REPRESENTATIONAL COMPLEXITY AND MEMORY RETRIEVAL IN LANGUAGE COMPREHENSION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Philip Hofmeister
January 2008
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Ivan A. Sag) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Thomas Wasow)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Joan Bresnan)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Daniel Casasanto)

Approved for the University Committee on Graduate Studies.
Abstract

How an object or event is described depends upon numerous factors that have been identified by prior research in linguistics, pragmatics, and psychology. What this prior research has not established, however, is what consequences the resulting choice of linguistic form has on subsequent comprehension processes. This dissertation begins with the assertion that the quantity of information encoded in a linguistic expression has an impact on the efficiency of subsequent retrieval processes that target the corresponding mental representation. In particular, I hypothesize that information processed during the encoding phase of comprehension facilitates the retrieval process. Consequently, more informative linguistic expressions are predicted to enhance memory retrieval.

A series of self-paced reading tasks show these predictions to be true in a variety of filler-gap constructions with a variety of phrase-types in English, e.g. argument wh-phrases, temporal adjunct phrases, indefinites, and definites. In all these experiments, reading times are faster when the memory representation associated with a more informative filler-phrase is targeted. Crucially, these “informativity effects” consistently appear at likely retrieval sites. Moreover, in nearly all of the experiments, there is no effect of informativity in the word regions prior to the retrieval site. This fact argues in favor of treating the informativity effects as intrinsically linked to memory retrieval.

The underlying relationship between informativity and retrieval is also explored in depth within this dissertation. That is, I address the question of why greater syntactic and semantic complexity facilitates retrieval. There are multiple, plausible reasons for why informativity attenuates retrievability, including encoding time differences and differences in plausibility. In the end, I conclude that this relationship reflects what happens cognitively during the encoding process and the retrieval process. Specifically, the process of building a more complex syntactic representation requires repeatedly accessing that representation
to modify it. This effectively amounts to increasing the activation associated with the corres-
ponding mental representation. The resulting representational complexity also presents mul-
tiple retrieval paths to offset the difficulty of retrieving any single item in memory. Fur-
thermore, semantic enrichment of a representation causes a memory item to become more
unique in memory, thereby limiting the effect of interference from other items in memory.

The activation of discourse representations, according to this research, fluctuates through-
out discourse as a function of constraints on working memory, as well as the history of ac-
cessing those representations and the quantity of information used to guide those retrievals.
Representation complexity partially offsets the hindrances to sentence processing posed by
decay and interference, because the process of building the complex representation creates
a highly activated memory network and because the ensuing cognitive structure allows for
multiple retrieval strategies.

The implications of this research extend to theories of discourse and reference, but also
to theories of syntax. Controlled acceptability studies demonstrate a sensitivity to
manipulations of informativity in a way that parallels the results of the reading time studies.
Hence, variation in acceptability judgments can be modeled on the basis of processing-based
differences. This type of explanation provides an economical way of accounting for variation
in grammaticality judgments without introducing any new machinery. Combined with prior
theories of reference and models of discourse, the research in this dissertation thus provides
a more comprehensive view of discourse and shows for the first time how linguistic form
can both reflect and attenuate retrievability. By presenting a theory of the consequences of
informational complexity for subsequent language comprehension, this research constitutes
a first step toward understanding how the way we talk about things now impacts how we
talk and think about them later.
Acknowledgements

This isn’t what I expected. Meaning that nearly every point along the way to where I am now was not what I predicted. A couple years ago I didn’t know anything about psycholinguistics, sentence processing, or running experiments. A couple years before that I wasn’t even sure that I would do linguistics for long. I certainly didn’t expect this to be the first book I ever wrote. Nor did I expect this dissertation to be so hard that it would feel like my brain had been torn out and turned inside-out and danced upon by a thousand pissed-off monkeys brandishing flaming truncheons that shoot exploding comets of shrapnel straight into my head. I suppose the point of saying this is partly that I’m not particularly adept at predicting what’s going to happen next, but also that knowing that something is going to happen is not a precondition for being pleased about it afterward. Which is to say, that despite the sensation of having just unexpectedly given birth to a hippopotamus, I’m quite taken with the little beast, so I think I’ll just keep it and see what happens next.

The research in this dissertation grew out of a series of collaborations with my adviser, Ivan Sag. In a very real sense, Ivan has contributed to much of the ideas and analysis that appear in the following pages. His interest in scrutinizing and reassessing empirical data that had been taken for granted for decades sparked my own interest in the matter. At all points during the research leading up to this dissertation and throughout the writing process, Ivan has remained almost stupefyingly attentive, engaged, and supportive, even when it meant sacrificing time for his own work. Beyond his inspiration and contributions, Ivan has also been an unfailing guide and friend in all matters.

Tom Wasow was also a critical part of the support and advice I received during the writing of this dissertation. He always made himself available and listened seriously to

---

1Well, it’s not really the first book I ever wrote, because there was that book I wrote when I was twelve about the human colonists that landed on an alien planet and evolved into a species without eyeballs but who nevertheless had an abiding interest in fencing. But, um, that doesn’t count.
whatever proposal I had, no matter how ridiculous it actually was. Just as importantly, the
constant encouragement that came from Tom was deeply appreciated.

Invaluable advice also came from Daniel Casasanto who was a great ally in deciding
issues of experimental design and who always seemed to see some of the most penetrating
questions to ask and connections to make. As the psychologist on my committee, Daniel
identified relationships and implications for my research that might never have been made
otherwise. Joan Bresnan provided another source of inspiration in spearheading a move-
ment at Stanford University to apply rigorous statistical methods to the understanding of
language data. Many of the ideas formulated here owe their status to a refining process
that truly began in Joan's laboratory syntax seminar.

I also wish to thank Peter Sells for helping me in numerous ways during my time at
Stanford. Peter has the gift of teaching and personal instruction that I have encountered
only very rarely during my entire education. I like to think that some of the clarity he
brings to linguistic issues rubbed off on me and hopefully that shows up in the dissertation.

The empirical research in this dissertation also reflects collaborations with a number
of my fellow students at Stanford University, particularly members of the WH-Research
group. Neal Snider was part of most of the studies on syntactic islands and was a constant
font of knowledge on statistical matters and natural language corpora, and endured hearing
about my research way too many times for his own good. Additionally, Inbal Arnon, Bruno
Estigarribia, and Jeanette Pettibone all played a significant role in helping launch this
research and setting its direction. A special debt of gratitude is due to Florian Jaeger, who
besides being one of the real moving forces behind the earliest studies on wh-dependency
processing at Stanford, was the source of some of the most useful and insightful comments
with regard to my research. Many of the statistical methods I have become familiar with
were learned (or re-learned) via Florian and my knowledge of language experimentation
reflects years of collaboration with and advice from Florian.

Presenting this material at a number of conferences and universities further helped me
to refine and develop the ideas in this dissertation. Audiences at the Linguistic Society
of America in Anaheim and Stanford, the Linguistic Evidence conference in Tübingen, the
Chicago Linguistics Society, the Brain and Cognitive Sciences Department at the University
of Rochester, and the Department of Linguistics at the University of California-San Diego all
provided critical feedback that has improved this dissertation. Special thanks go to Robert
Kluender, Marta Kutas, Hannah Rohde, Jeff Runner, Mike Tanenhaus, Ev Fedorenko and
Ted Gibson for their expertise and advice.

I can’t really express here how much I will miss Stanford. A lot of that relates to the people who were around me and listened to me wax on about my research, especially Andrew Koontz-Garboden, Judith Tonhauser, Laura Staum, Hal Tily, Casey Lew-Williams, and Scott Grimm. Most important of all, Lis Norcliffe rescued me many times from the worst periods of the dissertation process. Besides listening patiently on an almost daily basis about the dreary and tedious details of writing, Lis took me away from linguistics just when I needed to be taken away and repeatedly renewed my energy for this project with her own passion for language and life. Her presence made this a place I wanted to be. The sad thing is that she has effectively ruined other people for me, because there’s no one else quite like her.

Finally, where I am now is truly due to Jolie who has seen me through every step of my education. She has endured all the swings of energy and attention and stuck with me all the same for over eleven years. My mind reels at the thought of what I would be doing if I had not met you. Saying ‘thank you’ seems so incommensurate as to be almost offensive, so I’ll just say that it’s going to take some time for me to make it up to you.
# Contents

Abstract iv

Acknowledgements vi

1 Introduction 1
  1.1 Memory Retrieval in Language Comprehension ...................... 1
  1.2 Outline of Dissertation .............................................. 13

2 Motivating the Memory Facilitation Hypothesis 15
  2.1 Acceptability Paradigms .............................................. 16
    2.1.1 A processing account of Superiority-violations ............... 24
  2.2 Activation Enhancement ............................................... 29
  2.3 The Elaboration Model of Recall .................................... 34
  2.4 Summary ................................................................. 39

3 Processing Islands 41
  3.1 Gradience in Islands .................................................. 41
    3.1.1 Processing Perspectives on Island Effects ................... 51
  3.2 Methodology ............................................................. 58
    3.2.1 Removal of statistical outliers .............................. 60
    3.2.2 Interpreting reading time results ............................ 62
  3.3 Experiment I: Complex Noun Phrases ................................ 63
    3.3.1 Materials .......................................................... 65
    3.3.2 Participants ....................................................... 66
    3.3.3 Results ............................................................ 66
    3.3.4 Discussion ......................................................... 70
# List of Tables

2.1 DP classification from Cinque (1990); adapted from Chung (1994) . . . . . 18
2.2 Example of stimuli from Bradshaw & Anderson (1982) . . . . . . . . . . . . 36
3.1 Effect of informativity by region inside embedded clause in experiment I . . 67
3.2 Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment II. Each measure is followed by its mean standard error. . . . . . 80
3.3 Effect of informativity by region inside embedded clause in experiment II. . 81
4.1 Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment IV. Each measure is followed by its mean standard error. . . . . 106
4.2 Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment V. Each measure is followed by its mean standard error. . . . . 118
5.1 Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment VI. Each measure is followed by its mean standard error. . . . . 134
5.2 Effect of informativity by region inside embedded clause in experiment VI . 135
5.3 Effect of informativity by region inside embedded clause in experiment VII. 139
5.4 Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment VII. Each measure is followed by its mean standard error. . . . . 140
5.5 Corpus results from Gigaword for typ & atyp conditions from experiment VIII: mean frequencies of adjectives, the mean conditional probability of the adjective preceding the head noun \(P(w|w + 1)\) and the mean conditional probability of the head noun following the adjective \(P(w|w - 1)\). . . . . . . . . . . . 148
5.6 Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment VIII. Each measure is followed by its mean standard error. . . . 156
B.1 MinF’ values by region in experiment I ........................................... 236
B.2 MinF’ values by region inside embedded clause in experiment II for contrast between WHICH and BARE conditions .................................................. 236
B.3 Bonferroni corrected values by region in experiment II ...................... 237
B.4 Bonferroni corrected values by region inside embedded clause in experiment III .......................................................... 237
B.5 MinF’ values by region inside embedded clause in experiment III ........ 237
B.6 Bonferroni corrected values by region in experiment IV .................... 238
B.7 MinF’ values by region in experiment IV ............................................ 238
B.8 MinF’ values at retrieval site in experiment V .................................... 239
B.9 Bonferroni corrected values at retrieval site in experiment V ............ 239
B.10 MinF’ values by region inside embedded clause in experiment VI ...... 239
B.11 Bonferroni corrected values by region inside embedded clause in experiment VII .......................................................... 240
B.12 MinF’ values by region inside embedded clause in experiment VII ...... 241
B.13 MinF’ values by region inside embedded clause in experiment VIII ..... 241
B.14 Bonferroni corrected values by region inside embedded clause in experiment VIII .......................................................... 242

C.1 Model comparison of reading times at first six words of embedded clause in experiment I. Model A: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + \beta_3\text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + (\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$ .................................................. 243
C.2 LME analysis of reading times for first six words of embedded clause in experiment I .......................................................... 243
C.3 Model comparison of reading times at verb + three-word spillover region in experiment II. Model A: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + \beta_3\text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + (\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$ .................................................. 243
C.4 LME analysis of reading times at verb + spillover region in experiment II . 244
C.5 Model comparison of reading times at first word in embedded clause in experiment III. Model A: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{BASELINE}_i + \beta_2 \text{LONG}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \]
Model B: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{BASELINE}_i + \beta_2 \text{LONG}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \]

C.6 LME analysis of reading times at first word in embedded clause in experiment III

C.7 Model comparison of reading times at retrieval site in experiment IV. Model A: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{WHICH}_i + \beta_2 \text{COMPLEX}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \]
Model B: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{WHICH}_i + \beta_2 \text{COMPLEX}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \]

C.8 LME analysis of reading times at verb + next word in experiment IV

C.9 Model comparison of reading times at retrieval site in experiment V. Model A: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{MID}_i + \beta_2 \text{SIMPLE}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \]
Model B: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{MID}_i + \beta_2 \text{SIMPLE}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \]

C.10 LME analysis of reading times at verb in experiment V

C.11 Model comparison of reading times at retrieval site in experiment VI. Model A: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{TYPE}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \]
Model B: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{TYPE}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \]

C.12 LME analysis of reading times at verb + next word in experiment VI

C.13 Model comparison of reading times in regions 8 through 11 in experiment VII. Model A: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{TYPE}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \]
Model B: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{TYPE}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \]

C.14 LME model of reading times at regions 8-11 in experiment VII

C.15 LME analysis of reading times at regions 9-11 in experiment VII

C.16 Model comparison of reading times at retrieval site in experiment VIII. Model A: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{ATYP}_i + \beta_2 \text{TYP}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \]
Model B: 
\[ y_{ij} = \mu + b_i + \beta_1 \text{ATYP}_i + \beta_2 \text{TYP}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \]

C.17 LME model of reading times at verb in Experiment VIII

C.18 LME analysis of reading times at first word after verb in experiment VIII

C.19 LME analysis of reading times at verb + next word in experiment VIII
# List of Figures

1.1 Schematization of balance between the discourse function and processing cost of information (horizontal dimension), and the “pay now-pay later” trade-off (vertical dimension) ................................................................. 10

2.1 Log-normalized acceptability ratings of three kinds of Superiority-violations, varying by type of fronted *wh*-phrase, reported in Hofmeister et al. (2007) . 26

2.2 Mean residual reading times at verb plus spillover region in SUVs, based on results reported in Hofmeister et al. (2007) ................................. 27

3.1 Residual reading times at complementizer in experiment I ...................... 68

3.2 Residual reading times at auxiliary + verb in experiment I ..................... 69

3.3 Mean normalized acceptability ratings of complex noun phrase constructions 72

3.4 Mean normalized acceptability ratings of complex noun phrase constructions 72

3.5 Residual reading times in experiment I, ranging from first word after *wh*-phrase to three words after retrieval site. Error bars show one standard error. 78

3.6 Acceptability ratings of embedded *wh*-island violations: **Bare** = bare *wh*-phrase; **Which** = *which*-N’ phrase ....................................................... 83

3.7 Residual reading times in experiment III, ranging from first word after *wh*-phrase one word after retrieval site. Error bars show one standard error... 88

3.8 Residual reading times at first word after complementizer in experiment III 89

3.9 Residual reading times for the embedded clause in experiment III ......... 90

3.10 Mean normalized acceptability ratings of adverbial island violations ....... 95

3.11 Mean normalized acceptability ratings of subject condition violations .... 96

3.12 Mean normalized acceptability ratings of FGDs without island violations . 97

4.1 Residual reading times at subcategorizing verb in experiment IV ........... 107
4.2 Residual reading times at verb plus next word (e.g. *discussed*) in experiment IV ................................................................. 108
4.3 Aggregated residual reading times for two words after *wh*-phrase (region 1) contrasted with reading times at the retrieval site and subsequent word (region 2) ............................................................... 109
4.4 Histogram of reading times prior to removal of outliers ................................................................. 113
4.5 Histogram of reading times after removal of outliers ................................................................. 114
4.6 Reading time as a function of item sequence for subjects in experiment IV. ................................................................. 116
4.7 Residual reading times at subcategorizing verb in experiment V ................................................................. 119
4.8 Residual reading times in experiment V, beginning with the first word after the indefinite. ................................................................. 120
4.9 Residual reading times at first word after the indefinite NP (region 1) contrasted with reading times at the verb (region 7) in experiment V ................................................................. 121

5.1 Residual reading times at first word after *wh*-phrase in experiment VI ................................................................. 132
5.2 Residual reading times, word by word in experiment VI, ranging from the first word after the *wh*-phrase to the penultimate word in the sentence. ................................................................. 133
5.3 Residual reading times at matrix verb (e.g. *ordered*) in experiment VI ................................................................. 133
5.4 Mean residual reading times for regions 9 to 11 in experiment VII ................................................................. 141
5.5 Residual reading times at verb and four-word spillover region in experiment VII ................................................................. 143
5.6 Mean residual reading times at the first word after the *wh*-phrase (region 1) compared with aggregated residual reading times for post-retrieval spillover region (region 2) in experiment VII ................................................................. 144
5.7 Mean z-scores of likelihood ratings per item from norming study ................................................................. 149
5.8 Residual reading times, word by word in experiment VIII, ranging from the first word after the object NP to the word after the embedded verb ................................................................. 151
5.9 Mean residual reading times at verb in experiment VIII ................................................................. 153
5.10 Residual reading times as a function of item sequence in experiment VIII ................................................................. 154
5.11 Mean recognition accuracy of verb-noun pair in experiment IX ................................................................. 161

6.1 Relationship between anaphoric form choices and three hypothetical activation profiles ................................................................. 171
6.2 Visualization of consequences of anaphoric form choices on subsequent activation ........................................... 172
6.3 Example chunk of intransitive verb *dozed* in declarative memory ........................................... 182
Chapter 1

Introduction

...when once you’ve said a thing, that fixes it, and you must take the consequences.—The Red Queen in Through the Looking Glass

1.1 Memory Retrieval in Language Comprehension

The same object can be described at any given moment in many different ways, perhaps even infinitely many ways. This document, for instance, can be referred to legitimately as this document, this dissertation, this, it, the dissertation that you’re reading, my dissertation, my doctoral dissertation, my first and last doctoral dissertation, this Stanford University doctoral dissertation, etc. The choice of how to refer to an object depends upon a number of factors, including the context or setting in which it is used, whether it has already been mentioned before, and what kinds of other objects have recently been discussed. Some descriptions (“anaphors”) refer back to other descriptions (“antecedents”) previously introduced in the discourse, e.g. the previous sentence contains two anaphoric references (“it”) that refer back to an antecedent (“an object”). A large amount of work in linguistics, psycholinguistics, and psychology has been devoted to defining the conditions under which one anaphoric form is chosen over another (Clark & Wilkes-Gibbs, 1986; Sperber & Wilson, 1986; Mangold & Pobel, 1988; Gernsbacher, 1989; Ariel, 1990; Gordon, Grosz, & Gilliom, 1993; Gundel, Hedberg, & Zacharski, 1993; Grosz, Weinstein, & Joshi, 1995; Brennan & Clark, 1996; Arnold, 1998; Almor, 1999; Ariel, 2001; Almor & Nair, 2007; Arnold & Griffin, 2007). This research generally aims to look at the circumstances preceding a linguistic form choice and determine from that information why a particular referring form is chosen.
Basically, this work inspects discourse features prior to a linguistic form choice and what drives speakers to make these choices.

In this dissertation, I look at what might be thought of as the mirror image. That is, the focus of this dissertation is what effect choices of linguistic form have on subsequent language comprehension. In particular, I set out here to find if the choice of referring form for a particular discourse entity has an effect on the comprehension of subsequent references to that very same entity. This amounts to asking whether the information contained in a linguistic reference affects how that reference is re-accessed later. The answer that this dissertation supplies is, quite simply, yes. More specifically, this dissertation presents evidence that the information encoded in a linguistic description has significant consequences for how swiftly and accurately the corresponding mental representation can subsequently be retrieved from memory.

Memory is an essential component of language comprehension. Retrieval of stored linguistic information from memory is necessary to do things like understand references to previously mentioned individuals or concepts, to follow the topic thread of a discourse, to answer a question, etc. In fact, memory retrieval is likely happening at nearly every word you see or hear. To ensure successful comprehension, each word in an utterance has to be integrated with the words and phrases that came before it. As automatic and effortless as this process might seem, this integration process cannot happen without a preceding retrieval process. Take, for instance, the following sentence of English:

(1) There was a meteor shower last night that I hoped that you got a chance to see before it disappeared.

To properly understand this sentence, the phrase a meteor shower must be interpreted as the object of see, despite the fact that this object NP appears eleven words before its verbal head. The linguistic relationship between these two sentence constituents is hence typically referred to as a long-distance dependency and, more particularly, a filler-gap dependency (Fodor, 1978). Filler-gap dependencies link two phrasal constituents, one of which (the filler) is displaced from its canonical linear position, leaving an empty structural gap. Hence, the displaced noun phrase a meteor shower must be maintained in memory until it can be associated with the object position of the verb see. Evidence that the information in a filler-phrase is retrieved at this final verb comes from a variety of empirical methods, including probe recognition tasks, cross-modal priming, reading time studies, and neurophysiological
techniques (Tanenhaus, Carlson, & Seidenberg, 1985; Nicol & Swinney, 1989; Kluender & Kutas, 1993a; Osterhout & Swinney, 1993; Swinney, Ford, Bresnan, & Frauenfelder, 1988; McElree, 2000). Due to the rapid and incremental nature of language, retrieval from memory storage of the right information at the appropriate time must occur in an efficient and timely manner, in order to prevent compounding processing difficulties. A delay in the successful memory retrieval of the filler in (1), for instance, could add to the cognitive challenge of processing incoming constituents, including the pronominal reference whose interpretation also depends upon retrieval of the information encoded in the filler.

There are, of course, limitations and constraints on human memory. Most researchers agree that language production and comprehension access a short-term, limited-capacity buffer for storage and processing of verbal information, commonly known as working memory (Baddeley & Hitch, 1974; Baddeley, 1992; Miyake & Shah, 1999; Andrade, 2001; Baddeley, 2007). Thus, linguistic information contained within working memory is rapidly lost if overt rehearsal is prevented, and the amount of information that can be successfully recalled is extremely limited.

Ultimately, numerous dynamic and interacting factors bias working memory in language processing (Lewis, Vasishth, & Van Dyke, 2006). For instance, the more time between the initial encoding of some stimulus and the retrieval of that stimulus from memory, the more difficult the retrieval will be (Gibson, 1991, 1998, 2000; Grodner & Gibson, 2005). This means that increasing the distance between some encoded target and the retrieval point impairs the necessary retrieval process. So the further away a word or phrase is from the point where it needs to be re-accessed for comprehension, the more difficult that retrieval from memory is.

1Neurophysiological, neuroimaging, and double dissociation data provide compelling arguments for treating working memory as neurologically distinct from long-term memory (Warrington & Shallice, 1969; Vallar & Baddeley, 1987; Baddeley, Papagno, & Vallar, 1988; Hanley, Young, & Pearson, 1991). While there remains to this day a healthy amount of disagreement about the precise architecture of working memory and whether it contains resources for language processing not shared by other cognitive functions (Caplan & Waters, 1999; Just & Carpenter, 1992; Fedorenko, Gibson, & Rohde, 2006, 2007), there is less debate about the general existence of a short-term memory system whose limited resources place constraints on language processing (but see Ward (2001) for a dissenting opinion). Consequently, I use the term 'working memory' here without a commitment to any specific architecture of working memory.

2Distance here can be measured in distinct ways. Gibson and colleagues alternately measure distance in terms of discourse referents and the number of words between encoding and retrieval. Lewis et al. (2006) also raise the point that the relevant measure may simply be time.
In addition, the number of times some stimulus has been seen or heard also affects the efficiency of the retrieval process (Anderson, Budiu, & Reder, 2001; Lewis & Vasishth, 2005). In this regard, repeated mentions of some discourse entity will enhance retrieval. More plainly, something talked about a lot will be easier to retrieve than something rarely discussed.

Furthermore, the presence of other items in memory that share features with a particular memory target worsens retrieval of the target (similarity-based interference), according to a number of theorists, as measured by reading times at retrieval points and other behavioral measures (Lewis, 1996; Gordon, Hendrick, & Johnson, 2001; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). In theories of cue-based parsing (McElree, Foraker, & Dyer, 2003; Lewis & Vasishth, 2005), these interference effects emerge from the dynamics of memory retrieval in sentence parsing: “Retrieval is accomplished by a simple type of associative access: content-based retrieval, where the retrieval cues are a subset of the features of the item to be retrieved” (Lewis et al., 2006, p. 448). On this account, the mental representation of some discourse entity encoded in memory includes a set of features, including syntactic and semantic information. At a retrieval point (e.g. the verb in (1) above), a set of retrieval cues provided by the linguistic context guides the retrieval process. These retrieval cues, however, may overlap with features of memory representations that are not the true memory target. The extent to which these cues overlap with features of other representations determines the severity of the similarity-based interference and the likelihood of successful retrieval from memory.

This dissertation argues for another type of memory bias operative in language comprehension that concerns the choice of linguistic form. In particular, I present here a memory facilitation hypothesis for language comprehension:

(2) The Memory Facilitation Hypothesis

Given two linguistic expressions that can each be felicitously used to describe some discourse entity $e$, the expression that encodes more syntactic and semantic information will facilitate the retrieval process initiated at all subsequent (overt or covert) mentions of $e$, all else being equal.

