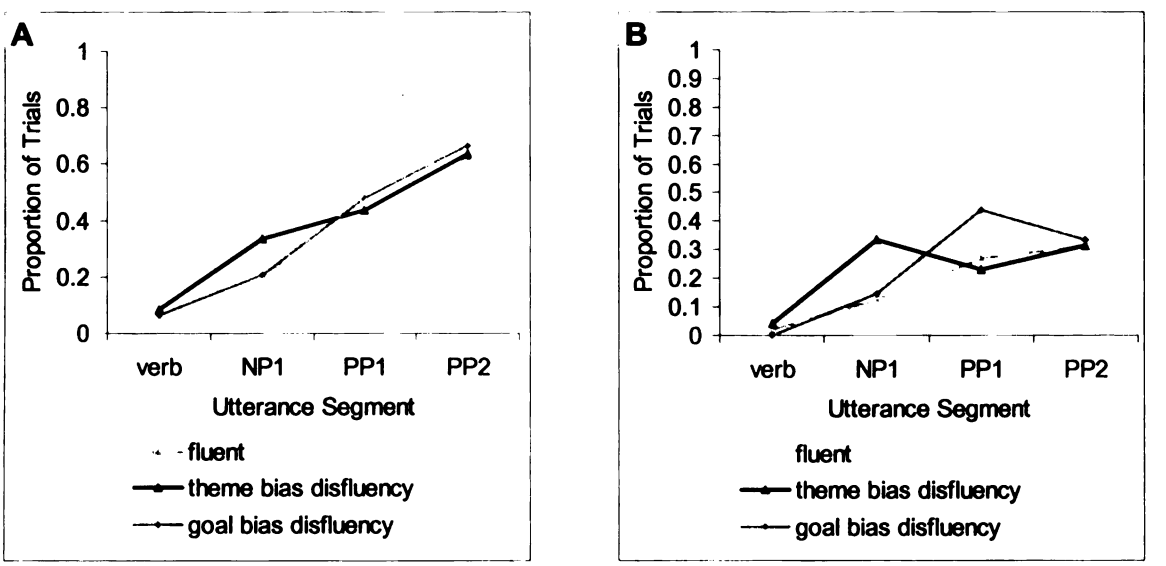


Figure 12. Proportion of trials with a fixation in (A) and with a saccade to (B) the target object for each segment of the utterance.

Interpretation of the disfluencies in this experiment as cues to referential ambiguity rather then syntactic complexity might also explain other effects noted in the data collected during Experiment I. First, the probability of launching a saccade to the target object (Figure 12B) mirrors the pattern of data described in saccades to the distractor/alternative objects (Figure 10). Interestingly, this pattern is not paralleled by the probability of fixation Figure 12A) for the target object region. This may be because the target object, once fixated, continues to be fixated until the parser is satisfied that it is the correct target object. Thus, those

fixations that are the result of saccades to the target object during NP1 continue through and contribute to the counts of fixations during PP1. At the same time, new saccades are launched to the target object and result in new fixations.

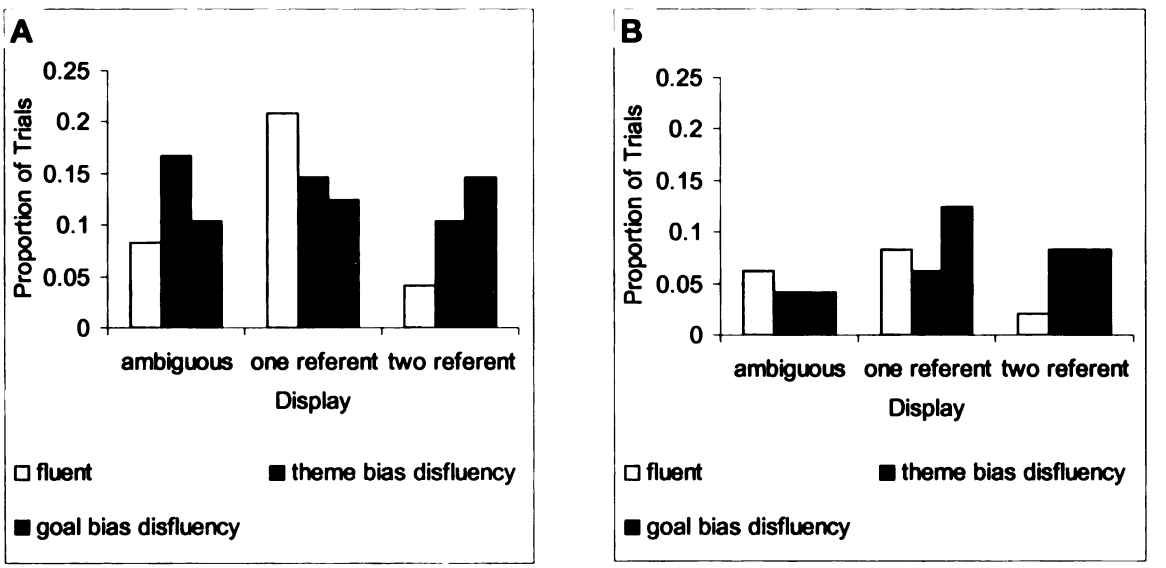


Figure 13. Proportion of trials with a fixation in (A) and with a saccade to (B) incorrect or early goal during PP1 segment of the utterance. Proportions for the fluent control utterance are included for comparison.

A referential ambiguity explanation might also explain the lack of an immediate effect of disfluency location on fixations on the incorrect/early goal (Figure 13) in all display conditions: No difference is present due to disfluency location ($p > 0.1$ for all tests). Looks to the incorrect goal have been interpreted as the presence of garden pathing due to the context set up by the one referent display (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002; Chambers, Tanenhaus, & Magnuson, in press) and thus we might expect to find that disfluencies affect the strength or chance of a garden path, at least in the one referent. This is especially true given that Bailey and Ferreira (2003) used an offline measure, grammaticality judgment, that has traditionally been taken to be a good index of the presence of garden path effects.

An effect of display was noted for fluent utterances ($\Delta G^2(1, \underline{N} = 96) = 7.52$, $p < 0.01$), and replicates the effect found in previous studies where an effect interpreted as a garden path (indexed by the probability of fixating the incorrect goal) is found in the one referent, but not the two referent display. No such effect was found for the probability of launching a saccade to the incorrect goal ($\Delta G^2(1, \underline{N} = 96) = 1.20$, $p > 0.1$), suggesting that the pattern of fixations interpreted as a garden path may not be due to the processing of the utterance during PP1. It is worth noting that an analysis of the probability of saccade launch has not previously been performed in this type of visual world paradigm.

The results from this initial application of the visual world paradigm to the study of disfluency processing, then, were mixed with respect to both the standard effects found in the visual world paradigm and its application to the study of disfluencies. The majority of effects that were found occurred immediately after the referential expression corresponding to the region of interest. This suggests that the longer time windows used in previous studies are unnecessary; the visual world paradigm can, in fact, be used to study online processing of spoken language. Another difference between the current study and previously reported findings concerns the distractor objects in the display, and suggests that there may be many cases when participants' eye movements are not tightly locked to language comprehension. This study was able to partially replicate the garden path type effects reported in the literature; however, the effects were only present in the analysis of fixations and not saccades, suggesting that the effect may be due to fixations on the incorrect goal in the one referent display that occur prior to the onset of PP1 (but continue through it). One reason

that these early saccades might be made is that NP1 reference resolution has been completed and an eye movement has been generated in anticipation of a goal. That is, the early saccades to possible goal locations may be exactly the type of anticipatory saccades described in studies of verb processing (e.g. Altmann & Kamide, 1999; Altmann & Kamide, 2004).

It also appears that the visual world paradigm may not be a good technique for examining syntactic issues (as has been previously assumed) in cases where a referential ambiguity can mask any syntactic effects, and that the display or task may be mitigating any effects of disfluency on eye movement patterns that reflect syntactic parsing. The visual world paradigm may instead be highly sensitive to the resolution of reference. However, given the presence of extra distractors in this experiment, it is necessary to examine a simpler visual world before drawing conclusions as the presence of additional distractor objects may have affected the relationship between the internal memory representation of the visual world and the movement of attention through the external visual world. This link is key to the interpretation of eye movement patterns in the visual world (Altmann, 2004).

One major reason for concern about the basic eye movement patterns in the visual world paradigm is the failure to find effects of disfluencies later in the utterance. The effects of disfluency on parsing described by Bailey and Ferreira (2003) concerned the processing of garden path utterances. Thus, it was expected that it was at the point of a possible garden path that effects of disfluencies would be found most strongly in the visual world paradigm. However, the only effects of disfluencies present were in the initial proportion of looks to the target referent in

the displays where the target referent was initially temporarily ambiguous. This may indicate that as the utterance is disambiguated by the display, the effects of the location of disfluency are overwhelmed by earlier ambiguity in the display and the difficulty of reference resolution.

In summary, then, there are some indications that the visual world paradigm may be valuable for the study of the processing of disfluencies (and the processing of spoken language in general), but first we need to have a solid understanding of the effects of ambiguities within the display and the demands of the instruction-following task on the processing of fluent utterances. In other words, one of the main observations that we can draw from the results of this initial study is that not enough is known about the visual world paradigm itself. The next few experiments focus on the processing of fluent utterances in the visual world paradigm in order to better establish a basic understanding of eye movement behavior in this task.

THE VISUAL WORLD PARADIGM: TASK EFFECTS

Experiment II

One possible reason for the mixed results in Experiment I may be differences in the way the task was presented in this study as compared to other studies that have used this paradigm. The possibility of task effects in visual cognition paradigms has been known since the descriptive studies conducted by Buswell (1935) and Yarbus (1967). In addition, Cooper (1974) found that specific instructions to listen to stories in preparation for a comprehension test led to an increase in the tendency for participants to hold their gaze steady relative to when participants did not know they would be tested.

Likewise, Trueswell et al. (1999) reported eye movement behavior unrelated to linguistic processing (i.e. point fixation) in a pilot study using the type of visual world task in Experiment I. A suggested explanation for this behavior (which was not reported in other studies; cf. Tanenhaus, et al., 1995; Spivey, et al., 2002) was that participants were aware that their responses in a seemingly simple task were being compared to the responses of children, and thus they were concentrating on not making errors. This is a similar response to that described by Cooper (1974), also under conditions of high concentration. Several changes were made to the experimental paradigm by Trueswell and his colleagues in order to elicit behavior more like that reported in early studies that utilized this type of visual world (Tanenhaus, et al., 1995). One of these changes was the introduction of time pressure to the task in the form of repeated instructions to "move as quickly as possible". The results obtained in that study (Trueswell, et al., 1999) under these modified conditions were similar in most

respects to those obtained in other studies that did not employ time pressure (Tanenhaus, et al., 1995; Spivey, et al., 2002), and so it would appear that the time pressure manipulation does not affect the resulting patterns of eye movement (except to the degree that it prevented the adult participants from engaging in point fixation behavior).

However, this conclusion might be premature. It is quite possible that time pressure can increase the rate at which attentional shifts take place, as per the model of eye movement control described in Chapter 1. Attention, in this model, shifts attention as the linguistic information is input. A shift of attention triggers the planning of an eye movement, which is executed once planning completes. There would be several possible effects of an increase in the rate of shifts in attention. First, shifts in attention that are driven by very short term internal representations might be reflected to a lesser degree in the eye movement record, because these eye movements might be cancelled before motor planning has been completed. Second, the probability of anticipatory saccades might increase, as participants attempt to fill all of the requirements of a verb (e.g. theme and goal for the verb "put") as soon as possible. Lastly, fixations should be shorter in duration as participants should move their eyes to other regions of the display (in order to solve the next referential ambiguity) as soon as they have resolved the current referential ambiguity. None of these effects, with the possible exception of the increase in cancelled eye movements, would have been identified by the analyses conducted in previous studies, because the use of frame by frame graphs, relatively long time spans for analysis, and probability of fixation measures obscure the locus of shifts in attention. Thus, an important test of the

effects of task instructions on the eye movement patterns generated in the visual world paradigm still remains to be conducted.

The current experiment directly pits the instructions from Experiment I (no time pressure conditions: "try to follow the instructions as accurately as possible"; see also Tanenhaus, et al., 1995; Spivey, et al., 2002) against the instructions introduced by Trueswell, et al. (1999; time pressure conditions: "move as quickly as possible"). If the instructions given to participants do in fact affect eye movement behavior, we would expect the pattern of data in the time pressure conditions to differ from that in the no time pressure conditions in exactly those ways described above: An increase in anticipatory saccades to unmentioned but possibly relevant referents, and a faster decrease in fixations on possible referents after referential ambiguity resolution.

In addition to this examination of the effects of task instructions on the performance of the task, this experiment also allows for a full replication of the design of this type of visual world experiment. In Experiment I, only fluent ambiguous controls were used. In previous studies employing this type of visual world, fully ambiguous utterances like (24) have been compared with unambiguous versions that are syntactically disambiguated such that the object of the sentence in unambiguously modified as in (25). Participants typically view the types of display shown in Figure 14.

(24) Put the apple on the towel in the box.

(25) Put the apple that's on the towel into the box.

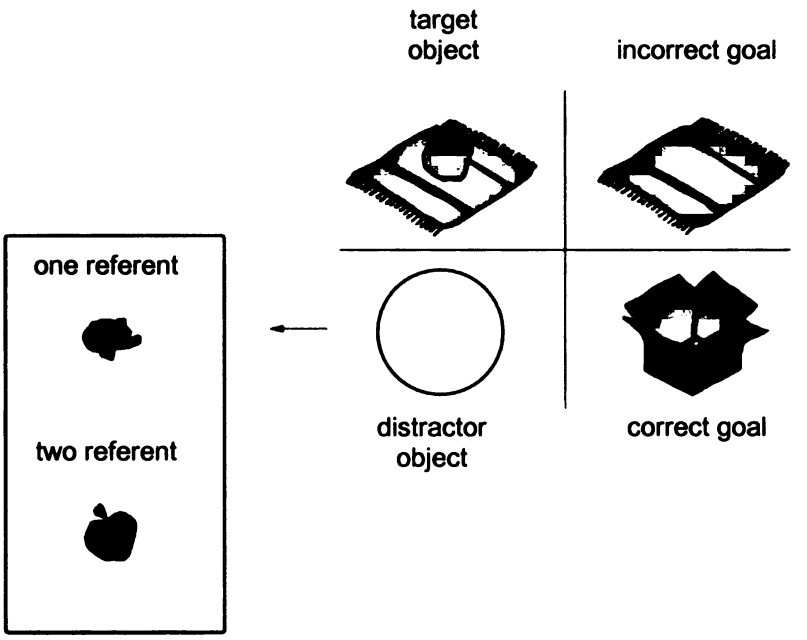


Figure 14. One referent and two referent displays used in Experiment II were similar to those described in Trueswell, et al. (1999).

The typical finding in these studies is that fixations on an incorrect goal (i.e. a towel by itself, which is not the correct goal, but might be interpreted as such after hearing "put the apple on the towel...") are more numerous in response to (24) than (25), but only when there is not another possible NP1 referent in the display – that is, effects that are interpreted as evidence for garden pathing are only present when there is only one possible referent for "the apple" in the display and modification of NP1 is unnecessary in order to complete NP1 reference resolution (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002; Chambers, Tanenhaus, & Magnuson, in press). A pattern of results that might be indicative of this garden path effect was found in Experiment I; however, the syntactically disambiguated controls were not included as a control in that experiment, and we thus cannot rule out the possibility that these effects

were simply due to the number of possible referents for the NP1 in the display. Although this effect has been replicated several times, eye movements in previous experiments (as indicated by frame by frame graphs (Trueswell, et al., 1999; Spivey, et al., 2002) were initiated more slowly and were less frequent overall than the eye movements made by participants in Experiment I. In fact, participants in Experiment I made twice as many eye movements in some conditions as would be predicted by the typical pattern of fixations described in the Tanenhaus, et al. (1995) study.

The purposes of this study, then, are twofold, and are concerned with better understanding the type of visual world experiment used in Experiment I. First, the current experiment directly compares the two types of task instruction used in previous studies in order to examine how these instructions might have affected eye movement behavior. Second, this experiment also attempts to replicate previously reported garden path effects in this type of visual world.

Material and Methods

Participants. Twenty four participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($7.00). All participants were native speakers of English, and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported in this dissertation.

Materials. Twenty four critical utterances were constructed using the nouns in Table 5. Utterances were recorded and digitized using the Computerized Speech Laboratory (Kay Elemetrics) at 10 kHz, and then converted to wav format.

Each utterance was recorded as a fluent unambiguous utterance with the same structure as (25). Following the procedure in Spivey, et al. (2002), the word *that's* was then excised, creating a corresponding ambiguous utterance (24). Each participant heard only one version (either ambiguous or unambiguous) of all 24 utterances in the course of an experiment (Table 7).

Table 7. Utterance types used in Experiment II. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance (segments for analysis indicated by subscripts)
Ambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
Unambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ that's on the towel /$_{PP2}$ in the box./

Forty eight filler utterances were also recorded and grouped with the 24 critical utterances into trials of three utterances each. A further 72 utterances were recorded to create 24 trials composed of only fillers. The types and proportions of syntactic structures used in the filler utterances were identical to those used in the Spivey, et al. (2002) study. Filler and critical trials were also interleaved, as in the Spivey, et al. (2002) study.

Displays consisted of a 2 by 2 grid, and objects were set up according to the description provided in Spivey, et al. (2002), with the exception that, depending on the height and posture of a given participant, 10-15° of visual angle separated the objects. Spivey and his colleagues do not report the angular distance between objects, but the objects appear to have been placed much farther apart in that study. In experimental trials, the possible target objects (the target and distractor objects) were always on the left, and were each placed equally in both the proximal and distal positions across trials. The possible goal objects (correct and incorrect) were always on the right, and likewise were each

placed equally in both the proximal and distal positions across trials. The locations of targets and goals for filler utterances were equally likely to occur in any of the four positions. In all, 48 displays were created, one for each set of three utterances. Of the 24 critical displays seen by any participant, 12 were two referent displays and 12 were one referent displays (Figure 14). A new random ordering of trials adhering to the interleaving requirements was created for every fourth participant in this experiment in order to maintain a balance of trials.

Apparatus. The eyetracker used in this experiment was an ISCAN model ETL-500 head mounted eyetracker (ISCAN Incorporated). This tracker functions in the same way as the tracker described in Experiment I, except that the eye and scene cameras are located on a visor, rather than on a headband. Participants were able to view 103° of visual angle horizontally, and 64° vertically. No part of the object display grid was occluded at any time by the visor. A plastic shield on the visor (which is used to block infrared radiation when tracking the eyes out of doors) was removed, as it affects color vision.

Procedure. The procedure for this experiment was identical to that used in Experiment I, with the following modifications. These changes were made to make it more like the procedure described in Spivey, et al. (2002) and Trueswell, et al. (1999). First, instead of rotating the objects into the view of a participant immediately before beginning the trial, the experimenter set up the objects in front of the participant. This gave participants an additional 20-30 seconds to view the objects prior to the onset of the first utterance in the trial. In addition, for half of the trials, participants were instructed to follow the instruction as quickly as possible (time pressure manipulation). They were reminded of this

instruction after eight trials had been completed. For the other half of the trials, participants were instructed to follow the instructions carefully, and to take as much time as they needed in order to make the correct movement (no time pressure manipulation). The task instruction manipulation was blocked, and the order of blocks was counterbalanced so that half of the participants received the time pressure instruction first, and half received it second.

Design. The two utterance types (ambiguous and unambiguous; Table 7) were combined with the two displays (one referent, and two referent; Figure 14) and two types of task instruction (time pressure and no time pressure) to create eight unique conditions for this experiment. Three trials in each condition were presented to each participant, for a total of 24 critical trials. Each display occurred in each condition an equal number of times.

Results and Discussion

The eye movement data gathered in this experiment were analyzed using the same procedure as in Experiment I (see Chapter 2, and Chapter 3, Data Analysis). On a trial by trial basis, strings representing the eye movement record for each trial were separated into segments corresponding to key phrases in the utterance. As in Experiment I, the utterance was divided into four segments (Table 7). The number of trials with a fixation (or saccade) during each segment was then calculated. These frequencies were submitted to a 2 (task instruction) by 2 (utterance type) by 2 (number of possible referents in display) multiway frequency analysis.

An analysis of fixations on the target object suggested that time pressure instructions can affect the pattern of eye movements in a visual world

experiment. During the NP1 segment, a significant effect of display was present in both the proportion of trials with fixations ($\Delta G^2(1, \underline{N} = 576) = 4.43$, $p < 0.05$; Figure 15) and with saccades ($\Delta G^2(1, \underline{N} = 576) = 6.90$, $p < 0.01$; Figure 16). This is likely due to the fact that there are two possible referents for NP1 in the two referent display and thus some initial fixations in the two referent display are on the distractor object. This is confirmed by significantly more fixations on ($\Delta G^2(1, \underline{N} = 576) = 48.29$, $p < 0.001$; Figure 17) and saccades to ($\Delta G^2(1, \underline{N} = 576) = 23.32$, $p < 0.001$; Figure 18) the distractor object region in the two referent display conditions during NP1. There was also a significant interaction between utterance type and display ($\Delta G^2(1, \underline{N} = 576) = 8.81$, $p < 0.01$) in fixations on the target object, with ambiguous utterances eliciting more fixations than unambiguous utterances in the one referent display and the reverse being true in the two referent display. It is not clear what might be causing this effect, as the ambiguous and unambiguous utterances are identical up to this point. Likewise, it is not clear what is causing the three-way interaction between task, utterance, and display in the proportion of trials with a saccade to the target region ($\Delta G^2(1, \underline{N} = 576) = 7.05$, $p < 0.01$). Similar three way interactions are present in fixations on ($\Delta G^2(1, \underline{N} = 576) = 20.89$, $p < 0.01$) and saccades to ($\Delta G^2(1, \underline{N} = 576) = 15.86$, $p < 0.01$) the distractor object.

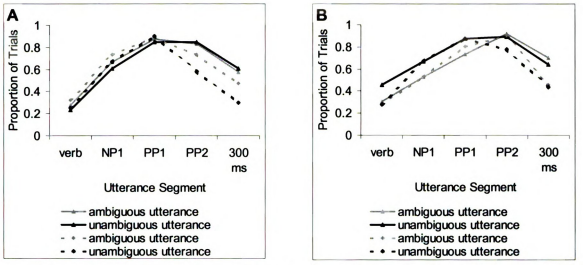


Figure 15. Proportion of trials with a fixation on the target object for each segment of the utterance in one referent (A) and two referent (B) display conditions In Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions.

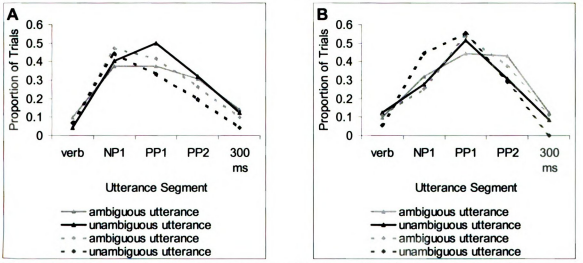


Figure 16. Proportion of trials with a saccade to the target object for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions.

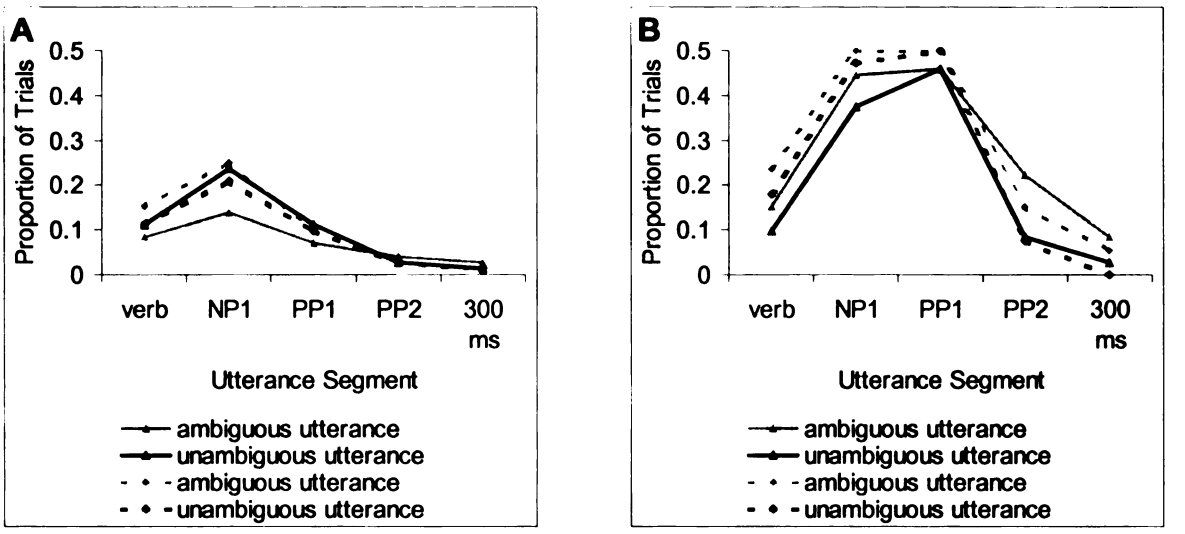


Figure 17. Proportion of trials with a fixation on the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions.