This hypothesis thus says that, given two expressions capable of describing the same discourse entity, the more informative of the two will lead to a more efficient retrieval process whenever that discourse information needs to be re-accessed. The relative informativity of
two expressions is determined by the following principle:

(3) An expression \( x_1 \) is more informative than an expression \( x_2 \) if the semantic and syntactic information encoded by \( x_2 \) is a proper subset of the information encoded by \( x_1 \).

This differentiates phrases that are semantically identical (or nearly so) but have different syntactic complexities, e.g. *somewhere* vs. *some place*, *who* vs. *which person*, *which* vs. *which one of them*, *Rome’s destruction* vs. *the destruction of Rome*, etc.

Intuitively speaking, vague descriptions of individuals and events and other discourse objects make recall more difficult because they produce memory representations that are underspecified, and hence non-distinctive and easily confused with other memory representations. Consider, as an example, the two noun phrases *the American novelist Cormac McCarthy* and *Cormac McCarthy* both of which refer to the same individual. Clearly, the first description includes information that the latter does not. According to the Memory Facilitation Hypothesis (MFH), the use of the longer and more informative description should result in faster and more accurate recall, as compared to the less informative proper name.\(^3\) Thus, in the filler-gap dependencies below, the definite description is expected to improve processing at (or shortly after) the most embedded verb (the position of the gap is identified with underscores):

(4) a. It was Cormac McCarthy that many critics felt that the public snubbed \_\_\_\_ for too long.

\(^3\)Precisely what information gets retrieved during comprehension remains an unanswered question. Theoretically, retrieval processes may target phonological form, in addition to syntactic and semantic features. The rhyme-priming results of Tanenhaus et al. (1985) are sometimes perceived as evidence for this position: they found that lexical decision latencies after gap sites were shorter when the filler phrase (e.g. *which beer*) rhymed with the target (e.g. *fear*). As Tanenhaus et al. put it: “Gap filling is thus a case of literal replacement” (p. 388). An alternative take is that only syntactic and semantic information is automatically retrieved during language comprehension, while any phonological priming stems from associative access. The Tanenhaus et al. results, however, can be explained simply on the basis of whether a rhyming word appeared in the sentence at all; furthermore, they themselves warn against strong interpretations of their results and admit that the response time differences were not significant when items were the random factor in the statistical analysis. The experimental evidence that appears throughout this dissertation argues against the idea that phonological form is necessarily retrieved from working memory during comprehension. Under such a hypothesis, longer phonological forms ought to result in a slower retrieval process, but the evidence accumulated here shows the exact opposite. This issue is raised again in the sixth chapter, where I consider models of retrieval consistent with my experimental findings.
b. It was the American novelist Cormac McCarthy that many critics felt that the public snubbed for too long.

Behavioral measures of retrieval should consequently reflect a processing facilitation at the retrieval point in constructions like (4b), as compared to ones like (4a). This assumes, however, that the retrieval process is not operating at optimal efficiency. In other words, a facilitation effect can only be noticed if the relevant process can actually be improved. Any process operating at its peak performance level cannot be made better. Hence, if sentence processing challenges are at a minimum and the retrieval process is unencumbered, then any strategy meant to improve memory retrieval will likely have little to no observable effect.

Because of this consideration, this dissertation focuses on the retrieval processes that occur in filler-gap dependencies. These linguistic constructions are known to impose numerous processing costs (Wanner & Maratsos, 1978; Hawkins, 1999): the information encoded by the filler must be maintained in memory, while other constituents along the filler-gap path must be simultaneously processed, and the correct “gap site” must be located (i.e. the filler must be associated with the appropriate grammatical position), in spite of the absence of any phonological cues as to its presence. Cumulatively, the consequence of all these responsibilities is a sentence that is relatively high in processing cost, as compared to a minimally different sentence without a filler-gap dependency. The cognitive challenges linked to filler-gap dependencies thus raise the chances for consistently finding significant facilitation effects due to informativity.

Note that the MFH does not specify what kinds of linguistic phrases are relevant for the retrieval-based principle. While the above-cited research on anaphoric form choices confines itself to a discussion of referential noun phrases, the MFH is theoretically meant to apply to phrases of any sort, referential or non-referential, anaphoric or non-anaphoric, NP or VP, etc. Thus, many of the experiments detailed here test the predictions of the MFH with respect to so-called wh-phrases, e.g. who, what, which child, where, etc. Typically, these phrases are not considered to be referential, in the sense that they do not pick out a uniquely identifiable individual in the real world. Throughout this dissertation, however, I make the uncontroversial assumption that wh-phrases introduce a discourse entity, as evidenced by the fact that anaphoric forms can refer to this entity, as in I don’t know who drank my coffee, but he’s going to be sorry he did. Just like the noun phrases in (4), wh-phrases can differ in informativity. The MFH consequently indicates that retrieval should be faster at the verb admired in (5b) than in (5a):
(5)  a. I forget who Harold said that he admired since he was a little boy.

b. I forget which superhero Harold said that he admired since he was a little boy.

The Memory Facilitation Hypothesis is also phrased in such a way that it makes no predictions about the effect of descriptions that refer to distinct discourse entities. Even if two phrases qualitatively appear to differ in terms of their informativity, the MFH offers no opinions about how the choice of one of these phrases over the other should affect memory retrieval. For instance, the accusative pronominal form *him* and the definite description *the fifteen year-old scotch sitting on the counter that comes from the Speyside area of Scotland* clearly express different quantities of information and the latter description highly restricts the set of possible referents. However, since these descriptions differ in meaning and undoubtedly require distinct discourse contexts to felicitously use them, it would be impossible to single out the role of informativity in subsequent retrievals. The MFH thus does not speculate on how such seeming contrasts impact the speed at which the corresponding mental representations can be re-accessed.

The magnitude of informativity effects is also expected to vary with factors other than the quantity of information. That is, other aspects of the discourse context are likely to modulate the strength of informativity effects. For instance, the usefulness or relevance of information in a linguistic description depends upon the contents of the preceding discourse. Information about a discourse entity can be highly pertinent to the discourse or completely irrelevant. Undoubtedly, relevant information facilitates comprehension at multiple stages of processing, as compared to irrelevant information. Furthermore, the strength of informativity effects may well vary across phrase types. An extra adjective or prepositional modifier may therefore affect the processing and retrieval corresponding to an indefinite NP differently than it does a definite NP. So, even though the MFH posits that the informativity of a linguistic description of any type bears on the speed and accuracy of memory retrieval, it does not preclude differences across linguistic types or the contribution of other interacting factors.

In normal conversation, of course, we don’t always choose to use the most informative description possible for some entity. In fact, we probably never do, as the most informative description possible could theoretically take the life of the universe to utter. What we observe in reality is a balance between saying too much and saying too little, depending

---

4 As demonstrated by the empirical results of the third chapter, the informativity of a filler-phrase in an FGD may also impact cognitive processes besides just memory retrieval.
upon the circumstances of the discourse. This point echoes the idea of Grice’s Maxim of Quantity (Grice, 1975), which claims that speakers should make their contributions as informative as possible, but not more informative than necessary. While Grice’s Maxim of Quantity has generally been used to account for the amount of information used in an utterance or during a conversational turn, it applies equally well to descriptions of discourse entities. Previous researchers, in fact, have suggested that linguistic form reflects a cooperative desire to maximize communicative success (Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989; Vonk, Hustinx, & Simons, 1992). Less informative forms like pronouns, therefore, signal a continuation of the discourse topic and indicate the speaker’s belief that the referent is easily identifiable. In contrast, highly informative forms serve the purpose of introducing new information, to shift discourse topics, or to simply increase the salience of some discourse entity. On this view, the balance between too little and too much information reflects the cooperative spirit of dialogue, where speakers and listeners strive for maximal clarity (i.e. successful communication) and depend on referential form to structure and update the discourse.

Closely following ideas expressed in Sperber and Wilson (1986), Almor (1999), and Almor and Nair (2007), I assume that this balance between over- and underspecification emerges from cognitive constraints on working memory and the dynamics of memory retrieval, on the one hand, and the discourse function of linguistic form, on the other hand. In short, processing cost is balanced by communicative function.\(^{5}\) Processing longer and more informative descriptions necessitates the expenditure of more working memory resources, since the incoming information must be identified, categorized, and integrated and the discourse model must be updated with any information that is new. The cost of processing this information pays for the discourse function of adding new information to the common ground or accessing a previously mentioned discourse entity, i.e. antecedent identification. To put it slightly differently, additional information raises the chances of successful communication.

There are thus two sides to informativity for comprehending a linguistic description.\(^{5}\) Note that the research on referential form choice is centered on the interacting processing and discourse factors that affect production (although production choices may reflect altruistic intentions that aim to help comprehension). In contrast, the present account is strictly dedicated to uncovering the factors that guide comprehension processes.
Information eats up processing resources, but at the same time it facilitates communication. Linguistic form choice consequently reflects a balance between these considerations. Accordingly, information that serves no discourse purpose incurs a processing penalty. Evidence of this penalty has surfaced in various studies looking at the consequences of repeated name anaphors (Gordon, 1993; Gordon et al., 1993; Gordon & Scearce, 1995). Gordon et al. (1993) found that sentence reading times slowed when a repeated proper name, as compared to a pronoun, was used to refer back to a focused antecedent. However, focused antecedents are already highly accessible discourse entities, so using a repetitive linguistic anaphoric form has no discernible purpose. The cost of processing the repeated information consequently serves no discourse function. In turn, this imbalance results in a pragmatic anomaly, leaving comprehenders to try and figure out what this perturbation of the balance is intended to signal (if anything).

Previous researchers, however, have generally neglected the import of linguistic form choice on subsequent comprehension. In this regard, there is another resource balance to consider and that is what might be thought of as the “pay now-pay later” balance. In short, there is a cost associated with retrieval whose value is regulated by properties of working memory. This price tag can be paid at different points in sentence processing. Obviously, the cognitive cost can be paid at the retrieval site, but it can also be paid \textit{in advance}. To the extent that the cost is paid ahead of time, fewer resources need to be expended at the retrieval site.

To understand how the cost can be paid early, it is necessary to consider how the cost accumulates in the first place. The basic idea is that inherent features of working memory place limitations on memory retrieval. As touched upon earlier, information stored in working memory rapidly decays—that is, the activation level of the memory trace decreases naturally over time. It is widely agreed that retrieval from working memory depends upon activation level, so activation decay must increasingly impair retrieval (Anderson et al., 2001; Anderson, Bothell, Byrne, Douglass, & Lebiere, 2004; Vasishth & Lewis, 2006, \textit{inter alia}). Activation decay can thus theoretically explain the distance-based effects reported in Gibson (1998, 2000) and similar research, as well as findings that processing an anaphoric expression becomes more difficult as the distance to its antecedent increases (Sanford & Garrod, 1981).

In addition to activation decay, retrieval in sentence processing is also subject to similarity-based interference effects. Memory representations, therefore, that exclusively contain
CHAPTER 1. INTRODUCTION

Figure 1.1: Schematization of balance between the discourse function and processing cost of information (horizontal dimension), and the “pay now-pay later” trade-off (vertical dimension)
non-distinctive information which other representations in memory likely share should lead to increased retrieval difficulties. As an example, consider a word or phrase that encodes only a minimal amount of information, such as *someone* or *the thing* or *it*. These descriptions carry very little in the way of distinctive information and present only a small number of target semantic features for retrieval, as they specify only animacy and humanness—features that nearly every noun phrase encodes. So in a discourse context where one of these phrases needs to be retrieved, any mental representation in working memory that shares the same feature-value with the true target will interfere with the retrieval process. Thus, these two aspects of working memory—decay and interference—both represent impediments for successful retrieval.

More informative antecedent expressions can be seen as counteracting these natural roadblocks for retrieval. I propose within this dissertation that increasing the information content of a linguistic description raises the activation level of the corresponding discourse referent, while simultaneously presenting a greater number of target features for cue-based memory retrieval. The activation “boost” mitigates the effects of activation decay, while the increased availability of target features reduces the probability of similarity-based interference. In contrast to vague and uninformative references, a highly specific description can encode multiple, distinctive features which can help focus the retrieval process (relative to their uniqueness in the discourse). So, while information processing requires an up-front investment of cognitive resources, it has the reward of facilitating memory retrieval. Less information processing conversely translates to a more difficult retrieval process. The limitations on informativity thus ultimately come from the cost of information processing and the necessity for information to have a discourse function.

It is worth stressing the point that the “pay now-pay later” principle does not embody a choice made during comprehension. This principle merely expresses complementary processing alternatives that speaker or writer choices of linguistic form force upon the comprehender. This raises the question of whether speakers plan their form choices with an awareness of these considerations. I think it is quite intuitive that linguistic form choices are not made with conscious predictions about the processing difficulty of future memory retrievals; however, speakers may intend to make certain discourse entities highly salient or

6Highly specific information therefore aids retrieval particularly when no other representations in memory share that information. That is, I expect that the retrieval benefits provided by an informationally rich description will be modulated by the presence of other discourse entities with the same feature-values.
topical via a highly informative linguistic form, which has the consequence of aiding future retrievals. Ultimately, the role of the MFH in production remains an open question and this dissertation will not attempt to address the issue of how informativity effects influence production choices.

This dissertation experimentally tests the claim that the quantity of information encoded in a linguistic description guides the retrieval process, but some previous empirical research already supports these assertions.\(^7\) Cowles and Garnham (1995), for example, report faster reading times for anaphoric expressions (e.g. *vehicle*) when the antecedent is more specific and informative (e.g. *hatchback* vs. *car*).\(^8\) As discussed extensively in the next chapter, Gernsbacher (1989) also suggests that full names activate their referents more than pronouns, as measured by word recognition of an antecedent. Indeed, other research in linguistics and psychology contributes further indirect evidence that the information content of linguistic entities plays an important role in subsequent linguistic processes and memory retrieval, as I detail in the next chapter. This dissertation, however, presents novel evidence showing that the quantity of information in a linguistic description has a significant influence on subsequent retrievals of the corresponding mental representation. The “pay now-pay later” principle is substantiated by findings which show that initial processing costs are correlated with the cost of retrieval. In particular, where information processing costs increase, retrieval costs decrease.

What I hope to convince you of in this dissertation is that linguistic form choices not

\(^7\)At first glance, some research may seem to present contradictory evidence to the claims made here. Upon closer inspection, however, these results do not pose any problems for the current hypothesis. De Vincenzi (1991, 1996) has been cited as providing proof that *which-N'* questions expend more processing resources than bare *wh*-questions. Her results, however, merely identify a preference in Italian for interpreting bare *wh*-words as subjects in structurally ambiguous questions, but no such preference for *which-N'* phrases, e.g. *which girl*. Other researchers have also identified processing differences between bare *wh-* and *which-*questions (Hickok & Avrutin, 1995; Kaan, Harris, Gibson, & Holcomb, 2000), but none of this research provides directly conflicting evidence to the predictions of the MFH.

\(^8\)The relevant measure for Cowles & Garnham was not informativity but conceptual distance. Thus, *hatchback* is considered more conceptually distant from *vehicle* than *car* is. The authors encounter some difficulty, though, with reconciling their results with other experimental findings (Garrod & Sanford, 1977) that show faster processing of an anaphor (e.g. *bird*) after more typical antecedents (e.g. *robin* vs. *goose*); however, the current hypothesis has no trouble reconciling these results in a content-addressable model of retrieval: the informativity of the antecedent explains the Cowles and Garnham results, while associative strength underlies the Sanford and Garrod findings. Simply put, these results do not conflict at all, as they illustrate the effect of two distinct factors.
only reflect, but also attenuate retrievability. The information content of a linguistic description should therefore not be considered the last stop in a series of cognitive processes. Instead, decisions of form and information content have notable repercussions on subsequent comprehension. This conclusion can offer a vital component to discourse models that strive to capture the dynamic processes underlying comprehension. In addition, my goal here is to show that the MFH expresses a general cognitive principle that can contribute to the understanding of an array of linguistic and psycholinguistic phenomena from theories of syntax to treatments of filler-gap processing. The implications of the discoveries detailed in the following pages are consequently relevant for various types of linguistic investigations and not just theories of sentence processing.

1.2 Outline of Dissertation

In the next chapter of this dissertation, I review research from a variety of sources that motivates the claims of the MFH. Evidence from linguistics, psycholinguistics, and psychology shows that information content can have a significant effect on subsequent linguistic processes. These sources do not directly link the informational “richness” of an expression with its subsequent retrieval; however, they each present phenomena that can be understood as reflecting the cognitive principle introduced in the MFH. In this sense, therefore, they argue for additional investigation of the MFH, while also evidencing the applicability of the MFH to a wide array of linguistic processes.

Chapter Three tests the predictions of the MFH in the context of so-called syntactic islands. These syntactic configurations have been labeled violations of putatively universal constraints on FGDs; however, accumulating evidence suggests that at least some of these constraints may derive from considerations of processing difficulty. Within this chapter, three self-paced reading experiments are described which all look at how the informativity of filler-phrases affects processing at potential retrieval points in syntactic island violations. In each experiment, the evidence reveals a processing facilitation for the experimental conditions involving longer and more informative filler-phrases.

The fourth chapter of this dissertation continues to evaluate the MFH, but this time in the context of filler-gap dependencies which do not violate any proposed syntactic constraints. The empirical data from the two experiments detailed in this chapter indicate the general applicability of the MFH, as the results verify that the informativity effects observed
in syntactic island settings extend to other FGDs that do not conflict with any proposed constraints on dependencies. Furthermore, the second experiment in this chapter supplies critical evidence that the MFH applies to other kinds of phrases besides just *wh*-expressions.

Chapter Five is devoted to examining the functional motivations for the relationship between informativity and memory retrieval. That is, I explore alternative explanations for why more information in a linguistic description aids future retrieval. One possibility for the correlation between informativity at encoding and “retrievability” is that more informative expressions typically contain more words and morphophonemic material—therefore, the retrieval process may depend on the amount of time originally spent constructing a targeted mental representation. In addition, I also present evidence from an incidental memory task that corroborates the role of retrieval in the informativity effects observed in the self-paced reading studies.

Chapter Six is a discussion of the MFH’s role in discourse and its relationship to current models of sentence processing. In particular, I return to many of the points raised in this chapter concerning the balance between over- and underspecification. Linguistic form is argued to both reflect and attenuate retrievability from memory, so that theories of discourse and activation “fluctuation” (caused by decay and boosts) must display the impact of informativity. In this light, I show in this chapter how these informativity effects connect naturally with cognitive models that view sentence processing as a type of skilled memory retrieval that exploits cue-based retrieval. In particular, I show how the current findings align with the predictions of the ACT-R model of cognition (Anderson & Lebiere, 1998; Anderson et al., 2001).

The final chapter of this dissertation reviews the empirical evidence presented in this dissertation and discusses some outstanding questions for future research. The generality of the MFH means that virtually any complex linguistic process that involves memory retrieval should be affected by the complexity of the relevant memory targets. So, while this dissertation dwells empirically on the evaluation of how informativity modifies the ease of filler-gap processing, the conclusions drawn here are equally pertinent for other sentence processing tasks such as anaphor resolution and subject-verb integration. Finally, I consider the implications of the MFH for other areas of investigation in linguistics, including research in pragmatics and syntactic theory.
Chapter 2

Motivating the Memory Facilitation Hypothesis

This chapter is intended to deliver some preliminary support for the memory facilitation hypothesis (MFH) by highlighting a range of previous research that, directly or indirectly, points to the role of informativity on linguistic performance and memory retrieval. The purpose of this chapter is thus to give credit to earlier work that furnishes the empirical backdrop for the current research. At the same time, however, the evidence derived from this previous work verifies the plausibility of the MFH and thus encourages further investigation. Research from a variety of fields, including linguistics, psycholinguistics, and psychology, touches upon phenomena that are easily explicable by the MFH. In fact, these seemingly disparate phenomena have a single, unified explanation under the MFH. The material in this chapter therefore attempts to illustrate the breadth and importance of the hypothesis by showing the wide range of linguistic (and non-linguistic) phenomena to which it applies.

While this chapter specifies the shared characteristics of the MFH and previous theories, the underlying goal is to clarify how the MFH differs from all these accounts. I review theories that posit that informativity influences (1) acceptability in certain kinds of filler-gap dependencies, (2) coreference processes (via retrievability), as well as (3) propositional recall. None of these alternatives, however, links the informativity of linguistic constituents to their own subsequent retrieval in language comprehension.
2.1 Acceptability Paradigms

The hypothesis that the information content of a description predicts subsequent retrieval has a resemblance to claims made within generative linguistics about the graded acceptability of sentences containing certain kinds of filler-gap dependencies (FGDs). The authors of this literature do not speak directly about sentence processing or memory retrieval, but they do make important observations about acceptability. And at least partly, the acceptability of syntactic structures depends upon the processing difficulty they involve—as demonstrated for the case of complex center-embeddings, strong garden path sentences, and numerous other constructions (Chomsky & Miller, 1963; Bever, 1970; Kluender & Kutas, 1993b; Fanselow & Frisch, 2004). In particular, increased processing difficulty typically leads to lower ratings of acceptability (although there are some exceptions; see Gibson and Thomas (1999); Fanselow and Frisch (2004)). Especially in cases of serious processing difficulty, the cognitive demands are likely to be responsible for a perception of ungrammaticality. Notably, the linguistic context of FGDs poses a relatively challenging processing domain, as reviewed in depth in the next chapter. I also demonstrate there how the gradience in acceptability emerges as a result of processing differences. For now, however, the goal is only to establish the effect of informativity on the acceptability of sentences with FGDs, as well as to clarify how previous theories have attempted to reckon with this fact.

Indeed, various authors within the generative literature have acknowledged that the content or specificity of an extracted element affects its “extractability”—the acceptability of the (sentence containing the) filler-gap dependency (Rizzi, 1990; Cinque, 1990; Chung, 1994). The origins of this observation go back to the contrast between extraction of arguments and adjuncts, which appear to be subject to different conditions on extraction (Huang, 1982; Chomsky, 1986). Specifically, extraction of an argument is arguably better than extraction of an adjunct in examples like (6) below (where ‘≥’ stands for ‘sounds more acceptable than’):

(6) a. Which tune did Jim wonder whether to sing ____ next? ≥
    b. How did Jim wonder whether to sing ____ next?

The acceptability of some _wh_-dependencies further depends on whether the extracted element is a _which-N_ phrase or an interrogative pronoun, as demonstrated in the following minimal pair (Maling & Zaenen, 1982):
(7)  a. Which article don’t you remember who wrote?
    b. What don’t you know who wrote?

Rizzi (1990), in order to account for these facts, suggested a division between NPs (or DPs) that are “intrinsically referential” versus those that are inherently non-referential, defining the former as those arguments which “refer to specific members of a set in the mind of the speaker or pre-established in discourse.” In an exploration of extraction asymmetries in Italian, Cinque (1990) adopts this distinction between types of NPs, calling for this division to extend to a range of quantifier phrases, as shown in Table 2.1. This broader generalization is based on Italian filler-gap dependency data that exhibit an acceptability contrast between seemingly referential versus non-referential arguments. Included among the list of so-called referential argument phrases are *which*-N’ phrases, while *who* belongs to the class of non-referential arguments, thus presumably accounting for the increased acceptability of (8b) over (8a):

(8)  a. A chi non ti ricordi quanti soldi hai dato?
    ‘To whom don’t you remember how much money you have given?’
    b. A quale dei tuoi figli non ti ricordi quanti soldi hai dato?
    ‘To which one of your kids don’t you remember how much money you have given?’

Cinque (again, following Rizzi) believed that being intrinsically referential obviated the requirement for traces to be antecedent governed, which effectively removes some major theoretical obstacles for phrasal movement. In contrast, for adjuncts and non-referential arguments, the requirement of antecedent government still applies, such that movement of these sorts of phrases necessarily proceeds in a fundamentally different way, i.e. successive-cyclic movement. Ultimately, then, the concept of phrasal movement developed a relativized flavor, as the conditions on movement depended critically on the type of phrase being moved. This division thus has had quite a significant impact on theories of syntax.

Similar arguments also appear in the discussion of acceptability contrasts involving multiple *wh*-questions. In English, there is a strong preference for the order shown in (9a) over the one in (9b):

(9)  a. Who got what?
    b. What did who get ___?
CHAPTER 2. MOTIVATING THE MEMORY FACILITATION HYPOTHESIS

"nonreferential" DPs | "referential" DPs
---|---
chi? ‘who?’ | definite DPs
ogni NP ‘every NP’ | quale? ‘what?’
nessun NP ‘no NP’ | tutti DP ‘all DP’
qualcosa ‘something’ | molti NP ‘many NP’
qualcuna ‘someone’ | qualche NP ‘some NP’
qualsiasi NP ‘whatever NP’ | alcuni NP ‘some NP’
chiunque ‘whoever’ |

Table 2.1: DP classification from Cinque (1990); adapted from Chung (1994)

Since Kuno and Robinson (1972), the variant in (9b) has been considered ungrammatical due to constraints on wh-movement. Chomsky (1973) referred to the relevant constraint as the Superiority Condition:

\[(10) \text{No rule can involve X, Y in the structure} \]
\[\ldots X \ldots [a \ldots Z \ldots WYV \ldots]\]
\[\text{where the rule applies ambiguously to Z and Y and Z is superior to [m-commands] Y.}\]

For multiple wh-questions, this condition means that if the grammar requires a wh-phrase to move, then the structurally lower wh-phrase cannot fulfill the requirement. As it turns out, however, the Superiority condition (as well as similar variations on the same theme) has some notable exceptions. In this regard, Pesetsky (1987, 2000) explores the acceptability difference between examples like (11a) and (11b), following earlier observations by Karttunen (1977) and Bolinger (1978):

\[(11) \]
\[\begin{align*}
a. \text{Which medication did which patient get?} \\
b. \text{What did who get?}
\end{align*}\]

Pesetsky explains this contrast via a mechanism he terms D(iscourse)-linking, claiming that which-phrases instantiate D-linked phrases and that the normal Superiority effects are void.
in the presence of D-linked phrases. Although a formal definition of D-linking is never offered, even in the secondary literature, the following provides the basic intuition of what is meant:

“Context sets previously mentioned in the discourse qualify a phrase as D-linked, but so do sets that are merely salient (e.g., which book, in a context where speaker and hearer both know that reference is being made to a reading list for a course) and sets whose salience is culturally determined (e.g. what day of the week, which sign of the zodiac)” Pesetsky (2000, p. 16).

Consequently, individual lexical items and phrases are not inherently specified for D-linking. Instead, according to the text, this property emerges from the context or conventionalized interpretations. The underlying link between discourse and conditions on movement as in (10) remains opaque, as Pesetsky himself admits immediately after the preceding quote:

“A reliable rule of thumb is that if a wh-word in a multiple question can be felicitously paraphrased with an expression of the form which of the X, it can cause the Superiority effect to disappear. The reason for this link between semantics and syntax is obscure, and will remain obscure even at the end of this book” Pesetsky (2000, p .16).

D-linking therefore ultimately stands as a descriptive label, recognizing a relationship between the content of a wh-phrase and the acceptability of certain kinds of dependencies, but it fails to establish or motivate why such a relationship should exist in the first place. It is, in fact, difficult to see why mere contextualization should be able to invalidate an otherwise ostensibly universal constraint such as Superiority, particularly given the fact that most, if not all, natural language is actually contextualized to some extent.

Despite its ubiquitous presence in the syntax and psycholinguistics literature (Frazier & Clifton, 2002; Reinhart, 1998; Boeckx & Grohmann, 2004; Diaconescu & Goodluck, 2004; Erteschik-Shir, 2004; Featherston, 2005; Kennedy, 2005), D-linking thus offers no clear rationale for why wh-phrase content or its relationship to the discourse context should affect the acceptability of certain kinds of FGDs. Indeed, the original observations from Pesetsky that superiority-violating wh-orders occur only with D-linked which-phrases appears fundamentally incorrect, in light of data reported in Arnon, Hofmeister, Jaeger, Sag, and Snider (2005), such as the following example (and many others like it) found via simple web searches:
(12) I must have missed something. What did who do to Pierre Salinger?¹

Pesetsky must also reckon with the acceptability of interrogatives involving three or more *wh*-phrases like those in (13), which, according to him, seem acceptable in a way that superiority-violating *wh*-orders with just two *wh*-phrases do not.

(13) What did who do to whom?

As a result of the categorical nature of D-linking, a separate mechanism must be invoked to explain such counterexamples. In particular, Pesetsky asserts that Richards’ (1998) Principle of Minimal Compliance suffices to explain such *wh*-orders. Given covert movement of the subject *wh*-phrase, thereby fulfilling some set of minimal demands imposed by Superiority or a comparable constraint, the other *wh*-phrases can move into pre-subject positions without violating Superiority. Such an explanation, however, immediately begs the question as to why such covert movement fails to apply in cases of only two *wh*-phrases. The picture left by Pesetsky is that the possibility for extraction or movement depends upon conditions that vary from context to context. In one environment (just two *wh*-phrases), only D-linking can license movement of one *wh*-phrase over a structurally higher *wh*-phrase, but in another setting (more than two *wh*-phrases), other exceptions come into play.

Chung (1994) recognizes the fact that generative accounts like Pesetsky’s lack a clear rationale for why referentiality, specificity, or descriptive content should affect the acceptability of FGDs. Perhaps more so than anyone else, she seriously considers the following questions:

“Why should long movement be legitimized in just those cases where the trace ranges over a sufficiently restricted set? To put the question differently, what is it about the ability to narrow down the domain of *wh*-quantification ‘enough’ that makes it possible for strict locality to be violated?” (p.39)

Chung considers a range of possibilities in answer to these questions, from largely pragmatic explanations to deeply theoretical and framework-specific syntactic explanations. For instance, one suggestion in this matter comes from Kroch (1998) and Comorovski (1989), who say that the existential presuppositions created by *wh*-questions require identification of a corresponding entity in the discourse, and that this is impossible in cases where the

---

¹http://www.freerepublic.com/forum/a3b1c8a4d1847.htm
wh-phrase is non-referential or non-specific.2

In contrast, Rizzi (1990) attempts to solve this mystery by pointing to the connection between referentiality and bearing a referential index. Specifically, Rizzi assumes that only referential arguments possess referential indices. These indices act as a means for establishing a binding relation with the trace of the referential argument. Since binding relations remain unconstrained by syntactic boundaries, this analysis justifies “long movement” of referential arguments. Unlike this case of variable binding via the referential index, “nonreferential” arguments or adjuncts, in the view of Rizzi, require successive-cyclic movement in order to antecedent govern their traces, i.e. to form a legitimate, grammatical dependency with their traces.

Unfortunately, as Chung notes, this reasoning diverges from what bound variable anaphora patterns tell us, as DPs falling into Rizzi and Cinque’s non-referential class can clearly antecede a bound variable pronoun. That is, Rizzi’s argumentation essentially requires a separate indexing system for cases of movement, in a way that differs substantially from the traditional system devised for anaphora, which seemingly breaks down the strength of the argument from binding:

“. . . Rizzi’s referential indices cannot be identified with the indexing mechanism that is a mainstay of current approaches to anaphora (see Frampton 1991:39-42 for similar conclusions). But if that is so, then we are left wondering whether the use of indices in this theory amounts to more than a diacritic to distinguish the DPs that allow long movement from those that do not” (Chung, 1994, p. 33).

Chung also addresses the semantic accounts of Szabolcsi (1992) and Frampton (1991), who argue that traces of long movement must have interpretations that range over sets of individuals. Such an explanation seems to merely push back the question one more layer: why should traces in certain syntactic contexts and not others have restrictions on their interpretation? To put it another way, why should trace interpretation depend upon the

2See also the account of Erteschik-Shir (2004), where the grammaticality of extraction is said to depend upon information-structure constraints, rather than syntactic constraints. Erteschik-Shir adopts this perspective largely due to the influence of context on acceptability and notes that accounts like Cinque and Pesetsky’s lack “an explanation of why referentiality should interact with syntactic constraints, such as subjacency” (p. 318-9).
clause and phrase boundaries which a dependency crosses? In the end, Chung admits that all of these explanations fail to fully capture the relevant data.

Part of the problem with all these explanations relates to the categorical distinctions being made with respect to kinds of noun phrases. Each noun phrase must fall into a referential or non-referential class, or a specific or non-specific class. Yet the boundaries between these two classes are vague at best. To accommodate seeming exceptions, the definitions of the categories must be extended further and further until the categories virtually lose all meaning. Moreover, due to the categorical divisions being imposed, these accounts fail to predict any differences among extracted elements of the same class that nevertheless differ in informativity, e.g. which person vs. which patient or a politician vs. a corrupt politician.

Secondly, these accounts typically only consider a subset of the relevant data. As mentioned above, Kroch (1989) and Comorovski (1989) present arguments about why set restriction matters for wh-questions, but at best their answer suffices to explain a correlation between specificity and the acceptability of extraction in that one particular domain. Such an explanation ignores the same type of correlation in other kinds of extractions not involving wh-questions that Rizzi, Cinque, and Chung discuss. Moreover, these stories tend to focus on extraction in a very particular set of constructions such as syntactic islands. Consequently, the rationalizations that emerge for the observed extractability or movement effects make specific reference to the syntactic context, thereby restricting the generality of these accounts.

In contrast to these essentially grammar-internal explanations of the gradience in these FGDs, the MFH offers a radically different way of understanding why the content of an extracted element influences acceptability judgments. Acceptability judgments represent the culmination or end-product of distinct cognitive processes, including the processing effort required to construct the relevant linguistic entity as mentioned at the outset of this section. Seen in this light, the MFH offers up an attractive means for explaining the gradience in extractability. According to the hypothesis, retrieval cost decreases as the quantity of information used to encode the retrieval target increases. For filler-gap dependency processing, this means that retrieval at the gap site happens more efficiently and accurately, thus ameliorating difficulty after the gap and raising the likelihood of a successful parse. Cumulatively, this could amount to a perception of higher acceptability
in contexts of more informative encodings.\(^3\)

The advantage of such an approach, as compared to the generative accounts considered above, is that it does not make any categorical distinctions among phrase types, i.e. informative vs. uninformative. All determinations of informativity are relative to the informativity of some other phrase. Plus, given that memory retrieval is necessary for phrases of all sorts, and not just in \(wh\)-dependencies, the processing-based hypothesis makes sense of the ubiquity of the information-based contrasts by appealing to broad cognitive principles that are essentially blind to the specific, grammatical characteristics of linguistic units. This translates into a more satisfying explanation of why the content of an extracted element has such an impact upon its extractability in an array of FGDs.

Besides the generality of the retrieval-based story, such a route also obviates the need for complicated, modular principles and classification systems whose functional purpose is entirely missing. Taking the theory of D-linking as an example, it is mysterious how anyone could acquire knowledge of such an “escape hatch” (or how such an exception could develop evolutionarily), particularly given the rarity of multiple \(wh\)-questions. The processing account, in contrast, has more obvious potential functional motivations: more informative descriptions call more attention to the entities they denote and will be increasingly distinct from other representations. When used anaphorically, longer and more informative descriptions also signal the need for more cognitive effort on the part of the comprehender to access the relevant representation (Ariel, 1990, 2001), and experimental results within this dissertation will reveal that the same is true for new discourse entities. In other words, the sheer amount of extra effort that goes into processing a more informative phrase forces attention to increase. Reflecting then upon the question of when a more informative linguistic form is preferred by a speaker, the desire to heighten the comprehender’s attention to a particular discourse entity must be one of the primary motivations for the choice of a more explicit form over a less explicit form. This heightened attention would essentially facilitate memory retrieval for subsequent references to the same entity. Thus, when a

\(^3\)This line of reasoning suggests that a “pay now” strategy is preferable to a “pay later” strategy for filler-gap processing (as well as other linguistic contexts). The preference to pay prior to the retrieval site follows from the fact that cognitive costs accumulate throughout sentence processing, so more processing resources are likely to be available earlier in the sentence (especially so in filler-gap processing). Furthermore, given the difficulty of efficiently retrieving the correct representation at a gap site with no phonological manifestation at a time when processing resources are already taxed, an early investment of available processing resources seems well worth the trouble.
speaker uses a complex linguistic form to introduce some topical discourse entity, that topic can be quickly retrieved throughout the conversation due to its high degree of salience or activation, thereby raising the likelihood of successful communication. There appear to be, consequently, some easily identifiable functional reasons for the relationship between informativity and memory retrieval, although which of these are actually operative remains an empirical question.\footnote{Chapters Five and Six contain a more extensive discussion of why longer and more informative expressions facilitate retrieval.}

This alternative explanation for the acceptability contrasts, which connects acceptability and processing effort, clearly relies on the assumption that the cognitive cost involved in sentence comprehension has a profound effect upon judgments of acceptability. Throughout this dissertation, and particularly in the next chapter, I provide empirical support for this stance, as well as arguments in favor of an underlying processing rationale for the acceptability data. In the next subsection, I begin this process by reviewing evidence from a study of Superiority-violations (SUVs) that aims to evaluate the relationship between processing and acceptability.

### 2.1.1 A processing account of Superiority-violations

Building upon the observations of Pesetsky, Bolinger, and Karttunen described above, Hofmeister, Jaeger, Sag, Arnon, and Snider (2007) again examine how the acceptability of superiority-violating multiple \textit{wh}-questions depends upon the linguistic form of the \textit{wh}-phrases. Hence, their empirical investigations essentially look once more at the acceptability contrasts involving filler-phrases of different informational complexity. Besides using systematic means for gathering acceptability judgments (i.e. magnitude estimation Bard, Robertson, and Sorace (1996))\footnote{Magnitude estimations allows participants to develop a continuous acceptability scale, which enables participants to make as many distinctions as desired.}, the authors also explore how the manipulation of the same non-syntactic factor predicts processing difficulty, and to what extent the processing results correspond with acceptability judgments. Based upon the results, they conclude that the more complex \textit{which}-N\textsuperscript{1} phrases make processing easier, which, in turn, raises judgments of acceptability.

In accord with the intuitions of linguists, Hofmeister et al. (2007) report the results of
a magnitude estimation study, finding that “complex” (i.e. multi-word) wh-phrases significantly improve acceptability in SUV contexts. Figure 2.1 thus shows that embedded SUVs where the fronted object wh-phrase consists of either a what-N’ phrase (14b) or a which-N’ phrase (14c) are judged better than minimally different examples where the fronted wh-phrase is a bare wh-word like what (14a). This study also reflects the fact that improved acceptability in SUVs does not exclusively originate with which-N’ phrases, as might be inferred from the text of Pesetsky (1987; 2000), but actually accompanies informationally richer wh-phrases of any type.

    b. Ted revealed what device who invented.
    c. Ted revealed which device who invented.

Similarly, in a separate magnitude estimation study, we also found that the informational complexity of the subject wh-phrase in SUVs has a significant impact on acceptability. Subject which-N’ phrases lead to significantly higher ratings of acceptability than the bare wh-item who. Thus, SUV formulations with which-N’ phrases in both the object and subject position end up receiving the highest ratings of acceptability, while those with bare wh-words in both positions get the lowest ratings.

Given that these multiple wh-questions can be judged more acceptable with longer and informationally richer wh-phrases, Hofmeister et al. (2007) further examine the role of processing differences in these varying judgments of acceptability. In a self-paced reading time study, we explored how manipulating this same factor found to influence acceptability would affect processing at the subcategorizing verb.

(15) a. Ashley disclosed what who signed after receiving permission from the president.
    b. Ashley disclosed which agreement who signed after receiving permission from the president.
    c. Ashley disclosed what which diplomat signed after receiving permission from the president.
    d. Ashley disclosed which agreement which diplomat signed after receiving permission from the president.

---

6Pesetsky’s original observations concerning D-linking actually focus on the status of this subject wh-phrase, despite the fact that all the SUV examples in Pesetsky (1987) have which-N’ phrases in both positions.
Paralleling the acceptability judgments, less informative \textit{wh}-phrases in both object and subject position were followed by slower reading times at the verb and spillover region encompassing the subsequent three words, as shown in Figure 2.2. On average, reading times at the verb for examples like (15a) were slowest, while sentences like (15d) produced the fastest times. The other two conditions, each with one bare \textit{wh}-word and one \textit{which-N'} phrase, were intermediate between these two extremes. In other words, the reading time results show main effects of the kind of \textit{wh}-phrase used in both positions of the SUVs.\footnote{We do point out, however, that question-answer accuracies in the self-paced reading experiment reflect an interaction, essentially produced by the fact that the form of the fronted object \textit{wh}-phrase had no effect on accuracy, while the form of the in-situ subject did. This result parallels another magnitude estimation study we report, where we observed an interaction due to the same pattern of results.}

In essence, the results of the acceptability tasks and the comprehension-based experiment align with one another. The condition with the fastest processing times also received the highest mean rating of acceptability. Conversely, the hardest condition for processing was judged the worst in terms of acceptability. Again, the intermediate conditions in terms of processing yielded similarly intermediate scores in terms of acceptability. On the basis of
these results, Hofmeister et al. (2007) conclude that processing difficulty ultimately is the causal factor determining acceptability in these multiple wh-questions, which theoretically excludes the need for any constraint such as Superiority located in the grammar.

Hofmeister et al. (2007) explain this processing difference, and the associated variability in grammaticality judgments, in terms of the notion of accessibility. Acknowledging that the choice of linguistic form marks salience or activation at the time of utterance (Ariel, 1990), we also contended that this choice partially determines subsequent activation—a property they refer to as “future accessibility” (Ariel, 2001) (see next section). In the case of wh-phrases, they suggest that increasing the explicitness or specificity of the wh-phrase effectively boosts the activation of the associated representation, causing the relevant information to be more accessible at the point of reading the subcategorizing verb:

“If the difficulty of processing a head is a function (among other things) of the activation levels of its arguments, then the form preferences for both wh-questions and referential NPs emerge as a preference for high argument activation at the point when the head is processed.”
CHAPTER 2. MOTIVATING THE MEMORY FACILITATION HYPOTHESIS

Using the which-N’ phrases in both argument positions of the SUV consequently raises the activation level for both arguments. This “boost” then improves processing at a critical site where processing costs run high, thereby inciting higher ratings of acceptability.

The MFH developed here expands upon this analysis by identifying the cognitive operation which is sensitive to these differences in activation level: memory retrieval. The role of retrieval is partly implicit in the account of Hofmeister et al. (2007), as we linked the processing differences to an operation which occurs at the selecting head. Memory retrieval, though, is not explicitly named as the operative mechanism whose speed and efficacy is contingent upon the information in the wh-phrases. The current investigation further seeks to find more widespread evidence for the link between informativity and subsequent activation by examining a wider range of FGDs. The hypotheses in Hofmeister et al. (2007) are presented strictly as theories about the processing of wh-dependencies, but the MFH theoretically applies to any sort of phrasal constituent, including VPs and PPs. Lastly, part of the purpose of the research documented in this dissertation is to address why such a link exists in the first place, as mentioned at the conclusion of the last section.

The evidence presented in Hofmeister et al. (2007) nevertheless foreshadows many of the results in this dissertation. Not only do more informative filler-phrases facilitate faster processing at the retrieval site, but they also have the consequence of producing higher judgments of acceptability. If this were only an isolated finding, it might be considered coincidental, but given the number of times it recurs, the evidence argues against such an interpretation. Noting some of the shortcomings of theories like D-linking, as well as the explanatory economy of a processing-based account discussed in the preceding section, Hofmeister et al. (2007) choose to interpret the results as reflecting an influence of processing on perceptions of well-formedness, rather than vice versa. This way of interpreting the results avoids postulating highly specific grammatical principles and exceptions, which must be tweaked or adapted to fit different types of constructions, and appeals instead to cognitive principles that have a wide range of application. Thus, this study of SUVs provides the empirical starting point for thinking that the extractability contrasts noted previously really echo performance-related factors.
2.2 Activation Enhancement

What I have proposed here in terms of memory retrieval facilitation also draws support and inspiration from previously made claims concerning enhancement of mental activation for referential entities, as explored in Gernsbacher (1989, 1995) and discussed further in Ariel (2001) and Gernsbacher and Shroyer (1989). These sources argue that increasing information in an anaphoric expression promotes or enhances the activation of the corresponding antecedent referent, thereby facilitating coreference processes. As information increases in an anaphoric expression, the number of retrieval cues increases, creating a larger and more reliable set of features to identify the antecedent: “accessibility (through retrieval) is a function of the similarity between a retrieval cue and a memory trace” (Gernsbacher, 1989, p. 104). At heart, then, this is a theory about what makes for a better memory trigger. 

In the view of Gernsbacher and her colleagues, comprehenders select antecedents for anaphors by accessing the most activated referent. The content of an anaphor thus guides interpretation by (partly) determining the activation of the true antecedent (via retrieval cues). But by increasing the activation of an antecedent for establishing coreference, this leaves a high activation level even after the coreference process is completed. By implication, this means that on subsequent anaphoric references, less information or fewer retrieval cues should be necessary to access the antecedent. Ariel (2001) talks about this in terms of “future accessibility,” contrasting accessibility prior to a referring expression with accessibility after the use of a given referential form:

“More explicit expressions (lower accessibility markers, proper names, for example) boost the activation of their mental representations faster and more than higher accessibility markers. In effect, the same accessibility marking scale reflects accessibility enhancing (and suppressing): the lower the accessibility marker used, the more enhanced the discourse entity coded by it will become.

. . . the same accessibility markers code a specific current degree of accessibility

---

8In this sense, Gernsbacher’s analysis goes back to ideas expressed as early as Bower (1967) and Ratcliff (1978). There is, in fact, a wealth of results from experimental psychology that confirm that increasing retrieval cues facilitates recall. For instance, the presence of more verbal retrieval cues improves the recall of patients with senile dementia (Diesfeldt, 1984), olfactory cues associated with a stimulus can facilitate memory retrieval (Maylor et al., 2002), and visual cues aid in infant object recognition (Bahrick, Hernandez-Reif, & Pickens, 1997). These are just a few of the results that establish that the speed and accuracy of memory retrieval depends upon the number of retrieval cues.
(say, low), but at the same time, they contribute (at least partly) to the opposite degree of future accessibility (high). This can explain why speakers shift to lower accessibility markers from time to time, even when they continue to discuss the same discourse entity” (Ariel, 2001, p. 68).

Gernsbacher (1989) addresses the possibility that the informational content of an anaphoric expression has an impact on the associated degree of activation. In a series of so-called probe-word recognition experiments, she looks at how the activation of a referent varies following the choice of different anaphoric forms. In these probe-word tasks, participants read a sentence word by word at a predefined rate and, at a particular point, a probe word is presented and the subject must indicate whether that word has appeared previously in the sentence. This task has been employed widely to investigate the time course of coreference and anaphoric interpretation (see Gordon, Hendrik, and Foster (2000) for a critique of this methodology).

Gernsbacher examines, in particular, how the choice of a repeated name versus a pronoun affects the ability to recognize the actual antecedent (e.g. Bill in (16a) or a non-antecedent (e.g. John in (16b)). In both conditions, the probe word was either presented immediately before or after the repeated name or third person pronoun, in order to evaluate the degree of activation change (probe words appear in capitalized letters below).

(16) a. Bill lost a tennis match to John. Accepting the defeat, (BILL) Bill/he (BILL) walked slowly to the showers.

b. Bill lost a tennis match to John. Accepting the defeat, (JOHN) Bill/he (JOHN) walked slowly to the showers.

The results from her experiments indicate that non-antecedents (e.g. John) were less recognizable after the repeated name. Gernsbacher terms this effect “suppression,” because the increased activation of one entity supposedly diminishes the activation strength of other entities. This effectively promotes discourse topics and reduces the salience of peripheral entities, ostensibly limiting any pronoun ambiguity.

In addition to an effect on reaction times due to suppression (MacDonald & MacWhinney, 1990; Nordlie, Dopkins, & Johnson, 2001), Gernsbacher observes effects due to what she calls activation enhancement. When an antecedent probe was used, according to the results from Gernsbacher (1989), recognition was faster when the subject in the second sentence was the proper name, e.g. Bill. Another way of putting this is, quite simply, people
recognize a word faster when they’ve just seen it, as opposed to a word that only bears an anaphoric relation to that word. Gernsbacher interprets these results, however, as proof that coreference is facilitated by the use of a repeated name because it provides more retrieval cues than a pronoun. These cues, in turn, “resonate” with the memory trace of the antecedent, causing it to be more accessible (easier to retrieve). She goes on to postulate that these results ultimately form part of the evidence for an explicitness principle:

“The more explicit the concepts, the more likely they are to trigger the suppression of other concepts, and, when used anaphorically, the more likely they are to enhance their antecedents” (Gernsbacher, 1989, p. 135).

These experimental results, despite the conclusions that Gernsbacher draws from them, merely indicate that a repeated name makes the name more recognizable than an anaphor with a different morphological form. How the form of an anaphor affects the activation of the antecedent more generally is not addressed. Moreover, as Gordon et al. (2000) point out, response times in probe-word recognition tasks with repeated name anaphors are “influenced by strategies aimed at keeping track of the words that are likely to be probed.” Especially in the experiments detailed in Gernsbacher (1989), where probe words consist only of proper names, participants may have simply relied on the construction of memory lists composed of proper names. In fact, Gordon and company offer compelling evidence that this is exactly what happened: subjects perform in the exact same way even when the sentences are scrambled randomly, i.e. the positions of the words (except the anaphor) were scrambled.

Aside from the concern of memory strategies involved in probe-word recognition tasks, repeated names are not the ideal type of NP to consider when studying either coreference timing or activation boosts. Exact repetition of a referential form raises the substantial confound of lexical priming and form recognition. No one would seriously question the fact that subjects can recognize a probe word faster when they have just seen the exact same word.\footnote{Gernsbacher attempts to defuse such concerns by pointing out that the probe word and proper name in the text differed visually: the probe words were capitalized, while the anaphors were in lower case. While this mollifies some of the concern around simple visual object recognition, it does not alleviate the worry about word recognition.} Informativity, of course, may have an effect above and beyond that contributed by the priming, but it is impossible to verify this. A different kind of comparison, therefore, between the pronoun and some appropriate definite description (e.g. “the loser”) would
have provided more telling evidence about the differing activation potentials of various NP types (see also Almor (1999) for further discussion of processing repeated names).