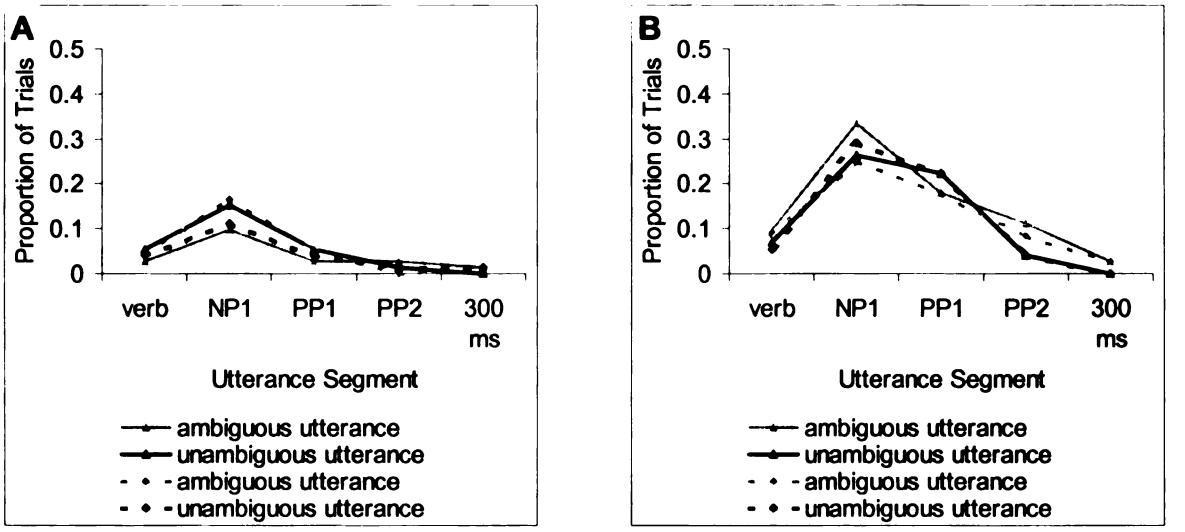


Figure 18. Proportion of trials with a saccade to the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions.

An effect of display is also present in fixations on the target object region during the PP1 segment ($\Delta G^2(1, \underline{N} = 576) = 13.19$, $p < 0.001$); however, it is in the opposite direction of the effect of display that is present in the saccade analysis ($\Delta G^2(1, \underline{N} = 576) = 7.49$, $p < 0.01$). In the former, there are more fixations in the

one referent display (Figure 15); in the latter, there are more saccades to the two referent display (Figure 16). This is a good example of the possible dissociation between fixations and saccades that has been discussed by Altmann and Kamide (2004). Because participants are fixating the target object at this point in the utterance on almost all one referent trials, there are less one referent trials than two referent trials in which it is possible to make a saccade to the target object at this point in the utterance. In the same vein, the significant interaction between task and display type ($\Delta G^2(1, \underline{N} = 576) = 5.79$, $p < 0.05$) and the marginal interaction between task and utterance type ($\Delta G^2(1, \underline{N} = 576) = 3.80$, $p < 0.1$) in the analysis of saccades to the target object suggest that participants are shifting attention during PP1 to the target object less often in conditions where early disambiguation is possible and time pressure is present. Thus, in the one referent display and the unambiguous utterance conditions, there are fewer saccades to the target object during the PP1 segment. The dissociation occurs, then, because fixations can occur only after a saccade, but can continue over many segments of analysis, while a saccade may only be launched from a single segment of analysis.

In the analysis of fixations on the target object, main effects of task ($\Delta G^2(1, \underline{N} = 576) = 13.12$, $p < 0.001$) and utterance type ($\Delta G^2(1, \underline{N} = 576) = 43.05$, $p < 0.001$) were also present during PP1. These effects were likely due to the presence of an additional possible referent in the two referent display, as the probability of fixation was greater with time pressure instructions and unambiguous utterances, both of which should have encouraged rapid NP1 reference resolution. Further effects of the presence of a distractor object that could serve as a possible NP1 referent are suggested by a marginal interaction between task

and display type ($\Delta G^2(1, \underline{N} = 576) = 2.99$, $p < 0.1$) and again by significantly more fixations on ($\Delta G^2(1, \underline{N} = 576) = 154.73$, $p < 0.001$; Figure 17) and saccades to ($\Delta G^2(1, \underline{N} = 576) = 37.02$, $p < 0.001$; Figure 18) the distractor object region during the PP1 segment. NP1 reference resolution, then, takes longer to complete in the two referent display condition because of the presence of a distractor object that is identical to the target object. Participants' eye movement patterns reflect this because participants continue to launch saccades to both the target object and distractor object as they attempt to resolve the referential ambiguity. During PP1, a syntactic disambiguation does occur on unambiguous utterance trials, and this is reflected by a marginal decrease in the proportion of trials on which a saccade is launched to the distractor object on unambiguous trials relative to ambiguous trials ($\Delta G^2(1, \underline{N} = 576) = 3.82$, $p < 0.1$; Figure 18).

A different pattern of results is present in the PP2 segment for both fixation (Figure 15) and saccade (Figure 16) analyses to the target object. For the fixation analysis, significant main effects of task ($\Delta G^2(1, \underline{N} = 576) = 15.82$, $p < 0.001$), utterance type ($\Delta G^2(1, \underline{N} = 576) = 5.76$, $p < 0.05$), and display type ($\Delta G^2(1, \underline{N} = 576) = 15.82$ $p < 0.001$) were all in the opposite direction of the effects observed during PP1. This likely represents the fact that the peak of fixations on the target object was during PP1 for those conditions where reference resolution was likely to occur earlier (i.e. time pressure instructions, unambiguous utterances, one referent displays), while the peak of fixation did not occur until PP2 in the other conditions. The same pattern is visible in the saccade analysis, with a significant main effect of display type ($\Delta G^2(1, \underline{N} = 576) =$

5.07, $p < 0.05$) and marginal effects of task ($\Delta G^2(1, \underline{N} = 576) = 2.78$, $p < 0.1$) and utterance type ($\Delta G^2(1, \underline{N} = 576) = 3.47$, $p < 0.1$). At this point, very few saccades were made to the distractor object (Figure 18). Of the fixations that remained on the distractor region, significantly more occurred in the two referent display ($\Delta G^2(1, \underline{N} = 576) = 23.24$, $p < 0.001$) where the distractor object was identical to the target object and in the ambiguous utterance type ($\Delta G^2(1, \underline{N} = 576) = 6.53$, $p < 0.05$), where no syntactic cue to disambiguation was present (Figure 17). A significant interaction between utterance type and display type was also present ($\Delta G^2(1, \underline{N} = 576) = 7.57$, $p < 0.01$) in the analysis of fixations on the distractor object.

Very few saccades were made to the target object region during the 300 milliseconds following utterance offset, likely because attention had shifted to possible goal locations (Figure 16). However, the target object was still being fixated on almost 50% of all trials (Figure 15). A significant main effect of task ($\Delta G^2(1, \underline{N} = 576) = 31.5896$, $p < 0.001$) was present, with fixations being less frequent in the time pressure condition during the 300 milliseconds following utterance offset. This is perhaps indicative of an tendency to disengage attention from a region more rapidly when under time pressure. A significant interaction of task, utterance, and ambiguity is again present ($\Delta G^2(1, \underline{N} = 576) = 6.85$, $p < 0.01$), and is likely driven by the least ambiguous condition (one referent display, unambiguous utterance) when performing under time pressure.

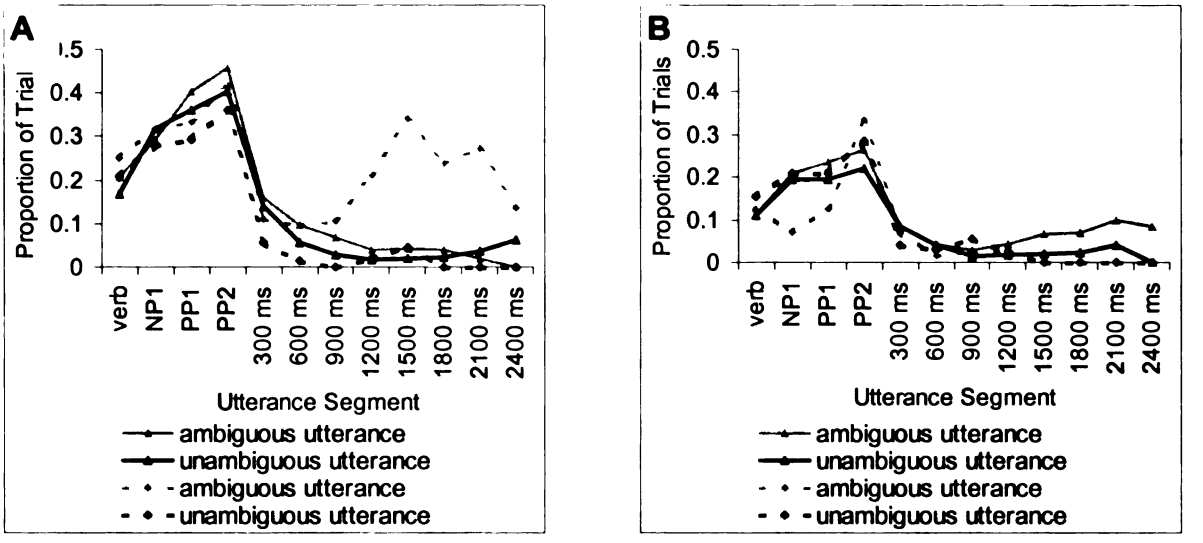


Figure 19. Proportion of trials with a fixation on the incorrect for each segment of the utterance and 300 millisecond windows after utterance offset in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions.

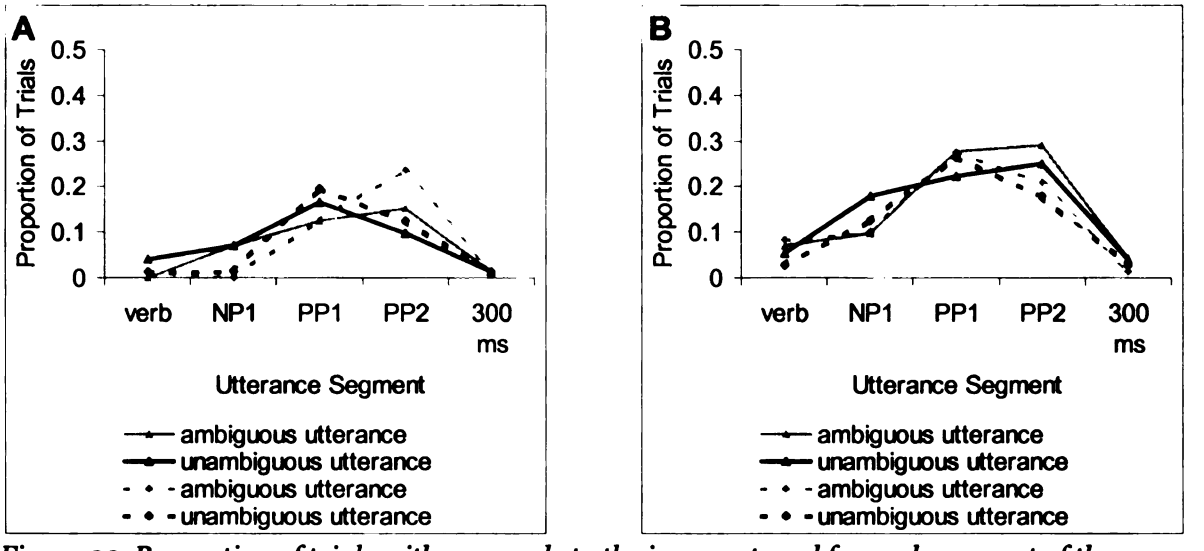


Figure 20. Proportion of trials with a saccade to the incorrect goal for each segment of the utterance in one referent (A) and two referent (B) display conditions in Experiment II. Solid lines represent no time pressure conditions, and dashed lines represent time pressure conditions.

Fixations on the incorrect goal ($\Delta G^2(1, \underline{N} = 576) = 19.57$, $p < 0.001$; Figure 19) and saccades to ($\Delta G^2(1, \underline{N} = 576) = 11.04$, $p < 0.01$; Figure 20) the incorrect goal were greater in the one referent than the two referent displays during PP1, as would be expected if a garden path was occurring in one referent displays alone. However, the interaction between utterance type and display type that has been previously reported (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002) was not found in this experiment , as the fixations on ($\Delta G^2(1, \underline{N} = 576) = 0.35$, $p > 0.1$) and saccades to ($\Delta G^2(1, \underline{N} = 576) = 2.35$, $p > 0.1$) the incorrect goal in response to ambiguous and unambiguous utterances were not significantly different across display types. A three-way interaction between task, utterance type and display type was present in the fixation analysis during PP1 ($\Delta G^2(1, \underline{N} = 576) = 5.56$, $p < 0.05$), which might be evidence that the expected pattern of results was obtained under only one set of task instructions. This does not seem to be the case, however: Neither task condition has the expected pattern of results – one where only the ambiguous utterance, one referent display condition has elevated levels of fixations relative to the other three conditions.

This pattern was repeated during the PP2 segment. Again, only a main effect of utterance type is present in both fixations on ($\Delta G^2(1, \underline{N} = 576) = 12.31$, $p < 0.001$; Figure 19) and saccades to ($\Delta G^2(1, \underline{N} = 576) = 6.50$, $p < 0.05$; Figure 20) the incorrect goal. Participants are marginally more likely to make a saccade to the incorrect goal in the ambiguous condition than in the unambiguous condition ($\Delta G^2(1, \underline{N} = 576) = 3.56$, $p < 0.1$) during PP2, as would be predicted if garden pathing was occurring; however, this effect is in the segment that follows the

location of syntactic disambiguation. If saccade launch is indicative of incremental processing, then we should have seen such effects in the PP1 segment as well, but they are clearly not present. This suggests that the marginal effect of utterance type seen in the PP2 segment may not be entirely due to syntactic disambiguation. A significant interaction is present between task and display type in saccades launched to the incorrect goal ($\Delta G^2(1, \underline{N} = 576) = 4.08$, $p < 0.05$), driven by an increase in saccades in the two referent display condition under time pressure. This increase is consistent with the evidence from fixations on the target object that show a more rapid decline under time pressure as attention is more rapidly deployed elsewhere (elsewhere, in this case, being towards the goal location). A marginal three-way interaction is again present in the fixation analysis ($\Delta G^2(1, \underline{N} = 576) = 3.33$, $p < 0.1$), but the pattern does not conform to what would be predicted if garden pathing was taking place.

One possible explanation for the lack of garden path like effects is the increase in the total number of eye movements made in this experiment (and in the previously reported Experiment I) relative to other studies that used this paradigm. The relatively small visual angles between objects in this experiment may have made participants more likely to fixate an object upon hearing a possible referent to that object, as long as the eye movement system was not busy with some other task. That is, when objects were close together, participants were more likely to use the visual world in place of (or at the same time as; Altmann, 2004) short term memory (Ballard, Hayhoe, & Pelz, 1995), and thus when memory was searched for possible referents following word recognition, the eyes were moved to those possible referents. There is no reason to believe that the

word recognition effects that have been described using related visual worlds (Allopena, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Dahan, Tanenhaus, & Chambers, 2002; McMurray, Tanenhaus & Aslin, 2002; Tanenhaus, Magnuson, Dahan, & Chambers, 2000) cease to exist simply because researchers are interested in studying syntactic phenomena. Thus, the increase in looks to the incorrect goal in the one referent display might simply be the result of word recognition processes. In this account, looks to the incorrect goal would not occur to the same degree in the two referent condition because the eye movement system is still occupied with the resolution of the earlier referential ambiguity presented by NP1 (e.g. "the apple").

Furthermore, previous experiments may have found an interaction between utterance type and display type simply because of the angular distance between objects in the visual displays. An increase in visual angle between objects in the display would have two major effects on the patterns of eye movements made in response to an utterance. First, participants would be much slower to initiate their first eye movement. The delay may have been exacerbated by an explicit instruction on every trial to fixate a center region (Spivey, e t al., 2002). As a result of this delay, participants would not be able to direct attention to the incorrect goal during the ambiguous PP1 segment, and thus any immediate effects (e.g. word recognition, anticipatory saccades, easily revised garden path structures) would not be present in the record. Slow reaction times for first fixations are, in fact, present in previously reported visual world studies

(Trueswell, et al., 1999; Spivey, et al., 2002). Thus, the garden path effects reported may be garden path effects, but they do not reflect online processing (recall the analysis time windows lasting the length of the utterance in some experiments). A second issue concerns the cost to participants of making an eye movement. When objects are far enough apart, participants are less likely to use the visual world as an external memory store and instead will prefer to search their internal memory (Ballard, Hayhoe, & Pelz, 1995) only. As a result, the amount of activation necessary to trigger an eye movement may be much larger when the objects are farther apart and the resulting series of eye movements will be a limited record of internal processing. Overall, a decrease in the total number of eye movements will occur. Such a decrease can be seen when previous studies are compared to the current experiment. In previous studies, a single couplet of fixations on the target object and the correct goal was reported in response to the unambiguous utterance conditions (e.g. Spivey, et al., 2002); the correct goal was not fixated until just before the target object was moved to the goal location. In this study, however, a common response pattern involved at least two couplets of fixations. One couplet occurred during the concurrent utterance, while the other was associated with the motor response. In addition, fixations on the incorrect goal and even the irrelevant distractor (in the one referent conditions) were interspersed. Thus, a combination of display and task factors may have yielded a spurious garden path like effect.

For this explanation of the results reported in previous studies to be true, it might be necessary to propose that looks to the incorrect goal in the one referent display are the result of a combination of anticipatory saccades, garden

pathing, and word recognition that may differ from condition to condition, depending on the difficulty of previous reference resolution. While it is not possible to determine whether participants are garden pathed in this particular visual world experiment (although the perseveration in looks to the incorrect goal long after utterance offset in the one referent display, ambiguous utterance condition under time pressure suggests that this might be the case; Figure 20A), there is some evidence for the presence of anticipatory saccades. Participants are significantly more likely to launch a saccade to the correct goal in unambiguous utterance conditions ($\Delta G^2(1, \underline{N} = 576) = 10.01$, $p < 0.01$; Figure 21) than ambiguous utterance conditions during the PP1 segment. Note that this segment is prior to the segment that directly references the correct goal (PP2). What does occur during this segment (or prior to it in the one referent display conditions) is the conclusive resolution of the referential ambiguity introduced at the NP. Thus, the eye movement system becomes available at some point during PP1 and then can be directed to the next available task, which is dictated by the requirement of a goal argument for the verb "put". The result is the sort of verb-driven anticipatory saccades reported elsewhere in the literature (e.g. Altmann & Kamide, 1999). Consistent with this suggestion, utterance type interacts with both task ($\Delta G^2(1, \underline{N} = 576) = 9.62$, $p < 0.01$) and display ($\Delta G^2(1, \underline{N} = 576) = 5.33$, $p < 0.1$). The former interaction reflects the tendency for time pressure to encourage the language comprehension system to satisfy all of the required arguments for the verb as soon as possible. The interaction with display type is indicative of the fact that disambiguation likely occurs later in the two referent display, ambiguous utterance conditions than in any other condition, as all other

conditions are disambiguated before the first word in PP1 by either the presence of only a single possible referent or a syntactic cue (i.e. "that's").

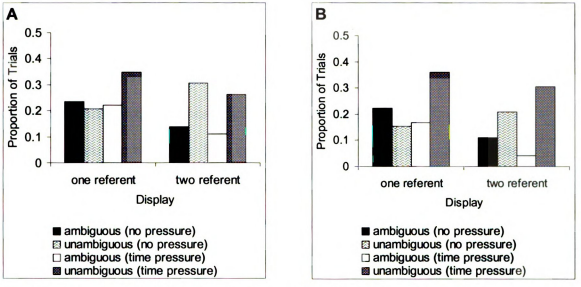


Figure 21. Proportion of trials with a fixation in (A) and with a saccade to (B) the correct goal during the PP1 segment of the utterance in Experiment II. "Ambiguous" and "unambiguous" refer to utterance types.

In summary, then, the results of Experiment II suggest that task instructions can have significant, though subtle, effects on eye movement patterns in the visual world paradigm. While the general patterns of results are relatively stable when time pressure is either present or absent, the tendency for participants to continue fixating a region that has already been matched to a referent in the concurrent utterance is affected by time pressure, as is the tendency to make anticipatory saccades. While patterns of eye movements to target and distractor objects matched those from previously reported experiments, the current experiment failed to fully replicate the previously reported garden path like effects in looks to the incorrect goal. Instead, the results from this experiment are consistent with an explanation that implicates

word recognition, anticipatory saccades, and the rate of reference resolution, in looks to objects referenced later in the utterance.

While the results of this experiment were analyzed in great detail, it is clear that for this particular type of visual world, and with these prepositional phrase ambiguities, the PP1 segment of the utterance is the locus of the majority of effects of interest, and those that illuminate the process of language comprehension in this task. This is not surprising, as the ambiguity in these utterances concerns whether PP1 should be interpreted as a modifier of NP1 or an argument of the verb. As the visual world paradigm was designed for the purpose of examining online language comprehension, it is encouraging to see that the PP1 segment, which is the site of syntactic ambiguity, is also the site of eye movement patterns that reflect the processing of that ambiguity, even if the syntactic effects observed (in the anticipatory saccades to the correct goal) do not show signs of a display-induced garden path. As a result, further studies using this particular syntactic ambiguity and visual world will mainly focus on eye movements during the PP1 segment.

This experiment is also useful in highlighting several issues concerning the design and reporting of visual world experiments. First, the results reported here may indicate that more care needs to be taken in designing experiments that are geared toward identifying syntactic effects, as these seem to be overwhelmed by a variety of other effects, not the least of which is the presence of an early referential ambiguity. It is also possible that not only the objects that are in the display, but the distances between those objects can greatly affect the pattern of eye movements that is elicited. It has not been common practice to date for

psycholinguists to carefully describe or construct the displays used in visual world studies, a state of affairs that is somewhat surprising given the care taken in designing reading studies (Rayner, 1998) that are not too distant relatives of visual world experiments. For example, it should become common practice to report relevant visual angles. The measures used are also important, as it appears that the probability of a saccade is a more sensitive online measure than is the probability of a fixation, at least in some circumstances. At the same time, the probability of fixation can index the point at which attention is shifted away from an object, and thus is useful as well, although under different circumstances.

THE VISUAL WORLD PARADIGM: DISPLAY AMBIGUITY

The results of Experiment II suggest that the difficulty of reference resolution imposed by ambiguity in the display can greatly affect the patterns of eye movements that are elicited in the visual world. Specifically, it appears that eye movements that are related to reference resolution late in an utterance are delayed or absent when ambiguities in the display make earlier reference resolution difficult. On the other hand, when early reference resolution is not difficult, anticipatory saccades may be seen. The three experiments in this chapter manipulate various components of the visual display in order to increase or decrease display ambiguity. Experiment IIIA is concerned with the effects of greater or lesser ambiguity concerning the NP1 referent on eye movements in response to both that referent and later referents. Experiment IIIB examines the pattern of eye movements in a fully ambiguous display in order to attempt to determine whether multiple final interpretations are possible in this paradigm. Finally, Experiment IIIC addresses the issue of whether garden path like effects are due to anticipatory saccades or attempted reference resolution by examining eye movement patterns made while interacting with unambiguous displays.

Experiment IIIA

In Experiments I and II, participants were able to identify the referent of NP1 more rapidly in the one referent display because only one object corresponding to that NP was present in the display. As a result, in the one referent display conditions, participants were more likely to fixate the incorrect goal and to make anticipatory saccades to the correct goal. The same processes may have occurred in the two referent condition, but these effects may have been

masked by the presence of a second possible referent for NP1 (e.g. "the apple" in "put the apple on the towel in the box"). That is, the eye movement system may have been occupied with NP1 reference resolution and thus is not available to make eye movements to other objects during the utterance.