In spite of the limitations and clear methodological shortcomings of Gernsbacher (1989), it presents an intuitive idea: linguistic elements with more information present more retrieval cues for accessing previously mentioned discourse entities. Beyond this account of memory trigger quality, though, the greater degree of information actually supplies another service besides aiding retrieval. It has the residual effect that activation levels remain elevated after the retrieval and coreference process. Again, if establishing coreference depends upon finding the most activated referent, a higher activation level stemming from the explicitness of one anaphor means that the next anaphor (referring to the same referent) should require less informational content. To illustrate, consider the two minimally different texts below:

(17) a. General William Tecumseh Sherman succeeded Grant as the Union commander in 1864. General Sherman proceeded to lead his troops to the capture of the city of Atlanta, which contributed to the re-election of President Abraham Lincoln. His subsequent march through Georgia and the Carolinas further undermined the Confederacy’s ability to continue fighting.

b. General William Tecumseh Sherman succeeded Grant as the Union commander in 1864. He proceeded to lead his troops to the capture of the city of Atlanta, which contributed to the re-election of President Abraham Lincoln. His subsequent march through Georgia and the Carolinas further undermined the Confederacy’s ability to continue fighting.

The second overall reference, or first anaphoric reference, to Sherman in these examples differs in form. According to the explicitness principle, the more informative anaphor, General Sherman, should aid the coreference process more. At the same time, the choice of the more explicit form also raises the activation level associated with General Sherman. Consequently, at the following anaphoric reference in the final sentence, his subsequent march, coreference should be established more easily in (17a) than in (17b). More informative anaphors may thus aid retrieval immediately, but they also facilitate interpretation of subsequent anaphors with the same referent.

To be clear, however, these points are merely implied by Gernsbacher’s research based on what activation boosting means. To the best of my knowledge, she never experimentally
investigated the impact of anaphoric form choice on subsequent anaphors and coreference processes. Gernsbacher also never explicitly states the repercussions of activation boosting (via explicitness) on subsequent retrievals, although Ariel (2001) makes the consequences somewhat clearer. Nonetheless, given the knowledge that higher activation rates facilitate subsequent retrievals (Anderson et al., 2001; Lewis & Vasishth, 2005), these points fall out incidentally. In some ways, therefore, the current proposal can be reframed in terms of activation: more informative descriptions boost the activation level of the associated discourse entity, which effectively results in faster retrieval.\footnote{From this perspective, which I expand upon in Chapter Six, increasing information is just one of numerous ways of raising activation and thus improving retrieval. Other tactics that focus attention on the future retrieval target and cause it to be more salient may thus have very similar effects to encoding more information.}

Several major differences distinguish, however, the claims made about activation enhancement in Gernsbacher’s work and my own hypothesis. First of all, Gernsbacher’s analysis focuses on anaphors, in order to capture how the content of an anaphor guides coreference.\footnote{Gernsbacher makes clear, however, that her structure-building framework, which utilizes the mechanisms of suppression and enhancement, applies more generally to a range of cognitive experiences inside and outside the linguistic domain. Moreover, Gernsbacher and Shroyer (1989) explicitly discuss how certain NPs create stronger expectations for further references. The reason for these stronger expectations, however, do not seem to be related to the explicitness principle of Gernsbacher (1989) or the informativity of the noun phrase. Instead, certain forms conventionally mark a high probability of a further mention of the relevant discourse entity, such as indefinite this. Most importantly, the explicitness principle, as stated in Gernsbacher (1989), only makes reference to antecedent-anaphor relations.} The language of the MFH, in contrast, is purposefully oriented to avoid reference to restrictions on category or discourse function. Thus, the MFH, although compatible with the predictions of the explicitness principle and how it applies to anaphors, also predicts how the informativity of newly introduced discourse entities should impact retrievability (or accessibility, in the terms of Ariel and Gernsbacher).

Secondly, increasing information in an anaphor enhances activation, in the view of Gernsbacher, to the extent that that information matches up with the previously encoded information about the antecedent: “the greater the similarity between an anaphor and its antecedent—in other words, the more explicit the anaphor is—the more powerfully the anaphor should trigger suppression and enhancement” (Gernsbacher, 1989, p. 101). By this principle, informativity is useful in so much as it resonates with some other memory trace. It is the relationship between the linguistic form of the anaphor and its antecedent
that determines the degree of activation boosting. Again, resonance or degree of similarity between two linguistic forms, while likely an important factor in retrieval, is not a part of the principle proposed in this dissertation. I am suggesting that, holding the form and content of the retrieval cues for a discourse entity constant, the more information encoded in the representation of that entity, the faster the retrieval. Thinking about this in terms of antecedent-anaphor relations, my concern is exclusively with the informativity or explicitness of the antecedent, while Gernsbacher focuses on properties of the anaphor itself. My focus on properties of the antecedent is realized in this dissertation by looking exclusively at FGDs, where there is no anaphor, i.e. no second phonological manifestation, but merely a gap.

The explicitness principle and the MFH are in agreement about the existence of a relationship between information and retrieval. Both stories concur that, all other things being equal, informationally richer descriptions facilitate memory retrieval. They differ, however, with respect to how informativity aids retrieval. One focuses on how explicitness makes for better memory triggers, while the other focuses on how informativity makes subsequent retrievals easier. These points of view, of course, are not at all mutually exclusive. In many ways, they naturally complement one another. Together, they form a cohesive picture of memory in language comprehension, wherein informativity plays a consistent role in retrieval processes.

2.3 The Elaboration Model of Recall

A third source of support for the MFH comes from an array of experimental findings demonstrating that successful memory retrieval depends upon characteristics of the encoding process (Reder, 1980; Bradshaw & Anderson, 1982; Anderson & Reder, 1979; Anderson, 1983; Wiseman, MacLeod, & Lootsteen, 1985; Reder, Charney, & Morgan, 1986; McDaniel, Dunay, Lyman, & Kerwin, 1989). The kind and quantity of information encoded during the presentation of some stimulus, this line of research shows, has significant consequences for subsequent recall. In contrast to much of the research documented in this dissertation, the research on encoding is primarily concerned with longer-term recall of episodic memories or propositional recall, as opposed to the retrieval of linguistic information for the purpose of natural language comprehension.

As an example, Bradshaw and Anderson (1982) provide evidence that sentence recall
improves when the encoding of a proposition occurs along with other causally related propositions. In their experiments, they presented subjects with a set of short texts to study for a subsequent recall task. These texts consisted of a set of trivia facts about historical figures. Subsequently, they tested recall accuracy for the first sentence in those texts. Some of those texts contain only a single sentence, i.e. the test sentence, while other texts contained two extra causally related sentences after the target sentence. Another condition included two additional sentences that presented facts that were basically unrelated to each other and to the proposition expressed by the first sentence. A sample item is reproduced in Table 2.2. What they found was that subjects’ proposition recall was better in those cases where the initial sentence was elaborated upon by the additional, related sentences—the ‘caused-by’ or ‘resulted-in’ conditions below. In contrast, the addition of the unrelated sentences failed to improve recall. Alongside similar arguments in Reder et al. (1986), Bradshaw and Anderson construe this as evidence that only causally linked elaborations enhance recall.

Bradshaw and Anderson explain these results more generally by postulating that the addition of propositional information that relates to some core concept increases the number of possible retrieval paths to the target concept. They conceptualize the addition of information as expanding or creating a network of information that links together certain traces in memory. When retrieval of any one memory trace is licensed by a memory trigger, the linked or networked memory traces provide multiple retrieval paths, or inferential processes that depend upon the successful recall of part of that network of associated information. These inferential paths are critical because there is a chance, for any given memory retrieval, that the target concept cannot be directly retrieved. Despite the limited retrievability of any one particular propositional trace, the recall of at least one node in a network of information should be significantly higher. Connected information thus makes retrieval easier and more likely to be successful:

“Any particular encoded proposition is fragile. There is a significant chance that the subject will not be able to activate that proposition at test. However, if the subject generated a memory episode which encoded a set of multiple propositions that were partially redundant with the to-be-remembered information, he or she would have a much better chance of recalling it at the time of test. There are two basic ways that a more elaborate encoding can lead to better memory. One, which we call network redundancy, involves using alternate retrieval paths
<table>
<thead>
<tr>
<th>Condition Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caused-by condition</td>
<td>Newton became emotionally unstable and insecure as a child. This fact was caused by: Newton’s father died when he was born. Newton’s mother remarried and left him with his grandfather.</td>
</tr>
<tr>
<td>Resulted-in condition</td>
<td>Newton became emotionally unstable and insecure as a child. This fact resulted in: Newton became irrationally paranoid when challenged by colleagues. Newton had a nervous breakdown when his mother died.</td>
</tr>
<tr>
<td>Unrelated condition</td>
<td>Newton became emotionally unstable and insecure as a child. This fact is unrelated to: Newton was appointed Warden of the London mint. Newton went to Trinity College in Cambridge.</td>
</tr>
</tbody>
</table>

Table 2.2: Example of stimuli from Bradshaw & Anderson (1982)
created by the elaborations. The second possibility, which is more appropriate to sentential material than single word material, is what we call inferential redundancy. This refers to the fact that the subject may be able to infer the material he studied from remaining elaboration” (Bradshaw & Anderson, 1982).

Expanding upon this topic, McDaniel et al. (1989) find that it is not just simple elaboration upon a topic or a proposition that improves the chances for successful retrieval. Not only do they find additional evidence that elaboration aids recall, they also reason that the distinctiveness of the target information plays a further significant role in the magnitude of the effect. Distinctiveness here refers to the uniqueness of the relationship between concepts in a proposition, as opposed to how much that proposition stands out from others in the context (cf. isolation effects (Restorff, 1933; Eysenck, 1979)). On the view presented by McDaniel et al. (1989), encoding unique relationships “would favor memory performance because it would reduce interference from related encodings” (p. 358). On one level, then, any connected or related information adds to the distinctiveness of some core concept in memory: a memory trace with absolutely no connections to other information should suffer the most interference from related encodings.

The power of elaboration upon memory recall seems to extend even beyond the realm of language processing. Minimally, the retrieval targets need not be linguistic entities. In a study of photograph recall, for instance, Wiseman et al. (1985) establish that successful recognition of a photograph varies with the amount of verbal description that accompanies it. In other words, just as recall of propositions improves with the addition of related propositions, recall of visual information improves with the addition of verbal material that elaborates upon the relevant image. The influence of elaboration or informativity on memory retrieval, therefore, is arguably a general cognitive principle, not specific to language comprehension at all.

Where the MFH diverges or expands upon the elaboration model is in the consideration of real-time language comprehension. The experiments and theories in the psychology literature outlined above deal exclusively with the accuracy of long-term recall of information or propositions. In contrast, the MFH speaks to the speed of retrieval processes initiated not long after the original target presentation, constrained by the pressures of real-time language comprehension. To put it another way, working memory processes are the focus here, rather than long-term memory storage and retrieval. I am making the prediction, therefore, that elaboration has identical, or at least quite similar, effects on retrieval from
both working memory and long-term memory.

The other area in which the elaboration model and my hypothesis differ is the type of representations targeted for retrieval. As mentioned above, the research on elaboration deals almost exclusively with the recall of propositions or facts, which themselves contain linguistic subconstituents. According to the elaboration research, linking related propositions has a beneficial effect upon subsequent recall of one or more of these propositions. Hence, the crucial factor is not how much information the sentence targeted for retrieval itself contains, but how many other sentences are coherently linked to it. What I am investigating, however, is the retrieval of a discourse representation corresponding to a syntactic constituent in the course of processing a sentence. The dependent factor in this case is the number of syntactic and semantic features specified about this entity during the encoding process. It is thus an empirical question as to whether features about a discourse entity create networks of information that have the same consequences as they do for the long-term recall of propositions.

While long-term memory and working memory are often considered to be separate memory systems with different neurological centers, subject to distinct principles and constraints, it is perfectly reasonable to think that they share at least some features.\textsuperscript{12} Certainly, many of the same biases operative on longer term recall, such as repetition and distinctiveness, also have an effect on the speed and accuracy of working memory. It would not be surprising, therefore, if the amount of information linked to a memory trace had an effect on both types of memory systems.

There is therefore an intriguing relationship between the elaboration model and the MFH, one that hints that they may both be addressing a single underlying cognitive principle. Both directly point to a connection between the encoding process and subsequent retrieval of the encoded item. Once information is linked together during the encoding process, accessing information within that linked network becomes easier. The elaboration model of retrieval thus extends the motivation for the basic premise of the MFH, as a similar supposition is necessary to account for aspects of longer-term recall.

\textsuperscript{12}I should note again that the hypothesis of this dissertation does not bring with it a commitment to any particular vision of working memory. Within cognitive psychology, a fair amount of dispute exists about what the term working memory encompasses and what sort of phenomena it should account for (see, for instance, discussions in Andrade (2001)). My use of the term is only intended to refer to its ‘every-day’ sense as a temporary storage system with a limited-capacity buffer.
2.4 Summary

The MFH has historical antecedents from different research fields that all can be seen as implicating the role that the informativity of a linguistic entity has on predicting the efficiency of subsequent processing involving that entity. This relationship has far-ranging consequences for the understanding of phenomena in quite separate areas of study. It offers a way of rationalizing the gradience in grammaticality seen in a range of extraction constructions, where no unifying explanation had previously been identified. In fact, as we shall see in the next chapter, there are even more linguistic constructions for which this hypothesis is relevant. The MFH also bears on theories of anaphor resolution and the choice of anaphoric form, although it has nothing to say about how the form of anaphor determines resolution. Instead, it suggests that the choice of linguistic form for one anaphor has a predictable effect on the form choice for the next anaphoric reference to the same referent. Long-term recall of propositional information also appears to be subject to the same (or a similar) correlation. In fact, there are still other linguistic and psycholinguistic theories which can potentially be reinterpreted in light of the MFH. For instance, the anti-locality effects described in Vasishth and Lewis (2006), which show up at retrieval points in various constructions, can partly be viewed as the product of linking together information about one discourse individual. Vasishth & Lewis construe the effects as the consequence of re-accessing a syntactic category during modification, which results in a heightened level of activation. The predictions of their theory and the MFH are nevertheless the same, i.e. faster retrieval when the discourse referent is characterized with more detail.\footnote{As with the other theories described in this chapter, there are some important differences between the main thesis of this dissertation and the anti-locality work of Vasishth and Lewis, particularly with regard to the predictions that the two theories make, although the two theories are quite compatible. Chapter Six considers the processing model of Vasishth & Lewis in some detail, as well as how the MFH fits into that theory.} Cumulatively, these distinct phenomena from distinct fields add credence and plausibility to the basic assumptions behind the MFH. This support from multiple sources thus warrants further investigation into the effects of informativity on retrieval processes.

None of these earlier research programs, however, makes the claim that the quantity of information used to encode a discourse entity is predictive of retrieval ease \textit{during} language comprehension. The elaboration model perhaps comes closest to this position, but
it is concentrated exclusively on long-term recall of propositions. That is, it focuses on retrieval after sentence processing. There are unique time-sensitive pressures within language comprehension that may cause informativity effects to drastically differ from their appearance in free recall or cued-recall tasks. Moreover, elaboration refers to the (co-)presence of related propositions, rather than the encoding of additional information within one proposition. Thus, the MFH introduces a novel way of looking at how the functional requirement of retrieving linguistic material for the purpose of integrating it with incoming material is tempered by the quantity of information about the retrieval target specified during the encoding process. The remainder of this dissertation is dedicated to finding evidence for this hypothesis and to exploring the functional motivations behind it.
Chapter 3

Processing Islands

The predictions of the facilitation hypothesis are tested in this chapter in the context of so-called syntactic islands. Islands constitute the linguistic starting point of this investigation due to the apparent high processing cost they impose on comprehension. Multiple factors seemingly contribute to the cumulative difficulty of many island constructions, based on evidence from a wide array of psycholinguistic studies. In a given environment, if retrieval can take place in the minimum amount of time that the cognitive architecture will allow, then any strategy meant to facilitate retrieval will not be able to lower retrieval times any further. In contrast, islands present a suitably challenging setting for language processing, making it more likely that advantages for sentence processing will be visible.

3.1 Gradience in Islands

While linear or structural distance does not theoretically constrain filler-gap dependencies, linguists have argued since Chomsky (1962) that the acceptability of FGDs hinges on structural constraints located in the competence grammar. For example, the filler-gap dependencies in constructions like those in (18)-(20) sound unacceptable to most speakers of English:

(18) Sawyer admitted who committed this atrocity.
What did Sawyer admit $[s$ who committed ___]?

(19) The journalist wondered whether the president planned the attack.
What did the journalist wonder $[s$ whether the president planned ___]?
(20) We met the mathematician who solved the puzzle.

This was the puzzle that we met the mathematician who solved ---.

The typical explanation for the markedness of these examples is in the form of supposedly universal restrictions on extraction. These restrictions are members of a larger class of restrictions on syntactic movement that go under the term “island” constraints (Ross, 1967; Chomsky, 1973, 1977, 1981, 1986), due to the fact that certain kinds of phrases act as metaphorical islands for movement, preventing any constituents from escaping them. More precisely, these so-called islands obstruct or block leftward and/or rightward movement of any internal constituent.

Over the past forty years, the explanations for the unacceptability of these constructions have varied from strictly syntactic constraints on movement to accounts incorporating syntactic, semantic, and pragmatic considerations. Originally, Chomsky (1962) proposed a relatively simple and presumably universal linguistic rule to account for the behavior of filler-gap dependencies:

1

(21) **A-over-A condition**

An element of category A cannot be extracted out of a phrase of category A.

While general in flavor, and able to account for a basic set of extraction facts, the A-over-A constraint incorrectly predicts ungrammaticality in a variety of examples, as noted by Ross (1967) with examples like (22) and (23), and fails to predict the poor acceptability of other types of filler-gap dependencies (24):

(22) Who would you approve of [\textit{NP} my seeing ---]?

(23) Which author did you read [\textit{NP} a book about ---]?

(24) Which dignitaries do you think [[Sandy photographed the castled] and [Chris visited ---]]?

Due to these apparent exceptions to the A-over-A condition, Ross proposed a variety of much more specific principles or constraints, such as the complex noun phrase constraint and the

1Chomsky revised this principle on numerous occasions, eventually producing a version of this principle that definitively allows extraction of a phrase of type A out of a phrase of category A (Chomsky, 1973). In fact, Chomsky clearly knew of the counterexamples to the original formulation before Ross addressed them, as Chomsky composed a revision to the 1962 paper that mentions some of the problematic examples for the condition.
right roof constraint. But while these replacement conditions cover a larger empirical area
than the A-over-A condition, the constraints introduced by Ross lacked a clear uniformity.\footnote{Ross, however, does attempt to unify these constraints with the principle of A directly over A at the end of his thesis.}

Subsequent treatments of the topic, therefore, aimed to identify some more general
property that could unite these seemingly disparate constraints. The classical unifying
analysis of these constraints comes from the notion of “Subjacency” (Chomsky, 1973).

(25) No rule may move a phrase from position Y to position X (or conversely) in:

$$\ldots X \ldots [\alpha \ldots [\beta \ldots Y \ldots] \ldots] \ldots X \ldots$$

where $\alpha$ and $\beta$ are cyclic nodes.

In English, IP and NP (or DP) are cyclic nodes, meaning that complex noun phrases like
[\textit{NP the rumor that [IP they started a new company]}] and subject noun phrases, as in [\textit{IP[NP the attempt to find the fountain of youth] ended in failure]}], represent islands to movement. Hence, examples like the following are ruled out:

(26) a. It was a new company that Simon spread [\textit{NP the rumor that [IP they started \ldots]}].

b. What did [\textit{IP[NP the attempt to find \ldots} end in failure\ldots]?}

This theory of islands thus places a limit on how many cyclic nodes a filler-gap dependency
can cross; however, it is possible to construct passable examples where three or more such
nodes are crossed, as demonstrated by Deane (1991):

(27) a. Nixon was one president that [\textit{IP they had [NP no trouble finding [NP votes for [NP the impeachment of \ldots]]]}].

b. The chief purpose is [\textit{IP to avoid any sentences that [IP our informants report [NP significant variations in [NP their judgments about \ldots]]]}}].

Hence, this theory posed a more general, unified analysis of island phenomena, but still
remained limited in terms of its empirical coverage.

The notion of subjacency was eventually replaced by the concept of “barriers” (Chom-
sky, 1986). According to this theory, certain XPs act as barriers to movement or extraction,
specifically XPs that are not theta-governed (or L-marked, as in the definition below) by a
lexical category, i.e. phrases not selected by a governing lexical head. Within this frame-
work, movement cannot cross two different barriers, which are characterized in terms of
“blocking categories”:

\[(28) \gamma \text{ is a BC [blocking category] for } \beta \text{ iff } \gamma \text{ is not L-marked and } \gamma \text{ dominates } \beta\]
\[
\gamma \text{ is a barrier for } \beta \text{ iff (a) or (b):
}\]
\[
a. \gamma \text{ immediately dominates } \delta, \delta \text{ a BC for } \beta;
\]
\[
b. \gamma \text{ is a BC for } \beta, \gamma \neq \text{ IP.}
\]

This means that any ungoverned phrase, except for IPs, will pose barriers to movement.
Furthermore, any YP immediately dominating a blocking category is categorized as a bar-
rier by inheritance. Unlike the account of subjacency, the Barriers treatment explicitly
acknowledged different degrees of acceptability, depending upon the number of barriers
crossed:

\[(29) \beta \text{ is } n\text{-subjacent to } \alpha \text{ iff there are fewer than } n + 1 \text{ barriers for } \beta \text{ that exclude } \alpha.\]

Dependencies which cross zero boundaries are thus 0-subjacent and should sound perfectly
acceptable, “1-subjacency” should translate to marginal acceptability, but anything higher
“should yield a considerable decrement in acceptability” (ibid., p. 30).³

As an illustration, examples like (30a) & (30b) involve movement across two barriers,
according to Chomsky (1986). One barrier is posed by the CP, which acquires this status
since it immediately dominates a blocking category—the most embedded IP. The second
barrier is actually the IP itself, despite the exception statement in (28) that suggests IPs
cannot be inherent barriers. But to account for the apparent difference between extraction
out of a tensed island like those in (30a) & (30b), Chomsky assumes that the most deeply
embedded tensed IP constitutes an inherent barrier to movement.⁴ The examples in (30c)

³Based on the implications of the text, subjacency violations beyond 1-subjacency should be indiscrimi-
nable in terms of acceptability, because “to specify n-subjacency for higher values of n requires counters.”
In other words, a dependency that cross two barriers should incur the same degradation in acceptability as
a dependency that crosses three or four barriers.

⁴This assumption is not unproblematic, as discussed in Kluender (1991). It implies that movement out of
the most embedded tensed IP in a sentence should result in a subjacency violation on a par with movement
out of the sentential complement of a noun.
& (30d) only present a single barrier each to movement: the CP that again earns its barrier status via inheritance from the IP. In these cases, the IP is only a blocking category and not a barrier, since the IP is not tensed:

(30)  a. What did Simon spread \[NP \text{ the rumor } [CP \text{ that } [IP \text{ they started } __ ]]\]?

b. What did Harold wonder \([CP \text{ whether } [IP \text{ they had ruined } __ ]]\)?

c. What did the captain give \([NP \text{ the command } [CP [IP \text{ to start } __ ]]\]?

d. Who did Adele wonder \([CP \text{ whether } [IP \text{ to invite } __ ]]\)?

The latter two examples should thus sound better than the first two, if the predictions of the Barriers account are accurate. This theory accordingly recognizes some amount of gradience in the island data, albeit limited to only three distinct levels.

Together with the analyses of Huang (1982) and Rizzi (1990), Barriers lays the groundwork for a distinction among types of islands. A two-way division of the class of islands was made between so-called ‘strong islands’ and ‘weak islands.’ This classification capitalized upon the apparent differences in acceptability among subjacency violations, although the distinguishing characteristic for these two kinds of islands is not in terms of the number of subjacency violations. The standard motivation for assuming the dichotomy is the belief that strong islands forbid extraction of any phrase type, but that weak islands block extraction of only a limited set of phrases. Cinque (1990), for instance, offers the standard view that weak islands allow a PP-gap, but strong islands can at best contain a DP-gap, e.g. About which topic did John ask whether to talk? vs. About which topic did John ask who was talking? So, while islands came to be seen as a divided class of constraints, the division was not stated in terms of acceptability or in the number of violations the structures incurred.

Eventually, semantic and pragmatic explanations for subjacency effects developed as well, especially for the case of weak islands like tenseless \textit{wh}-islands, factive islands, and negative islands (Erteschik-Shir & Lappin, 1979; Kroch, 1998; Comorovski, 1989; Rizzi, 2000; Szabolcsi & Zwarts, 1993; Szabolcsi, 2006; Oshima, 2007; Truswell, 2007). In general, the semantic and pragmatic accounts operate from the starting point that not all islands are equally bad. Moreover, the acceptability differences do not seem explicable in strictly syntactic terms. Szabolcsi and Zwarts (1993) argue, for instance, that it is not syntactic constraints which govern extraction out of weak island contexts, but rather the semantic concept of individuation (see also Frampton (1991)). That is, the domain of an operator (= the filler-phrase) must range over individuals, rather than amounts. From this position,
they argue that the D-linking effects described in the previous chapter actually stem from turning a “non-individuated domain” into a set of “discrete individuals.” Kroch (1998), too, rejects a “pure” syntactic explanation for island phenomena. He reasons that some island effects, like the ban on moving adjunct phrases out of \textit{wh}-islands and negative islands, owe their status to pragmatic considerations. He points out that all direct interrogatives introduce existential presuppositions; however, some presuppositions are particularly hard to compute or identify. Given these assumptions, the unacceptability of an example like (31a) arises from the pragmatic oddity of the associated presupposition in (31b):

\begin{enumerate}
  \item How much money was John wondering whether to pay?
  \item There was a sum of money about which John was wondering whether to pay it.
\end{enumerate}

The presupposition in (31b) states that John wondered whether to pay some specific amount of money, but as Kroch explains, this presupposition is highly implausible: “John might plausibly wonder how much to pay, but then there is no unique sum with the property that John was wondering whether to pay IT.” In support of this analysis, Kroch adds that further contextualization which makes the presupposition easier to identify raises the acceptability of the construction. For instance, in a situation where John has been offered the opportunity to buy a car for six hundred dollars, a question like (31a) is more passable. In sum, syntactic constraints alone seem unable to adequately account for all the variation in acceptability. That said, these semantic and pragmatic accounts do not dismiss the validity of syntactic island constraints. Instead, they implicitly aim to preserve the essentials of the syntactic theories by claiming that problematic data arises from non-syntactic factors, thus redeeming the syntactic accounts.

Since the very inception of the notion of syntactic islands, non-structural variation has in fact been recognized as affecting the acceptability of syntactic dependencies into islands, as Kluender (1991, 1998) notes. In short, various linguistic factors that do not change or eliminate the number of barriers or cyclic nodes crossed by the dependency still have a seemingly profound effect on how good or bad the island constructions sound. Ross (1967), for instance, observed that tense significantly alters the acceptability of certain kinds of filler-gap dependencies, as well as modality introduced by periphrastic phrases like \textit{make the claim that} or \textit{have hopes that}.

\begin{enumerate}
  \item You are making the claim that the company squandered a large amount of money.
\end{enumerate}
b. How much money are you making the claim that the company squandered?

As discussed extensively in the previous chapter, the content of the extracted element (\textit{who} vs. \textit{which doctor}) also appears to be a determinant of acceptability in various island constructions.

(33) 
\begin{enumerate}
\item a. Which article don’t you remember who wrote?
\item b. What don’t you know who wrote?
\end{enumerate}

Lexical choices for complementizers and verbs also play in role in the acceptability of these filler-gap dependencies. Kluender and Kutas (1993b), for instance, present evidence that the choice of lexical complementizer and associated embedded clause type impacts acceptability:

(34) 
\begin{enumerate}
\item a. What did he wonder [if we should consider ___]? ≥
\item b. What did he wonder [when we should consider ___]? ≥
\item c. What did he wonder [who should consider ___]?
\end{enumerate}

Chomsky himself notes an asymmetry in acceptability between examples like those in (35), which differ only in the lexical realization of the complementizer (the grammaticality diacritics belong to Chomsky (1973)):

(35) 
\begin{enumerate}
\item a. What crimes does the FBI know how to solve?
\item b. *What crimes does the FBI know whether to solve?
\end{enumerate}

Additionally, the severity of CNPC violations appears to vary with the specificity or referentiality of the island-forming head noun.

(36) 
\begin{enumerate}
\item a. That’s the article that we need to find someone who understands ___ ≥
\item b. That’s the article that we need to find the reviewer who understands ___ .
\end{enumerate}

This tendency intriguingly matches the contrast observable in center-embeddings, where lessening the specificity of embedded subjects improves the examples:

(37) 
\begin{enumerate}
\item a. The host [someone [I knew ___] brought ___] left. ≥
\item b. The host [the boy [the girl [I knew ___] brought] left.
\end{enumerate}

As one last example, direct question formation (especially without prior context) as in (38b) is typically worse than relativization or extraction of a referential NP as in (38):
(38) a. That’s the article that we need to find someone who can understand anything ever.

b. What do you need to find someone who can understand?

This rampant gradience defies an explanation in strict syntactic terms, even when some notion of accumulating violations accompanies the syntactic theory. To begin with, the preceding examples do not differ with respect to the number of violations they incur (in terms of syntactic boundaries or other relevant constraints, such as the Empty Category Principle (Huang, 1982; Chomsky, 1986)). Preserving the syntactic explanation consequently requires the introduction of other grammatical principles. Also of importance, however, is the fact that any account of this gradience ultimately must deal with the number of possible acceptability distinctions. The Barriers analysis, for instance, acknowledges only three levels of grammaticality, yet there is no evidence that this number is sufficient to describe the gradience in the relevant dataset. Examples can easily be constructed which illustrate five or more levels of gradedness. Further levels of grammaticality could be stipulated in the grammar, of course, but it is not clear that any finite number would actually be adequate. And as Chomsky says, if languages do not employ “counters” for the purpose of tracking violations, it becomes difficult to justify the existence of some large number of grammaticality levels.

Lastly, the transformationalist accounts outlined above make the general prediction of the universality of these constraints. Yet, some languages such as Swedish (Allwood, 1976; Engdahl, 1982; Andersson, 1982), Danish (Erteschik-Shir, 1973), Icelandic (Maling, 1978), Norwegian (Taraldsen, 1982), Italian (Rizzi, 1982), French (Sportiche, 1981; Hirschbühler & Valois, 1992), Akan (Saah & Goodluck, 1995), Palauan (Georgopoulos, 1985, 1991),

5It should be noted, though, that certain island constraints are not predicted to be universal, at least according to some. Rizzi (1982) makes this point with respect to the wh-island constraint, while referring to counterexamples of the wh-island constraint in Italian: “Given the assumption that TSC[Tensed S Condition] and SSC[Specified Subject Condition] admit language specific exceptions (‘escape hatches’), as far as I can see nothing in this framework can exclude the existence of a grammar whose specific ‘escape hatch’ allows extraction from a wh-clause” (p. 50). In English, this escape hatch is a rule that allows movement from complementizer positions to subjacent complementizer positions (thereby allowing constructions like Who did you think Bill saw?), but this movement is theoretically blocked in wh-islands because a complementizer position cannot be doubly filled. Rizzi suggests, however, that other languages may possess unique ways around this problem. In addition, Rizzi argues that languages may differ with respect to the identity of bounding nodes. Thus, while S (IP) is a bounding node in English, S’ (CP) is a bounding node in Italian.
Malagasy (Sabel, 2002b), Chamorro (Chung, 1994), Bulgarian (Rudin, 1988), Greek (Alexopoulou & Keller, 2003), Yucatec Mayan (Elisabeth Norcliffe, p.c.), and likely others, exhibit clear counterexamples to this predictions:

(39) a. Den dår gamla skräpögen känner ja killen som köpte?
   That old piece junk know I the guy who bought.
   ‘That old piece of junk, I know the guy who bought (it).’

   b. Vilken bok kunde ingen minas vem som skrivit?
   Which book could no one remember who that had written?
   ‘Which book could no one remember who had written?’


(40) Petta er lagið, sem enginn vissi hver sandi.
   This is song-DEF that no one knew who wrote.
   ‘This is the song that no one knew who wrote.’

   Icelandic: Maling (1978)

(41) Pion anarothikes an that apolisoune?
   who-ACC wondering-2SG whether/if will fire-3PL
   ‘Who did you wonder whether they will fire?’


(42) Voilà la personne [que vous ne sauriez imaginer [avec quelle sauvagerie la police secrète a essayé de [faire parler]]].
   ‘This is the person that you can’t imagine with what brutality the secret police tried to get to talk.’

   French: Hirschbühler and Valois (1992)

(43) Hafa na kareta guäha mayulang ramienta in-isa pàra
   what? L car AGR.exist broken tools WH[OBJ].AGR-use FUT
   infa’maolik
   WH[OBJ].AGR-fix
   ‘Which car were there some broken tools that you used in order to fix?’

---

6To be clear, these languages are not identical when it comes to which island constraints are clearly violable. Some of these languages (e.g. Icelandic) seem to readily permit dependencies into wh-islands, while other island constructions in the same language are nevertheless degraded in acceptability.
(44) maax t-u tuklah Juan wah t-u ts-uts'-ah x-maria-o
who PERF-A3 think Juan quest.part PERF-A3 kissed-COMP FEM-Maria-DEM
‘Who did John wonder whether Maria kissed?’

Yucatec Mayan: Elisabeth Norcliffe, p.c.

As perhaps the most infamous source of counterexamples to island constraints, the Scandinavian languages are recognized for yielding constructed examples that sound passable, but for also exhibiting tokens of these violations in everyday speech. Besides the examples shown above, Swedish texts also attest to the use of constructions that violate other island constraints such as the bans on extraction from subject NPs and sentential subjects (Ejerhed, 1982). In fact, considering all the world’s languages, it is not clear that there is any one proposed island constraint which is universally inviolable. Even the Conjunct Constraint (part of Ross’ Coordinate Structure Constraint) has been proposed to be violable in certain languages (Johannessen, 1998).

The island data is thus characterized not only by gradience and variability in judgments, but also exceptions within and across languages. Historically, the major theoretical accounts of this data have attempted to present a unified syntactic analysis, but they have been

---

7Despite recognizing the cross-linguistic variability surrounding CNPC, subject, and wh-island violations, Sabel (2002a) asserts that adjuncts are universal islands and that “no language allows for extraction of adjuncts across wh-islands.” This assertion, however, seems too strong. A painless internet search uncovers a plethora of examples like the following in English:

(1) a. Among his most famous works are the Mass in C Minor and the Requiem, which he died before finishing.
   [www.dallasnews.com/sharedcontent/dws/dn/religion/stories/012806dnrelmozart.1f63dee2.html]

b. Reynolds completed Sayers’ translation of The Divine Comedy, which Sayers died before finishing.

c. It is one of the ingredients in the Science Diet dog food that my dogs got sick after eating.

Similar passable examples are not difficult to construct. In fact, the acceptability study reported at the end of this chapter confirms that adjunct island violations are likely subject to similar processing pressures as other island constructions. The empirical claim made by Sabel regarding adjunct extraction across wh-islands is suspect, as well, for reasons highlighted in the discussion of Kroch (1998). Given the proper contextualization, such filler-gap constructions sound reasonably acceptable, although perhaps still somewhat difficult to process.
repeatedly met with clear counterexamples and problematic subtleties of judgment. In turn, this has led to a virtual panoply of accompanying proposals that have aimed to preserve the core beliefs/principles of the syntactic theory while explaining away the “peripheral” data that counterexemplify the syntactic accounts. So while on some level there is consensus within the field that the full set of island data cannot be dealt with in purely syntactic terms, there is still an impetus to approach as much of the data as possible in syntactic terms. Virtually unrecognized in much of the literature reviewed here, however, has been the role that processing difficulty plays—the topic of the next section.

3.1.1 Processing Perspectives on Island Effects

In contrast to the accounts labeling island constructions as unacceptable due to categorical constraints located in the grammar itself, a number of sources trace the acceptability of structures involving island violations to the processing complexity of these constructions (Deane, 1991; Kluender, 1991; Kluender & Kutas, 1993b; Alexopoulou & Keller, 2003; Kluender, 1998, 2005; Alexopoulou & Keller, 2007; Hofmeister et al., 2007; Sag, Hofmeister, & Snider, 2007; Hofmeister, 2007). The motivation for these processing accounts partly stems from the clear gradience associated with these constructions, the numerous counterexamples to the proposed constraints both within and across languages, as well as the lack of a clear, unified principle that can handle the known data.

More importantly, island violations involve many factors known independently to contribute to processing difficulty. For starters, a sentence with a filler-gap dependency imposes a relatively high degree of processing difficulty, compared to a minimally different sentence without the dependency (Wanner & Maratsos, 1978; King & Just, 1991; Kluender & Kutas, 1993a; Hawkins, 1999).

“Filler-gap dependencies are difficult structures to process . . . Identifying the gap is not easy. It is an empty element with no surface manifestation and its presence must be inferred from its immediate environment. At the same time, the filler must be held in working memory, and all other material on the path from filler to gap must be processed simultaneously, and the gap must be correctly identified and filled” Hawkins (1999, pp. 246-7).

Beyond this baseline difficulty carried by all types of filler-gap dependencies, island violations typically require relatively long filler-gap dependencies. Despite the fact that filler-gap
dependencies are, theoretically speaking, unbounded in length, psycholinguistic research provides strong evidence that processing difficulty increases with dependency length. (Gibson, 1998; Gibson & Pearlmutter, 1998; Gibson, 2000; Hawkins, 1999; Grodner & Gibson, 2005; Fiebach, Schlesewsky, & Friederici, 2001). In the relative constructions below, for instance, the increased distance between the subject NP and gap site (identified by e) in (45b) causes increased processing difficulty around this retrieval point:

(45) a. The administrator who_i the nurse supervised e_i scolded the medic.
   
   b. The administrator who_i the nurse who was from the clinic supervised e_i scolded the medic.

Furthermore, the fact that the distance between the relative pronoun and the gap is longer in object relatives (e.g. The boy who Tom scolded ___ ran away), as opposed to subject relatives (e.g. The boy who ___ scolded Tom ran away), can explain why numerous measures of processing difficulty show increased effort in the object RCs (Wanner & Maratsos, 1978; King & Just, 1991). Similarly, the processing difference between nested and non-nested structures with identical meanings can be understood in terms of dependency length.

(46) a. [The scientist collaborated with the professor [who had advised the student [who copied the article]]]]].

   b. [The student [who the professor [who the scientist collaborated with] had advised] copied the article].

In the non-nested sentence, (46a), the distance between dependents is minimal, making the sentence relatively easy to process. But the nesting or center-embedding process evidenced in (46b) increases the distance between the dependents, which raises the retrieval and integration costs. Many types of island-constructions, such as complex noun phrase extractions and wh-island violations, of course, are characterized by particularly long filler-gap dependencies. This feature, therefore, makes them particularly susceptible to the influence of other compounding comprehension difficulties.

Besides this factor of dependency length, which by itself would hardly be sufficient to account for perceptions of ungrammaticality, examples of island violations in the syntactic literature also tend to contain a relatively high number of discourse-new characters, describe

---

There are, however, some well-noted exceptions to this principle. Generally, these exceptions involve material interceding between filler and gap that improves the predictability of the subcategorizing head or else improves the retrievability of the dependent phrase. See, for instance, Vasishth and Lewis (2006).
numerous events and sub-events, and have complex propositional meanings. Deane (1991) suggests that the ability to form a dependency is contingent upon the ability to attend to both the extracted element and the retrieval site simultaneously—in other words, the ability to retrieve an extracted element (or have it in some sort of attentional focus, in his terms) and integrate it at the correct gap site. These attentional demands are impaired by distractions that occur along the filler-gap path, according to Deane:

“Now, in a sentence involving extraction, we have hypothesized that the extraction site and the extracted NP must command attention. Of course, they would elicit attention most readily in the absence of competing elements (distractors). It would follow that the rest of the sentence should be relegated to the background, that is, presupposed or at least given . . . Failures of extraction come when these conditions are not simultaneously satisfied.” (p. 36)

Thus, the acceptability difference between (47a) and (47b) emerges as a result of the extra processing that goes on at the verb in the manner-of-speaking example, which ultimately distracts attention or expends cognitive resources otherwise spent on the filler-gap dependency.

(47)   a. Who did you say that John believes you saw?

       b. Who did you lisp that John believes you saw?

Deane claims that the informational focus of (47b) is at lisp, which violates a stated condition that the extraction site be focused (the most embedded complement clause, in this case). There are, however, other available processing-based explanations for this contrast, such as the possibility that the lower lexical frequency of the verb lisp contributes to the difficulty. Either of these specifications is nevertheless compatible with Deane’s overarching point that the more processing or attention needed for elements between the filler and gap, the harder it becomes to link the filler and gap.

It has also been shown experimentally that the linguistic form used to present discourse entities seriously affects the processing of filler-gap dependencies (Warren & Gibson, 2002, 2005). Among other results, Warren and Gibson (2005) provide reading time evidence that definite NPs and proper names that intervene along a filler-gap path cause slower reading times at the retrieval site than intervening personal pronouns do.
Basically, the explanation for this contrast amounts to the claim that certain types of NPs incur a higher processing cost than others. Warren & Gibson attribute this cost difference to accessibility, but whatever the underlying mechanism is, their results position informationally light descriptions like pronouns at one end and informationally heavier descriptions like definite NPs at the other end of the scale. Given the limited cognitive resources available for linguistic processing, expending those resources between a filler and gap increases the overall difficulty of processing the dependency at the retrieval and integration site, as the resources necessary to complete that dependency have been stressed by processing intervening constituents. The ability of this factor to modify acceptability judgments, in fact, has already been implicated in CNPC violations and center-embedding constructions like those in (36) and (37). For syntactic island constructions, these experimental results hence implicate the form of intervening NPs as another source of potential processing difficulty.

Beyond these factors, finiteness also enters into the equation of processing difficulty. Again, this point has its origins in comments made in Ross (1967). In short, dependencies into non-finite clauses are easier than dependencies into finite clauses:

(49)  a. What were you unsure how to file? ≥

   b. What were you unsure how you should file? ≥

   c. What were you unsure how they had filed?

Unlike some other factors, the impact of finiteness has actually penetrated the syntactic generative literature. The well-known characterization of \textit{wh}-islands, for instance, crucially depends upon the finiteness of the embedded clause (e.g. Chomsky (1973)). Embedded finite \textit{wh}-clauses typically represent strong islands, while non-finite \textit{wh}-clauses pose mere weak islands or else a syntactic context only mildly resistant to extraction. From a processing perspective, these differences are understandable as the result of either adding an event to the discourse representation or not (Gibson, 2000). Just as processing a referential noun phrase imposes a cost, since it requires adding that representation to the discourse model and tracking that participant, a finite clause also introduces an event that must be stored and integrated with the developing discourse model. Non-finite clauses, in contrast, do not add any events to the discourse context, which effectively means that less work is necessary when processing a non-finite clause.
Of course, many syntactic island violations also require a filler-gap dependency to cross a clause boundary. Kluender & Kutas (1993) argue that clause boundaries present a processing bottleneck, based on evidence from event-related potentials (ERPs) and acceptability judgments. They report that, even in yes-no questions where no filler is carried across the clause boundary, different complementizers elicit significantly different neurophysiological responses, as well as acceptability judgments. The complementizer *that* produces the highest rating of acceptability, while the bare *wh*-word *who* garners the lowest rating:

(50)  

(a) Has she forgotten [*that* he dragged her to a movie on Christmas Eve]?  

(b) Has she forgotten [*if* he dragged her to a movie on Christmas Eve]?  

(c) Has she forgotten [*who* he dragged ___ to a movie on Christmas Eve]?

This acceptability ranking is argued to be the result of processing differences at the complementizer. In particular, differences in N400 amplitude across conditions are interpreted as signaling significantly different degrees of referential processing load. The ERP results from Kluender & Kutas show larger N400 amplitudes in yes/no questions when the interrogative pronoun *who* is used rather than the complementizers *that* or *if*. Kluender & Kutas discuss the difference across complementizers in terms of referential processing and “lexical complexity”, appealing to the intuitive idea that these complementizers specify distinct ways of treating the embedding material. A *that*-clause, for instance, introduces just a single proposition, but an *if*-clause requires consideration of two contrasting alternatives. Interrogative pronouns like the one in (50c) trigger additional referential processing and the beginning of a filler-gap dependency. The difficulty of processing a complex embedded constituent is further compounded by the simultaneous processing of another filler-gap dependency, e.g.  

*What did she forget *who he dragged ___ to ___ on Christmas Eve?*  

Chen, Gibson, and Wolf (2005) also point out that clause boundaries initiate a new set of predictions for upcoming constituents, while the predictions for the preceding clause must be temporarily shelved in memory storage. The frequency of clause boundaries in island phenomena thus seems to be intimately tied up with the complex cognitive processes that automatically occur at those boundaries.

---

9 The N400 is an event-related brain potential (ERP) component characterized by a temporally localized negative (N) voltage deflection that appears 400 milliseconds post-stimulus presentation. Typically, the N400 component is associated with semantic anomalies and expectation violations. For the purposes of this discussion, however, the precise interpretation of the N400 is not critical; the important point is that higher N400 amplitudes mark anomaly detection, which is likely to lead to decreased processing efficiency.
Certainly, other linguistic choices may also have some part in the determination of processing difficulty within islands. Included in this list are such forces as similarity-based interference (Lewis, 1996; Gordon et al., 2001; Van Dyke & McElree, 2006), lexical frequency and predictability (Jurafsky, 2003; Levy, 2005), not to mention plausibility and coherence. These factors do not necessarily pose a problem for any given island construction; however, they can exacerbate the cognitive challenge that is inherent in islands. Many examples of islands in the literature, moreover, do reflect the imprint of these factors.\textsuperscript{10}

Furthermore, there are well-known acceptability differences associated with whether the extracted element is an argument or an adjunct phrase. Potentially, this too may echo processing-related differences, although data on this issue is quite limited.\textsuperscript{11} Intuitively, however, it seems reasonable to expect that adjunct \textit{wh}-phrases would create greater processing difficulties than argument phrases. Considering the arguments of Kroch (1998) alone, adjunct questions typically demand more effort for constructing the relevant existential presupposition and imagining an appropriate discourse context for the question. This consideration by itself hints that the extraction differences between arguments and adjuncts has some roots in processing issues.

The issue of contextualization (imagining a suitable discourse setting) just touched upon

\textsuperscript{10}For instance, as proof of a ban on extraction from an adjunct, Haegeman (1994) provides the example \textit{Where did Bill go to Rome to work?} The unacceptability of this example, and others like it, is clearly tied up with coherence. Furthermore, reading this question for the first time will almost undoubtedly lead down the garden-path. Modifying the example slightly, but preserving the basic structure, the acceptability of the FGD can be much improved: \textit{For which company did Bill go to Rome to negotiate?} or \textit{Where did Bill go in Rome to relax?}

A common problem for many of the cited examples is that they often contain possibilities for misparsing the sentence. As one further example of this, Haegeman also cites the following as a violation of the complex noun phrase constraint: \textit{Whom do you know the date when Mary invited?} Not only is this example extremely hard to situate contextually, but the verb \textit{know} presents a reasonable location to try and integrate the interrogative pronoun. This potentially leads to a misparsing, incurring heavy processing costs and causing the sentence to be deemed unacceptable.

\textsuperscript{11}In fact, there are no studies that I am aware of which explore the contrast between FGDs with adjuncts as fillers, as opposed to argument phrases. This is not surprising, though, since it would be impossible (as far as I can tell) to construct minimally different stimuli that did not differ drastically in meaning.
is the final processing factor I will consider.\textsuperscript{12} Just as adjunct questions require more contextualization than argument questions, comprehending a decontextualized direct interrogative undoubtedly requires a little more imagination than it does to parse a declarative. The “disembodied” direct questions used throughout the syntax literature have to be put in the context of an appropriate situation where such a question could be asked. Plus, going back once more to Kroch’s point, direct questions carry an existential presupposition with them that plain declaratives do not. Of course, declaratives usually sound better and are generally easier to process when they appear in context, too. Effectively, most linguistic examples have an automatic processing penalty against them for not having any supportive context.

Cumulatively, island effects look to be wrapped up with numerous challenges for sentence processing. From island to island, these cognitive challenges undoubtedly have different weights and effects. Indeed, certain island effects may be quite independent of processing difficulty. Suggesting that some island effects have their source in processing difficulty does not entail that \textit{all} island effects can be traced back to processing. Nevertheless, the evidence is compelling that many island constructions have characteristics which have been independently identified as adding to processing difficulty.

Given the high processing costs associated with syntactic island constructions, and the intuition that acceptability varies significantly with the form of the extracted element, islands therefore constitute an optimal setting for testing the predictions of the MFH. Simple, unary \textit{wh}-questions are presumably processed without much difficulty and thus memory retrieval processes may be operating at optimal efficiency. Syntactic islands, in contrast, pose various well-known difficulties which likely drive processing cost past some threshold where the sentences in question become uninterpretable or unacceptable exemplars of English. Removing one or more of these difficulties, therefore, may bring the cumulative processing difficulty below this critical threshold. In this sense, syntactic islands represent a suitable place for beginning this investigation.

In the remainder of this chapter, I consequently describe a series of experiments on \textit{wh}-dependency processing in island contexts. The first experiment investigates the effect of informativity upon the processing of complex noun phrase violations, while the next two experiments consider the same effect in the context of \textit{wh}-islands, looking at the extraction

\textsuperscript{12}Other processing considerations not mentioned here may also be relevant to the comprehension difficulty of certain islands.
of both argument and adjunct phrases. The results consistently show a processing advantage for using a more informative filler-phrase. It is critical to point out, however, that these processing results have potentially different explanations, depending upon the syntactic context. Because informativity impacts reading times prior to retrieval in complex noun phrase constraint violations, some other explanation besides memory facilitation must be invoked to account for the early reading-time differences.

3.2 Methodology

Self-paced reading tasks are employed in most of the experiments discussed in this thesis. In these comprehension experiments, subjects read sentences at their own pace on a computer screen (Just, Carpenter, & Woolley, 1982). Initially, they are presented with a screen of dashes separated by spaces, representing the words for that experimental item, so that they have some sense of the number of words in the sentence. With each press of a predefined key, a new word appears on the screen and the previous word disappears. An example state for such a method is given below, where the first three words have already been read and the fourth word in the string is in the process of being read.