It is clear, then, that introducing an ambiguity into a display (i.e. adding an additional possible referent) can affect the pattern of eye movements produced during the comprehension of an utterance. A related prediction is that the degree of that ambiguity might also have an effect on eye movement patterns. This is a key issue, because previous studies have varied in the point at which the NP1 referent can be disambiguated with respect to the two referent display. In some studies (as in Experiments I and II above) the distractor objects in both the one referent and two referent displays were alone in their quadrant of the display. Thus, when trying to determine which apple was the correct referent of NP1 in the two referent display, participants could rely on the word "on" in an utterance like "put the apple on the towel in the box" to differentiate between the two apples present in the display as only one apple is on another object. On the other hand, other studies (e.g. Spivey et al., 2002) have had both the target object and the distractor object on or in another object in the two referent display; thus, the correct referent would not be distinguished until the noun in PP1 (e.g. "towel") was encountered. Thus, the early disambiguation display (where the distractor object is by itself) of the type used in Experiments I and II could allow participants to more rapidly identify the NP1 referent than would be possible while viewing a late disambiguation display (where the distractor is on or in another object).

If the NP1 referent was disambiguated before hearing "towel", the eye movement system may have been freed from the task of determining what object to pick up and the processing of PP1 may have resulted in eye movements to the incorrect goal for reasons other than garden pathing. That is, a true garden path effect may have been swamped by additional eye movements that were only possible because the display allowed for early disambiguation. If disambiguation was delayed, anticipatory saccades and fixations due to word recognition may have been less likely to occur. This experiment tests the role of early disambiguation in affecting eye movement patterns by comparing eye movement patterns from displays where disambiguation can occur earlier to displays where disambiguation occurs later. The general patterns of pattern of eye movement data described in Experiment II for the time pressure task should be replicated in this experiment, with decreases in anticipatory or early looks to the incorrect goal and correct goal in the late disambiguation conditions.

Material and Methods

Participants. Sixteen participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($7.00). All participants were native speakers of English, and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported in this dissertation.

Materials. The materials created for Experiment II were used in this experiment. However, the 24 displays created for critical trials were further modified to create versions where the distractor object was located on or in

another object (Figure 22). Twelve such displays (six one referent and six two referent) were substituted for the displays used in Experiment II in each trial list created. Each display occurred in the early and late disambiguation forms an equal number of times in each experiment. Random trial lists were created according to the procedure described in Experiment II. Utterance types used were the same as in Experiment II (Table 8).

Table 8. Utterance types used in Experiment IIIA. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance (segments for analysis indicated by subscripts)
Ambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
Unambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ that's on the towel /$_{PP2}$ in the box./

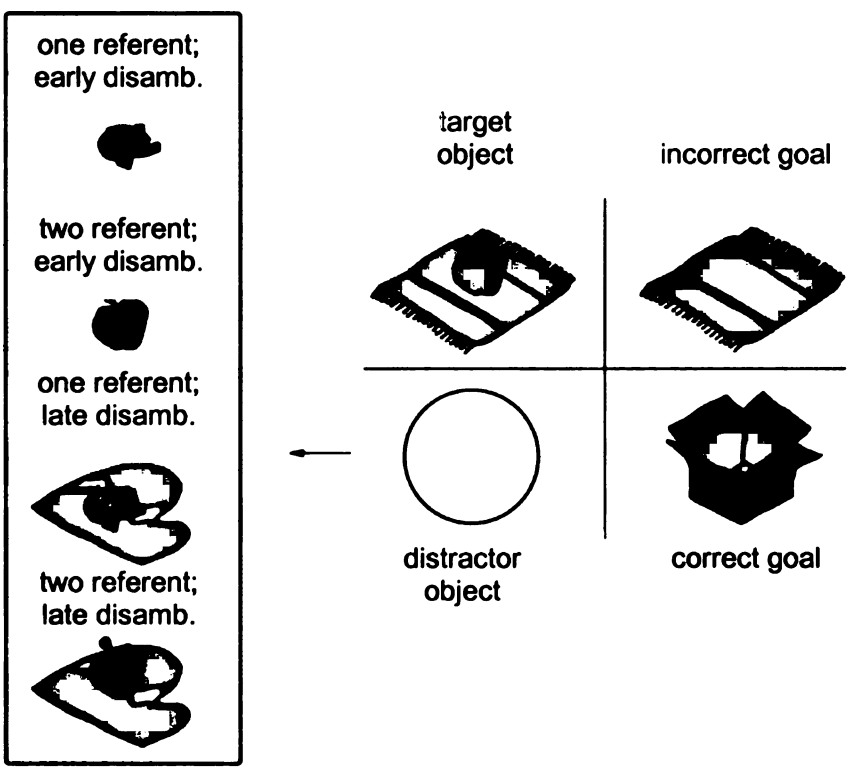


Figure 22. Early and late disambiguation versions of the one referent and two referent displays for Experiment IIIA.

Apparatus. The apparatus for this experiment was identical to that used in Experiment II.

Procedure. The procedure for this experiment was identical to that of the time pressure task condition in Experiment II, except that participants were not reminded to move as quickly as possible after the experiment had begun.

Design. Four different display conditions were created by crossing the different number of possible referents manipulation (one referent, and two referent) with the point of disambiguation manipulation (early or late disambiguation; Figure 22). The two utterance types (ambiguous and unambiguous; Table 8) were then combined with the four displays to create eight unique conditions for this experiment. Three trials in each condition were presented to each participant, for a total of 24 critical trials. Each display occurred in each condition an equal number of times.

Results and Discussion

As in previous experiments, the proportion of trials with a fixation on or saccade to a region of interest was calculated for each segment in the utterance (Table 8) on a trial by trial basis. These frequencies were then submitted to a 2 (point of disambiguation) by 2 (utterance type) by 2 (number of possible referents in display) multiway frequency analysis.

If the point of disambiguation manipulation in this experiment is effective, we should expect to see two different types of effects. First, we should see an increased and longer lasting tendency for participants to make eye movements to and to continue to fixate the distractor object in the late disambiguation conditions. This effect should be stronger in the two referent conditions where

both the target and the distractor object are possible referents of NP1. Second, we should expect to see more looks to the incorrect goal and more anticipatory saccades to the correct goal during the PP1 segment in the early disambiguation conditions. Again, we should expect to see differences depending on the number of possible NP1 referents in the display, this time with the one referent conditions eliciting more early looks to possible goals.

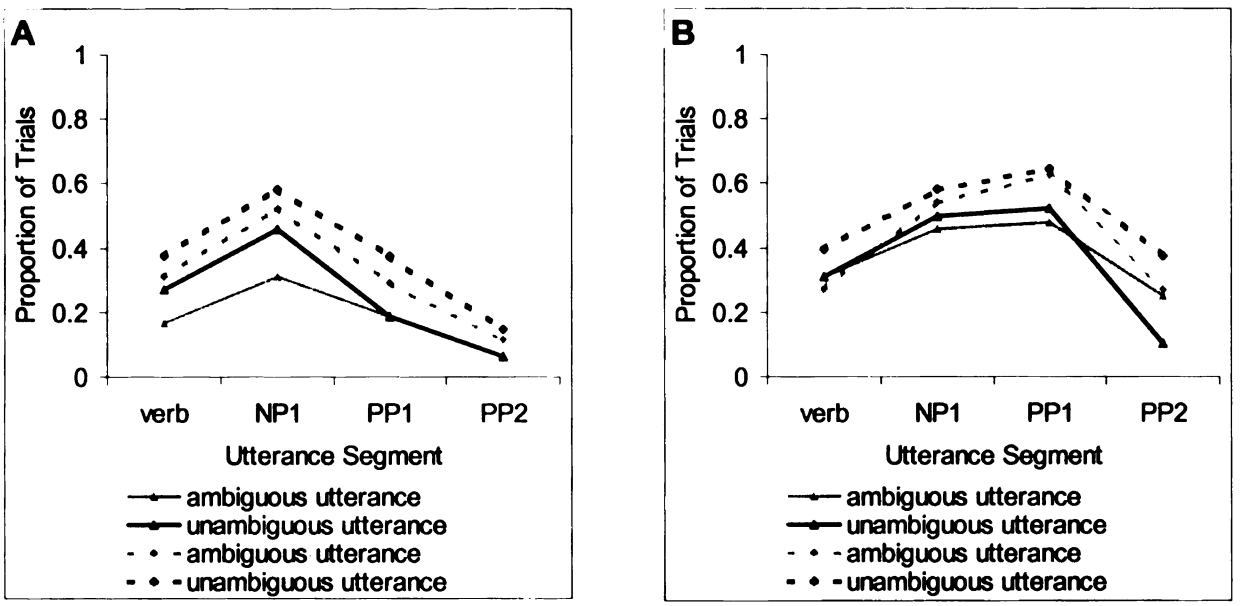


Figure 23. Proportion of trials with a fixation on the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions In Experiment IIIA. Solid lines represent early disambiguation conditions, and dashed lines represent late disambiguation conditions.

The pattern of looks to the distractor objects throughout the utterance indicates that the visual world paradigm is sensitive to the point in the utterance at which disambiguation occurs. During the NP1 segment itself, participants were significantly more likely to fixate on ($\Delta G^2(1, \underline{N} = 384) = 7.31$, $p < 0.01$; Figure 23) and marginally more likely to make a saccade to ($\Delta G^2(1, \underline{N} = 384) = 3.44$, $p < 0.1$; Figure 24) the distractor object in the late disambiguation conditions. There were no effects of the number of possible referents in the display during this segment, nor were there any interactions with the point of disambiguation (all $p > 0.1$).

Participants were, however, significantly more likely to fixate on ($\Delta G^2(1, \underline{N} = 384) = 48.97$, $p < 0.001$) and to make a saccade to ($\Delta G^2(1, \underline{N} = 384) = 26.47$, $p < 0.001$) the distractor object in the two referent displays during PP1. At the same time, participants continued to be more likely to both fixate on ($\Delta G^2(1, \underline{N} = 384) = 11.22$, $p < 0.001$; Figure 23) and saccade to ($\Delta G^2(1, \underline{N} = 384) = 8.42$, $p < 0.01$; Figure 24) the distractor object in the late disambiguation conditions. There were no significant interactions between the number of referents in the display and the point of disambiguation.

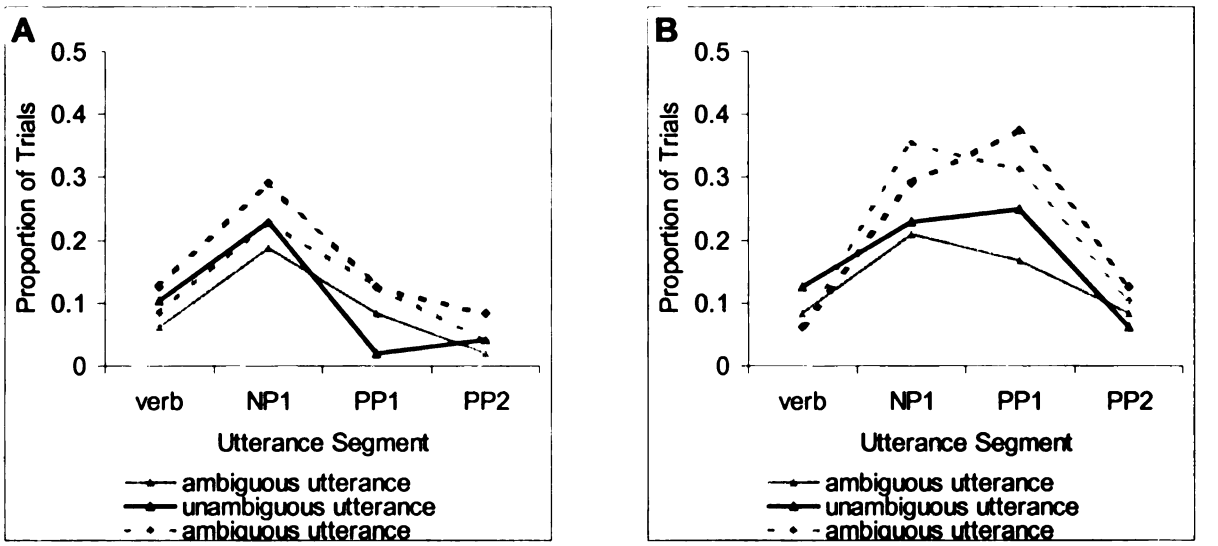


Figure 24. Proportion of trials with a saccade to the distractor object for each segment of the utterance in one referent (A) and two referent (B) display conditions In Experiment IIIA. Solid lines represent early disambiguation conditions, and dashed lines represent late disambiguation conditions

By the PP2 segment, very few saccades were made to the distractor object (Figure 24), and no significant main effects or interactions were present (all $p > 0.1$). However, participants continued to fixate the distractor object significantly more often in both the two referent ($\Delta G^2(1, \underline{N} = 384) = 16.97$, $p < 0.001$; Figure 23) and late disambiguation ($\Delta G^2(1, \underline{N} = 384) = 9.00$, $p < 0.01$) conditions. In addition, significant interactions between the point of disambiguation and

utterance type ($\Delta G^2(1, \underline{N} = 384) = 6.55$, $p < 0.05$) and between the point of disambiguation, utterance type, and the number of referents in the display ($\Delta G^2(1, \underline{N} = 384) = 5.49$, $p < 0.05$) were present. Both of these interactions were likely due to the fact that participants fixated the distractor objects less often in the early disambiguation, unambiguous utterance condition in the one referent display and especially in the two referent display. This is exactly the pattern of results that would be predicted if the ease of reference resolution was due to both the syntactic structure of the utterance and the ambiguity of the display.

It is clear, then, that participants were able to disengage from fixations on the distractor object earlier in the trial when utterance and display factors allowed for earlier NP1 reference resolution. Recall that we are assuming a model of eye movement control in the visual world that is consistent with evidence for incremental processing. According to this model, we might expect to see anticipatory eye movements launched prior to particular referents being encountered, as long as there is some basis in the utterance heard up to that point (such as verb requirements) for such an eye movement. Moreover, this type of model would posit that eye movements are not simply reactions to words in an utterance, but part of an interactive process of language comprehension that involves both processing of linguistic material and active internal and external search of the current context.

There is, in fact, evidence from looks to both the incorrect goal and the correct goal in this experiment that supports a model of anticipatory and early looks to possible goal objects in conditions where early NP1 reference resolution has occurred. Participants are more likely to saccade to ($\Delta G^2(1, \underline{N} = 384) = 15.39$,

$p < 0.001$; Figure 25B) and fixate on ($\Delta G^2(1, \underline{N} = 384) = 19.02$, $p < 0.01$; Figure 25A) the incorrect goal during PP1 in one referent displays, regardless of utterance type or point of disambiguation, as was found in Experiment II. This is not surprising, as the one referent displays allow disambiguation of NP1 reference during NP1 itself, while two referent displays require at least some part of PP1 to identify the referent of NP1. There is some influence of the point of disambiguation on the probability of launching a saccade to the incorrect goal, however. A significant three way interaction between the point of disambiguation, the number of referents in the display, and utterance type ($\Delta G^2(1, \underline{N} = 384) = 5.49$, $p < 0.05$) is present, and appears to be driven by a much higher probability of saccade launch in the one referent, early disambiguation, unambiguous utterance condition. This is the condition where NP1 reference resolution should be easiest, as the most cues to the correct referent are present.

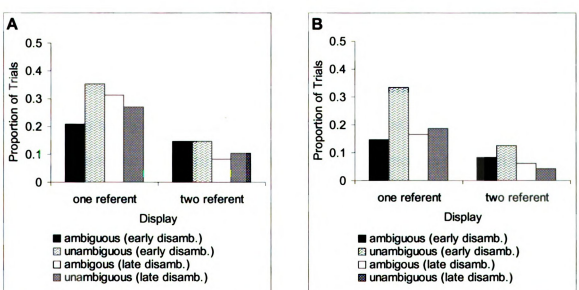


Figure 25. Proportion of trials with a fixation in (A) and with a saccade to (B) the incorrect goal during the PP1 segment of the utterance in Experiment IIIA. "Ambiguous" and "unambiguous" refer to utterance types.

A similar pattern of results is present in the fixations on the incorrect goal during PP1 and is supported by a marginal interaction between the point of disambiguation ($\Delta G^2(1, \underline{N} = 384) = 3.14, p < 0.1$) and a significant interaction between the point of disambiguation, the number of referents in the display, and utterance type ($\Delta G^2(1, \underline{N} = 384) = 5.41, p < 0.05$). These interactions suggests that the language comprehension system is at the same time conservative and liberal with respect to shifts of attention following reference resolution. Language comprehension processes are conservative in the sense that attention is not shifted away from an object immediately when the minimum amount of information necessary to identify a referent is present; if this was true, we should see saccades launched to possible goals (including the incorrect goal) equally often in each one referent display condition. On the other hand, language comprehension processes are liberal in that attention can shifted toward an object before the noun that refers to that object is heard. That is, shifts of attention still anticipate the minimum amount of information necessary to identify a referent. These saccades can be considered anticipatory in some sense, whether they are launched due to the syntactic or thematic requirements of verbs, or due to the cues such as the preposition that precedes the noun in PP1.

The presence of anticipatory saccades driven by verb constraints (rather than saccades driven by lexical processing of either the preposition or the noun in PP1) is supported by an examination of looks to the correct goal during PP1. These looks cannot be driven by lexical processing of either the preposition nor the noun in PP1 as the correct goal was always an object that was referred to by a different preposition that the preposition in PP1. Participants were significantly

more likely to fixate ($\Delta G^2(1, \underline{N} = 384) = 7.02$, $p < 0.01$; Figure 26A) and marginally more likely to saccade to ($\Delta G^2(1, \underline{N} = 384) = 2.91$, $p < 0.1$; Figure 26B) the correct goal during PP1 in the one referent display.

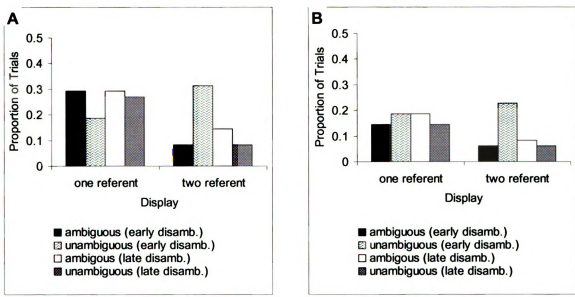


Figure 26. Proportion of trials with a fixation in (A) and with a saccade to (B) the correct goal during the PP1 segment of the utterance in Experiment IIIA. "Ambiguous" and "unambiguous" refer to utterance types.

Of interest is the relatively high probability of fixations on and saccades to the correct goal in the two referent, early disambiguation, unambiguous utterance condition (relative to the other two referent conditions). This pattern might be predicted because it would be easier to complete NP1 reference resolution in this condition than in any other two referent display condition. The difference between this condition and other two referent conditions is supported by marginal interactions between utterance type and the number of referents in the display in analyses of fixations on ($\Delta G^2(1, \underline{N} = 384) = 3.44$, $p < 0.1$) and saccades to ($\Delta G^2(1, \underline{N} = 384) = 2.73$, $p < 0.1$) the correct goal during PP1. A significant interaction between the point of disambiguation, the number of

referents in the display, and utterance type ($\Delta G^2(1, \underline{N} = 384) = 9.59, p < 0.01$) was also present in the fixation analysis.

In summary, Experiment IIIA indicates that the visual world paradigm is sensitive to the point at which a referent can be disambiguated. In establishing reference, the language comprehension system uses cues from the syntactic structure of the utterance and the visual display. This experiment again failed to find evidence of garden pathing; rather, the degree to which participants looked to the incorrect goal (and the correct goal) was related how quickly NP1 reference resolution could be completed. Shifts of attention did not occur immediately after the minimum amount of information was available, however, as the language comprehension appears to wait for confirmation from multiple sources of information before committing to a shift of attention.

Experiment IIIB

As has been noted in previous chapters, looks to the incorrect goal have been interpreted in the literature as evidence that participants were garden pathed by ambiguous utterances (e.g. "put the apple on the towel in the box") in the one referent display, but not in the two referent display. A garden path is said to occur because there is an ambiguous PP ("on the towel") and a goal location that is a possible referent of the noun in that PP. This garden path is avoided in the two referent display because the ambiguous PP must be used by the language comprehension system to resolve the referent of the first NP ("the apple"; Tanenhaus, et al., 1995; Spivey, et al., 2002).

However, it is not clear from the experiments presented earlier in this dissertation that a garden path is occurring. While participants are more likely to look to the incorrect goal during the ambiguous PP (PP1), there is no significant difference between ambiguous and unambiguous ("put the apple that's on the towel in the box") utterances. Moreover, a similar pattern occurs during PP1 in looks to the correct goal, despite the fact that the correct goal is not a possible referent of any portion of the utterance at that point. There is some evidence of a difference between the ambiguous and unambiguous utterances in looks to the correct goal, but it is in a direction that corresponds to the earlier release of the eye movement system from NP1 reference resolution in the unambiguous utterance conditions; looks to the correct goal during PP1 are more likely with unambiguous, rather than the ambiguous concurrent utterances. Thus, it is possible that garden path like effects are only appearing in the one referent display because the visual system is freed from other tasks (e.g. NP1 reference

resolution) and can be directed to the incorrect goal in anticipation of identifying the goal location.

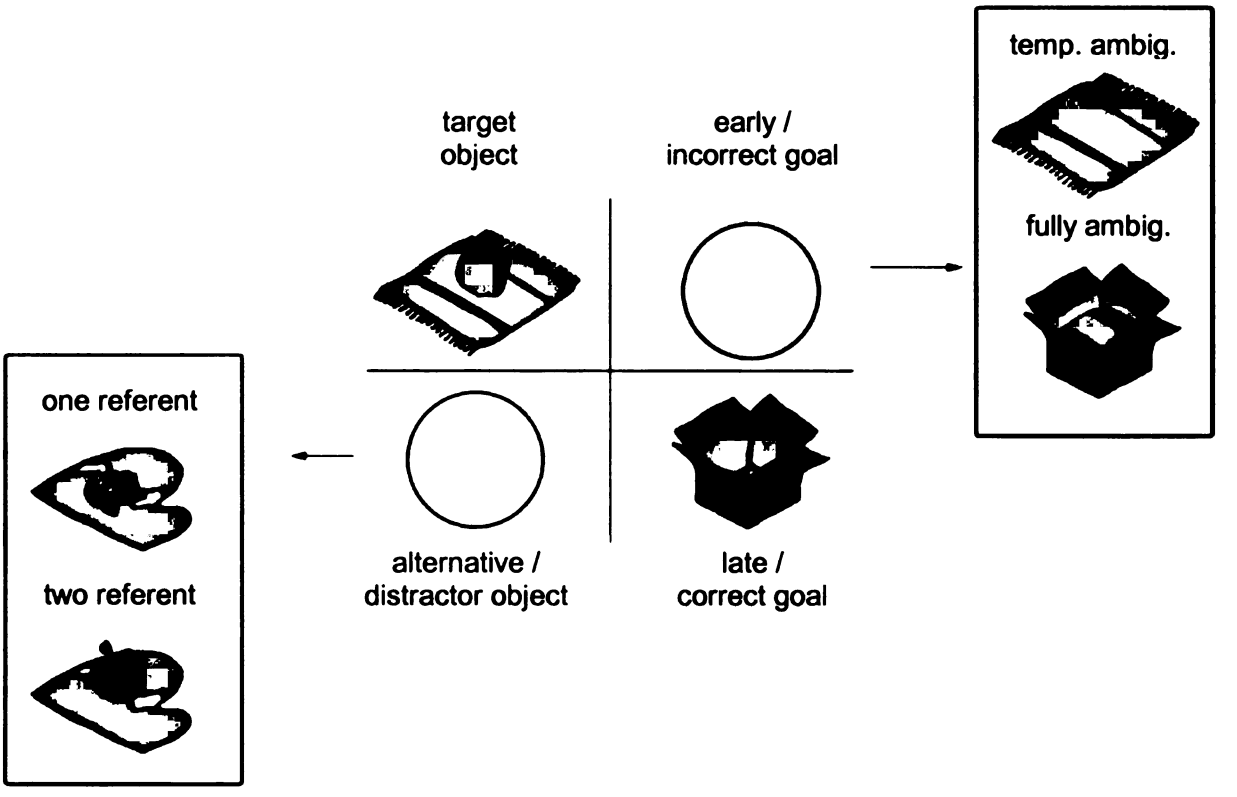


Figure 27. Displays created for Experiment IIIB. All four possible displays were seen an equal number of times by participants.

What cannot be determined from the experiments conducted so far is whether the looks to the incorrect goal are due solely to anticipatory eye movements, garden pathing, or some combination of the two. In order to determine whether a garden path is occurring, a fully ambiguous display (Figure 27) that allows "put the apple on the towel in the box" to be parsed as either (26) or (27) could be compared with the standard display. If a garden path is occurring, not only should similar patterns of eye movements be found, but the behavioral response of the participants would indicate their final interpretation of the utterance. Because there would be no cue to reanalyze the structure built

by the parser while viewing the fully ambiguous display, if a garden path occurs, participants should place the object at the early/incorrect goal on some significant proportion of the trials.

(26) Put the apple on the towel that's in the box.

(27) Put the apple that's on the towel into the box.

Material and Methods

Participants. Sixteen participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($7.00). All participants were native speakers of English, and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported in this dissertation.