(51) — — —— confirm ——- —- ——– – — ———

The amount of time between each press of the key is automatically recorded and tagged with relevant information, such as the identity of the lexical item, the order in which the particular sentence was presented amongst all items, the subject ID number, the item’s experimental condition (if not a filler), etc. When comparing two minimally different conditions, longer reading times at a particular word or region are interpreted as an indication of processing difficulty.

Instructions varied slightly from experiment to experiment. Generally, though, subjects were informed beforehand that they would be participating in a reading comprehension task and that their ability to understand small passages of English was being tested. Between six and eight practice examples preceded the real experimental items to acquaint the participants with this manner of reading and reduce the magnitude of order effects on items presented early on in the experiment. During the experiments proper, each participant saw each item in exactly one condition (Latin-square design). Blocking of the items into lists and randomization within each list was all automatically handled by the reading-time
software, LINGER, used for all the reading-time studies described here.

All experiments reported on here involved some sort of check on comprehension to ensure that subjects paid attention and understood the meaning of the items. In some cases, the comprehension questions were presented in the same word by word method. In others, the questions appeared on a separate screen with all words in the question simultaneously visible. After reading the question, subjects responded either by selecting the correct answer from a set of possible choices or by responding 'yes' or 'no' in the case of yes/no questions. Response times for answering the question, as well as question-answer accuracy for these answers, were also collected and reported here where relevant. If the mean question-answer accuracy for a subject was below 67%, the entire data set for that individual was dropped from the overall analysis. Moreover, if a subject’s global reading time average differed from the sample’s global average by more than 2.5 standard deviations, the results for that subject were excluded. Since reading times have a naturally imposed minimum time, i.e. it’s impossible to read a content word in fifty milliseconds, subjects with global reading averages under 200 ms were also excluded, even though this reading time might not be excluded by considering only standard deviations. Typically, subjects with reading times at such a rate also had a corresponding poor question-answer accuracy, cumulatively indicating they were not reading with comprehension.

Residual reading times were derived for each word on the basis of a linear regression equation that computes reading time for each individual as a function of word length. Thus, a subject’s residual reading time for a given word of length $n$ expresses how fast that word was read compared to that subject’s average reading time for other words of length $n$. So a negative residual reading time means a comparatively fast reading time. This practice of using residual reading times effectively reduces variability due to individual differences in reading times (Ferreira & Clifton, 1986). Where I report statistical comparisons, only the residual reading times are considered, rather than the raw reading times.

The results from all experiments were analyzed with repeated measures ANOVAs which take into account the random factors of subjects and items. In experiments where more than two treatment or condition levels are compared with one another, the results of planned t-tests are also reported. As repeated hypothesis testing with the same sample increases the likelihood of false positives (Type I error), the family-wise (FW) error rate must be corrected in such situations. For the purposes of this dissertation, I will exclusively employ the Bonferroni correction, which compensates for the higher FW error rate.
text itself, I report only the uncorrected p-values for these comparisons, although I note for each experiment what the relevant cut-off level for significance is after correcting for the increased error rate. In appendix B, I also include the corrected p-values for regions of interest. For all critical comparisons of reading times, I also include the min F’ values (a conservative estimation of the lower bound of the F’ statistic where both subjects and items are simultaneously treated as random variables) in the second appendix (Clark, 1973; Raaijmakers, Schrijnemakers, & Gremmen, 1999). Min F’ values for other dependent measures such as acceptability or response times are included directly in the text.

In several cases, I also include the results of linear mixed effects modeling on the residual reading times directly within the text. This method of statistical analysis not only avoids the loss of statistical power that comes with prior averaging, but as Baayen (2004) points out, it allows for a principled way of incorporating longitudinal effects and covariates into the analysis. Moreover, multilevel models are well-adapted to the consideration of nested random effects, such as effects of trial number (i.e. the position of the item in the experiment) that differ from subject to subject. Besides these considerations, it is also a rather trivial matter to evaluate the effect of including a particular random effect in the model. In many ways, therefore, multilevel models present a superior means for quantitative data analysis. For all experiments described here, therefore, I report the results of linear mixed effects modeling, including model comparisons, in Appendix C.

3.2.1 Removal of statistical outliers

Due to the high processing cost of many of the syntactic structures considered in this dissertation, reading times occasionally rose to extraordinary limits. These especially high reading times (sometimes over 2000 milliseconds for a single word) often reflect serious parsing errors, a lack of comprehension, or merely a random slowdown for irrelevant reasons such as being distracted. Inspection of the data reveals that, even considering only a single subject’s data, such data points lie far beyond the normal range of reading times. It is consequently necessary to remove such data points from further statistical analyses, as such extreme scores can heavily bias the results. As in many reaction time studies, the problem is that anomalous data points due to factors and elements beyond the purview of the experiment can be scattered amongst legitimate data points. In dealing with highly complex constructions, in fact, we expect a certain amount of reading times to show up past the boundaries of typical reading rate averages.
CHAPTER 3. PROCESSING ISLANDS

To determine which data points should be counted as statistical outliers and subsequently removed, numerous statistical strategies are open and can yield significantly different results, depending on the shape of the data (Ratcliff, 1993). For each data set in this dissertation, one of three strategies was pursued based primarily on the number and distribution of the outliers; however, all of these strategies focused only on the removal of extreme outliers that typically constituted less than 3% of the data—all “mild” outliers were preserved in the statistical analyses.\textsuperscript{13} In cases where each experimental condition produced comparatively few extreme outliers and roughly the same number, all residual reading times with a z-score more than 2.5 standard deviations from a subject’s mean were excluded. In some experiments, however, certain conditions led to a particularly high number of extreme outliers at a particular word region, while other conditions exhibited relatively few. To use the same criterion for outlier removal in this circumstance would likely eliminate legitimate data points and heavily distort the data. So to ensure that no such bias occurs, an equal number of extreme outliers were removed from each condition. Typically, this translated to the excision of the five or ten slowest readings times for each condition. For experiments like reaction time studies and reading-time studies, extreme outliers typically occur in the far right tail (i.e. slow response times), thus making this method for outlier removal particularly effective.\textsuperscript{14} The third strategy amounted to simply eliminating all extreme outliers for each condition from the data. This strategy was reserved for situations where the total number of data points was relatively small, meaning that extreme values carried excessive weight in the statistical analysis.

\textsuperscript{13}Extreme outliers are defined as data points that lie more than three times the value of the interquartile range (the difference between the third and first quartiles of the data) beyond one end of the interquartile range. As a quick example, if the interquartile range equaled 50 and the upper bound of the third quartile was 75, then any point greater than 225 (75 + 50 x 3) would be considered an extreme outlier. Data points that appear past an interquartile boundary by 1.5 to 3 times the interquartile range are typically considered “mild” outliers.

\textsuperscript{14}It should be noted, however, that (except for the few cases where the findings were only marginally significant) the particular outlier analysis chosen for the data does not seriously impact the results. For instance, in those experiments where I remove an equal number of outliers per conditions, post-hoc analyses revealed that nearly identical results ensue from eliminating data points with z-scores more than 2.5 SDs from the mean.
3.2.2 Interpreting reading time results

Self-paced reading studies provide a researcher with a means for identifying differences in reading times across experimental conditions. Unlike some other experimental methodologies used in linguistics (e.g. eye-tracking, event-related potentials, or acceptability results), a reading time at a particular word in a particular condition has an inherent interpretive value or significance, irrespective of the reading times in other conditions. This is because we have a good idea of how long an average individual should spend reading a word of a given length. Thus, if someone takes 800 milliseconds to read a relatively short, frequent word in English like *chair* or *walk*, it is reasonable to assume that something in the proximal reading context has made this word hard to recognize or integrate.

The time it takes to read a single word, of course, depends upon more than just features or properties of the word itself. To a large extent, reading times at any given point in a text are strongly influenced by reading times prior to that word. If someone spends a relatively long time reading a word, compared to some average reading rate for that individual in a similar reading context, the next word is likely to be read relatively faster. And if a word is read extraordinarily fast, then the probability increases that the subsequent word will be read at a relatively slower rate. Put more simply, the reading time at the previous word acts as a predictor for any given word’s own reading time.

Interpreting the reading time results for any one word region, therefore, requires some inspection of the surrounding word regions. One recurring feature of the reading time studies in this dissertation, for instance, is the manifestation of spillover effects. Spillover results from continued processing of a particular lexical item or region into subsequent regions. Since most language processing tasks typically are followed by another, the completion of one task (especially complex and challenging processing tasks) often overlaps with the processing of one or more subsequent tasks. In fact, the processing difficulty associated with some word region may not show up during that region at all; rather, it may exclusively surface on the following word region. Evaluating the significance of some factor at a particular word region thus often necessitates including subsequent word regions in the statistical analysis.

Along similar lines, reading times at preceding regions factor into the analysis of reading time differences. A reading time difference across conditions at a particular word region may not be due to any process strictly associated with that region. As I just mentioned, spillover effects have the potential to seriously affect reading times. So an observed difference at one word may actually be the consequence of a difference engendered by processes initiated
at an earlier point. These considerations of spillover effects are particularly important in processing investigations such as this one which are concerned with the localization of processing effects.

Finally, reading studies produce two other dependent measures that can be used to evaluate comprehension: response times and question-answer accuracy. To ensure that subjects are not simply blindly pressing a key as fast as they can to finish the experiment, comprehension questions are included after each item to check that the subject read for comprehension. By utilizing comprehension questions that ask about various aspects of the items, including the gist of the sentence, it is possible to compel subjects to read relatively closely. Subsequent analysis of these question-answer accuracies can therefore be used to determine the effect of systematic manipulations on understanding at a global level. Quite related to this, the time it takes to decide upon an answer is also recorded, which provides a third measure for measuring processing difficulty. It is, in fact, quite possible and reasonable that processing difficulty shows up in one of these dependent measures but not the others. Upon encountering some parsing difficulty, a comprehender may choose to deal with that difficulty immediately, which translates to an impact on reading times, or postpone consideration of that difficulty until after the entire sentence has been parsed. This latter scenario would equate to an effect on response time or question-answer accuracy. For this reason, all three measures are reported in this dissertation for each experiment.

### 3.3 Experiment I: Complex Noun Phrases

The original formulation of the Complex Noun Phrase Condition in Ross (1967) places the following limitation on filler-gap dependencies:¹⁵

(52) The Complex NP Condition

No element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation.

As detailed above, this condition on movement was eventually subsumed under the umbrella of Subjacency and its own replacements. But, as I pointed out, numerous distinct linguistic factors intuitively enter into the calculation of acceptability for these constructions. Chung & McCloskey (1982), for instance, observe that the specificity of the extracted element

¹⁵The experiment described in this section was part of a joint research project with Ivan Sag and Neal Snider.
affects the overall acceptability of the construction: extraction of an indefinite like a paper is preferred over extraction of a wh-phrase, while a which-N’ phrase is better than an adjunct phrase and an interrogative pronoun like what. Of course, a strictly syntactic formulation of the CNPC would fail to predict any difference in these examples (absent any evidence that these structures are syntactically different).

(53) a. This is a paper that you really need to find someone you can intimidate with. 
   ≥
   b. Which paper do you really need to find someone you can intimidate with? ≥
   c. How many papers do you really need to find someone you can intimidate with? ≥
   d. What do you really need to find someone you can intimidate with?

In addition to the referentiality or specificity of the filler-phrase, the acceptability of an extraction out of a complex noun phrase also appears to intuitively vary with the type of the island-forming NP. Kluender (1992) orders the acceptability of CNPC violations in the following way from best to worst:

(54) a. This is the paper that we really need to find someone who understands. ≥
   b. This is the paper that we really need to find a linguist who understands. ≥
   c. This is the paper that we really need to find the linguist who understands. ≥
   d. This is the paper that we really need to find his advisor, who understands. ≥
   e. This is the paper that we really need to find John, who understands.

As Kluender notes, “a referentially specific head noun in an embedded complex NP necessitates the mental identification of an extra referent in addition to the logical subject of the entire predication” (239). Essentially, more complex constituents intervening between a filler and its gap consume processing resources that could otherwise be applied to processing the filler-gap dependency.

The first experiment reported here thus inspects the role of two independent factors—the informativity of the extracted element and the type of island-forming NP—in the processing of CNPC violations. The MFH makes the clear prediction that processing should be faster when the extracted element encodes more information. Moreover, it predicts that this effect should be visible at the retrieval site. And while the MFH says nothing about how the type
of island-forming NP should bias processing, the concomitant consideration of this factor allows for a side-by-side comparison of two factors thought to influence the acceptability of CNPC violations.

3.3.1 Materials

The stimuli in this experiment consisted of 36 embedded complex noun phrase violations, which varied with respect to the information content of the filler-phrase and the island-forming NP type. For each item, each participant saw one of seven experimental conditions (2 x 3 + 1). Across conditions, the fronted wh-phrase was either a bare wh-item (=BARE condition), such as who or what, or a comparatively more informative which-N' phrase (=WHICH condition). Precisely half of the items presented animate wh-phrases, while the other half contained inanimate wh-phrases. The other factor considered in this experiment was the effect of NP type on subsequent sentence processing. Subjects read one of three kinds of island-forming NPs: a definite NP (DEF), an indefinite plural (PL), or an indefinite singular (INDEF). Additionally, a baseline for each item was included that lacked the island-forming NP. (55) shows a sample experimental item with all seven conditions:

(55) **BARE-DEF:** I saw **who** Emma doubted the report that we had captured in the nationwide FBI manhunt.
**BARE-PL:** I saw **who** Emma doubted reports that we had captured in the nationwide FBI manhunt.
**BARE-INDEF:** I saw **who** Emma doubted a report that we had captured in the nationwide FBI manhunt.
**WHICH-DEF:** I saw **which convict** Emma doubted the report that we had captured in the nationwide FBI manhunt.
**WHICH-PL:** I saw **which convict** Emma doubted reports that we had captured in the nationwide FBI manhunt.
**WHICH-INDEF:** I saw **which convict** Emma doubted a report that we had captured in the nationwide FBI manhunt.
**BASELINE:** I saw **which convict** Emma doubted that we had captured in the nationwide FBI manhunt.

All reading-time results were analyzed in a 3 x 2 repeated measures ANOVA design, which also evaluated interactions between the kind of filler and the definiteness of the head noun.
Post-hoc comparisons were also made to the baseline as necessary. Due to poor question-answer accuracy in some conditions of some items (7 out of 252 cells, or 2.77% of the item analysis data set), the missing data for those cells were replaced with the linear trend for that point using the Replace Missing Values command in SPSS 15.0. This method of data imputation inserts the predicted values of empty data cells by regressing the existing series on an index variable scaled 1 to n. For this experiment, outliers were removed via data trimming—in this case, eliminating the five most extreme reading times from each condition—which affected 4.9% of the total data, after responses on incorrectly answered stimuli were removed.

3.3.2 Participants

Twenty-five Stanford University undergraduates, all native speakers of English, were paid $15 to complete this experiment along with a short off-line survey that was unrelated to this experiment. Combined, the two experiments lasted between twenty and thirty-five minutes. For the entire experiment, global average reading time per word was 398.9 milliseconds (SD = 83.2). Average question-answer accuracy for all items, including fillers, equaled 79.45%. There were no main effects of condition in the accuracy scores (BARE-DEF: 81.10%, SE = 3.49; BARE-INDEF: 80.47%, SE = 3.52; BARE-PL: 75.00%, SE = 3.84; WHICH-DEF: 80.77%, SE = 3.47, WHICH-INDEF: 75.20%, SE = 3.82; WHICH-PL: 83.72%, SE = 3.26, BASELINE: 76.74%, SE = 3.73).¹⁶

³.3.3 Results

Immediately after the wh-phrase, informativity of the wh-phrase does not have a significant impact on reading times (NP type is, of course, irrelevant as the NPs have not been reached yet). Yet, on the second word after the wh-phrase, there is a main effect for wh-phrase informativity, such that the less informative BARE conditions lead to significantly faster reading times early on in the sentence processing (F1(1,24) = 8.335, p < .01; F2(1,35) = 6.152, p < .05).

Beginning at the complementizer that, however, the WHICH conditions generate significantly faster reading times than the BARE conditions (F1(1,24) = 13.776, p = .001; F2(1,35) = 13.776, p = .001; F2(1,35) =

¹⁶There was a marginal interaction observed in the accuracy scores, however, which stems from the fact that the WHICH-PL condition elicited significantly higher response accuracies than the BARE-PL condition, as revealed by a post-hoc comparison. All other comparisons were not significant.
Table 3.1: Effect of informativity by region inside embedded clause in experiment I

<table>
<thead>
<tr>
<th>Region</th>
<th>Test</th>
<th>df</th>
<th>$F$</th>
<th>$p$ - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complementizer</td>
<td>Subject</td>
<td>(1,24)</td>
<td>13.776</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,35)</td>
<td>18.953</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Embedded Subject</td>
<td>Subject</td>
<td>(1,24)</td>
<td>5.028</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,35)</td>
<td>11.972</td>
<td>.001</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>Subject</td>
<td>(1,24)</td>
<td>16.283</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,35)</td>
<td>8.245</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Verb</td>
<td>Subject</td>
<td>(1,24)</td>
<td>7.013</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,35)</td>
<td>5.249</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Two-word spillover region</td>
<td>Subject</td>
<td>(1,24)</td>
<td>11.003</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,35)</td>
<td>9.635</td>
<td>.004</td>
</tr>
</tbody>
</table>

In all three NP-type conditions, as shown in Figure 3.1, the WHICH version is processed faster than the corresponding BARE version. This processing advantage for the WHICH conditions extends beyond just the complementizer—it covers the subsequent pronominal subject, the embedded auxiliary and verb (see Figure 3.2), as well as the regions after the verb where the presence of the gap is confirmed. In other words, from the complementizer until several words after the subcategorizer, the WHICH conditions are read faster than the corresponding BARE conditions.

Looking at the embedded pronominal subject, for instance, the statistical analysis reveals a main effect for the informativity of the *wh*-phrase ($F_{1}(1,24) = 5.028, p < .05; F_{2}(1,35) = 11.972, p = .001$). Likewise, at the auxiliary, where retrieval of the stored filler-phrase may begin, reading times are fastest in the WHICH conditions ($F_{1}(1,24) = 16.283, p < .001; F_{2}(1,35) = 8.245, p < .01$). In line with the predictions of the facilitation hypothesis, the more informative WHICH conditions also produce faster reading times at the subcategorizer ($F_{1}(1,24) = 7.013, p < .05; F_{2}(1,35) = 5.249, p < .05$). At the subsequent word where readers receive overt evidence of the missing constituent, this main effect continues ($F_{1}(1,24) = 11.097, p < .01; F_{2}(1,35) = 3.665, p = .064$), and even shows upon the next word ($F_{1}(1,24) = 4.653, p < .05; F_{2}(1,35) = 12.132, p = .001$). It is only at the third word after the subcategorizer that the main effect of informativity disappears, although
post-hoc comparisons show that the WHICH-DEF condition remains significantly faster than the BARE-DEF condition.\footnote{Due to this fact, there is a significant interaction at this word, but no main effects of informativity or NP type.}

Also of importance is the fact that, throughout the most embedded clause, the average reading time for the WHICH conditions does not significantly differ from that of the syntactically simpler baseline. Figures 3.1 and 3.2 reflect this point: reading times for the baseline are comparable to the average reading time of the WHICH conditions. This stands in stark contrast to the BARE condition, which remains consistently slower than the baseline throughout the embedded clause.

At only two word regions do the results indicate an effect of the type of island-forming noun phrase. First, at the complementizer, there is a significant effect of NP type. A closer inspection of the data reveals that the effect at this particular region largely stems from drastically slowed reading times in the BARE-DEF and BARE-PL conditions, as shown in Figure 3.1, which produced a significant effect of NP type at the complementizer ($F_1(1,19) = 9.415$, $p = .001$; $F_2(1,34) = 11.619$, $p < .001$; min$F' (1,46) = 5.201$, $p < .05$). This effect recurs at the first word after the verb, e.g. the word in in (55). In this case, the

![Figure 3.1: Residual reading times at complementizer in experiment I](image-url)

\[ 	ext{Mean Residual Reading Time} \]

\[ 	ext{COND} \]

\[ 	ext{Error bars: +/- 1 SE} \]
reduced reading times associated with the definite NPs are responsible for the effect, as the bare-def and which-def conditions produced the slowest reading times of any condition.

In sum, the reading results identify an initial, but brief, slowdown after processing the more informative *wh*-phrases, but beginning with the complementizer, this disadvantage reverses and becomes a highly significant processing advantage. This advantage lasts until several words after the subcategorizing verb which the *wh*-phrase must be associated with. Effects due to the type of the island-forming NP are much more localized (and relatively weaker) and appear only at the complementizer and at the verb, plus the first word after the verb.

Reaction times to the comprehension questions do not indicate any statistically significant effects of informativity or NP type. There is a trend for faster reaction times in the baseline condition, as compared to the mean reaction times for the bare conditions, but this difference does not reach standard levels of significance (bare: 2715.49, SE = 91.43; baseline: 2482.38, SE = 129.28; which: 2629.52, SE = 69.70).

---

**Figure 3.2:** Residual reading times at auxiliary + verb in experiment I
3.3.4 Discussion

Reading-times in this study show a strong influence of the informational content of the filler-phrase, in accordance with the predictions of the MFH. The results, however, specify that the processing advantage for the more informative WHICH conditions actually begin several words earlier than the retrieval site. Hence, this experiment indicates the processing differences across conditions cannot be attributed solely to differences in retrieval speed or accuracy. At the same time, however, these results are entirely compatible with the predictions of the MFH: an effect of informativity is observed at the retrieval site, although the results obscure to what extent this effect derives from differences in retrieval.

Some other mechanism is required to explain the processing contrasts that emerge prior to the retrieval site. Given the localization of these effects and the nature of the stimuli, one possibility is that the more complex and informative WHICH-N’ phrases reduce the chances for garden-pathing the filler-gap dependency. Along the filler-gap path in these complex noun phrase constructions, another verb appears whose complement is the complex noun phrase out of which the wh-phrase is extracted. In some of the stimuli, such as the example appearing in 56, the bare wh-item could easily be interpreted as a nominal argument of this intervening verb, e.g. announced. In contrast, a lexically richer WHICH-N’ phrase like WHICH STRUCTURE is unlikely to be interpreted as a nominal argument of the verb, as propositions and facts get announced, but not concrete objects.

(56) a. He published which structure Brittany announced plans that we would build to replace the condemned building.

b. He published what Brittany announced plans that we would build to replace the condemned building.

In the case of the bare wh-phrase, however, the filler can be associated with either a propositional- or an individual-denoting gap. This temporary ambiguity, therefore, can lead to an early misparsing in the conditions with a bare wh-phrase, whereby subjects initially assume the bare wh-item to be an argument of the first verb encountered. Upon receiving counterevidence to this initial analysis (the head noun of the complex noun phrase), a reanalysis takes place that naturally requires extra processing resources. Potentially, this resource expenditure may incur a cost over a relatively lengthy period of time. Such a scenario would thus explain the finding that wh-phrases of varying levels of informativity result
in processing differences beginning at the complementizer, which is just after receiving evidence in the form of the island-forming NP that the \textit{wh}-phrase is not a nominal argument of the first verb encountered along the filler-gap path.

Whatever the explanation for it is, an early effect does not preclude the presence of a second informativity effect at the retrieval site. The reading time results may consequently reflect two related effects, which nevertheless arise from different processing considerations. Parsing problems may be responsible for the slowdown in the \textit{bare} conditions at the complementizer and embedded subject, but retrieval difficulties may kick in at the auxiliary and spill over onto the post-gap region. Hence, informationally “heavier” filler-phrases may have more than one advantage for filler-gap processing. The MFH identifies one of these advantages as a faster and more efficient memory retrieval process. Thus, these results support the general predictions of the MFH, although further evidence is needed to isolate the effect of informativity on retrieval costs.

As opposed to the filler effects which start early and remain significant throughout nearly the entire filler-gap dependency (and even further), manipulating the type of island-forming NP generates only small and temporary effects during the processing of the filler-gap dependency. The first position where NP-type influences processing is at the complementizer, where, as mentioned before, the singular indefinite phrases lead to faster processing. The effect of NP type shows up again four words later, immediately following the retrieval site. In this case, the main effect originates with the slower reading times associated with the \textit{def} conditions. Because of the sporadic and temporary nature of these effects, it is difficult to extrapolate very much about the significance of NP type on the processing of the filler-gap dependencies. What can be said, however, is that NP type plays less of a role in processing than filler-informativity, at least in extractions out of complex NPs.

To complement the self-paced reading experiment, we also conducted a controlled acceptability study using the exact same stimuli, including the same fillers. As with the self-paced reading task, items were randomized and distributed across lists such that each subject saw only one condition per item. Sixteen subjects (none of whom had participated in the previous reading time study) were asked to rate the sentences on a scale of 1 to 7 for naturalness and were specifically instructed not to rate the sentences according to prescriptive grammar rules. The subjects in this study were given course credit for their participation.
Figure 3.3: Mean normalized acceptability ratings of complex noun phrase constructions

Figure 3.4: Mean normalized acceptability ratings of complex noun phrase constructions
As shown in Figure 3.3, the acceptability judgments yield the same main effect of \( \text{wh} \)-phrase informativity as the reading time study. Just as \( \text{which} \)-N’ phrases improve processing of the CNPC violations, they also raise ratings of acceptability (\( t_1(15) = 6.394, p < .0001; \)
\( t_2(35) = 5.263, p < .00001; \) \( \min F’(1,48) = 16.510, p < .001 \)).\(^{18}\) The baseline in the acceptability study, however, does receive significantly higher ratings than the \( \text{which} \) conditions (\( p < .001 \)). In the reading time study, however, the baseline did not consistently produce significantly faster reading times inside the critical embedded clause.\(^{19}\) This suggests that acceptability surveys can be sensitive to factors which do not noticeably impact on-line performance measures such as reading times. In particular, reading times may not reflect the effects of semantic, pragmatic, and/or discourse processing that extend beyond the first-pass reading of the sentence. Indeed, we find that question response times for the baseline case in datasets like (55) are about 150 milliseconds faster than those of the \( \text{which} \)-conditions (reaction time means: \( \text{Bare} = 2715.49 \text{ ms}, \text{SE} = 91.43; \) \( \text{Baseline} = 2482.38 \text{ ms}, \text{SE} = 129.28; \) \( \text{Which} = 2629.52, \text{SE} = 69.70 \)). Thus, in comparing reading times and acceptability judgments, some differences are definitely expected on the assumption that acceptability judgments constitute the end-result (or involve the input) of a number of cognitive processes, including sentence processing effort.

Overall, though, the sentence processing and acceptability results show a strong resemblance to each other. While filler-informativity plays a highly significant and comparable role in each, the type of island-forming NP has only a minimal impact on each set of results. Assuming that the parallel findings across methodologies are not coincidental, the likelihood of a causal relationship between the two increases. I will postpone a discussion of this matter, however, until the penultimate section of this chapter, where the relationship between processing difficulty and gradient acceptability is further discussed.

In sum, this experiment provides a compelling first look at the role of informativity in filler-gap processing, as well as acceptability judgments. Substantial reading time differences emerge within CNPC violations as a function of the quantity of information encoded in an extraposed \( \text{wh} \)-phrase. A greater amount of information within the \( \text{wh} \)-phrase ameliorates the processing difficulty of CNPC violations to the extent that reading times are not

\(^{18}\)Even with the Bonferroni adjustment for multiple comparisons, all p-values are less than .001.

\(^{19}\)There is a very marginal effect of NP type in the acceptability results. In particular, as pictured in Figure 3.4, plural indefinites in the \( \text{which} \) condition produced slightly better sounding CNPC violations, as compared to singular indefinites and definite NPs. All other comparisons of NP type were found to be non-significant.
different from those for minimally different sentences without the island boundary. Effectively, removing just a single processing burden has a major effect on how these FGDs are processed. This empirical investigation also offers the first piece of experimental evidence in support of the MFH: reading times at the retrieval area are faster when a which-\textit{N'} phrase constitutes the target for retrieval, rather than a bare \textit{wh}-item. As we saw, however, there is some question about how much of the effect can be attributed to retrieval as opposed to some other processing consideration. To continue the search for unequivocal evidence of informativity’s influence on retrieval, I now turn to a second island environment: \textit{wh}-islands.

### 3.4 Experiment II: Wh-Islands

The \textit{wh}-island constraint essentially prohibits questioning or relativizing an element out of an indirect question. On the Subjacency account, the strong version of this constraint applies to tensed clauses (because of the Tensed-S Constraint of Chomsky (1973)), while only a weaker version of it applies to non-finite clauses (\textit{whether to invite}).

\begin{equation}
\text{(57)} \quad \text{No rule can involve X, Y in the structure:}
\end{equation}

\[
\ldots X \ldots [\alpha \ldots Z \ldots \neg WYV \ldots \ldots
\]

where (a) Z is the specified subject of WYV  
or (b) Y is in COMP and X is not in COMP  
or (c) Y is not in COMP and \(\alpha\) is a tensed S.

So a typical \textit{wh}-island violation like \textit{Who did he wonder whether Jeff ignored?} is ruled out according to (57), on the assumption that the bare \textit{wh}-phrase is base-generated in the subordinate clause, but cannot be moved out of that clause because (i) the complementizer position is already filled by \textit{whether}, thus cutting off the escape hatch, (ii) the embedded clause is a tensed \textit{S} and (iii) the embedded clauses contains a specified subject, \textit{Jeff}. In contrast, a minimally different example like \textit{Who did he wonder whether to ignore?} is considered marginally acceptable as the embedded clause is not a tensed \textit{S}, nor does it contain a specified subject. Of course, the precise explanation for the markedness of these FGDs has varied considerably throughout the various incarnations of transformationalist syntax. But while non-finite \textit{wh}-islands have elicited semantic characterizations, tensed
wh-islands have consistently been diagnosed in syntactic terms.

The second experiment in this dissertation again addresses how the quantity of information in an extracted element affects subsequent processing. This experiment, however, avoids one of the potential confounds contained in the CNPC study. Specifically, the verbs intervening between filler and gap exclusively select for complement clauses, while only animate wh-phrases are used as filler-phrases. This avoids the possibility of early processing differences engendered by reanalysis costs in one condition, but not the other. As in the previous study, the MFH makes the prediction that processing differences should emerge at the retrieval site. More particularly, it predicts that a filler-phrase which encodes a greater quantity of information should cause faster processing at the retrieval site.

### 3.4.1 Materials

All experimental items consisted of wh-islands, presented as main clause interrogatives, as in (58). Subjects initially saw a declarative background sentence, after which a basic comprehension question was presented. The real stimuli of interest were the comprehension questions themselves; the initial context sentences merely justified the presence of these questions which would be entirely unnatural without any background. The wh-phrase always corresponded to the animate object of the embedded verb. The stimuli varied in terms of whether the sentence-initial wh-phrase was the bare wh-item who (bare) or the more complex and informative which-phrase (which).\(^{20}\) An additional condition was included to serve as a baseline against which the results could be compared. This condition differed only in terms of the complementizer and, because this complementizer was always

\(^{20}\)In addition, the collocational frequency of the subcategorizing verb and the head noun in the which-N phrases was systematically varied. For each experimental item, two verbs of roughly equivalent lexical frequency were selected to use with the wh-phrases. The verbs were selected such that one appears commonly with the head noun of the which-phrase, while the other verb is followed far less often by the noun. Due to an experimental confound, however, collocational frequency had little or no effect on reading times. The earlier presentation in the background sentence of both the high frequency verb-object pairs and low frequency verb-object pairs effectively eliminated any differences when subjects encountered the pair the second time (in the comprehension question). Consequently, that factor is removed from all subsequent analyses and the statistics reported here only represent the result of contrasting two levels in the repeated measures analysis. In fact, including the factor of collocation frequency in the repeated measures design, along with the interaction of collocational frequency and informativity, yields even stronger effects for informativity, particularly at the subcategorizing verb, where mild differences due to collocational frequency account for some of the observed variation.
that, these questions do not represent instances of syntactic islands.

(58) Albert learned that the managers dismissed the employee with poor sales after the annual performance review.

**Bare:** Who did Albert learn whether they dismissed after the annual performance review?

**Which:** Which employee did Albert learn whether they dismissed after the annual performance review?

**Baseline:** Who did Albert learn that they dismissed after the annual performance review?

Twenty-four experimental items and forty-eight fillers constituted the materials for this study. Twelve of the fillers were also *wh*-islands, such that half of the overall items were *wh*-islands and the other half were not syntactic islands. The additional twelve *wh*-islands were included to mask some of the more salient properties of the true experimental items. For instance, while the twenty-four experimental items always asked about the object of the embedded verb, some fillers asked about a prepositional object, e.g. *Which legislation did Andrew report whether they reached a compromise over?*. Additionally, half of the island-type distractors contained inanimate *wh*-phrases to contrast the animate *wh*-phrases in the experimental items. Across the entire item set, an equal number of questions began with *who*, *what*, and *which*-phrases to remove any experiment-internal bias for processing one type of interrogative faster than another. After reading the question, subjects answered the question by selecting from multiple choices. Of the three possible answers presented to them, one was correct (*the employee with poor sales*), another was lexically and syntactically similar (*the employee with poor hygiene*), and the third option differed drastically (*the cashier who stole money*).

With respect to the data itself, only correctly answered stimuli are considered here. Due to the small number of outliers found in the results and the fact that a roughly equal number of outliers occurred across conditions, all reading times more than 2.5 standard deviations from the mean were removed from the analysis. This affected less than 2% of the data. Removing these outliers changed the skewness statistic from 4.074 to -0.077 and also significantly altered the kurtosis value from 47.463 to 1.428 for the embedded verb
and the associated spillover region. Repeated measures ANOVAs were carried out on the bare and which conditions at each word region, and post-hoc t-tests with the baseline were made as necessary.

### 3.4.2 Participants

Twenty subjects participated in this reading-time experiment. All participants were Stanford University undergraduates over the age of 18 who received course credit in a linguistics class for their participation. The experiment was run immediately before or after another reading time experiment on multiple *wh*-questions. Each experimental session lasted about 45 minutes. Average question-answer accuracy for the entire experiment (including fillers) was 91.2%. The global average reading time was 311.4 ms ($SD = 71.7$). Question-answer accuracies, however, did not vary significantly across conditions within the experimental items (bare = 89.6%, $SE = 2.21$; which = 88.5%, $SE = 2.31$; baseline = 90.6, $SE = 2.99$).

### 3.4.3 Results

The results verify that more informative *wh*-phrases lead to faster reading times at the subcategorizing verb combined with the subsequent spillover region. At the embedded verb itself, there is only a numerical trend for faster reading times in the which condition ($t_1(19) = 1.619, p = .122, t_2(23) = 2.048, p = .052$). When spillover regions are considered, however, the effect of phrase informativity is observable and statistically significant. Considering the verb and the first word of the subsequent modifying material, the difference between which- N’ phrases and who is significant by both subjects and items ($t_1(19) = 2.127, p < .05$; $t_2(23) = 2.31, p = .03$).

---

21Skewness measures the degree of asymmetry in the distribution of a random variable. Thus, negative skew values indicate that the frequency distribution is asymmetrically long to the left of the mean, i.e. the left tail is longer than the right; positive values mean asymmetrically long right tails. A perfectly symmetrical distribution would have a skewness value of 0. Reading time data, of course, is almost always right-skewed given the natural limitations on minimum reading times.

Kurtosis is another kind of measure of the distribution of a random variable; however, it assesses the effect of infrequent extreme deviations. High kurtosis values mean that the distribution has a sharp “peak”, with the distribution densely clustered around the mean, but it also can signal the heaviness of the distributional tails, or the relative frequency of extreme values as compared to the normal distribution. In terms of reading time, therefore, a high kurtosis value specifies that a relatively large amount of the variance ultimately stems from infrequent and high reading times.
Figure 3.5: Residual reading times in experiment I, ranging from first word after *wh*-phrase to three words after retrieval site. Error bars show one standard error.

t_2(23) = 2.615, \ p = .015). Indeed, this trend continues for three words after the verb, as pictured in Figure 3.5, such that reading times in the bare-*wh* condition remain slower until sentence-final wrap-up effects obscure the difference between conditions. Taking the verb together with the next three words shows a highly significant effect of informativity (t_1(19) = 3.600, \ p = .002; t_2(23) = 3.669, \ p = .001). This suggests that at least some (if not a large proportion) of the difficulty associated with retrieving the *wh*-object continues after the verb is processed. According to post-hoc comparisons, the processing facilitation associated with the more informative *which*-N’ phrases also eliminated any substantial difference between the island-violating *which* condition and the baseline condition, which does not violate any supposed constraints on extraction.

In regions before the embedded verb, a greater degree of informativity in the *wh*-phrase does not significantly quicken reading times. To a limited extent, in fact, the reading times reflect a relatively reduced reading rate after the *which*-N’ phrases. Immediately after the *wh*-phrase, at the dummy auxiliary *did*, the *which* condition results in slightly slower reading times, when compared against the other two conditions, which are still lexically
identical at this point (see Table 3.2). Similarly, at the subsequent proper name, the two conditions beginning with the bare *wh*-item exhibit shorter reading times than the *which* condition. By the matrix verb, however, the differences across conditions have more or less equalized. For the first several words after the *wh*-phrase, therefore, the reading times actually slow down in the *which*-condition, as compared to the other conditions that both begin with a single *wh*-word.

At the complementizer, which differed lexically in the baseline and island conditions, the raw reading times look nearly identical (baseline = 246.56 ms, SE = 7.66; which = 248.53 ms, SE = 5.41; bare = 251.51 ms, SE = 5.27); however, due to the word length differences between *whether* and *that*, the residual reading times indicate a slower reading rate for the baseline. At the subject of the embedded clause, there is a slight difference between the bare condition and the *which* condition. But this small effect is not significant by either subjects or items, as shown in Table 3.3. Finally, at the last two words of the comprehension questions, reading times again equalize across conditions with the end-of-sentence wrap-up effects.

Reaction times to the question showed a strong advantage for the *which* condition (F1(1,19) = 13.664, p = .002, F2(1,23) = 10.778, p = .003; minF’(1,42) = 6.025, p < .025). This bias, however, is likely unrelated to the ease of retrieving the *wh*-phrase from memory. Given a *which*-N’ phrase, as opposed to a bare *wh*-item, subjects see a more focused question containing more information to answer the question. This does not necessarily serve as evidence that more complex *wh*-phrases are easier to recall, since *wh*-phrase retrieval is no longer relevant at that stage of processing. Additionally, the likelihood of lexical priming is quite strong given the layout and presentation of the stimuli. The presence of the head noun in the *which*-conditions and in the answers permits a simple recognition operation to take effect, since the head noun appears in the answer as well. The facilitation hypothesis defended here is meant to be distinct from simple priming and/or recognition effects.

### 3.4.4 Discussion

Filler-gap processing, according to the evidence, improves when the filler-phrase encodes more information. The processing advantage, however, does not begin immediately after the filler-phrase. In fact, the more complex and informative *wh*-form initially causes slower reading, likely due to the added task of integrating the information from the head noun of the *which*-N’ phrase—essentially, the cost of building a more complex representation.
### Table 3.2: Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment II. Each measure is followed by its mean standard error.

<table>
<thead>
<tr>
<th>Region + Example</th>
<th>RWRT</th>
<th>BASE</th>
<th>WHICH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (did)</td>
<td>RWRT</td>
<td>256.66 (9.35)</td>
<td>261.65 (6.21)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-40.26 (8.39)</td>
<td>-32.47 (4.89)</td>
</tr>
<tr>
<td>Region 2 (Albert)</td>
<td>RWRT</td>
<td>242.61 (7.47)</td>
<td>260.14 (6.16)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-73.68 (8.20)</td>
<td>-51.02 (5.71)</td>
</tr>
<tr>
<td>Region 3 (learn)</td>
<td>RWRT</td>
<td>251.86 (9.16)</td>
<td>255.45 (6.04)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-65.21 (8.56)</td>
<td>-60.78 (5.41)</td>
</tr>
<tr>
<td>Region 4 (whether)</td>
<td>RWRT</td>
<td>246.56 (7.66)</td>
<td>248.53 (5.41)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-54.83 (7.52)</td>
<td>-70.47 (5.52)</td>
</tr>
<tr>
<td>Region 5 (they)</td>
<td>RWRT</td>
<td>238.95 (6.23)</td>
<td>243.10 (4.72)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-65.523 (5.47)</td>
<td>-56.88 (4.09)</td>
</tr>
<tr>
<td>Region 6 (dismissed)</td>
<td>RWRT</td>
<td>231.93 (6.87)</td>
<td>242.63 (4.53)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-96.72 (8.65)</td>
<td>-82.05 (5.22)</td>
</tr>
<tr>
<td>Region 7 (after)</td>
<td>RWRT</td>
<td>255.60 (8.50)</td>
<td>248.77 (5.05)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-45.14 (7.96)</td>
<td>-49.56 (4.66)</td>
</tr>
<tr>
<td>Region 8 (the)</td>
<td>RWRT</td>
<td>242.75 (6.03)</td>
<td>242.22 (4.76)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-57.87 (5.98)</td>
<td>-59.48 (4.62)</td>
</tr>
<tr>
<td>Region 9 (annual)</td>
<td>RWRT</td>
<td>248.27 (7.57)</td>
<td>239.24 (4.44)</td>
</tr>
<tr>
<td></td>
<td>RZRT</td>
<td>-64.69 (7.64)</td>
<td>-69.83 (4.92)</td>
</tr>
<tr>
<td>Region</td>
<td>Test</td>
<td>df</td>
<td>t</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------</td>
<td>----</td>
<td>-------</td>
</tr>
<tr>
<td>Preceding Word</td>
<td>Subject</td>
<td>19</td>
<td>1.366</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>1.716</td>
</tr>
<tr>
<td>Verb</td>
<td>Subject</td>
<td>19</td>
<td>1.619</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>2.048</td>
</tr>
<tr>
<td>Verb plus next word</td>
<td>Subject</td>
<td>19</td>
<td>2.127</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>2.615</td>
</tr>
<tr>
<td>Next word</td>
<td>Subject</td>
<td>19</td>
<td>2.369</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>2.244</td>
</tr>
<tr>
<td>2nd word after verb</td>
<td>Subject</td>
<td>19</td>
<td>3.479</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>3.574</td>
</tr>
<tr>
<td>3rd word after verb</td>
<td>Subject</td>
<td>19</td>
<td>4.411</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>3.421</td>
</tr>
<tr>
<td>Verb + 3-word spillover region</td>
<td>Subject</td>
<td>19</td>
<td>3.600</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>23</td>
<td>3.669</td>
</tr>
</tbody>
</table>

Table 3.3: Effect of informativity by region inside embedded clause in experiment II.
Shortly thereafter, processing of the *wh*-phrases has no substantial residual effects, until the subcategorizer or gap site. Using the more informative *which*-N’ phrase in *wh*-island constructions results in faster reading times at the verb and subsequent spillover regions. This falls into line with the predictions of the MFH. Decreased reading times emerge significantly starting at the retrieval and integration site and intensify during the immediately subsequent spillover regions where the presence of the gap is confirmed.\(^22\)

The data also show that processing improves so much that the overall difference between the non-island baseline and the *which* condition disappears. This finding aligns with the intuitions expressed in the theoretical syntax literature that extraction out of island contexts improves with the specificity or descriptiveness of the extracted element. The absence of a difference between the baseline and *which* conditions is notable for another reason: some small but significant decrease in processing difficulty may be viewed as immaterial if processing difficulty remains at an extremely high level. Given the fact that reading times in the *which* condition are essentially equivalent to those of the baseline, however, the evidence argues for a meaningful interpretation of the facilitation. Due to the fact that the baseline and *which* conditions produce similar reading times, both significantly faster than the bare condition, the perception of unacceptability for *wh*-island violations involving vague or non-specific filler-phrases may reasonably have its origins in processing-related differences. The evidence at hand, therefore, suggests an alternative to the (pseudo)-grammatical mechanisms used to explain the previously unsubstantiated contrasts. Specifically, to the extent that acceptability is derived from considerations of processing difficulty, the observed processing contrast due to the informativity of the extracted *wh*-phrase may be playing a large role in the perception of the acceptability of the entire sentence.

\(^{22}\)Methodologically speaking, the reading-time experiment does suffer from a few undesirable features. As mentioned before, there is a great amount of lexical overlap between the initial background sentence subjects saw and the following comprehension question, which contained the real areas of interest. This led to quite fast reading times in the comprehension question, as reflected by the abundance of negative residual reading times in the reported results. Subsequent experiments rectified this problem by minimizing lexical repetitions and using synonyms and alternative formulations where reasonable. Such fast reading times also happen because participants consistently read comprehension questions faster than they do the preceding texts, whether there is lexical repetition or not. This is partly a position effect: participants in self-paced reading tasks typically read words more quickly as they progress through a sentence (with the exception of the last word or two). Secondly, since their understanding is being tested on the preceding declaratives and not on the questions, the latter probably receive slightly less attention.
To test the hypothesis that acceptability improves where processing difficulty decreases, a separate acceptability study was run using the same stimuli as were used in the reading-time task with a few variations. To remove the pragmatic oddity of decontextualized questions, the $wh$-islands were presented as embedded questions, as in the following modified version of (58):

(59)  

a. Only a few individuals repeated who Albert learned whether we dismissed after the annual performance evaluations.

 b. Only a few individuals repeated which employee Albert learned whether we dismissed after the annual performance evaluations.

All stimuli began with vague, quantified NP subjects, such as *some people* or *no one*, where no particular type of individual is named, under the assumption that such NPs would incur fewer processing costs than naming a specific individual or type of individual. Participants were instructed to rate the sentences for how natural the sentences sounded as examples of English on a scale of 1-7 (7 being perfectly natural). As depicted in Figure 3.6, the more informative $wh$-phrases significantly improve judgments of acceptability (F1(1,14)
This acceptability study therefore verifies that acceptability judgments fall precisely where processing difficulty significantly increases. The correlation between acceptability and processing difficulty in \textit{wh}-islands also parallels the results of the CNPC studies just reviewed. Of course, the interpretation that this indicates that processing difficulty is responsible for the acceptability judgments can be easily turned on its head. That is, this evidence could be taken as proof that grammaticality predicts processing difficulty. This issue of how to interpret the correlation between acceptability and processing results is still postponed until after the consideration of a final island context.

Overall, exploring informativity effects in \textit{wh}-islands has supplied strong evidence for the MFH: a higher quantity of information in the filler-phrase facilitates processing starting at the retrieval site. Due to the absence of any prior informativity effects, these results unequivocally link the processing differences to operations that occur at the retrieval site. These results consequently also lend weight to the proposal that the observed effects in the CNPC study at least partly reflect differences in retrieval.

### 3.5 Experiment III: Adjunct Extraction from \textit{Wh}-Islands

The third and final experiment on the processing of islands focuses on the referentiality of the extracted element. More precisely, this island study supplements the previous investigations by examining whether the same informativity effects are observable in the case of adjunct extractions. This type of extraction has long been deemed to be fundamentally different than argument extractions, based on acceptability contrasts like those mentioned in the second chapter (Huang, 1982). Examples like the one below putatively violate a constraint (the Empty Category Principle) that \textit{wh}-island violations like those considered in the previous section do not:

\begin{equation}
\text{(60) How long did Gregor decide whether the monster slept?}
\end{equation}

Standardly, these FGDs are ruled out on the assumption that extraction of adjuncts leaves behind a trace that is not properly governed. The non-referentiality or non-specificity of these adjunct phrases prevents them from entering into any other kind of binding relation with the covert variable.

\footnote{As with the CNPC experiment, this experiment constitutes collaborative work with Ivan A. Sag and Neal Snider.}
According to the predictions of the MFH, however, informativity should affect retrieval of adjuncts in a way that is qualitatively similar to how it affects the retrieval of arguments.\textsuperscript{24} This hypothesis is purposefully worded to avoid reference to the kind of phrase for which it is relevant, as it ostensibly reflects a general cognitive principle that is insensitive to linguistic particulars. In the same way, as referentiality is not a prerequisite for memory retrieval, the MFH predicts informativity effects for non-referential entities in general. Thus, while FGDs with adjunct phrases may cumulatively present more difficulty than an FGD with an argument phrase, increasing the quantity of information in the adjunct filler-phrase should still facilitate memory retrieval. This represents an important test case for the MFH, because while other semantically- or pragmatically-based theories might expect differences between \textit{which}-N\textsuperscript{'} phrases and bare \textit{wh}-words, they do not predict any differences among adjunct phrases.\textsuperscript{25}

### 3.5.1 Materials

Twenty-four adjunct extractions from \textit{wh}-islands were employed in this study. Stimuli were systematically varied in terms of the amount of information contained in the extracted temporal adjunct phrase. In the \textsc{bare} condition, the temporal adjuncts contained only two words, either \textit{how long} or \textit{how often}. The \textsc{long} condition had longer temporal adjuncts of at least three words and as many as eight, as exemplified in 61 below. As in the previous experiment, a baseline condition containing a different lexical complementizer, \textit{that}, was also included in the study as a means of evaluating differences between the two main conditions of interest.

In contrast with the previous two experiments, however, subjects were instructed not to answer the comprehension question, but to indicate whether the question \textit{could be} answered, given the information stated in the context sentence. Half of the total items appearing in this experiment, including half of the twenty-four adjunct extractions, presented questions

\textsuperscript{24}Again, this does not mean informativity is expected to impact the processing and retrieval of all kinds of phrases in a quantitatively identical way. The addition of a given syntactic or semantic feature to a phrase may thus yield retrieval effects of different magnitudes, depending upon the phrase type.

\textsuperscript{25}I suppose one could theoretically argue that informationally-richer adjunct phrases are D-linked, but this runs into the same problems that were covered in the last chapter, e.g. the lack of a precise formulation of D-linking, the absence of a rationale for the connection between discourse salience and conditions on movement, etc. Besides which, as I detail within the materials section, the adjunct phrases have the exact same preceding contexts, so the salience or availability of “context sets” is identical across conditions.
which could not be reasonably answered given the preceding text. For instance, in the
item shown in (62), none of the three versions of the question can be answered, since the
preceding text makes no mention of how long the children played during the afternoon
recess. Thus, after each question, the subjects saw the prompt, *Is it possible to answer
the question?*, and were instructed to provide a negative response when the preceding text
did not contain the necessary information to answer the question. Subjects were informed
during the training session of this experiment that each item did have a correct answer, and
consequently received negative feedback if they answered the question incorrectly. Prior to
the presentation of the actual experimental stimuli, subjects became familiar with this task
via a practice session with eight items.