Materials. The late disambiguation materials created for Experiment III were used in this experiment. However, the 24 critical displays were modified so that the incorrect goal was now a possible goal location (referred to here as the early/incorrect goal as it is not an incorrect goal in a fully ambiguous display) composed of the nouns in both in NPs (Figure 27). Twelve such displays (six one referent and six two referent) were substituted for the late disambiguation (referred to henceforth as temporarily ambiguous) displays in each trial list created. Each display occurred in fully and temporarily ambiguous forms an equal number of times. Random trial lists were created according to the procedure described in Experiment II. An additional object was added to half of the filler displays in order to make them more similar in appearance to the fully ambiguous displays.

Apparatus. The apparatus for this experiment was identical to that of Experiment II.

Procedure. The procedure for this experiment was identical to that of Experiment IIIA. However, because of the full ambiguity involved in the initial utterance in the fully ambiguous display conditions, it was possible on a small number of trials for participants to make a movement that would also complete one of the filler instructions prior to that filler instruction being heard. Thus, participants were additionally told that if they heard an instruction that was impossible to complete because it had already been completed by an action on a previous trial, they did not need to do anything. This instruction only affected filler utterances, as all critical utterances were the initial instruction in the trial. This situation occurred on less than three filler trials for any one subject.

Design. Four different display conditions were created by crossing the different number of possible referents manipulation (one referent, and two referent) with the display ambiguity manipulation (fully or temporarily ambiguous; Figure 27). The two utterance types (ambiguous and unambiguous; Table 9) were then combined with the four displays to create eight unique conditions for this experiment. Three trials in each condition were presented to each participant, for a total of 24 critical trials. Each display occurred in each condition an equal number of times.

Table 9. Utterance types used in Experiment IIIB. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance
Ambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
Unambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ that's on the towel /$_{PP2}$ in the box./

Results and Discussion

The number of trials with fixations on and saccades to regions of interest were calculated for each of the segments in Table 9. The frequencies for each segment and region of interest were then submitted to a 2 (display ambiguity) by 2 (number of possible referents in the display) by 2 (utterance type) multiway frequency analysis. The box on the right in Figure 27 indicates the display ambiguity conditions, the box on the left the number of referents conditions, and Table 9 indicates the utterance type conditions.

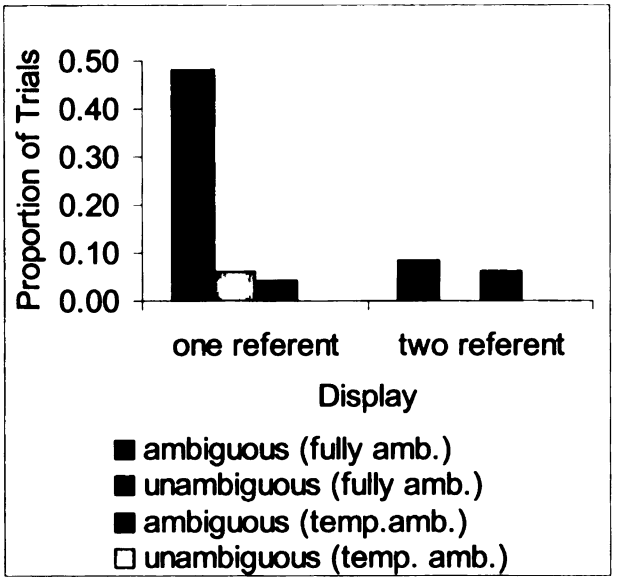


Figure 28. Percent trials in which a target object is moved to the early/incorrect goal. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types.

In only condition in which subjects should move an object to the early/incorrect goal location was the one referent, fully ambiguous display, ambiguous instruction condition. An examination of the other trials indicates why this pattern of behavior ought to occur. All trials with a concurrent unambiguous instruction necessitate a movement to the late/correct goal. In the temporarily ambiguous display conditions, the early/incorrect goal is not a

possible goal location because a full and correct parse of the utterance only allows PP1 to be a modifier of NP1. In any two referent display condition, PP1 must be used as a modifier in order to disambiguate the NP1 referent. It is only when PP1 is not constrained as a modifier by the display or the utterance type that either syntactic structure corresponding to the meanings in (26) or (27) is licensed. The pattern of behavior obtained in this experiment matches these predictions. Only the one referent, fully ambiguous display, ambiguous utterance condition resulted in a large percentage of trials (48%; Figure 28) where participants moved a target object to the early/incorrect goal. This result indicates that PP1 is interpreted as a goal location on some percentage of ambiguous utterance trials, and thus the looks to the early/incorrect goal may reflect garden pathing to some degree. Alternatively, looks to the early/incorrect goal may be anticipatory and precede any syntactic garden path, but the likelihood of continuing to fixate the region may be determined by whether or not a garden path occurs. Because any garden path will be brief in temporarily ambiguous displays, there may, in fact, be no evidence of garden pathing in the eye movement record, as attention will be shifted as soon as the disambiguating PP2 is encountered.

The fully ambiguous display, then, allows participants to continue to retain an interpretation where the object denoted by the noun in PP1 is the eventual goal location throughout the utterance up to the final behavioral response. The temporarily ambiguous display, on the other hand, requires that this interpretation be abandoned when participants encounter PP2. Thus, at PP1, both display types can support the interpretation that the early/incorrect goal is the eventual goal, but by PP2, only the fully ambiguous display can support this

interpretation. Thus, we should not expect to see effects of display ambiguity in looks to the early/incorrect goal at PP1. At the same time, it is possible that the introduction of a new object into the display to create the fully ambiguous display will affect both the low level characteristics and internal memory representation of the display, in which case effects of display ambiguity might be widespread. In Experiment IIIA, we have already seen that even minor changes in the components of a display can effect eye movement patterns.

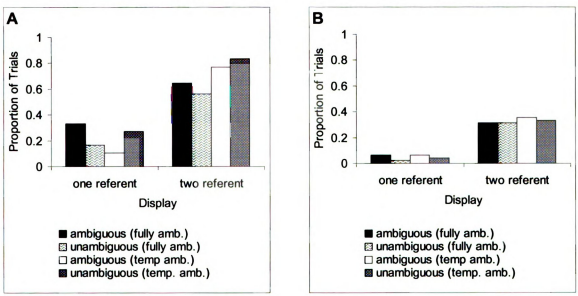


Figure 29. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor object during the PP1 segment of the utterance in Experiment IIIB. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types.

An analysis of looks to the alternative/distractor object during the PP1 segment suggests that the latter hypothesis may have some merit. Saccades to the alternative/distractor object showed only a main effect of the number of referents in the display ($\Delta G^2(1, \underline{N} = 384) = 61.77$, $p < 0.001$; Figure 29B), with more saccades occurring in two referent displays. This is same pattern of results that has been noted in previous experiments, both in this dissertation and in other

published studies (Trueswell, et al., 1999; Spivey, et al., 2002). However, the analysis of fixations on the same region, in addition to indicating a main effect of number of referents ($\Delta G^2(1, \underline{N} = 384) = 101.68$, $p < 0.001$; Figure 29A), also shows significant interactions between display ambiguity and utterance type ($\Delta G^2(1, \underline{N} = 384) = 11.49$, $p < 0.01$) and between display ambiguity and the number of referents in the display ($\Delta G^2(1, \underline{N} = 384) = 14.54$, $p < 0.01$). Because the same interactions are not significant in the analysis of saccades (all $p > 0.1$), it would appear that many of these fixations were initiated prior to the onset of the PP1 segment. It is not clear why the display ambiguity should interact with other variables in this way, but this pattern of results may again indicate that there are unforeseen consequences of altering displays.

With regards to the predictions about looks to the early/incorrect goal, the same increased likelihood in all one referent conditions of fixations on ($\Delta G^2(1, \underline{N} = 384) = 44.75$, $p < 0.001$; Figure 30A) or saccades to ($\Delta G^2(1, \underline{N} = 384) = 62.96$, $p < 0.001$; Figure 30B) this region was seen in this experiment. However, participants were also significantly more likely to fixate on ($\Delta G^2(1, \underline{N} = 384) = 16.25$, $p < 0.001$) or saccade to ($\Delta G^2(1, \underline{N} = 384) = 4.14$, $p < 0.05$) the early/incorrect goal in the fully ambiguous display. This tendency was stronger in the two referent condition, as evidenced by a significant interaction between display ambiguity and the number of possible referents in the display in the fixation analysis ($\Delta G^2(1, \underline{N} = 384) = 4.21$, $p < 0.05$) during the PP1 segment. The same interaction was marginally significant in the saccade analysis for the PP1 segment ($\Delta G^2(1, \underline{N} = 384) = 3.64$, $p < 0.1$). This pattern of results suggests that

either the fully ambiguous display caused more initial garden paths, or that the two object (e.g. towel in box) location created for the fully ambiguous display was more likely to draw anticipatory saccades. In either case, the fully ambiguous display has immediate effects on the likelihood of looks to the early/incorrect goal. An examination of the probability of fixation following the offset of the utterance (Figure 31) indicates that subjects perseverate in their fixations on the early/incorrect goal in ambiguous utterance conditions. In the one referent, fully ambiguous display, ambiguous utterance condition, perseveration is likely because the object was moved to this region on about half the trials. However, similar perseveration is visible in the two referent, fully ambiguous display, ambiguous utterance condition, and the one referent, temporarily ambiguous display, ambiguous utterance condition, although it occurs to a lesser degree.

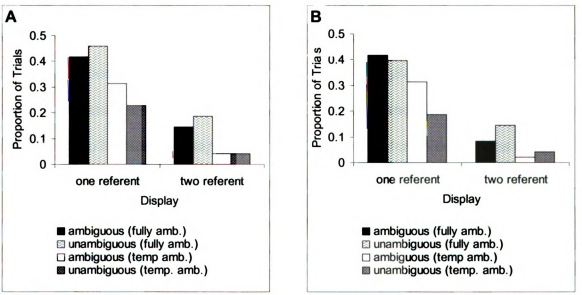


Figure 30. Proportion of trials with a fixation in (A) and with a saccade to (B) the early incorrect goal during the PP1 segment of the utterance in Experiment IIIB. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types.

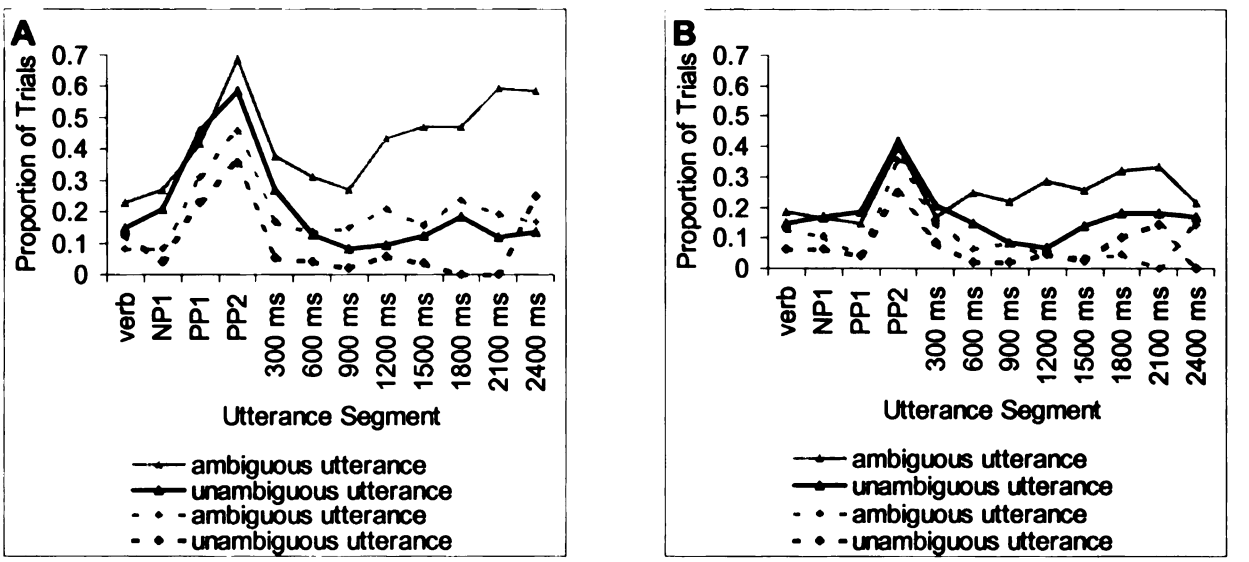


Figure 31. Proportion of trials with a fixation on the early/incorrect goal during and after the utterance in one referent (A) and two referent (B) display conditions In Experiment IIIB. Solid lines represent fully ambiguous display conditions, and dashed lines represent temporarily ambiguous display conditions.

Analyses of anticipatory saccades to the late/correct goal during PP1 show no significant effect of display ambiguity, although a marginal effect was present in the fixation analysis ($\Delta G^2(1, \underline{N} = 384) = 3.04$, $p < 0.1$). The main effect of the number of referents in the display was present in the fixation ($\Delta G^2(1, \underline{N} = 384) = 11.03$, $p < 0.001$; Figure 32A) and saccade ($\Delta G^2(1, \underline{N} = 384) = 7.89$, $p < 0.01$; Figure 32B) analyses during PP1, as in Experiments II and IIIA. Participants were more likely to look to the late/correct goal in the one referent conditions than the two referent conditions. They were also more likely to saccade to the late/correct goal in the unambiguous utterance conditions ($\Delta G^2(1, \underline{N} = 384) = 6.15$, $p < 0.05$), indicating again that the point at which disambiguation occurs can affect whether or not anticipatory saccades are made. This tendency was marginally more likely to occur in the fully ambiguous display than the temporarily ambiguous display ($\Delta G^2(1, \underline{N} = 384) = 3.65$, $p < 0.1$).

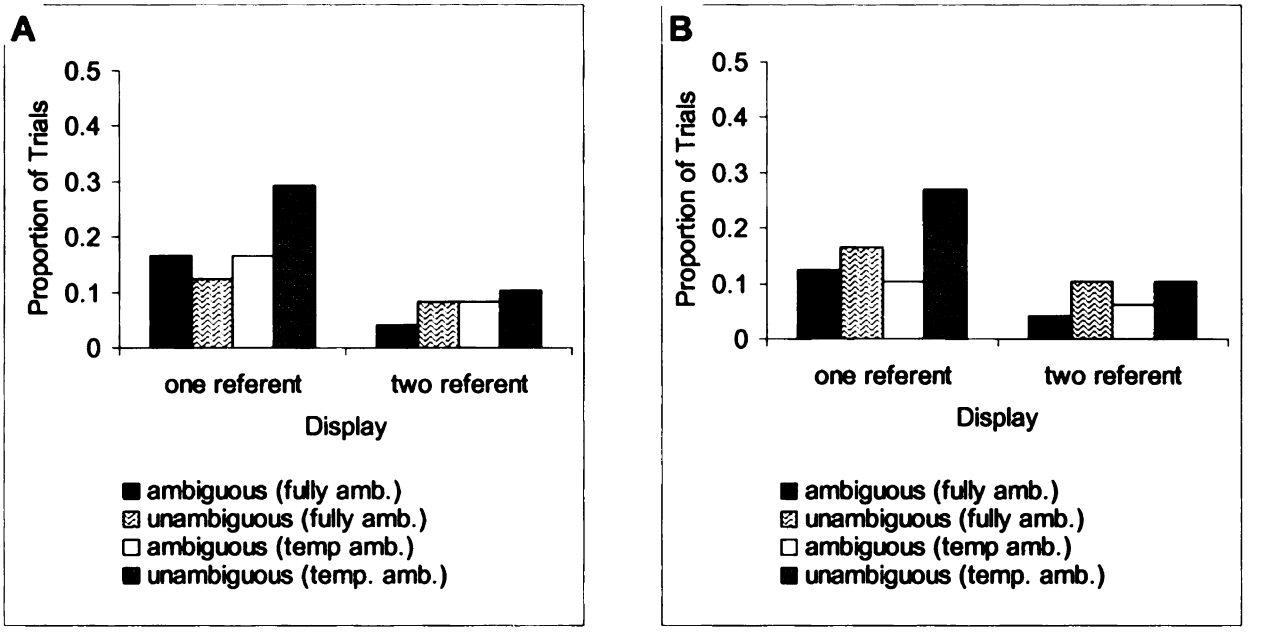


Figure 32. Proportion of trials with a fixation in (A) and with a saccade to (B) the late/correct goal during the PP1 segment of the utterance in Experiment IIIB. "Ambiguous" and "unambiguous" refer to utterance types and "fully ambiguous" and "temporarily ambiguous" to display types.

The behavioral evidence from Experiment IIIB, then, indicates that both possible interpretations of an ambiguous utterance are activated at some point during language comprehension, as the condition that allows for both final interpretations elicited behaviors consistent with both interpretations. Eye movement evidence is consistent with this conclusion. Participants make eye movements consistent with either a garden path or with the anticipation of a goal argument, and perseveration suggests the presence of a garden path on not only one referent, but also two referent trials. Some unexpected effects of the fully ambiguous display did occur, suggesting that the effects of the fully ambiguous display were widespread and not merely confined to looks to the early/incorrect goal region. Thus, Experiment IIIB suggests that a garden path can occur during these trials, but reinforces the earlier claims that this paradigm may not be

robustly sensitive to syntactic effects and that difference between displays can cause unexpected effects.

Experiment IIIC

Experiment IIIB demonstrated that the interpretation of an ambiguous utterance such as "put the apple on the towel in the box" can be consistent with either of its syntactically licensed interpretations as long as the display does not constrain the final interpretation of the utterance. This allows for the possibility of garden pathing during comprehension of the utterance, although unambiguous utterance conditions continue to show garden path like effects during the utterance. Several experiments presented here have also shown anticipatory saccades to the correct goal before either a preposition or a noun referred to that goal. Based on these anticipatory saccades, I have argued that fixations on the incorrect goal are also anticipatory, and that a garden path may occur at the same time, or even after these anticipatory saccades have been launched. Thus, there is no online evidence for garden pathing, only evidence from the eye movement record that the language comprehension system has completed NP1 reference resolution.

This hypothesis can be tested directly by observing participants' eye movement patterns while they view unambiguous displays. If participants are presented with a display that contains no temporarily ambiguous referent to the incorrect goal, and eye movements are driven by either lexical processing of the noun in PP1 or by a syntactic garden path, the incorrect goal should never be fixated. On the other hand, if participants are making anticipatory saccades in those cases where they have completed NP1 reference resolution, we might expect to see saccades to the incorrect goal, even though it has not yet been referred to by any portion of the utterance.

Material and Methods

Participants. Sixteen participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($7.00). All participants were native speakers of English, and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported in this dissertation.

Materials. The materials created for Experiment IIIA were used in this experiment. However, the 24 displays created for critical trials were again modified so that the incorrect goal was replaced with an object that was never mentioned (Figure 33). This object was identical to the object on which the distractor object was resting; thus, prior to the onset of the utterance, the ambiguous and unambiguous displays were indistinguishable from the point of view of a participant. Twelve unambiguous displays (six one referent and six two referent) were substituted for the temporarily ambiguous displays used in Experiment II in each list created. Fillers were adjusted if necessary to refer to the new incorrect goal. Each display occurred in the temporarily ambiguous and unambiguous forms an equal number of times. Random trial lists were created according to the procedure described in Experiment II.

Apparatus. The apparatus for this experiment was identical to that of Experiment II.

Procedure. The procedure for this experiment was identical to that of Experiment IIIA.

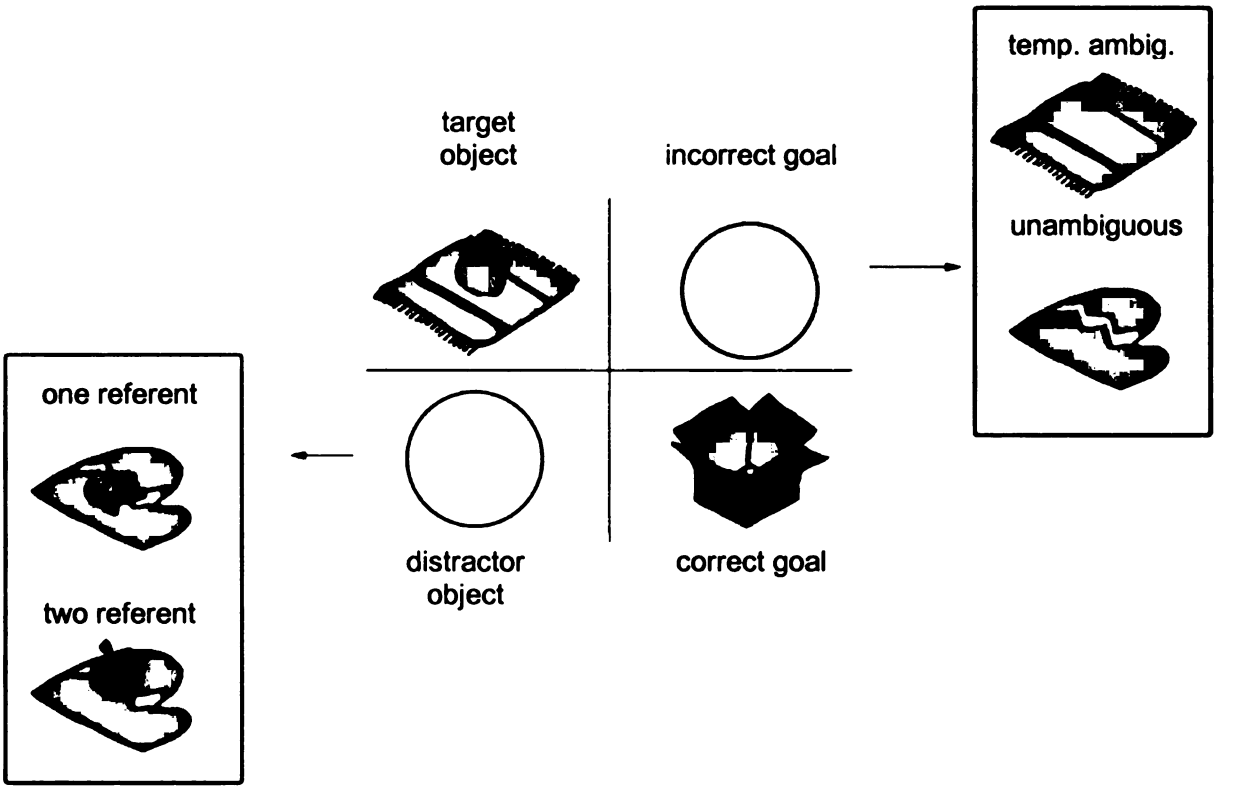


Figure 33. Displays created for Experiment IIIC. All four possible displays were seen an equal number of times by participants.

Design. Four different display conditions were created by crossing the different number of possible referents manipulation (one referent, and two referent) with the display ambiguity manipulation (temporarily ambiguous or unambiguous; Figure 33). The two utterance types (ambiguous and unambiguous; Table 10) were then combined with the four displays to create eight unique conditions for this experiment. Three trials in each condition were presented to each participant, for a total of 24 critical trials. Each display occurred in each condition an equal number of times.

Table 10. Utterance types used in Experiment IIIC. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance (segments for analysis indicated by subscripts)
Ambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
Unambiguous	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ that's on the towel /$_{PP2}$ in the box./

Results and Discussion

The number of trials with fixations on and saccades to regions of interest were calculated for each of the segments in Table 10. The frequencies for each segment and region of interest were then submitted to a 2 (display ambiguity) by 2 (number of possible referents in the display) by 2 (utterance type) multiway frequency analysis. The box on the right in Figure 33 indicates the display ambiguity conditions, the box on the left the number of referents conditions, and Table 10 indicates the utterance type conditions.

One possibility that has not been discussed concerning visual world experiments is that participants may implicitly or explicitly be learning the types of utterances and display ambiguities that occur in this paradigm during the course of an experiment. There are several reasons to believe that this may be the case. The same verb is used on each trial in these experiments, and so subjects are likely expecting a target object and goal on each trial. In addition, the fillers in this study and in other studies have been biased so that the first PP encountered in an utterance is the correct goal over 70% of the time. Participants make relatively few nonstandard responses in these experiments (a non standard response to a temporarily ambiguous display would involve moving the target object to the incorrect goal and then to the correct goal, for example), but when such errors are made, they tend to be earlier in the experiment (Engelhardt,

Bailey, & Ferreira, 2004). Although there are relatively few utterances in a visual world experiment, evidence (Bienvenue & Mauner, 2003) indicates that participants can learn a syntactic structure relatively quickly over the course of an experiment.

Most of the above observations have to do with subjects' expectations concerning the structure of the utterance. However, there is also the possibility that participants are learning the relationship between the distinctive displays used in these experiment and the ambiguity of the utterances. If participants are capable of doing this, they may attempt to anticipate which objects will be targets and goals and this may bias their eye movement pattern. The current experiment allows for a direct test of whether subjects are forming expectations about the target and goal objects in the display prior to the onset of the utterance. Because the unambiguous and temporarily ambiguous displays are indistinguishable prior to the onset of the display, we might expect subjects in this experiment to show no difference in looks to the one referent and two referent displays early in the utterance as they will not be able to easily predict which object will be the correct target object.

This, in fact, is the pattern of results during the NP1 segment for fixations on and saccades to the distractor object. No significant effects or interactions were present (all $p > 0.05$). This supports the hypothesis that subjects may be generating expectations about the critical displays, as lexical processing and reference resolution should not be driving eye movements to the distractor object early on in the one referent conditions.

The tendency to look to the distractor object in the one referent conditions does not last long, as participants are more likely to fixate on ($\Delta G^2(1, \underline{N} = 384) = 76.54$, $p < 0.001$; Figure 34A) and saccade to ($\Delta G^2(1, \underline{N} = 384) = 31.97$, $p < 0.001$; Figure 34B) the distractor object in the two referent condition during PP1. Participants were also significantly more likely to launch a saccade to the distractor object in the unambiguous utterance conditions ($\Delta G^2(1, \underline{N} = 384) = 5.21$, $p < 0.05$), a tendency that was more prevalent in the two referent, unambiguous display, as evidenced by an interaction between all three variables ($\Delta G^2(1, \underline{N} = 384) = 5.49$, $p < 0.05$). Thus, whatever effects are due to expectations about the display appear to be quickly replaced by effects due to the processing of the utterance in the context of the display.

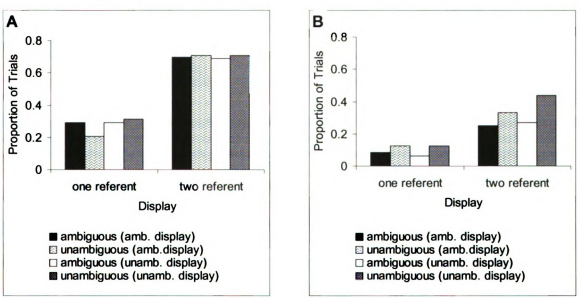


Figure 34. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor object during the PP1 segment of the utterance in Experiment IIIC. "Ambiguous" and "unambiguous" refer to utterance types and "ambiguous display" and "unambiguous display" to display types.

The most important predictions made about the processing of unambiguous displays relative to ambiguous displays concerns whether or not

participants will look towards the incorrect goal during PP1. If they do not , it is likely that the looks to the incorrect goal are generated by the lexical processing of the noun in PP1, and by syntactic processing leading to a garden path. If, on the other hand, subjects do look at the incorrect goal in the unambiguous display, this behavior is likely more akin to the anticipatory saccades made to the correct goal during PP1 in previous experiments.

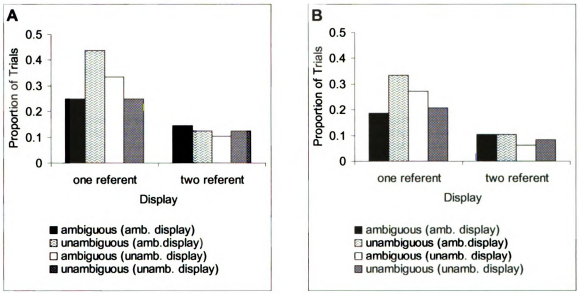


Figure 35. Proportion of trials with a fixation in (A) and with a saccade to (B) the incorrect goal during the PP1 segment of the utterance in Experiment IIIC. "Ambiguous" and "unambiguous" refer to utterance types and "ambiguous display" and "unambiguous display" to display types.

Participants are, in fact, more likely to fixate on ($\Delta G^2(1, \underline{N} = 384) = 22.97$, $p < 0.001$; Figure 35A) and saccade to ($\Delta G^2(1, \underline{N} = 384) = 19.66$, $p < 0.001$; Figure 35B) the incorrect goal in the one referent conditions than in the two referent conditions, but are not significantly more likely to do so in either the temporarily ambiguous or the unambiguous display (fixations: $\Delta G^2(1, \underline{N} = 384) = 0.86$, $p > 0.1$; saccades: $\Delta G^2(1, \underline{N} = 384) = 0.53$, $p > 0.1$). There is a three way interaction between display ambiguity, the number of referents in the display,

and utterance type in the fixation analysis ($\Delta G^2(1, \underline{N} = 384) = 5.49$, $p < 0.05$) that may be due to a combination of the ease of NP1 reference resolution, and lexical processing in the temporarily ambiguous display.

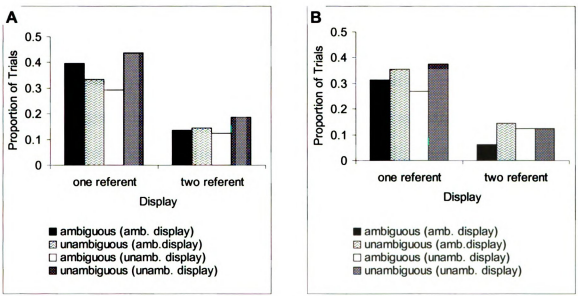


Figure 36. Proportion of trials with a fixation in (A) and with a saccade to (B) the correct goal during the PP1 segment of the utterance in Experiment IIIC. "Ambiguous" and "unambiguous" refer to utterance types and "ambiguous display" and "unambiguous display" to display types.

A similar pattern of results is present in looks to the correct goal during PP1. Participants were again more likely to fixate on ($\Delta G^2(1, \underline{N} = 384) = 25.44$, $p < 0.001$; Figure 36A) and saccade to ($\Delta G^2(1, \underline{N} = 384) = 27.98$, $p < 0.001$; Figure 36B) the correct goal despite a lack of reference to that region in the utterance up to that point. Marginal interactions between display ambiguity and utterance type in the fixation analysis ($\Delta G^2(1, \underline{N} = 384) = 2.94$, $p < 0.1$) and between display ambiguity, the number of referents in the display, and utterance type in the saccade analysis ($\Delta G^2(1, \underline{N} = 384) = 3.25$, $p < 0.1$) suggest that to some degree the conditions where NP1 reference resolution was easier elicited more anticipatory saccades. The results are consistent with the hypothesis that looks to

the incorrect and correct goals are the result of anticipatory processes that occur once NP1 reference resolution has completed and the referent of NP1 (e.g. the correct apple) has been identified. These anticipatory processes are likely driven by the requirement of the verb "put" that a theme and goal both be present in an utterance.

In summary, then, Experiment IIIC, like the other two experiments in this chapter, suggests that garden path like effects that have appeared in other studies may be due to anticipatory saccades rather than syntactic processing. A garden path can occur in these experiments (as was indicated by the behavioral and post utterance eye movement data from Experiment IIIB), but likely does so after anticipatory saccades have been launched and thus only results in the continued direction of attention to the incorrect goal until the disambiguating PP2 is encountered. Eye movement patterns, then, are likely not sensitive to garden paths because fixations due to garden pathing and those due to anticipation are underadditive.

Nevertheless, Experiments II and IIIA-C allow us to identify three replicable effects that may be diagnostic of the processing of utterances in this particular visual world paradigm. First, looks to the distractor object during PP1 indicate the relative rapidity with which NP1 reference resolution can be completed in various conditions. In conditions where NP1 reference resolution is relatively easy (usually because there are multiple cue that either identify or rule out possible referents), participants are less likely to fixate on and saccade to the distractor object during PP1. When the eye movement system is released from processes related to NP1 reference resolution, anticipatory saccades are seen to

the incorrect goal and correct goal during PP1. More saccades are usually made to the incorrect goal, because of lexical processing of the preposition and noun in PP1, both of which are semantically related to the incorrect goal. The effect of syntactic disambiguation can be seen in looks to the correct goal during PP1 (which is prior to any possible reference to the correct goal) as unambiguous utterance conditions are more likely in some circumstances to elicit anticipatory saccades.

These diagnostic patterns of eye movement behavior can thus be applied to the study of disfluencies to determine whether the presence of a disfluency can change language processing behavior. I will return to questions concerning the processing of disfluencies in the next chapter.

COMPREHENSION OF DISFLUENT SPEECH: FILLED PAUSES

Experiment IV

Although Experiment I was conducted under the assumption that the visual world paradigm was sensitive to syntactic processing, evidence from Experiments II and IIIA-C suggest that it is instead sensitive to the relative difficulty of resolving the referent of the first NP and the requirements of the verb "put". Nevertheless, we might still expect to find some effects of disfluencies. In Experiment I, there were indications that disfluencies were affecting early processing (i.e. NP1 reference resolution), but that there were few effects later in the utterance. However, Experiment I also raised several questions about the visual world paradigm itself. After testing several of the factors that may have contributed to the failure to find effects later in the utterance (namely the effects of the point of disambiguation, and the degree to which garden pathing and anticipatory saccades were responsible for effects that were initially interpreted as garden path effects), it is clear that the type of visual world used in Experiment I is sensitive to language comprehension processes, albeit processes that are somewhat different than those previously assumed to be at work.

Important questions still remain to be asked about the effects of disfluencies on these processes, however. First, to what degree do disfluencies cue participants to the presence of an ambiguity in the display? Does the presence of the disfluency, in addition to the time taken up by the disfluency increase the rate of reference resolution? Do disfluencies make participants more likely to immediately shift their attention to locations that could be the source of difficulty for the speaker? Do disfluencies affect the overall tendency for

participants to make saccades in anticipation of satisfying all of the requirements of the verb? This last question is very similar to the cuing hypothesis described in Experiment I, but instead of cuing a particular syntactic structure, it simply proposes that disfluencies will make the language comprehension system more likely to attempt to anticipate how the speaker will continue the utterance.

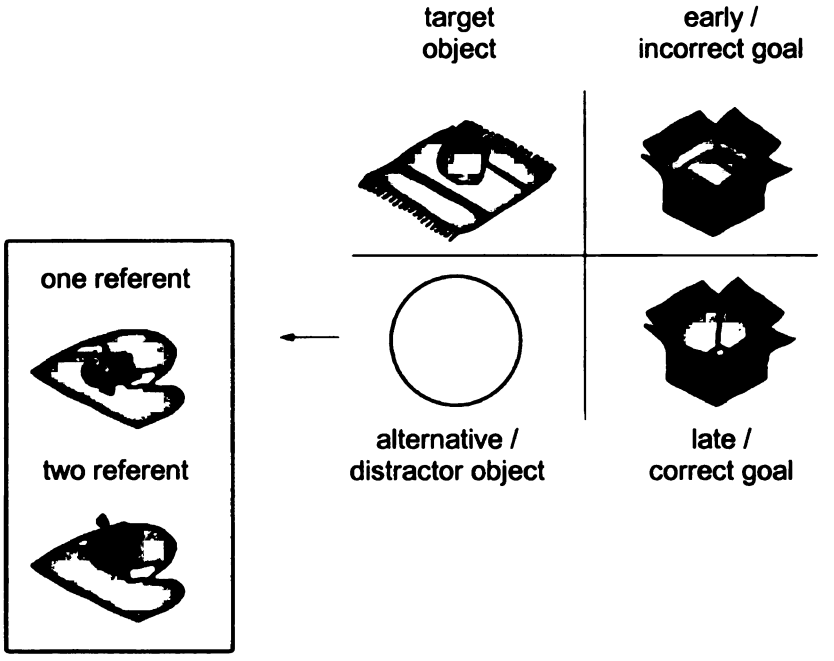


Figure 37. Fully ambiguous displays used in Experiment IV.

To begin to answer these questions, a modified version of Experiment I was conducted, with changes to the display and utterance types. Based on the results of Experiments IIIA-C, the fully ambiguous display was selected for use in this experiment. This display (Figure 37) provides two (in the one referent version) or three (in the two referent version) referential ambiguities and thus allows for a variety of points within the utterance where the presence of a

disfluency might have some effect. Critically, a one referent version of the fully ambiguous display will be used in this experiment (only a two referent fully ambiguous display was used in Experiment I). The behavioral results of Experiment IIIB indicate that this is the only display that does not constrain the interpretation of a concurrent ambiguous utterance. In this chapter, I will refer to the object represented by the towel in the box in Figure 37 as the early goal (that is, the goal that would be selected by interpreting PP1 as a goal and PP2 as a modifier) and the object represented by the box by itself as the late goal (the goal that would be selected by interpreting PP1 as a modifier and PP1 as a goal).

Disfluencies in either NP1 or PP1 were again used (Table 11). Although these disfluencies were no longer predicted to bias the syntactic parse of the utterance by cuing particular syntactic structures, they may focus attention on a particular reference resolution problem and thus bias the shift of attention to the regions that are possible referents. As a result, differences in final syntactic structure may still result but will be due to the timing of shifts of attention, rather than to the expectations of the syntactic parser. This is not to say that disfluencies cannot cue the syntactic parser, but rather that lexical and reference resolution effects are also possible (see, for instance, Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Altmann, et al., 2003), and in this case are much more likely to be tapped by this particular visual world experiment. Unambiguous control utterances that disambiguate the utterance in favor of either NP1 or PP1 modification were also included. Disfluent utterances formerly referred to as "theme bias" and "goal bias" will now be referred to as NP1 disfluencies and PP1 disfluencies respectively, based on their location in the utterance.

Material and Methods

Participants. Sixteen participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($7.00). All participants were native speakers of English, and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported in this dissertation.

Materials. Disfluent utterances were created for this experiment in the same manner as was described in Experiment I (Table 11). The ambiguous displays used in Experiment IIIB were used. Half of all fillers were also rerecorded with disfluencies in a variety of syntactic positions.

Table 11. Utterance types used in Experiment IV. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance (segments for analysis indicated by subscripts)
NP1 disfluency	/$_{VERB}$ put /$_{NP1}$ the uh apple /$_{PP1}$ on the towel /$_{PP2}$ in the box./
PP1 disfluency	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the uh towel /$_{PP2}$ in the box./
NP1 modification	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ that's on the towel /$_{PP2}$ in the box./
PP1 modification	/$_{VERB}$ put /$_{NP1}$ the apple /$_{PP1}$ on the towel /$_{PP2}$ that's in the box./

Apparatus. The apparatus for this experiment was identical to that of Experiment II.

Procedure. The procedure for this experiment was identical to that of Experiment IIIB.

Design. The four utterance types (NP1 and PP1 disfluencies and NP1 and PP1 modifiers; Table 9) were combined with the two displays (one referent and two referent; Figure 37) to create eight unique conditions for this experiment.

Three trials in each condition were presented to each participant, for a total of 24 critical trials. Each display occurred in each condition an equal number of times.

Results and Discussion

A previous experiment using the fully ambiguous displays used in this experiment (Experiment IIIB) found that the final interpretations of ambiguous utterances were split approximately evenly between the two possible interpretations of the utterance in the one referent condition. In the two referent condition, as with PP1 modifier unambiguous utterances, almost all (> 90%) behavioral responses involved the movement of the target object to the late goal. Note that in the current experiment, all ambiguous utterances were also disfluent.

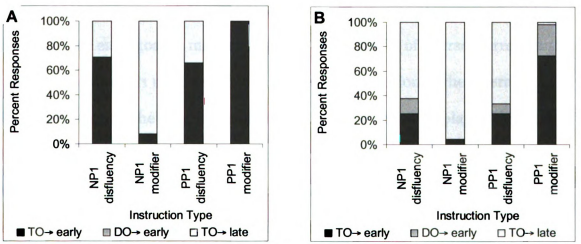


Figure 38. Percent trials in which a target object is moved to the early and late goals in one referent (A) and two referent (B) displays. "TO" and "DO" refer to target object and distractor object respectively.

The behavioral responses to disfluent instructions in the current experiment were classified as either early goal directed, or late goal directed and were submitted to a 2 (number of referents) by 2 (location of disfluency) multiway frequency analysis. Unambiguous controls were not included in this

analysis as subjects moved an object to the appropriate goal on over 90% of trials in each condition. In the disfluent trials, participants were more likely to move a target object to the late goal in the one referent display than in the two referent display ($\Delta G^2(1, \underline{N} = 192) = 25.64$, $p < 0.001$; Figure 38). The location of the disfluency had no effect ($\Delta G^2(1, \underline{N} = 192) = 0.51$, $p > 0.1$; Figure 38) and there was no significant interaction between the number of referents in the display and the location of disfluencies. The proportion of trials on which an object was moved to the late goal was higher than was found with the fluent ambiguous utterances in Experiment IIIB for both one and two referent displays.

One other interesting behavioral response was greater proportion of trials in the two referent display, PP1 modifier condition in which the target object was moved to the early goal in comparison to trials where the alternative object was moved to the early goal. A movement to the early goal is, of course, correct in this condition, as the PP1 modifier constrains the interpretation of the utterance. However, because the alternative object in the two referent display is identical to the target object, participants must decide which of the objects to move in some way. The final syntactic parse of the utterance does not allow them to use "on the towel" to determine which apple should be moved. A partial parse of the utterance ("put the apple on the towel..."), on the other hand, does allow this. At the onset of PP2 ("...that's in the box"), this analysis should be revised, of course; nevertheless, participants appear to use the PP "on the towel" to both resolve the NP1 ("the apple") reference <u>and</u> denote the goal. This is surprising in some ways, because there is a behavioral option that would allow subjects to avoid using "on the towel" twice; namely, moving the alternative object to the early goal. Such a

response would require some reanalysis of previously completed reference resolution, of course, and this may be beyond the scope of reanalysis processes.

Disfluencies then, had no effect on the final interpretation of the utterance in this study as indicated by behavioral responses. While numerically it would appear that the mere presence of a disfluency in an utterance made a move to the late goal less likely (the late goal being the preferred goal location in the two referent ambiguous conditions), this comparison cannot be easily made as the fluent baseline was obtained in an experiment (IIIB) with a completely different design. In order to ascertain whether disfluencies had any online effects on language comprehension, the number of trials with fixations on and saccades to regions of interest were calculated for each of the segments in Table 11. The frequencies for each segment and region of interest were then submitted to a 2 (cue location; NP1 or PP1) by 2 (number of possible referents in the display) by 2 (cue type; disfluency or modifier) multiway frequency analysis. The box on the left in Figure 37 indicates the one referent and two referent display conditions, and Table 11 indicates the utterance type conditions.

If the presence of a disfluency was interpreted as a cue that the speaker was attempting to overcome some planning problem, we might expect the language comprehension system to attempt to anticipate the problem and produce possible continuations during the disfluent interruption of the delivery. The flavor of this process can be obtained from the anecdotal example of listeners completing the utterances of disfluent speakers, although the process proposed here does not require conscious awareness of possible completions. In a grammaticality judgment experiment of the type described in earlier chapters

(Bailey & Ferreira, 2003), where no context is present, the language comprehension system might anticipate continuations by activating possible syntactic structures, as there is no basis for guessing the upcoming word. In the visual world paradigm, however, it is likely that the next word in a disfluency that begins "...the uh -..." will be the name of one of the objects in the display. Thus, we might expect a disfluency to bias participants to shift their attention more rapidly in the temporal vicinity of a disfluency. If this is the case, we would expect to see more saccades to possible goals in the NP1 disfluency than the PP1 disfluency during NP1, which is prior to the disfluency in the latter utterance type. Moreover, this tendency should be greater in the disfluent than the modifier conditions as the modifier conditions are fluent and ambiguous during NP1 with respect to NP1 reference resolution.

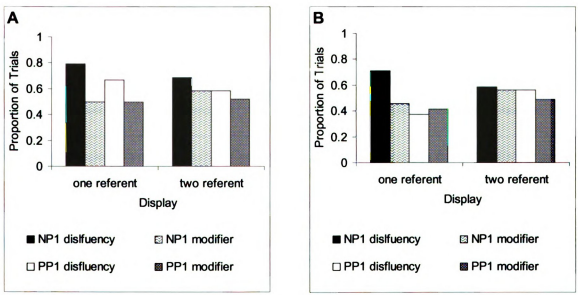


Figure 39. Proportion of trials with a saccade to the target object (A) and distractor object (B) during the NP1 segment of the utterance in Experiment IV.

A pattern of results that supports this account was found for saccades to both the target object and the alternative/distractor object regions. Participants

were more likely to launch a saccade to the target object in the disfluent conditions ($\Delta G^2(1, \underline{N} = 384) = 10.11$, $p < 0.01$; Figure 39A) during NP1. A marginal interaction between cue location and cue type was also present ($\Delta G^2(1, \underline{N} = 384) = 2.84$, $p < 0.1$). More saccades were launched to the alternative/distractor object in conditions where the cue was associated with NP1 (e.g. NP1 disfluency and modifier conditions; $\Delta G^2(1, \underline{N} = 384) = 5.41$, $p < 0.05$; Figure 39B). This effect was largely driven by the high proportion of trials with saccades to the distractor goal in the one referent, NP1 disfluency condition.

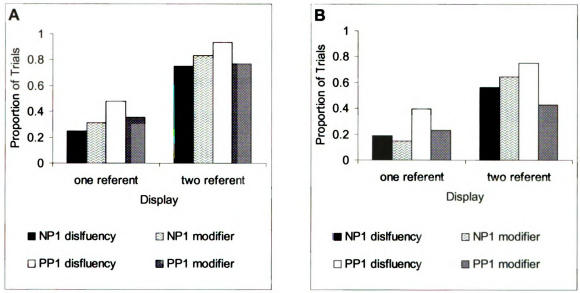


Figure 40. Proportion of trials with a fixation in (A) and with a saccade to (B) the distractor object during the PP1 segment of the utterance in Experiment IV.

During the PP1 segment, the familiar main effect of number of referents in the display appeared, with more saccades to ($\Delta G^2(1, \underline{N} = 384) = 60.29$, $p < 0.001$; Figure 40B), and more fixations on ($\Delta G^2(1, \underline{N} = 384) = 101.77$, $p < 0.001$; Figure 40A) the alternative/distractor object in the two referent conditions. There were more saccades to the alternative distractor object in disfluent trials ($\Delta G^2(1, \underline{N} = 384) = 6.24$, $p < 0.05$) and more fixations on the alternative/distractor object in

trials were the cue was associated with PP1 ($\Delta G^2(1, \underline{N} = 384) = 5.58, p < 0.05$). The predicted drop off in looks to the alternative/distractor object in the NP1 disfluency conditions and increase in looks in the PP1 disfluency conditions relative to NP1 was reflected in significant interactions of cue location and cue type in both the saccade ($\Delta G^2(1, \underline{N} = 384) = 11.64, p < 0.001$) and fixation ($\Delta G^2(1, \underline{N} = 384) = 8.39, p < 0.01$) analyses. A marginal interaction between the number of referents in the display and cue location ($\Delta G^2(1, \underline{N} = 384) = 3.83, p < 0.06$) and a significant interaction between all three variables ($\Delta G^2(1, \underline{N} = 384) = 6.25, p < 0.05$) was present in the analysis of saccades to the alternative/distractor goal during PP1, and reflected a low likelihood of saccade launch in the one referent display conditions where the cue was associated with NP1.

Predictions for looks to possible goals during PP1 segment are a little more difficult as a disfluency will have occurred in both disfluent utterance types by the end of that segment. However, the PP1 disfluency will be more recent, as it occurs during PP1. If we assume that disfluencies are cues to shift attention and anticipate the speaker's continuation, then we might expect that these cues will decrease in strength as time passes and predictions are either confirmed or rejected. Thus, the PP1 disfluency should be a better cue to anticipate the speaker's continuation during PP1 than should the NP1 disfluency, although the latter may still have an effect. This is the same type of pattern discussed earlier in the looks to the alternative/distractor object during PP1.

Looks to the early goal also showed the typical pattern of more looks in the one referent than in the two referent displays, reflecting the earlier NP1 reference resolution in the former display type. This main effects was significant for both

saccade ($\Delta G^2(1, \underline{N} = 384) = 59.54$, $p < 0.001$; Figure 41B) and fixation ($\Delta G^2(1, \underline{N} = 384) = 61.18$, $p < 0.001$; Figure 41A) analyses. A main effect of cue type was also present, with disfluent conditions eliciting more saccades ($\Delta G^2(1, \underline{N} = 384) = 19.23$, $p < 0.001$) and fixations ($\Delta G^2(1, \underline{N} = 384) = 16.41$, $p < 0.001$) than unambiguous fluent controls. Finally, there were marginal main effects of cue location in fixation on ($\Delta G^2(1, \underline{N} = 384) = 3.09$, $p < 0.1$) and saccades to ($\Delta G^2(1, \underline{N} = 384) = 3.52$, $p < 0.1$) the early goal during PP1. Looks are more likely in the conditions where the cue is associated with PP1, as would be predicted if the recency of the disfluency was important.

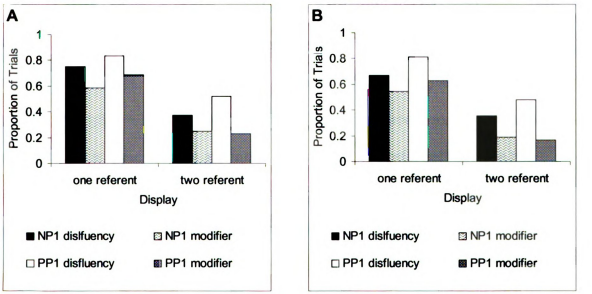


Figure 41. Proportion of trials with a fixation in (A) and with a saccade to (B) the early goal during the PP1 segment of the utterance in Experiment IV.

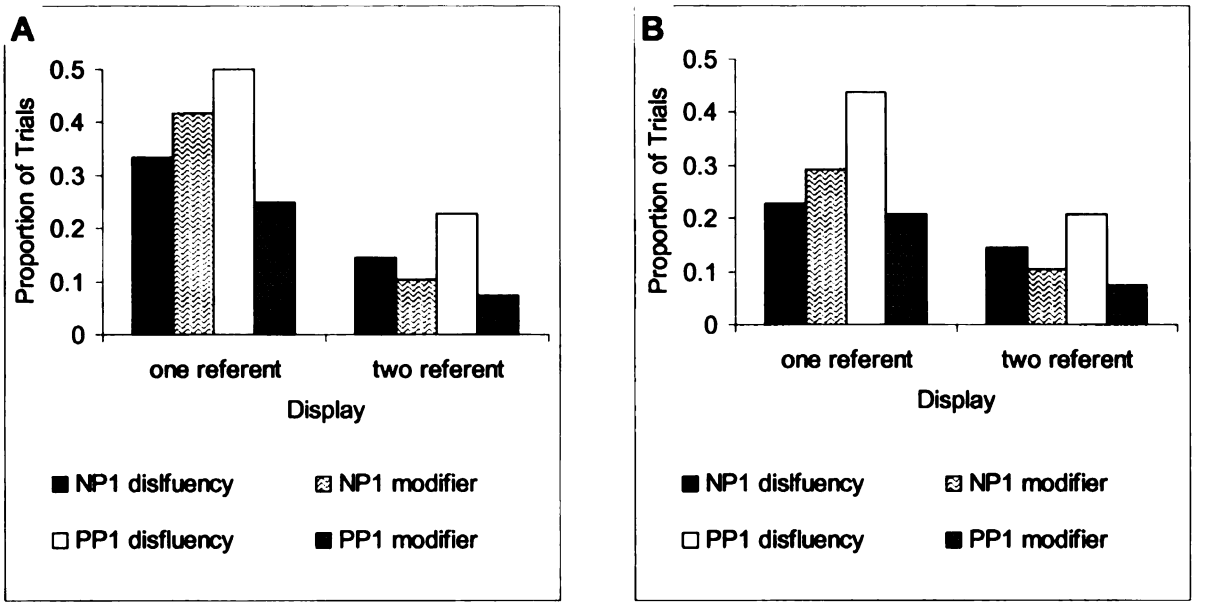


Figure 42. Proportion of trials with a fixation in (A) and with a saccade to (B) the late goal during the PP1 segment of the utterance in Experiment IV.

A similar pattern is seen in anticipatory looks to the late goal during PP1. Again, a main effect of the number of referents in the display is present for both fixations on ($\Delta G^2(1, \underline{N} = 384) = 33.51$, $p < 0.001$; Figure 42A) and saccades to ($\Delta G^2(1, \underline{N} = 384) = 16.17$, $p < 0.001$; Figure 42B) the correct goal during this segment. As is typical, there are more looks in the one referent than the two referent displays. The same main effect of cue type seen in the looks to the early goal, with disfluencies eliciting more looks than modifiers, was present in the fixation ($\Delta G^2(1, \underline{N} = 384) = 5.04$, $p < 0.05$) and saccade ($\Delta G^2(1, \underline{N} = 384) = 4.69$, $p < 0.05$) analyses for the correct goal during PP1. Finally, an increased number of looks in the PP1 disfluency condition, relative to the NP1 disfluency condition, in the one referent display was present as a marginal interaction between cue type and location in the saccades to the correct goal during PP1 ($\Delta G^2(1, \underline{N} = 384) = 3.18$, $p < 0.1$) and a significant interaction between the same two variables in the fixation on the correct goal ($\Delta G^2(1, \underline{N} = 384) = 8.61$, $p < 0.01$).

Disfluencies in the visual world paradigm, then, are interpreted by the language comprehension system as an indication of trouble in producing the correct descriptor to unambiguously identify an object in the display. Because of the limited number of objects in a visual world display, is possible for the language comprehension system to make predictions about how the disfluent speaker will continue the utterance. These predictions are realized as anticipatory saccades. The likelihood of making this type of saccade decreases with time. Thus, the failure to find the sorts of syntactic effects of disfluencies that have been reported in offline studies (Bailey & Ferreira, 2003) are likely due to the presence of a concurrent visual context in the visual world paradigm, the ease of parsing that context, and the necessity of rapidly resolving NP reference. If we assume that the language comprehension system deal with disfluencies by making predictions about the speaker's continuation in whatever way is possible during the disfluent interruption, these are the results that we would expect. Rather than consider this behavior to be a flaw in the visual world paradigm, I will take advantage of it to ask questions about the online processing of a variety of types of disfluencies in the next chapter.

COMPREHENSION OF DISFLUENT SPEECH: REPEATS AND REPAIRS

Experiment V

Experiments I and IV suggested that disfluencies may increase the tendency of the language comprehension system to attempt to predict how the speaker will continue the utterance. In visual world experiments, these predictions are realized as an increased likelihood of making an anticipatory saccade to a possible referent object as a result of a shift in attention to that object. Anticipatory saccades, are, of course, also made in the absence of disfluency, and seem to be driven by the semantic and syntactic features of material early in the utterance, especially verbs (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Altmann & Kamide, 2004).

Naturally, the filled pause disfluencies used in Experiments I and IV are not the only type of disfluencies with which the language comprehension system must deal. Recall the disfluency schema described earlier (Figure 1). Every disfluency is made up of, at minimum, the suspension of an original delivery, an optional hiatus, and a resumption of fluent delivery. In some types of disfluencies, a portion of the original delivery prior to the suspension point may be revised in the resumption; this portion is referred to as the reparandum. In a filled pause disfluency, the hiatus is filled by an "um" or "uh" and there is no reparandum. The resumption simply continues the original delivery. Thus, filled pause disfluencies do not present problems for the parser beyond the delay and filler during the hiatus. No previously built structure must be revised as a result of the resumption.

Additional processing difficulty might be introduced by a resumption that involves lexical material that is related to the reparandum in some way. In the simplest case, a resumption might repeat the reparandum; this is referred to as a repetition disfluency. Repetition should also be relatively easy for the language comprehension system to deal with, as it simply involves the matching of the repeated material to reparandum material that has already been incorporated into the structure built by the parser (Ferreira, Lau, & Bailey, in press; Ferreira & Bailey, 2004). More difficult to process might be resumptions that revise material in the reparandum, known as repair disfluencies. There is evidence that revision during disfluency processing sometimes incomplete: Previously constructed syntactic and semantic structures occasionally linger (Ferreira, Lau, & Bailey, in press; Ferreira & Bailey, 2004). The overlap and relationship between the resumption and the reparandum, then, present a processing challenge to the language comprehension system.

While it seems reasonable that repair disfluencies should be more difficult to process than repetition disfluencies, and that repetition disfluencies may be slightly more difficult to process than pauses due to the presence of extraneous lexical material, research on the processing of disfluencies is failed to yield consistent results. Some research has found that repair disfluencies have an associated processing cost (Fox Tree, 1995), and that repetition disfluencies are, at worst, no more difficult to process than fluent utterances. Other studies, however, have found that not only do repair disfluencies not seem to have an associated processing cost, they may, in fact, aid processing in some circumstances (Brennan & Schober, 2001). Complicating matters even further,

studies of the processing of repair disfluencies involving verbs seem to indicate that repair disfluencies can either increase or decrease the likelihood of the parser arriving at a final, grammatical structure, depending on the circumstances (Ferreira, Lau, & Bailey, in press).

All of these studies, however, have used relatively offline tasks in order to try to measure the effects of disfluencies on language comprehension. As a result, these studies cannot tell us how disfluency processing unfolds over time. If, as has been proposed earlier, the language comprehension system attempts to anticipate the speaker's continuation during the disfluent interruption, we should expect to see two separate effects of processing of repair disfluencies. These effects should occur at particular points during the processing of the disfluent utterance. First, we should see evidence of anticipatory processes beginning with the suspension of fluent speech. Second, at the resumption, any overlap with or modification of the reparandum should result in a shift in attention in accordance with the information introduced in the resumption. If there are any lingering effects of the reparandum, however, the shift of attention will not occur in all cases, or there may be shifts of attention between the old and new referents.

The hypothesis that disfluencies result in the anticipation of upcoming material can explain, in part, the difference found in past studies of disfluency processing. In the experiments that found that repair disfluencies speeded response times (Brennan & Schober, 2001), the set of possible referents was very small (two or three objects). Thus, anticipatory processing could have resulted in attention already being directed to the correct referent before the resumption of the utterance. In experiments that found processing costs associated with the

presence of a repair disfluency (again using a reaction time measure; Fox Tree, 1995), there were no supporting contexts for the utterance; rather, utterances were sampled from a corpus of spontaneous speech, were unrelated to each other, and were presented in isolation. Thus, it would have been much more difficult for the language comprehension system to anticipate the speaker's continuation.

The current experiment takes advantage of the anticipatory effects describe in earlier experiments and in previous studies by other researchers (e.g. Altmann & Kamide, 1999) to examine the processing of various types of disfluencies involving verbs. Although disfluencies on verbs are not necessarily the most common type, they do occur in spontaneous speech. In addition, verbs have a great deal of linguistic information associated with them and thus are a good starting point for a study of the time course of disfluency processing.

In this experiment, I hope to answer three separate questions. Repetition disfluencies have received little attention because of their low rate of occurrence in the corpora from which disfluent utterances in previous experiments were selected, and because of task related confounds. However, repetition disfluencies provide an important comparison to repair disfluencies, because they introduce additional lexical material in the resumption, require the resumption to be compared with the reparandum, but do not require revision of the reparandum. Thus, we can begin by asking whether there is any effect associated with the mere comparison of material from the resumption and the reparandum, even when no revision is required. Second, we can ask whether revision of the representations generated at the reparandum can affect processing. Lastly, the type of revision

required in a repair disfluency can vary. Appropriateness repairs involve a change in the specificity or degree of some concept expressed in the reparandum (Levelt, 1983; Levelt & Cutler, 1983); while error repairs involve a complete change in concept. Do these two different types of repair themselves differ?

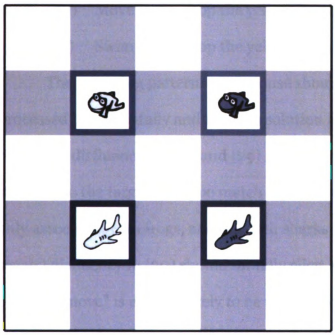


Figure 43. Visual world for Experiment V. The grid used in the actual experiment utilized dark background colors in order to produce better video images.

In order to attempt to answer these questions, a novel visual world was constructed for this experiment (Figure 43). In this visual world, four objects of two different types and two different colors are present. Thus, for any target object (e.g. a yellow frog) in the display, there are two objects that overlap on one feature (e.g. color: a yellow shark; type: a blue frog), and one object that shares no features (e.g. a blue shark). By placing the disfluency at the verb in the utterance, as in (28)-(31), and examining anticipatory eye movements to the four objects, it is possible to identify anticipatory processing, track the time course of

repetition and repair resolution, and monitor the degree to which incomplete disfluency revision persists over time.

(28) Hop- uh ... the yellow frog one square diagonally. (filled pause)

(29) Hop – uh – hop the yellow frog one square diagonally. (repetition)

(30) Move – uh – hop the yellow frog one square diagonally. (A-repair)

(31) Swim – uh – hop the yellow frog one square diagonally. (E-repair)

The following patterns of response should be present if the disfluency is processed incrementally and repair resolution is complete. Filled pause and repetition disfluencies ((28) and (29) respectively) should elicit anticipatory saccades to the target and type match objects (e.g. the frogs) because "hop" is only associated with frogs, and not with sharks. The appropriateness repair (A-repair) disfluency in (30) should initially elicit anticipatory looks to all four objects ("move" is equally likely to be associated with the frog and shark objects), and then only to the target and type match objects when the more specific second verb "hop" is encountered. The error repair (E-repair) in (31) should initially elicit anticipatory looks to the color match and unrelated objects, and then looks to the target and type match objects when the resumption replaces the reparandum. Once the NP identifying the object is encountered, participants should immediately direct their attention to the target object, regardless of disfluency type. Any deviation from this time course would be evidence of difficulties in processing or lingering representations.

The goal of this experiment, then, is to further examine the effects of disfluencies on language processing. Filled pause disfluencies used in previous experiments in this dissertation seemed to increase the already present tendency

for attention to be moved around the display in a manner consistent with the anticipation of upcoming information. The visual world paradigm seems quite sensitive to this sort of effect, and thus it is an ideal technique for studying the online processing of a variety of types of disfluencies. The novel visual world constructed for this experiment, allows filled pause, repetition, and repair disfluencies to be compared in a single experiment under controlled conditions.

Material and Methods

Participants. Sixteen participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or money ($7.00). All participants were native speakers of English, and had normal hearing and corrected to normal or normal vision. No participant was involved in any of the other studies reported in this dissertation.

Materials. Four versions of 20 critical utterances were constructed using five nouns (frog, shark, plane, car, soldier) and five related verbs (hop, swim, fly, drive, march). Repetition, A-repair, and E-repair utterances (Table 12) were recorded and digitized using the Computerized Speech Laboratory (Kay Elemetrics) at 10 kHz, and then converted to wav format. The initial verb in the A-repair and E-repair utterances was then excised and pasted over the initial verb in the repetition utterances to create the final version of the A-repair and E-repair utterances. Thus, all three utterance types were identical following the suspension point of their respective original deliveries. The repetition disfluency was selected as a baseline because repetition disfluencies tend to have a more neutral prosody (many repair disfluencies also have this unmarked prosody;

Levelt & Cutler, 1983). A filled pause disfluency utterance was also created by replacing the second verb in each repetition utterance with a silent pause of the same length (the filler "uh" was retained). Thus, repetition and filled pause disfluencies differed only in whether the second verb was present or was replaced by period of silence of equal duration. Each participant heard only one version (either filled pause, repetition, A-repair, or E-repair; Table 12) of all 20 critical utterances in the course of an experiment.

Table 12. Utterance types used in Experiment V. Segments for analysis are indicated by subscripts in the example utterances.

Utterance Type	Example Utterance (segments for analysis indicated by subscripts)
Filled Pause	/$_{VERB1}$ Hop /$_{UH}$ uh /$_{VERB2}$ /$_{OBJECT}$ the yellow frog /$_{LOCATION}$ one square diagonally./
Repetition	/$_{VERB1}$ Hop /$_{UH}$ uh /$_{VERB2}$ hop /$_{OBJECT}$ the yellow frog /$_{LOCATION}$ one square diagonally./
Appropriateness Repair	/$_{VERB1}$ Move /$_{UH}$ uh /$_{VERB2}$ hop /$_{OBJECT}$ the yellow frog /$_{LOCATION}$ one square diagonally./
Error Repair	/$_{VERB1}$ Swim /$_{UH}$ uh /$_{VERB2}$ hop /$_{OBJECT}$ the yellow frog /$_{LOCATION}$ one square diagonally./

Eighty filler utterances were also recorded and grouped with the 20 critical utterances into blocks of five utterances each. In addition to the six verbs used in the critical utterances (hop, swim, drive, fly, march, move), a seventh verb, "switch", was also used in the construction of fillers. Disfluencies occurred on 50% of fillers, and involved adjectives, nouns, and verbs an equal number of times. Retracing (repetition of part of the reparandum followed by the modification of the remaining proportion reparandum) also occurred on some fillers.

Twenty critical and 20 filler displays corresponding to the critical and filler utterances were created, each consisting of four objects placed on a 5 by 5 grid, as

shown in Figure 43. The five by five grid was constructed out of felt squares attached to a board by magnets. The squares were of a dark color that allowed objects to be better picked up by the scene camera on the head mounted eye tracker. At the beginning of each block of trials, objects were placed at the four central locations indicated in Figure 43. After every two or three trials, objects were then reset to their central positions and a substitution of one or more objects was made as necessary to change the set of objects (i.e. the display) to correspond to upcoming utterances. In addition to the two color, two object type critical displays, filler displays with two, three, four, or no matching colors and objects were included. 10-15° of visual angle separated the objects when they were at the central locations prior to a critical utterance. Each object type (target, color, match, type match, unrelated) occurred in each of the four central positions an equal number of times in the course of an experiment. A new random ordering of trials was created for every fourth participant in this experiment in order to maintain a balance of trials.

Apparatus. The apparatus for this experiment was identical to that of Experiment II.

Procedure. The procedure for this experiment was identical to that of Experiment IIIA. Additionally, participants were told that all of the objects had to be moved in a particular way (for instance, frogs always had to be hopped), and the experimenter demonstrated these movements. Participants then followed practice instructions, and the movements were demonstrated again, if necessary.

Design. Only one display type was used in this experiment, and thus the four utterance types (Table 12) comprised the four unique conditions for this

experiment. Five trials in each condition were presented to each participant, for a total of 20 critical trials. Each utterance occurred in each condition an equal number of times.

Results and Discussion

As in previous experiments, the number of trials with fixations on and saccades to regions of interest (i.e. the four objects) was calculated for each of the segments in Table 12. The frequencies for each segment and region of interest were then analyzed via three planned comparisons using multiway frequency analysis. The first comparison examined differences between filled pause and repetition disfluencies, the second, differences between repair disfluencies and repetition disfluencies, and the last, differences between A-repair and E-repair disfluencies. I will deal with each of these in turn.

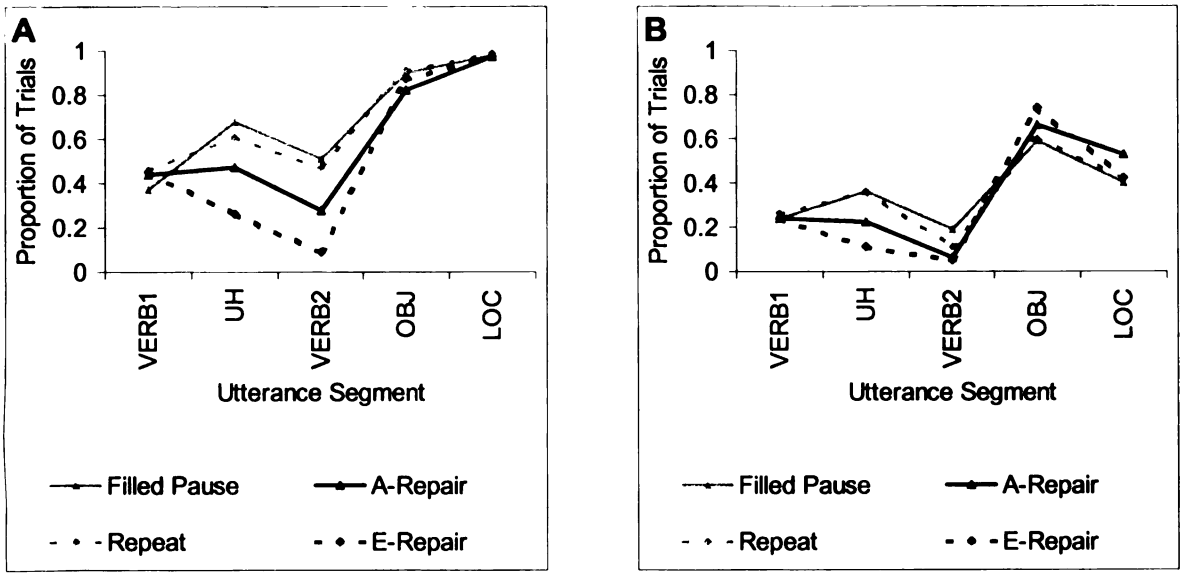


Figure 44. Proportion of trials with a fixation on (A) or saccades to (B) the target object for each segment of the utterance in Experiment V.

I will begin, however, by describing the general patterns of fixations and saccades to each object in the display. Over the course of the utterance, fixations

on the target object generally increased (Figure 44). There was a decrease in fixations on (and saccades to) the target object during the VERB2 segment of the utterance; this may reflect shifts in attention during the disfluent interval in order to anticipate the continuation of the utterance. E-repair and A-repair disfluencies elicited less fixations on the target object than did repetition and filled pause disfluencies during the early portion of the utterance, but fixations on and saccades to the target object increased rapidly during the OBJECT segment in all conditions.

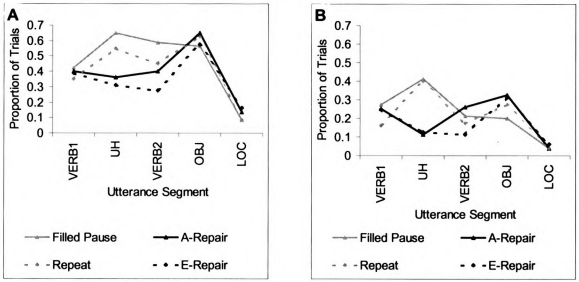


Figure 45. Proportion of trials with a fixation on (A) or saccades to (B) the type match object for each segment of the utterance in Experiment V.

The patterns of fixations on and saccades to the type match object (Figure 45) generally parallels the patterns for the target object up to and including the OBJECT segment. After the OBJECT segment, the probabilities of fixations on and saccades to this region quickly dropped to near zero, as the type match object is quickly rejected as a possible referent of the OBJECT segment because it is the wrong color. The similarity between looks to the target object and type match

object during the first part of the utterance is not surprising given that many of the fixations and saccades during the first part of the utterance are likely anticipatory and being driven by verb information. Thus, the likelihood of fixation on any object is driven by the relationship between the type (but not the color) of the object and the reparandum or resumption verb. This would also predict that looks to the color match and unrelated objects (which are both of the same type) should be similar during the first few analysis segments.

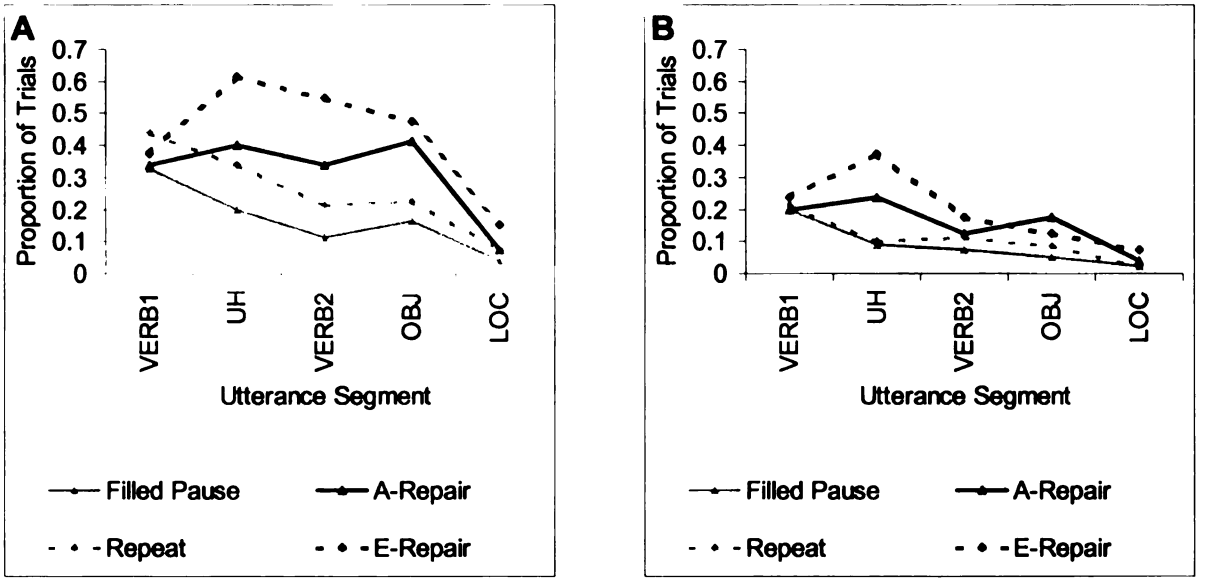


Figure 46. Proportion of trials with a fixation on (A) or saccades to (B) the color match object for each segment of the utterance in Experiment V.

In fact, this is exactly the pattern of data that is present. Fixations on and saccades to both the color match (Figure 46) and unrelated (Figure 47) objects show a higher probability of looks in the A-repair and especially the E-repair disfluency conditions early in the utterances. Recall that E-repair disfluencies contain a reparandum verb that is closely related to the color match and unrelated objects, while the A-repair disfluencies contain a verb ("move") that can refer to any of the objects in the display (and thus every object should be

equally likely to draw anticipatory saccades early on). Fixations on and saccades to the color match and unrelated objects quickly drop towards zero after the OBJECT segment, again likely reflecting reference resolution.

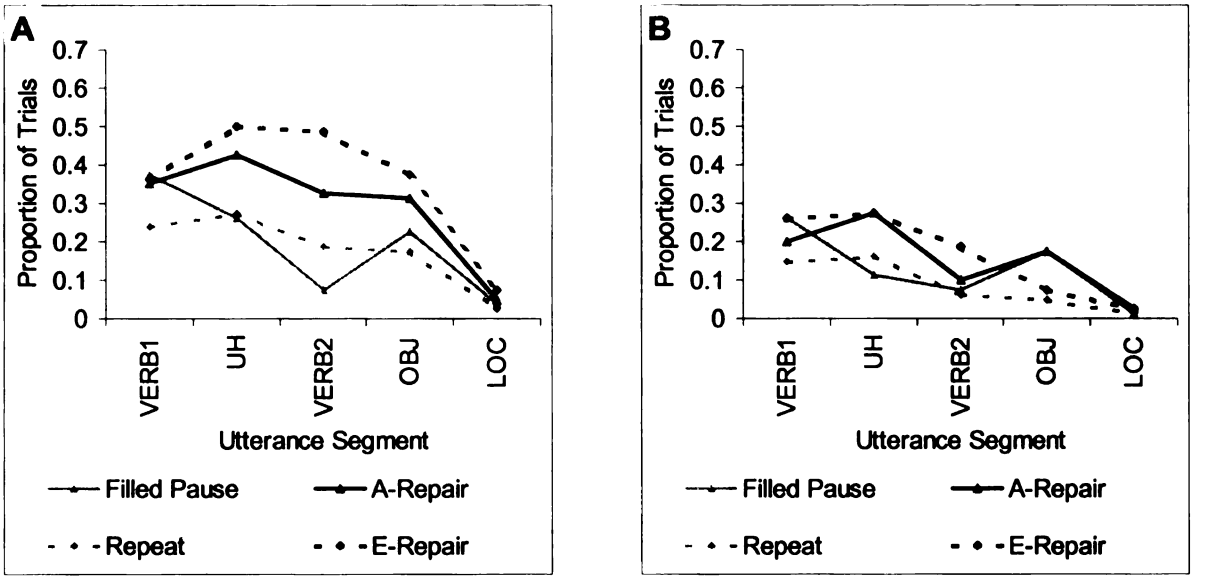


Figure 47. Proportion of trials with a fixation on (A) or saccades to (B) the unrelated object for each segment of the utterance in Experiment V.

The first planned comparison concerned the question of whether repetition disfluencies are as easy to process as filed pause disfluency or whether there is some cost in processing due to the presence of lexical material (i.e. the repeated words) that cannot be incorporated into the syntactic structure being built by the parser. Recall the filled pause disfluency stimuli in this experiment were created by simply replacing the second verb in the corresponding repetition disfluency utterance with a silent pause of equal length, resulting in a long filled pause. Thus, there are no differences in the two utterances, except during the VERB2 segment (Table 12). If repetitions are processed differently than filled pauses, and those differences lead to shifts in attention, then we should expect to see immediate effects in saccades launched during the VERB2 region. A tendency

for attention to remain in a single location, on the other hand, will be realized in perseveration in fixations on a particular region. If either repetitions or filled pauses make attention more likely to continue to be directed to a single location, we would expect to see effects at VERB2 and also possibly during later segments.

The only significant difference in saccades launched between repetition and filled pause disfluency conditions was during the OBJECT segment, where significantly more saccades were launched to the unrelated object during the filled pause disfluency condition than the repetition disfluency condition ($\Delta G^2(1, N = 160) = 7.12$, $p < 0.01$). As the OBJECT segment is the one where the target object is explicitly referenced, it is unlikely that these saccades are due to anticipatory processes. It is possible that this is an instance of a contrast effect which is occasionally seen in displays with competing referents after the correct referent has been identified (Sedivy, et al., 1999; Kamide, Altmann, & Haywood, 2003). The contrast effect consists of looks to a possible contrast object in the display after enough information to identify the referent is present, and is thought to reflect a shift of attention in order to double check the assignment of reference. Why a contrast effect would appear in looks to the unrelated object is not clear. An alternative explanation for these saccades is that reference resolution was more rapid in the filled pause condition because of the longer hiatus in this condition than in any other. Late looks to the unrelated object may thus be simply due to the fact that the eye movement system is free to explore the display in the filled pause condition. No other significant differences in saccades launched were noted (all $p > 0.1$).

Participants were marginally more likely to fixate on the color match object in the repetition disfluency condition ($\Delta G^2(1, \underline{N} = 160) = 3.42$, $p < 0.1$), and to fixate on the type match object in the filled pause condition ($\Delta G^2(1, \underline{N} = 160) = 3.45$, $p < 0.1$). The same differences are not present in the saccade analyses and it seems that these difference are likely due to the continuation of differences prior to the VERB2 segment. That is, saccades seem to be launched from the target and type match objects at about the same rate in both the repetition and filled pause disfluency conditions and so perseveration does not seem to be a likely explanation for these effects. Participants were significantly less likely to fixate the unrelated object during the VERB2 segment ($\Delta G^2(1, \underline{N} = 160) = 5.55$, $p < 0.01$) in the filled pause disfluency condition. This is likely to due the removal of attention from the unrelated object during the silent VERB2 portion of the filled pause disfluencies in anticipation of the speaker's continuation.

There seem, then, to be few differences between the processing of filled pause and repetition disfluencies in this experiment. This supports an account of disfluency processing where repetitions are dealt with easily because they exactly match the structures and representations already built by the language comprehension system. At the same time, the repetition of a lexical item does not seem to confer a direct benefit (for example, earlier reference resolution) either.

The second issue of interest was whether repair disfluencies were processed differently than repetition disfluencies, as the former requires revision of previously generated representations. Both types of disfluency have a reparandum of identical length, but they differ in that repetitions require no modification of the reparandum. These disfluencies do differ in the content of the

reparandum (the VERB1 segment, in this experiment), and thus should show major differences early in the utterance (during the UH segment, for instance). From the UH segment onwards, however, the stimuli were identical, allowing the eye movement record to reveal the time course of disfluency processing. If the replacement of the reparandum with the resumption material is immediate and complete in the same way that matching the reparandum and resumption material appears to be in repetition disfluencies, we should expect to see immediate anticipatory saccades to the target and type match objects, as well as an immediate decrease in fixations on the color match and unrelated objects at VERB2 (i.e. at the resumption). Continued fixation of the color match and unrelated objects might be evidence of delayed or incomplete processing of the repair. These issues were examined by a planned comparison between the two repair conditions (A-repair and E-repair) and the repetition condition.

As was noted earlier, the probability of fixating on and launching a saccade to each object in the display was significantly different in the repetition and repair disfluency conditions during the UH segment. The direction of the difference depended on the reparandum verb (VERB1): More saccades were made to objects that were related to the particular reparandum verb in each condition. Thus, there were more saccades to and fixations on the target object (saccades: $\Delta G^2(1, \underline{N} = 240) = 11.45$, $p < 0.001$; fixations: $\Delta G^2(1, \underline{N} = 240) = 14.25$, $p < 0.001$) and the type match object (saccades: $\Delta G^2(1, \underline{N} = 240) = 10.59$, $p < 0.01$; fixations: $\Delta G^2(1, \underline{N} = 240) = 26.97$, $p < 0.001$) in the repetition disfluency condition. Likewise, there were more saccades to and fixations on the color match object (saccades: $\Delta G^2(1, \underline{N} = 240) = 14.60$, $p < 0.001$; fixations: $\Delta G^2(1, \underline{N} = 160) =$

6.48, $p < 0.05$) and the unrelated object (saccades: $\Delta G^2(1, \underline{N} = 160) = 4.39$, $p < 0.05$; fixations: $\Delta G^2(1, \underline{N} = 160) = 8.70$, $p < 0.01$) in the repair disfluency conditions. These patterns continued through the VERB2 segment for fixations on the target object ($\Delta G^2(1, \underline{N} = 240) = 24.56$, $p < 0.001$), color match object ($\Delta G^2(1, \underline{N} = 240) = 13.58$, $p < 0.001$), and unrelated object ($\Delta G^2(1, \underline{N} = 240) = 12.81$, $p < 0.001$). A marginal difference in the expected direction was seen in looks to the type match object ($\Delta G^2(1, \underline{N} = 240) = 3.08$, $p < 0.1$). During the OBJECT segment, participants continued to be more likely to fixate on the color match ($\Delta G^2(1, \underline{N} = 240) = 11.88$, $p < 0.001$) and unrelated objects ($\Delta G^2(1, \underline{N} = 240) = 8.76$, $p < 0.01$) in the repair disfluency conditions.

A similar pattern was not seen in the analyses of saccades launched to any of the regions after the UH segment (all $p > 0.1$), except for the unrelated object region. Saccades were more likely to be launched to this region in the repair disfluency conditions than the repetition disfluency condition during both the VERB2 ($\Delta G^2(1, \underline{N} = 240) = 4.07$, $p < 0.05$) and OBJECT ($\Delta G^2(1, \underline{N} = 240) = 4.05$, $p < 0.05$) segments. Interestingly, this effect seems to be being driven by E-repair disfluencies at VERB2 and by A-repair disfluencies at the OBJECT segment. As the last planned comparison compared these two types of repair disfluencies directly, I will postpone a discussion of this pattern of data for the time being.

A comparison of repetition and repair disfluencies, then, indicates that attention is shifted to possible target objects at about the same rate in both repetition and repair disfluencies. However, there does seem to be a tendency in repair disfluency conditions for participants to perseverate in their fixations on the objects introduced by the reparandum verb (the color match and unrelated

objects). Whether the perseveration is due to a lingering interpretation or to slower processing is not addressed directly by this experiment, but the lack of difference between repetition and repair conditions in saccades to the target and type match objects at VERB2 suggest that it the former is more likely. However, it is important to note that the effects of lingering interpretations are likely reduced at the very end of the utterance because of the pressure in this task to perform the correct motor response, resulting in shifts of attention related to the performance of that motor task (i.e. to the target object) during the LOCATION segment (thus leading to a lack of significant effects in any analysis during this segment). If both the interpretation based on the reparandum and the interpretation based on the resumption are present in memory, it seems that the latter has priority, even if the former is causing some brief perseveration.

The issue of whether it is difficult to completely revise representations when processing a repair disfluency can also be examined by comparing the two repair disfluency conditions in this experiment. These two types of repair differ in the degree to which the reparandum verb (i.e. VERB1) is related to objects in the display. In A-repair disfluencies, the verb ("move") is generally related to all of the objects in the display. Thus, at the reparandum, no strong commitment can be made to any object or subset of objects, and shifts of attention are likely made to all objects in the display. In the E-repair disfluency, on the other had, the reparandum verb is strongly related to two of the objects in the display, specifically the color match and unrelated objects. Thus, a strong commitment is likely made to these objects, as evidenced by the anticipatory saccades made during the UH segment. Finally, the comparison of repair and repetition

disfluencies indicated that there was some tendency for participants to perseverate in fixating the color match and unrelated objects. If a stronger commitment is made in the E-repair disfluency condition, we might expect greater perseveration in that condition.

Based on a hypothesis that a strong commitment is made to a subset of the objects in the display in the E-repair disfluency condition, but not in the A-repair disfluency condition we should expect to find specific patterns of fixations and saccades. First, we should see more fixations on and saccades to the target type match objects early in the utterance in the A-repair than in the E-repair disfluency conditions as the reparandum verb (VERB1) in the A-repair disfluency condition is not related to any particular object or objects in the display. For the same reason, we should see more saccades to and fixations on the color match and unrelated objects in the E-repair condition. Lastly, we should expect to see a late tendency for looks to be directed to the color match and unrelated objects in the E-repair, but not the A-repair disfluency condition as resolving the E-repair disfluency requires the complete replacement of the representation of reparandum, while the A-repair disfluency requires only a modification of that representation. Responses in the E-repair and A-repair disfluency conditions were therefore examined in a third planned comparison.

The pattern of both fixations and saccades was generally consistent with the reparandum commitment hypothesis described above. During the UH segment, fixations on ($\Delta G^2(1, \underline{N} = 240) = 9.27$, $p < 0.01$) and a saccades to ($\Delta G^2(1, \underline{N} = 240) = 4.28$, $p < 0.05$) the target object occurred more often in the A-repair condition than the E-repair condition. The opposite pattern was seen in fixations

on ($\Delta G^2(1, \underline{N} = 240) = 7.66$, $p < 0.01$) and saccades to ($\Delta G^2(1, \underline{N} = 240) = 3.85$, $p < 0.05$) the color match object. No other effects were present during the UH segment ($p > 0.1$). This pattern continued during the VERB2 segment. A-repairs elicited significantly more fixations on the target object ($\Delta G^2(1, \underline{N} = 240) = 11.05$, $p < 0.001$) and marginally more fixations on the type match object ($\Delta G^2(1, \underline{N} = 240) = 3.05$, $p < 0.1$). E-repairs elicited significantly more fixations on the color match ($\Delta G^2(1, \underline{N} = 240) = 7.92$, $p < 0.01$) and unrelated objects ($\Delta G^2(1, \underline{N} = 240) = 4.97$, $p < 0.05$). Fixations on any object were not any more likely to occur in one repair type than the other in any segment after VERB2.

The pattern of fixation during VERB2 described above was also partially present in saccades launched to particular objects in the display. Participants were significantly more likely to saccade to the type match object during VERB2 in the A-repair disfluency condition ($\Delta G^2(1, \underline{N} = 240) = 6.66$, $p < 0.05$). A marginal effect in the other direction was seen in saccades to the unrelated object during the same segment ($\Delta G^2(1, \underline{N} = 240) = 3.02$, $p < 0.1$). During the OBJECT segment, however, participants were more likely to saccade to the unrelated object in the A-repair disfluency condition, rather than the E-repair condition ($\Delta G^2(1, \underline{N} = 240) = 4.38$, $p < 0.05$). This is the pattern that was briefly alluded to during the discussion of the differences between repair and repetition disfluencies. One possible explanation for this pattern of results is that E-repair disfluencies do result in lingering interpretations to a greater degree than do A-repair disfluencies. As a result, participants continue to make saccades to the unrelated goal through the VERB2 segment in anticipation of the unrelated goal being the correct goal. This does not occur in the A-repair disfluency conditions.

Instead, eye movements due to a contrast effect or exploratory processes result in looks to the incorrect goal. In fact, these saccades may be generated for the same reasons as the saccades to the unrelated object during the OBJECT segment in the filled pause disfluency condition. Indeed, the A-repair and filled pause disfluency conditions are similar in that the both contain only one verb that is semantically related to a particular subset of the objects in the display.

Thus, it appears that E-repairs do result in a greater commitment to a specific subset of the objects in the display than do A-repair disfluencies. While both disfluency types do elicit anticipatory saccades based on verb information, the targets of those saccades differ according to the reparandum verb (VERB1). There is some evidence as well that the strong initial commitment made in the E-repair disfluency conditions is more difficult to revise than the weak commitment made in the A-repair disfluencies. This may be due to either the strength of the initial commitment or the amount of revision required by the type of repair (A-repair likely requires less revision of memory representations than does E-repair).

In summary, then, the results of this experiment indicate that there is some cost associated with the processing of disfluencies. This cost is not related to the mere presence of lexical material in the resumption that must be coordinated with the reparandum, as repetition disfluencies are neither more nor less difficult to process than matched filled pauses. Rather, costs seem to be associated with the revision of representations when the resumption involves a repair. Moreover, the strength of the initial commitment at the reparandum and

the amount of revision involved in completely processing the repair appear to affect the degree to which processing the disfluency is costly.

In this experiment, any cost appeared to be local as participants' eye movements did not differ according to disfluency type by the end of the utterance, and participants were able to easily identify the target object and follow the instruction. However, the task in this experiment precludes drawing any conclusions about whether both the interpretations based on the reparandum and the resumption are present in memory, or if complete revision occurs after the OBJECT segment.

What is clear is that the version of the visual world paradigm developed here is sensitive to differences in disfluency processing at different points throughout the utterance. This is an important finding, as it provides the field of psycholinguistics with a tool for the online examination of the processing of disfluent speech. Thus, the visual world paradigm promises to be a valuable technique for beginning to examine the many unanswered questions about how the language comprehension system deals with disfluent speech.

GENERAL DISCUSSION

The study of disfluent speech presents some clear opportunities for the field of psycholinguistics. By studying the processing of disfluent speech, it may be possible to better understand some of the basic mechanisms involved in language comprehension. Repair disfluencies, for example, may illuminate the process of syntactic and semantic reanalysis. Repetition and filled pause disfluency can also be studied profitably to understand the effects of such factors as the passage of time on the parser and on language comprehension in general. The study of disfluent speech should allow us to develop models of language comprehension that can account for one of the most natural uses of language, spontaneous dialogue.

However, the study of disfluent speech is not without challenges. The greatest challenge is the lack, until recently, of a technique for studying the comprehension of spoken language as it takes place. While offline measures have been quite useful in advancing the understanding of how spoken language comprehension proceeds, there are limits to the conclusions that can be drawn from these studies. Moreover, offline measure cannot be used to describe the time course of disfluency processing. While several techniques exist for studying how comprehension of written texts unfolds over time (eye tracking, self paced reading), disfluencies exist solely in the domain of speech, and thus these techniques cannot be used.

An analogue of the eye tracking techniques used for studying written texts has, however, been recently rediscovered (Tanenhaus, et al., 1995). This technique makes use of eye movements to objects or locations in a visual display

as indices of attention and, by inference, language comprehension. Just as regressions (saccades back to the point of an earlier fixation) can be used to infer shifts of attention to previously processed information in order for reanalysis to occur, looks to the various objects in the display, when interpreted with respect to the concurrent utterance, may be used to infer the time course of spoken language processing.

The focus of this dissertation was the application of this technique, known as the visual world paradigm, to the processing of disfluent speech. Application preceded along two separate, but mutually informative lines. Studies of the processing of fluent speech under different display conditions were used to better understand the nature of eye movements in the visual world paradigm, and to identify the eye movement patterns that were characteristic of language comprehension processes. Complementary studies of the processing of disfluent speech yielded information about how the presence of disfluencies affected these eye movement patterns.

Summary of Results

A variety of eye movement patterns have been described in visual world experiments. Rapid reference resolution has been described in studies where a single referent must be identified on any given trial and is realized as a rapid increase in the probability of fixation on the correct target object or region (Allopena, Magnuson, & Tanenhaus, 1998). Garden path like effects have been described, where the makeup of the display seems to constrain the parse of the utterance (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002). When information early in the utterance restricts the set of possible referents,

anticipatory saccades to those regions have been described (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003).

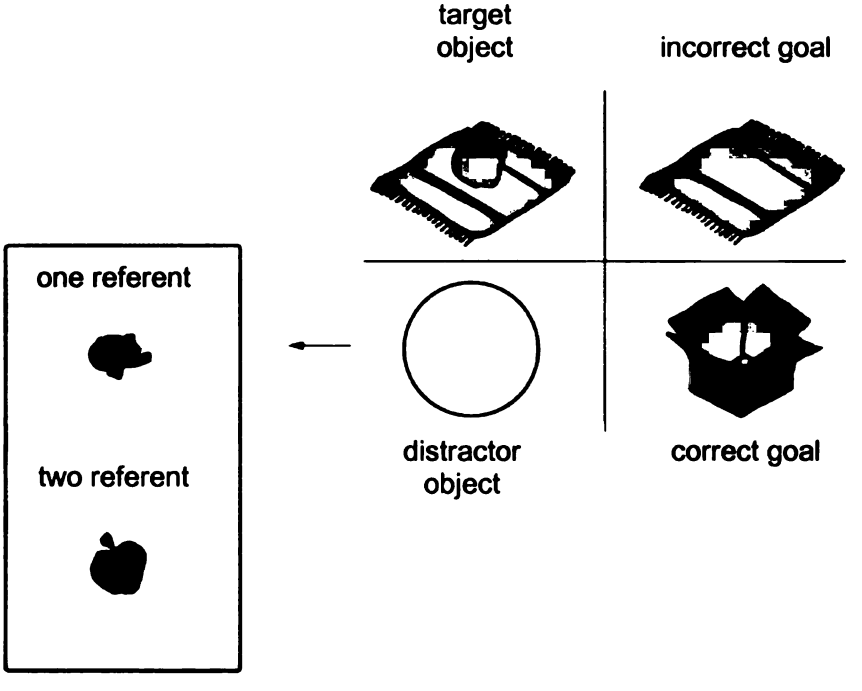


Figure 48. General format of visual world used in fluent utterance experiments.

The visual world experiments in this dissertation that examined fluent utterances found that reference resolution and anticipation were the main determinants of eye movement patterns. Participants in these experiments heard either ambiguous ("put the apple on the towel in the box") or unambiguous ("put the apple that's on the towel in the box") utterances. At the same time they viewed displays that were made up of either one set (two towels) or two sets (two apples and two towels) of identical objects, in addition to several singleton objects (Figure 48). The displays were described as "one referent'" or "two referent" depending on whether the first NP in the utterance had either one or

two possible referents in the display. Participants were instructed prior to the start of the experiment that their task was to follow the instructions they were about to hear. After listening to an utterance, participants executed a motor response.

In previous studies using this visual world (Tanenhaus, et al., 1995; Trueswell, et al., 1999; Spivey, et al., 2002), a garden path like effect was described in the one referent, but not the two referent display. In one referent display conditions, participants were more likely to look at the incorrect goal (the towel by itself) while listening to an ambiguous utterance than when listening to an unambiguous utterance. It was hypothesized that the components of the display were able to affect the syntactic parse of the utterance at an early stage. The two referent display required the ambiguous PP "on the towel" to be a modifier in order to disambiguate the reference of NP1, thus, it could never refer to a possible goal and the tendency to look to the incorrect goal was blocked. In the one referent conditions, however, no modification was necessary (except when an unambiguous utterance was heard and modification was obligatory), and thus the PP "on the towel" could be initially taken as a reference to a goal.

In the fluent utterance experiments conducted here (Experiments II and IIIA-C), no evidence was found for the prevention of a garden path by the two referent display, as there were no differences between ambiguous and unambiguous utterances in any of the experiments. Instead, looks to the incorrect goal were attributed to anticipatory processing and the differences between the one referent and two referent displays were attributed to the length of time necessary to complete reference resolution for the first NP. That is, in the one

referent condition, reference resolution was almost immediate, as there was only one possible referent in the display (i.e. only one apple). In the two referent display, on the other hand, reference resolution had to be delayed until the first PP because there were two possible referents present. Anticipatory saccades to possible goal objects could not thus be launched until reference resolution had been completed.

As supporting evidence for the claim that anticipatory processing was responsible for the looks to the incorrect goal, analyses of the correct goal in each fluent utterance experiment during the same time period found the same pattern of results. The correct goal (a box by itself) had not yet been referenced explicitly in the utterance at this point, either by noun or by preposition. Thus, any looks to the correct goal had to be generated in anticipation of those references being produced by the speaker.

Further evidence for this account of the eye movement patterns in this type of visual world came from the specific manipulations in each experiment. Experiment II examined the effects of task instructions that emphasized that responses to utterances should be carried out as quickly as possible. Time pressure of this sort had few effects on the pattern of results, but did result in participants launching saccades away from an identified referent to possible goals at an earlier point in the utterance. This is a pattern of results that would be predicted if eye movements in this paradigm were being driven by the anticipation of upcoming information once the current reference problem had been resolved.

Experiment IIIA manipulated the display so that the referent of the first NP could be identified either earlier or later in the utterance. This manipulation affected the likelihood of anticipatory saccades, again implicating the time course of upstream reference resolution as a limiting factor on downstream anticipatory saccades.

Experiment IIIB examined eye movement patterns and motor responses to utterances when copresent displays were fully ambiguous in order to ascertain whether or not a garden path was occurring, despite the lack of evidence for a garden path in the eye movement record. Displays were made fully ambiguous by placing the incorrect goal inside (or on top of) another token of the correct goal (e.g. putting the towel that was by itself into another box). Motor responses to the utterances indicated that a garden path was likely taking place on some proportion of trials, as about half of the trials in the only condition that allowed for either interpretation of the ambiguous utterance "put the apple on the towel in the box" resulted in a movement of the target object to the former incorrect goal (e.g. the towel inside a box). The other half of the motor responses were to the former correct goal (e.g. the box by itself). The pattern of eye movements, however, continued to show no evidence of an immediate garden path.

Experiment IIIC tested the hypothesis that anticipatory eye movements were responsible for looks to the incorrect goal by replacing the incorrect goal from earlier experiments (e.g. a towel) with a goal that was never mentioned in the utterance (e.g. a mitten). The same pattern of looks to the incorrect goal in both the ambiguous and unambiguous utterance conditions with a copresent one referent display resulted. These looks could not have been generated by a word

recognition process, as the incorrect goal in this experiment was never mentioned in the utterance; thus, the only catalyst for these eye movements could be anticipation of the incorrect goal as a possible goal, either based on the fact that the verb "put" always has a theme and a goal associated with it, or that the preposition "on" in this context identifies a goal. The presence of the effect in the unambiguous conditions where it is clear that "on" was not identifying a goal suggests that it is the requirement of the verb "put" that is responsible for these effects. Such an explanation is consistent with the anticipatory looks to the correct goal that were present in each fluent utterance experiment.

The widespread reference resolution and anticipatory effects in the eye movement record were used to test hypotheses about the processing of filled pause disfluencies (e.g. "uh"). An earlier study (Bailey & Ferreira, 2003) that used grammaticality judgments of disfluent garden path sentences had described a disfluency cueing effect where the presence of a disfluency seemed to be used by the parser as a cue that a complex syntactic structure was about to be produced. An initial study was conducted in this dissertation (Experiment I) in order to determine whether the location of a filled pause disfluency could affect the interpretation of an ambiguous utterance like "put the [uh] apple on the [uh] towel in the box". In a fully ambiguous display, the location of the disfluency was expected to change the interpretation of the utterance; in a temporarily ambiguous display (i.e. Figure 48), a disfluency was expected to reduce or intensify a garden path, depending on its location in the utterance. No effects of disfluency on the interpretation of the utterance were found in this initial study, however. Rather, it appeared that the presence of a disfluency simply made

saccades to possible referents of the currently ambiguous reference more likely. This effect was confined to the local temporal region of the disfluency, and decreased as the utterance continued.

A follow up experiment (Experiment IV) which used only ambiguous displays, but which compared the disfluent ambiguous utterances to unambiguous controls that were disambiguated to both possible interpretations of "put the apple on the towel in the box", found a similar pattern of results. Motor responses indicated that there were no effects of the location of a filled pause disfluency on the final interpretation of the utterance, although evidence of garden path effects were present. However, no garden path like effects were present in the eye movement record. Instead, participants were again more likely to launch a saccade to a possible referent of the currently ambiguous reference in the temporal vicinity of a disfluency. By comparing disfluent ambiguous utterances to fluent unambiguous utterances in this experiment, it was also possible to see that the presence of a nearby disfluency also increased the likelihood of an anticipatory saccade once the referent of the first NP had been established. Filled pause disfluencies in the visual world paradigm, then, seem to have been interpreted not as planning difficulty related to upcoming syntax, but planning difficulty related to describing objects in the world. Thus, it appears that the language comprehension system is sensitive to the possible causes of disfluent speech; in an experiment where there is no supporting context for a disfluent utterance and no evidence is available to identify possible lexical items that could continue the utterance (e.g. the type of study conducted by Bailey & Ferreira, 2003), the parser may be more likely to identify possible syntactic

structures that can continue the current utterance fragment. In the visual world, however, possible lexical items that can continue the utterance are easy to identify as the visual world has only a limited number of components. Thus, the parser may attempt to identify the specific lexical items, rather than specific structures that can continue the utterance.

The remaining disfluent utterance experiment (Experiment V) took advantage of the anticipatory processes present in both the visual world paradigm and in disfluent speech comprehension in order to examine the time course of processing various types of disfluencies. Filled pause disfluencies, repetition disfluencies, and two types of repair disfluency were compared using eye movements in a novel visual world. In this visual world, there were four objects, one of which was to be moved to a different location according to the concurrent utterance. The objects differed along two dimensions, type and color. Each object type had an associated mode of movement. The disfluencies in this experiment involved the initial verb in the utterance that described the mode of movement (e.g. "hop - uh – the yellow frog one square diagonally"). Anticipatory saccades were made immediately upon hearing the first verb in the utterance. There were few differences between filled pause (e.g. "hop – uh – the frog...") and repetition disfluencies (e.g. "hop – uh – hop the frog..."), suggesting that the mere presence of a lexical item that cannot be inserted into the syntactic structure currently being built is not costly to the parser.

Repair disfluencies, on the other hand, did elicit different eye movement patterns than did repetition disfluencies. Repairs in this experiment were of two types. Appropriateness, or A-repairs, involved the replacement of a more vague

verb with a verb denoting a specific movement (e.g. "move – uh – hop the frog..."). Error, or E-repairs, on the other hand, involved the replacement of a specific verb with a completely different specific verb (e.g. "swim – uh – hop the frog..."). The first verb in the utterance, of course, elicited anticipatory saccades to different subsets of objects in the repetition disfluency condition than the repair disfluency conditions, as the initial verb was different in each condition and was thus associated with different objects in the display (two of the objects in each display were related to the initial verb in the repetition disfluency condition, and two to the initial verb in the E-repair disfluency condition). However, the pattern of eye movements in the repair and repetition conditions remained significantly different through the segment of the utterance that explicitly identified the target object. Thus, the representation of the utterance was not revised immediately after the disfluent interruption upon encountering the second verb.

The two different types of repair disfluency were also found to elicit different eye movement patterns. As expected, the initial verb in the E-repair disfluencies resulted in a stronger commitment to a subset of objects (the verb in the A-repair disfluencies did not pick out a subset of objects) and, as a result, more perseveration in looks to incorrect objects occurred in E-repair disfluency conditions.

Eye Movements in the Visual World Paradigm

In the first chapter of this dissertation, I outlined a number of assumptions about eye movement control that were necessary in order to be able to interpret the patterns of looks to objects in the visual world paradigm.

Specifically, it was necessary to assume that the language comprehension system is somehow linked to eye movement control. In models of eye movement control in reading (Reichle et al., 1998; Reichle, Rayner, & Pollatsek, 1999; Henderson & Ferreira, 1990; Henderson & Ferreira, 1993; Henderson, 1992), attention has been implicated as a linking mechanism between language and eye movement control. The general principles underlying this account of eye movement control have been adapted here as a starting point for developing a model of eye movement control in the visual world paradigm.

At the beginning of any given fixation, the eye is assumed to be fixating the center of attention. As the internal model of the visual world is searched for possible entities that are referred to by the concurrent utterance, markers that tie those entities to locations in the visual world are activated and attention is directed to those locations (Altmann, 2004). When attention is shifted, motor planning for a corresponding eye movement is initiated. If that plan is not cancelled before a certain point (because of another rapid shift in attention, for instance), a saccade is launched, moving the center of fixation to the location currently being attended. This behavior seems to take place even when the corresponding visual world is no longer present (Altmann, 2004; Richardson and Spivey, 2000), and so it seems as if a model of eye movement control will need to account for many cognitive systems, including memory, attention, visual cognition, and language comprehension.

It was also assumed that the cost of eye movements in the visual world paradigm needed to be kept relatively low (Ballard, Hayhoe, & Pelz; 1995) by reducing the visual angle between objects in the display. This assumption was

confirmed to some small degree by the experiments reported here, as more eye movements were made in response to utterances than had been reported in past studies, and as a result immediate effects of the processing of those utterances were present in the eye movement record. As the visual world paradigm has been promoted as a method for studying the time course of spoken language processing, this is an important observation.

I also assumed that the processing of reference in the visual world paradigm was serial because of the serial nature of eye movements. This was supported to a large degree by the experiments reported here. Looks to possible goals in the fluent utterance experiments, for instance, did not occur until after the target had been identified and fixated. This is not to say that searches of the internal representation of the visual world cannot take place in parallel; however, whatever link exists between the internal search and the shift of attention in the external visual world appears to deal with reference in a serial manner.

Finally, the methods of data analysis described in Chapter 2 aided the observation of eye movement patterns by tying the eye movement record to the concurrent utterance throughout each trial. Such standardization has been lacking in the field, but reference to a model of eye movement control that links eye movements to language comprehension through memory and attention makes clear the need for a corresponding link when analyzing eye movement patterns to infer language comprehension processes.

Disfluencies and Language Comprehension

Based on the results of the disfluency experiments presented in this dissertation, it is necessary revisit some of the models that have been proposed to

explain how disfluencies can affect comprehension. Of course, few models exist that make predictions about where, when, and how disfluencies should affect comprehension, as the majority of studies to date have simply described the effects of disfluencies on some facet of language processing.

Bailey & Ferreira (2003) proposed two possible mechanisms by which disfluencies could affect language comprehension. The first, passage of time, comes into play because the disfluent interruption often delays the onset of later material in the utterance. Bailey & Ferreira (2003) proposed that during this delay previously constructed syntactic structure becomes more resistant to revision. Thus, if a disfluency occurs following an ambiguous phrase but before the disambiguating word, and if the ambiguous phrase is initially analyzed incorrectly, the presence of a disfluency will make reanalysis more difficult.

The second mechanism by which disfluencies can effect language comprehension that was proposed by Bailey and Ferreira (2003) was the cueing of upcoming structure. Disfluencies have a particular distribution with respect to the onset of complex constituents in an utterance. Specifically, disfluencies tend to cluster around the first word (usually a function word) in an utterance. Thus, the presence of a disfluency could be taken by the parser as evidence that a complex constituent is about to be produced. However, the results of the disfluent utterance experiments in this dissertation that examined the processing of ambiguous utterances such as "put the apple on the towel in the box" found no such cueing effects. Rather, disfluencies seem to cue listeners to the presence of an unresolved referential ambiguity or to verb requirements that had not yet been filled.

If the cueing hypothesis is modified such that a disfluent interruption serves as a cue for the listener to try to anticipate the speaker's continuation of the disfluent utterance, we can account for both of these effects. As I argued earlier, experiments where syntactic cueing was present differ from experiments where referential cueing was present in that the former lacks a well defined and copresent context. Thus, it is difficult to anticipate the particular lexical item or items that the speaker will produce when fluent delivery resumes. Instead, a more general anticipation of a syntactic constituent may be made. When a well defined context is available, however, as in the visual world, a prediction about the next lexical item is possible and relatively easy to make. Thus, referential cueing trumps syntactic cueing in studies that use the visual world paradigm.

This account also requires the reinterpretation of the mechanism by which the passage of time affects the comprehension of disfluent speech. Instead of the "firming up" of an already constructed syntactic parse, the passage of time allows for anticipations to be made. Thus the anticipation hypothesis ties the passage of time and cueing into a single, more parsimonious explanation. It also makes the unique prediction that the length of the hiatus component of a disfluency should affect the degree of anticipation of the speaker's continuation and further predicts that language production processes will come online when a disfluent interruption is encountered. Because anticipations must be made based on some evidence, the amount of evidence for or against particular structures or lexical items should affect the probability of that structure or item being anticipated. Finally, this account predicts that repair disfluencies will be easier to process when a short hiatus or some type of prosodic or suspension cues are present.

This account is also generally consistent with a model of disfluency processing that is based on a formal grammar (tree adjoining grammar; Ferreira, Lau, & Bailey, in press; Ferreira & Bailey, 2004). This model proposes a novel operation, Overlay, to account for repetition and repair disfluencies. Overlay operates by identifying matching root nodes in the current syntactic tree and the disfluent resumption and then overlaying the resumption material at the corresponding site in the syntactic tree. In repetition disfluencies, this process should be relatively easy because of the identity relationship between reparandum and resumption. In repair disfluencies, on the other hand, this may be more difficult and the reparandum may "show through" the overlaid resumption material. The results of Experiment V suggest that this is true not only for syntactic information (Ferreira, Lau, & Bailey, in press; Ferreira & Bailey, 2004), but also semantic information. Moreover, the reparandum is more likely to "show through" when there is a greater difference between the reparandum and the repair (in E-repair as compared A-repair, for instance).

The anticipation hypothesis and Overlay can account for a wide variety of effects of disfluency processing that have been reported elsewhere in the literature. The relative ease of repetition processing (Fox Tree, 1995; Brennan & Schober, 2001) was replicated by Experiment V and is clearly predicted. In addition, the apparently inconsistent cost (Fox Tree, 1995) and benefit (Brennan & Schober, 2001) account of processing repair disfluencies are also predicted by the anticipation hypothesis. In studies where there is no supporting context and utterances are unrelated, anticipation of upcoming material is difficult; repair disfluencies are costly when the resumption material cannot be predicted. On the

other hand, when a very circumscribed context is present, anticipation of the resumption is simple and repair disfluencies appear beneficial in the long term because of the delay during the hiatus. Thus, it is not necessary to propose that the language comprehension system is constantly monitoring for errors and error signals in order to predict a benefit in processing; rather, it is sufficient to assume that the language comprehension system constantly attempts to predict and anticipate upcoming structure and lexical items. A low level of anticipation may be present throughout the processing of fluent utterances, but a disfluent interruption allows time for the anticipation processes to progress further.

The anticipation hypothesis may also be able to account for given-new effects (Arnold, Fagnano, & Tanenhaus, 2003). In a given-new effects, listeners may anticipate that the speaker is having difficulty retrieving a lexical item corresponding to an as yet unmentioned entity. Because this effect was reported in the presence of a visual world consisting of only a few objects, it would have been easy for the listener to parse the display into given and new items.

A combination of the anticipation hypothesis and Overlay, then, may be a framework for accounting for those effects of disfluencies on language comprehension that have been previously reported.

Future Directions

The results of the experiments presented in this dissertation form the basis for a wide ranging program of future research. This program focuses on two main objectives. The first is the further development of a model of eye movement control in the visual world paradigm. If the visual world paradigm is to be used to better understand the processes involved in spoken language comprehension

and especially disfluency processing, it is essential to have a complete understanding of the link between language comprehension and eye movements. Carefully controlled visual displays and utterances must be combined into well designed experiments in order to continue to identify the characteristic eye movement patterns that reveal language comprehension processes. A series of studies using techniques such as differences in the onset of displays (Altmann, 2004), change blindness (Henderson & Hollingworth, 1999), and manipulation of parafoveal preview (using the same logic as was used in studies of eye movements and reading; Rayner, 1998) would shed light on some of the issues raised earlier in this dissertation.

The results of such experiments would allow for the creation of visual worlds in service of the second focus, the processing of disfluent speech. Several hypotheses and predictions about the effects of disfluencies on language comprehension can immediately be tested. For instance, the amount of delay necessary during the hiatus of a disfluency in order to elicit anticipation effects should be determined. Studies using the novel visual world created for Experiment V might also examine the effects of disfluencies that involve other syntactic categories, for example, nouns or adjectives. The relationship between repair disfluency processing and garden path reanalysis also needs to be studied. Likewise, the relationship between repetition disfluencies and the fluent repetition of words is currently unknown. Finally, questions remain as to how exactly the reparandum is identified, for instance, when overlapping reparanda of different sizes exist.

In summary, it is possible to study how disfluency processing unfolds over time. The visual world paradigm is a promising technique for studying the processing of disfluent speech because it seems to be sensitive to many of the mechanisms by which the language comprehension system deals with disfluent speech. As the visual world paradigm is better understood and a model of eye movement control in the visual world is developed this technique will continue to be more and more useful. Clearly, because so many facets of human cognition are involved in this paradigm, the cooperation of researchers from many areas of cognitive science is necessary.

Disfluencies appear to affect language comprehension because they provide opportunities for anticipation of upcoming structure and lexical items and because they require specific operations to deal with interruptions and lexical items that cannot be substituted directly into available structures. As a result of recent research, including the experiments presented in this dissertation, we have hopefully reached a point where viable models of disfluency processing can be developed, predictions can be generated, and a better understanding of human language comprehension during spontaneous dialogue can take place.

REFERENCES

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology, 21*, 93-113.

Abbott, A. & Hrycak, A. (1990). Measuring resemblance in sequence analysis: An optimal matching analysis of musicians' careers. *American Journal of Sociology, 96*, 144-185.

Abbott, A. & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology. *Sociological Methods & Research, 29*, 3-33.

Aldenderfer, M.S. & Blashfield, R.K. (1984). *Cluster Analysis*. Newbury Park, Ca: Sage.

Allopena, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1999). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition, 93*, B79-B87.

Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Altmann, G. T. M. & Kamide, Y. (2004). Now you see it, now you don't: mediating the mapping between language and the visual world. In J. M. Henderson and F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Altmann, G. T. M. & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research, 32(1)*, 25-36.

Arnold, J. E., Altmann, R. J., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluency effects in comprehension: The discourse-new bias. Paper presented at the *44th Annual Meeting of the Psychonomic Society*, Vancouver, B. C., Canada.

Bailey, K. G. D. & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language, 49*, 183-200.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*, 66-80.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), *Oralité et gestualité: Interaction et comportments mulitmodaux dans la communication* (pp. 597-600). Paris: L'Harmattan.

Beattie, G. & Butterworth, B. (1979). Contextual probability and word-frequency as determinants of pauses in spontaneous speech. Language & Speech, 22, 201-221.

Bienvenue, B. & Mauner, G. (2003). Effects of modest exposure to infrequent syntactic structures on sentence comprehension. Poster presented at the 44th annual meeting of the Psychonomic Society. Vancouver, BC, Canada.

Blackmer, E. R. & Mitton, J. L. (1991). Theories of monitoring and the timeing of repairs in spontaneous speech. *Cognition, 39,* 173-194.

Bortfeld, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44,* 123-147.

Brandt, S. A. & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9,* 27-38.

Branigan, H., Lickley, R., & McKelvie, D. (1999). Non-linguistic influences on rates of disfluency in spontaneous speech. In *ICPhS99 - Proceedings of the XIVth International Congress of Phonetic Sciences.* Berkeley, August 1999.

Brennan, S. E. & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44,* 274-296.

Brennan, S. E. & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34,* 383-398.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (in press). Actions and affordances in syntactic ambiguity resolution. *Journal of Memory and Language: Learning, Memory, and Cognition.*

Charniak, E. & Johnson, M. (2001). Edit detection and parsing for transcribed speech. *Proceedings of NAACL.*

Chomsky , N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA : The MIT Press.

Clark, H. H. (1996). *Using Language.* Cambridge, UK: Cambridge University Press.

Clark, H. H. & Clark, E. (1977). *Psychology and language: An introduction to psycholinguistics.* New York: Harcourt Brace.

Clark, H. H. & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84,* 73-111.

Clark, H. H. & Wasow, T. (1999). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201-242.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6(1),* 84-107.

Core, M. G. & Schubert, K. (1999). Speech repairs: A parsing perspective. In Satellite Meeting ICPHS 99, pp. 47-50.

Corley, M., & Hartsuiker, R. J. (2003). Hesitation in speech can... um... help a listener understand. In *Proceedings of the twenty-fifth meeting of the Cognitive Science Society.*

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser'. In D. Dowty, L. Karttunen and A. Zwicky (Eds.) *Natural Language Processing: Psychological, Computational and Theoretical Perspectives* (pp. 320-358). Cambridge: Cambridge University Press.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken word recognition: Evidence from eye movements. *Cognitive Psychology, 37,* 201-242.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507-534.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language, 42,* 465-480.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47,* 292-314.

Elzinga, C. E. (2003). Sequency similiarity: A nonaligning technique. *Sociologial Methods & Research, 32,* 3-29.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2004, March) "But it's already on a towel!": Reconsidering the one-referent visual context. Paper to be presented at the 17th Annual Meeting of the CUNY Conference on Human Sentence Processing, College Park, MD.

Ferreira, F., Anes, M.D., & Horine, M.D. (1996). Exploring the use of prosody during language comprehension using the auditory moving window technique. *Journal of Psycholinguistic Research, 25,* 273-290.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science,* 11, 11-15.

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., Jr., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken language comprehension: Evidence from the auditory moving window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 324-335.

Ferreira, F., & Henderson, J.M. (1991) Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 31,* 725-745.

Ferreira, F., & Henderson, J. M. (1999). Syntactic analysis, thematic processing, and sentence comprehension. In J. D. Fodor and F. Ferreira (Eds.), *Reanalysis in sentence processing.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Ferreira, F. Lau, E. F. & Bailey, K. G. D. (in press). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science.*

Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 798-827). MIT Press, Cambridge, MA.

Fox Tree, J. E. (1995) The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34,* 709-738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition, 29(2),* 320-326.

Fox Tree, J. E. & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62,* 151-167.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language, 40,* 280-295.

Frazier, L. (1987). Processing syntactic structures: evidence from Dutch. *Natural Language & Linguistic Theory, 5,* 519-559.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 4,* 178-210.

Gajewski, D. A., & Henderson, J. M. (in press). Minimal use of working memory in a scene comparison task. *Visual Cognition.*

Gusfield, D. (1997). *Algorithms on strings, trees, and sequences: Computer science and computational biology.* Cambridge: Cambridge University Press.

Hartsuiker, R. J. & Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology, 42,* 113-157.

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14,* 277-288.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and scene perception. In K. Rayner (Ed.) *Eye Movements and Visual Cognition* (pp. 260-283). Springer-Verlag.

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, 3,* 417-429.

Henderson, J. M. & Ferreira, F. (1993). Eye movement control during reading: Fixation measures foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology, 47,* 201 - 221.

Henderson, J. M., & Ferreira, F. (2004). *The interface of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 5,* 438-443.

Hindle, D. (1983). *User manual for Fidditch.* Technical memorandum 7590-142, Naval Research Lab. USA.

Joshi A. K. and Schabes, Y. (1997) Tree-Adjoining Grammars. In G. Rosenberg and A. Salomaa (Eds.), *Handbook of Formal Languages* (pp. 69-123). Springer.

Just, M. A., Carpenter, P. A., and Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111(2),* 228-238.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 113-156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32,* 37-55.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA:MIT Press.

Levelt, W., and Cutler, A. (1984). Prosodic marking in speech repair. *Journal of Semantics, 2,* 205-217.

Lickley, R. J. (1995). Missing disfluencies. In *Proceedings of the International Congress of Phonetic Sciences.* (Vol. 4, pp. 192–195). Stockholm.

Lickley, R. J., & Bard, E. G. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech, 41,* 203–226.

Lickley, R. J., & Bard, E. G. (1996). On not recognizing disfluencies in dialogue. In *Proceedings of the International Conference on Spoken Language Processing.* (pp. 1876–1880). Philadelphia.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676-703.

Maclay, H. & Osgood, C. (1959). Hesitation phenomenon in spontaneous English speech. *Word, 15,* 19-44.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology, 10,* 29-63

Masson v. New Yorker Magazine, Inc. et al., 501 U.S. 496, (1991).

McMurray, B., Tanenhaus, M.K., & Aslin, R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33-B42.

Nakatani, C. H. & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America, 95(3),* 1603-1616.

Oomen, C. C. E., & Postma, A. (2001). Effects of increased speech rate on monitoring and self-repair. *Journal of Psycholinguistic Research, 30,* 163-184.

Oviatt, S. L. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language, 9,* 19-35.

Pynte, J. (1978). The intra-clausal syntactic processing of ambiguous sentences. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 109–127). New York: Wiley.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124(3),* 372--422.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research, 39,* 4403-4411.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movment Control in reading. *Psychological Review, 105,* 125-157.

Rouanet H., Bernard JM, Lecoutre B. (1986) Non-probabilistic statistical inference: A set theoretic approach. *The American Statistician, 40,* 60-65.

Sankoff, D. & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison.* Reading, MA: Addison-Wesley.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57,* 1030-1033.

Schacter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology, 60,* 362- 367.

Sedivy, J.C., Tanenhaus, M. K., Chambers, C., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109-147.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies.* Unpublished Ph.D. thesis, University of California at Berkeley.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language,* 32, 25-38.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002) Effects of visual context in the resolution of temporary syntactic ambiguities in spoken language comprehension. *Cognitive Psychology, 45,* 447-481.

Stolcke, A., Shriberg, E., Bates, R., Ostendorf, M., Hakkani, D., Plauche, M., Tur, G. & Lu, Y. (1998). Automatic detection of sentence boundaries and disfluencies based on recognized words. In R. H. Mannell & J. Robert-Ribes (Eds.), *Proc. ICSLP, vol. 5* (pp. 2247-2250). Sydney: Australian Speech Science and Technology Association.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557-580.

Tanenhaus, M. K., Spivey-Knowlton, M. J. Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

Tanenhaus, M. K., & Trueswell, J.C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication, Second Edition* (Vol. 11, pp. 217-262). San Diego: Academic Press.

Trueswell, J. C. & Tanenhaus, M. K. (in press) *Approaches to processing world-situated language: Bridging the product and action traditions.* Cambridge, MA: MIT Press.

Traxler, M. J., Bybee, M., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology, 50A,* 481-497.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73,* 89-134.

Vokey, J. R. (1997). Collapsing multi-way contingency tables: Simpson's paradox and homogenisation. *Behavior Research Methods, Instruments, & Computers, 29,* 210-215.

Vokey, J. R (2003). Multiway frequency analysis for experimental psychologists. *Canadian Journal of Experimental Psychology, 57,* 257-264.

Wickens, T. D., *Multiway Contingency Tables Analysis for the Social Sciences,* Mahway NJ, Lawrence Erlbaum Associates, Inc., 1989.

Yarbus, A. L. (1967). *Role of eye movements in the visual process.* New York: Plenum Press.


MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 02504 1264