(61) Julie discerned that the survivor had managed to stay alive for eight days after the

  crash in the harsh conditions.

  **BARE:** How long did Julie observe whether the passenger had survived in the unbe-
  lievably harsh conditions?

  **LONG:** For what period of time after the crash did Julie observe whether the pas-
  senger had survived in the unbelievably harsh conditions?

  **BASELINE:** How long did Julie observe that the passenger had survived in the unbe-
  lievably harsh conditions?

(62) Andrew overheard the daycare staff discussing how they wanted to get away from

  the children for a few hours.

  **BARE:** How long did Andrew hear whether the children had played during the
daycare’s afternoon recess?

  **LONG:** How many hours did Andrew hear whether the children had played during
  the daycare’s afternoon recess?

  **BASELINE:** How long did Andrew hear that the children had played during the
daycare’s afternoon recess?

The primary reason for utilizing this particular methodology stems from the comparatively
fast reading-times evident in the comprehension questions in experiments like the *wh*-island
study. Subjects appear to consistently read comprehension questions faster than the pre-
ceding text, partly due to a relatively high degree of lexical overlap and the predictability
CHAPTER 3. PROCESSING ISLANDS

of upcoming constituents. The methodology employed in this experiment consequently
dissuades participants from relying on predictability and repetition in reading and answer-
ing the comprehension question itself, while also encouraging them to read the question
carefully.

Residual reading times with z-scores more than 2.5 standard deviations from each sub-
ject’s mean were excluded from the statistical analysis. This deleted less than 4% of the
total data. For the entire embedded clause (except the complementizer), this changed
the skewness value from 4.861 to .180 and the kurtosis value from 56.24 to 3.502. As in
the previous study, repeated measures ANOVAs were performed on the two conditions of
interest—BARE and LONG. The results of planned t-tests involving the baseline are reported
where relevant.

3.5.2 Participants

Twenty-eight Stanford University students were paid $10 for their participation in this
study. Despite a small numerical trend for better question-answer accuracies in the LONG
condition, there were overall no significant differences across conditions (LONG: 86.16%; SE
= 2.31; BARE = 84.82%; SE = 2.40; BASELINE: 84.82%, SE = 2.40).

3.5.3 Results

The data identifies two particular word regions where the experimental conditions create
significantly different reading-times: (1) at the complementizer and (2) at the word after
the complementizer. These effects appear to be genuinely distinct effects, as the differences
observable at the word after the complementizer bear no relation to the complementizer
effect. At this clause boundary, the complementizer whether, appearing in both the LONG
and BARE conditions, results in faster reading times than the complementizer that in the
BASELINE condition (BARE-BASELINE: t_1(27) = -2.930, p < .01; t_2(23) = -2.522, p = .019;
LONG-BASELINE: t_1(27) = -6.325, p < .001; t_2(23) = 3.200, p < .01). This contrast mimics
the much smaller trend that appeared in the wh-island study, which also reflected faster
processing with whether. One potential explanation for this strong effect concerns the
syntactic ambiguity of that, which can initially be parsed as either a determiner, an indexical
NP, or a complementizer. This syntactic ambiguity therefore may lead to a temporary slow-
down in parsing, but the relative difficulty of the BASELINE condition at the complementizer
immediately disappears at the next word and this condition ultimately leads to the fastest reading times throughout the embedded clause, as shown in Figure 3.7.

The second main effect that emerges from this study, more relevant to the predictions of the MFH, is found at the word after the complementizer, which was always the determiner *the*. Here, the BARE condition produces significantly slower reading times than the LONG condition by nearly 50 milliseconds ($t_1(27) = 3.484, p = .002; t_2(23) = 3.513, p = .002$). The baseline here patterns along with the LONG condition, such that reading times for the baseline at the first word of the embedded clause are substantially faster on average than those for the BARE condition.

In contrast with the previous two experiments, no effect of informativity is observed at the embedded verb. In fact, after the first word of the embedded clause, the LONG condition is not processed significantly faster than the BARE condition at any one particular word inside the embedded clause. However, considering the entire embedded clause (except for the final word of the clause where end-of-sentence wrap-up effects occur), the LONG condition produces significantly shorter reading times than the BARE condition ($t_1(27) =$
2.356, \( p = .026 \); \( t_2(23) = 1.648, \ p = .113 \). Of course, much of this effect stems from the first word itself.

Additionally, while reading times for the entire embedded clause are faster for the baseline, as compared to the bare condition \( (t_1(27) = 3.218, \ p < .01; \ t_2(23) = 2.518, \ p = .019) \), the baseline did not produce significantly faster reading times in this region than the long condition \( (t_s < 1) \). The implications of these results are dealt with in the discussion section below. Reaction times did not differ significantly by condition \( \text{bare} = 867.45, \ SE = 38.19; \ \text{baseline} = 843.83, \ SE = 43.46; \ \text{long} = 819.45, \ SE = 33.57 \).\(^{26}\)

### 3.5.4 Discussion

This experiment’s main purpose was to evaluate whether the informativity effects found in the previous experimental studies are specific to argument extractions, as opposed to adjunct extractions. At the same time, this experiment was designed to evaluate the effect of informativity in another unequivocally strong island context. Just as in the other studies

\(^{26}\)Note that these relatively fast response times are due to the fact the actual question prompt was the same for all items, which essentially obviates the need to read the question.
reported so far, it was found that encoding more information in the adjunct filler-phrase facilitated subsequent processing. The effect of informativity, therefore, seems to function independently of referentiality and also appears to be generally insensitive to the argument-adjunct distinction.

In one respect, the observed effects in this study differ notably from those seen in the previous *wh*-island study. In the first study, more explicit fillers generated a processing advantage at the verb where retrieval and integration of information from the filler-phrase clearly occurs. In the adjunct extraction study, however, it is at the first word of the embedded clause where the effect of informativity is apparent (and marginally so for the entire embedded clause when the reading times for all regions are aggregated). Each of these studies nevertheless shares the attribute that the more informative filler-phrases ultimately has the consequence of faster reading times inside the complement clause.

It seems, at first glance, that the effect discovered in the adjunct experiment does not stem from differences in retrievability.27 This interpretation depends, though, on the

---

27Note that the results would not support the conclusion that the informativity effects derive from parsing differences at the matrix verb, e.g. *hear*, since reading times in the BARE and LONG conditions are not significantly different at this region or the next.
assumption that retrieval of the adjunct information occurs at the same place where retrieval with argument phrases happens, i.e. the embedded verb. An alternative view would be that retrieval in the adjunct dependency takes place at the clause boundary—in other words, the first word of the embedded clause. Unfortunately, sentence processing research on filler-gap dependencies with adjunct extractions is exceedingly sparse; in fact, I am aware of no research which evaluates how, if at all, adjunct extractions are processed and where retrieval effects typically surface. It nevertheless seems plausible, from a theoretical perspective, that extracted adjuncts would lead to retrieval at the embedded clause boundary, rather than at the embedded verb itself. Certain adverbials modify not just the information encoded by the verb, but the entire clause. Included in this class of adverbs are sentence modifiers like possibly and undoubtedly. But temporal adjuncts, too, can modify whole clauses as in Yesterday, I saw Peter playing foosball. Given a cognitive pressure to discharge fillers from memory as soon as possible (Fodor, 1978; Clifton & Frazier, 1989), it is not hard to fathom that retrieval would happen once the embedded clause is reached. Since temporal modifiers can at least theoretically modify entire clauses, the start of the embedded clause represents a possible retrieval site, and hence some attempt at retrieval should take place. Of course, this reasoning calls for further empirical investigation into the processing of FGDs with adjunct-fillers.

While this study shows that adjunct extractions are subject to some of the same processing biases as argument extractions, it does not directly address why, intuitively speaking, adjunct extractions out of wh-islands seem harder and less acceptable than argument extractions. Unfortunately, the MFH has virtually nothing to say about this subject, because arguments and adjuncts obviously differ in important ways unrelated to informativity. As noted previously in this chapter, however, these differences may derive from considerations linked to the processing difficulty of imagining a suitable context and constructing the appropriate existential presupposition.

It is possible, however, that retrieval-based issues may also play a part in the argument-adjunct contrast. For instance, the retrieval cues triggered at retrieval sites may be better suited for recovering arguments than adjuncts. Which is to say that the subcategorization and thematic role information, as well as encyclopedic knowledge, provided by retrieval contexts in FGDs seem tailored for the retrieval of arguments, but not adjuncts. A verb missing its object argument, as an example, can specify grammatical, thematic, and plausibility information to guide the retrieval process of the appropriate object NP. Moreover,
while argument dependencies can yield some indirect evidence for the location of the retrieval site in the form of a missing obligatory constituent, adjunct dependencies offer no such clues: a clause or a verb phrase with an adjunct extracted out of it looks just like it would if there was no extraction. This amounts to saying that there is a paucity of available retrieval cues in adjunct dependencies. According to theories of memory retrieval that posit that retrieval in language comprehension is fundamentally cue-based, the unavailability of clear cues for retrieval should cause dependency processing to be substantially harder.

A third reason why adjunct extractions may differ in acceptability from argument extractions concerns the potential for misparsing or garden-pathing. The stimuli in this experiment were purposefully constructed so as to reduce the likelihood of attaching the adjunct to the first verb encountered (or at least to make this likelihood comparable across conditions). In general, however, it will often be the case that a fronted adjunct phrase can plausibly modify the first verb along a filler-gap path. Within the syntax literature on this topic, in fact, many examples of adjunct extractions out of \textit{wh}-islands involve a potential site for misparsing the dependency. For instance, in the example like the one in (63a) below, the adjunct phrase can reasonably be interpreted as modifying the verb \textit{wonder}, but in a minimally different version like (63b) that contains an argument \textit{wh}-phrase, the filler cannot be discharged at the matrix verb. This effectively obviates the need for any sort of syntactic reanalysis in the latter case, but makes it highly probable in the adjunct case:

(63) a. When did Gregor wonder whether the monster ate?
   b. Which villagers did Gregor wonder whether the monster ate?

Thus, there are generally more chances for misparsing an FGD with an adjunct-filler because of weaker selectional restrictions imposed on it. Cumulatively, then, adjunct dependencies present an assortment of processing difficulties that argument dependencies do not. Any grammatical theory that seeks to account for the differing behavior of these two types of phrases in FGDs should consequently first weigh the contributions of processing load before making categorical distinctions that potentially carry far-ranging and profound consequences for a grammatical system.

There are clearly a number of open areas of investigation for the processing of adjunct questions, but this third experiment verifies that adjunct dependencies are subject to similar processing constraints as argument dependencies. As in the first two experiments of this thesis, longer and more informative filler-phrases produce faster processing inside the
syntactic island. These results are not easily explicable under theoretical accounts that predict the quality of an FGD on the basis of referentiality or other categorical divisions amongst phrase types, but are entirely in line with the predictions of the MFH.

3.6 The Source of Gradience

These empirical investigations into the processing of island constraint violations converge on the conclusion that where processing difficulty increases, acceptability decreases. Assuming a causal relationship for this (negative) correlation, the following question arises: why claim that processing difficulty is responsible for the acceptability judgments, rather than vice versa? That is, why not conclude that differences in grammaticality engender the observed processing effects? At the very least, it does seem reasonable to think that marked constructions should be read more slowly than unmarked constructions.

There are, however, a number of points that favor the conclusion that processing effort underlies the acceptability differences, rather than the other way around. First among these, a processing-based explanation appeals to the existence of independently-motivated factors, some of which may not even be specific to language. Locality effects, referential processing load, and the like exist regardless of whether they are credited as having a role in island effects or not. As we shall see in the next chapter, too, the informativity effects observed already in island-violating FGDs have a widespread influence on language comprehension, i.e. these effects are not specific to island contexts.\(^{28}\) This type of explanation, therefore, is “cost-free” in the sense that no new or highly specific constraints have to be introduced. A grammar-internal characterization of the gradient acceptability, on the other hand, demands the appropriate linguistic machinery. This includes the constraints governing FGDs like Subjacency (or some earlier alternative), and a system for calculating fine-grained acceptability differences on the basis of grammatical violations. Plus, the grammar must somehow specify that certain kinds of constituents (e.g. D-linked phrases or phrases with referential indices) provide exceptions to the normal constraints.

A processing-based account accordingly preserves the simplicity of the grammar by

----

\(^{28}\)The presence of these informativity effects on FGDs in non-islands poses an additional problem for the grammar-internal treatment, given the absence of any clear grammar violation. If grammaticality explains the processing differences in the island dependencies, then some secondary explanation must be invoked to account for the same differences in syntactic contexts which do not contain violations of any known grammatical constraint.
explaining the variability in acceptability judgments via extra-grammatical forces. This translates to a grammar that places fewer restrictions on the establishment of filler-gap dependencies. Essentially, the grammar overgenerates FGDs, while processing constraints cull the set of possible filler-gap dependencies. Again, there may indeed be some island effects which warrant a purely syntactic or grammar-internal description. But identifying the source of some island effects in processing effort reduces the set of phenomena that require grammatical constraints. Minimizing the purview of syntactic island theory may indeed lead to a more transparent and homogenous set of constraints. As I detailed at the beginning of this chapter, even within the same theoretical syntactic treatment, different islands have been characterized in seemingly unrelated ways. Looking at these effects from the perspective of sentence processing effort, however, capitalizes upon the existence of distinct processing pressures across linguistic contexts. The observation that some island violations sound worse than others follows directly from the point that each linguistic context poses unique processing challenges.

A rationalization of graded acceptability based on processing considerations also has the capability to explain this gradience in a wide array of linguistic constructions. This applies equally to constructions or sentence-types that do or do not violate any proposed rules of syntax. Besides the island contexts already tested, informativity effects crop up in other island settings, such as adverbial islands (64) and subject condition violations (65). In further acceptability investigations, it was found that the quantity of information encoded in the filler-phrase has parallel effects in these two environments, as it does in the CNPC and wh-island environments.\textsuperscript{29} Specifically, in adverbial island violations, both \textit{which-N’} phrases (\textit{which}) and partitive \textit{which}-phrases (\textit{part}) significantly boost ratings of acceptability, as compared to a bare \textit{wh}-word (\textit{BARE-PART}): $t_1(15) = -5.340$, $p < .001$, $t_2(8) = -2.829$, $p < .05$; minF’ (1,13) = 6.248, $p < .05$; \textit{BARE-WHICH}: $t_1(15) = -4.073$, $p = .001$, $t_2(8) = -2.244$, $p = .055$; minF’ (1,13) = 3.863, $p = .071$).\textsuperscript{30}

\textsuperscript{29}We ran these acceptability surveys as pilot studies, which is the reason for the low number of items. Accordingly, these results are only intended to be suggestive, although they do produce highly significant contrasts.

\textsuperscript{30}Partitive phrases were introduced as another means of increasing the syntactic complexity of the filler phrases. As is evident from the results, this increase in syntactic complexity does not produce reliably different acceptability judgments. Notice, though, that the absence of an acceptability contrast does not preclude the possibility of a processing difference, as not all processing differences are expected to influence judgments in a noticeable way.
Figure 3.10: Mean normalized acceptability ratings of adverbial island violations

(64)  
a. I knew who my boss said she was calm before meeting ___ in the White House yesterday.

b. I knew which staff members my boss said she was calm before meeting ___ in the White House yesterday.

c. I knew which of the staff members my boss said she was calm before meeting ___ in the White House yesterday.

This pattern is repeated when looking at subject condition violations: sentences with which-\textit{N'} phrases and partitive which phrases are deemed better than comparable sentences with a bare \textit{wh}-word (BARE-PART: $t_1(15) = -2.553, p < .05; t_2(8) = -1.883, p < .1$; minF' (1,16) = 2.296, $p = .149$; BARE-WHICH: $t_1(15) = -4.477, p < .0001; t_2(8) = -3.928, p < .01$; minF' (1,19) = 8.719, $p < .01$).

(65)  
a. Brendan acknowledged who it was possible that our attempts to visit with had been misinterpreted by the press.

b. Brendan acknowledged which suspects it was possible that our attempts to visit with had been misinterpreted by the press.
c. Brendan acknowledged which of the suspects it was possible that our attempts to visit with had been misinterpreted by the press.

Interestingly, a similar effect arises when looking at wh-dependencies that cross a clause boundary, but do not violate any proposed island constraint (they could thus be characterized as weak Subjacency violations):

(66) a. Justin proved what the engineers lied that they had invented without any help or instruction.

b. Justin proved which devices the engineers lied that they had invented without any help or instruction.

c. Justin proved which of the devices the engineers lied that they had invented without any help or instruction.

A which-N' phrase elicits higher ratings of acceptability than a bare wh-word, although this effect is marginal (t_1(15) = -2.032, p = .060; t_2(8) = -2.178, p = .061; minF' (1,22) = 2.207, p = .151). What this shows is that, even in grammatical environments, the effect of informativity can surface in acceptability judgments.\(^{31}\) A strictly grammar-based account

\(^{31}\)The marginality of this effect undoubtedly derives from a kind of floor effect, as these dependencies do
would be hard-pressed to rationalize these differences, as none of these structures conflict with any proposed constraints on dependencies. As an explanation of this gradience, it would thus be meaningless to suggest that certain kinds of phrases are exempt for one reason or another from the normal constraints on FGDs.

Variation from speaker to speaker and across dialects is also naturally modeled on a processing-based account. Individual differences in working memory capacity can readily justify the fact that some individuals are much more accepting of island violations than others. Besides differences across individuals, the incumbent processing pressures on a single language user also vary widely throughout discourse or reading. Other on-going cognitive tasks, fatigue, or even just absent-mindedness may limit an individual’s ability to process an incoming string. On the other hand, repeated exposure to a particular kind of stimulus may facilitate processing. Disparate dialects, too, may impose significantly different processing costs.

Theoretically, something similar could be stated for a grammar-internal account: individuals possess distinct grammars and grammars also vary nominally across dialects. Not pose much difficulty. With additional subjects and items, however, this result may well turn out to be statistically significant.
However, the judgment data from a single individual can vary widely on tokens that are structurally identical but diverge with respect to the accompanying processing load. Similarly, speakers of the same dialect may have very different reactions and assessments of the exact same stimuli. This sort of variation becomes very difficult to account for on a grammar-internal rationale. Furthermore, while there is a huge literature on individual differences in memory and language processing ability (see King and Just (1991), among many others), there is far less understanding about the ways in which grammars belonging to members of the same speech community can differ from one another.

Additionally, the extreme view that grammaticality differences calculated on-line are responsible for the pattern of processing results exhibited here brings with it the following assumption: graded grammatical violations could be detected fast enough to influence processing effort at the first word where such a violation becomes apparent.32 As an illustration, a simple number marking problem as in *The boys is here for dinner* presents a violation at the verb. However, it is not at all apparent that such an error would slow down processing at the verb itself. It might be several words later before the effect of this violation surfaces. Data from event-related potentials indeed show that the neural correlates of syntactic and morphological violations do not surface until at least 500 ms after word onset (e.g. Coulson, King, and Kutas (1998)). Furthermore, speeded grammaticality tasks indicate that, with a variety of grammatical violations, people essentially perform at chance when given less than half a second to respond (McElree & Griffith, 1995). To be clear, perceptions of grammatical deviance undoubtedly enter into the equation of processing difficulty at some point during sentence processing. The outstanding question is whether judgments of acceptability at each word in a string could be developed fast enough to impact processing at the most recent input word.

Ultimately, additional evidence from other areas and methods of investigation like ERPs, functional magnetic resonance imaging, and speed-accuracy trade-offs can help decide such debates on a case by case basis. As one example of this, (Fiebach, Schlesewsky, Bornkessel, & Friederici, 2004), in an fMRI study, report two different sources of neural activation for complex grammatical sentences and ungrammatical structures in German. Such evidence, combined with the results from other experimental methodologies, can convincingly pinpoint

---

32 This perspective assumes that grammaticality judgments are regularly made at intermediary stages of sentence processing—that is, at multiple points before the end of the sentence. A grammar-based take on the processing results of this chapter need not make such an assumption.
the source of gradient acceptability in a variety of linguistic constructions.

At this point, though, there are a number of points which argue that some island phenomena ought to be treated as the byproduct of aggregated processing difficulties: (1) the resulting explanation appeals to independently motivated principles, which ultimately leads to a much simpler and parsimonious account of the data; (2) variation within and across island contexts can be understood as the result of distinct processing pressures; (3) manipulating factors known to affect the acceptability of island constructions also yields a similar variation in acceptability and processing within non-island contexts (4) variation across speakers, and even with the same speaker, can be naturally modeled on the assumption that speakers differ from one another in working memory capacity, and that the availability of working memory resources for a single speaker can also vary in significant ways throughout discourse.

3.7 Summary

This chapter has presented some initial evidence in support of the MFH. The clearest evidence emerged in the consideration of argument extractions out of tensed \textit{wh}-islands. In this experiment, \textit{which}-\textit{N}' phrases elicited faster reading times than bare \textit{wh}-words beginning at the embedded subcategorizer, where retrieval of information linked to the filler-phrase must take place. In the other two experiments, the evidence was also compatible with the MFH: a higher quantity of information in the filler-phrase also translated to faster processing times at potential retrieval sites.

The time spent looking at syntactic island violations was originally justified by pointing to the high processing cost associated with these constructions. These filler-gap dependencies exhibit numerous characteristics that peg them as difficult constructions to process for comprehension. The likelihood of inherent processing difficulties therefore makes it probable that a facilitating manipulation would have a measurable effect. At the same time, however, the motivation for this island exploration was to highlight the explanatory value the MFH can have for theories of grammar. The relevance of this hypothesis is thus not restricted to understanding forces that guide sentence processing. It was shown in the preceding chapter how numerous theoretical accounts have been developed over the past thirty years to cope with the extraction asymmetries that revolve around the nature of the filler-phrase. Yet these theoretical treatments of the problem could generally only deal successfully with some
subset of the data, and the underlying motivations for the crucial assumptions were opaque or altogether absent.

The MFH, in contrast, offers a simple way of understanding the connection between the informational properties of a filler-phrase and the acceptability of the sentence in which it occurs. It only requires the reasonable assumption that processing effort can have a significant impact on perceptions of acceptability. General evidence for this assumption reaches back more than forty years and specific evidence for the case at hand comes from the high degree of correlation between the processing and acceptability results. I have argued that processing effort is the causal force behind the acceptability of the island constructions considered here. The weight of this argument comes largely from the economy of the explanation. The potential exists, therefore, to greatly simplify these accounts by appealing to general cognitive principles like the MFH.

33Note that this assumption does not entail that increased processing effort always decreases acceptability, or even that significant differences in processing will have a noticeable impact on judgments of acceptability. Participants in acceptability tasks are often instructed to ignore plausibility and other factors such as string length that undoubtedly contribute to processing load. It is therefore not surprising that acceptability results show a greater sensitivity to processing factors that affect structural parsing, rather than the ease of interpretability. The strength of the argument for a processing-based explanation of the island phenomena is that processing difficulty repeatedly correlates with acceptability in the same way across various contexts.
Chapter 4

Beyond Islands

The experimental evidence from the preceding chapter shows how the amount of information encoded in a filler-phrase facilitates processing in a variety of syntactic island constructions. Of course, syntactic islands such as the *wh*-islands, superiority-violating multiple *wh*-questions, and complex noun phrases constructions do not represent how people normally talk or what people normally hear. These constructions pose too many difficulties for both speakers and hearers, and less complicated alternatives are usually available. Thus, the forced comprehension of these syntactic constructions thrusts participants into an unnatural situation.

Based on this consideration, the concern arises that the results from the foregoing experiments depend upon the atypical situation of encountering islands and the abnormal processing strategies that result to compensate. In order to allay these concerns, it is necessary to consider the same factor of informativity in other complex, but unquestionably grammatical contexts. Indeed, the facilitation hypothesis is not framed specifically for the comprehension of syntactic islands. It makes no claims, in fact, about the syntactic contexts in which memory facilitation via increased informativity should be observable, nor does it specify what sort of phrases this potential memory facilitation applies to. Again, the main motivations for beginning with an investigation of syntactic islands were the previous observations that (1) acceptability intuitively differs, depending on the content of the extracted element (Chung & McCloskey, 1983; Chung, 1994), and (2) such constructions impose a high processing load (Deane, 1991; Kluender & Kutas, 1993b; Alexopoulou & Keller, 2003).

In this section, I evaluate the predictions of the memory facilitation hypothesis in two grammatical contexts with two different types of extracted elements. In the first experiment,
I reconsider the case of *wh*-dependencies, but in the setting of nested dependencies such as in the example below:

(67) What did the visitor that Arthur invited ___ break ___ in a fit of violence?

A shorter dependency, between the *visitor* and *invited*, is here nested inside a longer dependency between *what* and *break*. This requires maintaining the two filler-phrases simultaneously for a short time in memory and typically creates a slow-down in reading times at the retrieval and integration sites. Particularly, the back-to-back placement of the verbs introduces a sizable processing difficulty, since such a sequence requires multiple retrievals in a short period of time (not to mention the fact that such sequences are unquestionably rare in English). Thus, like the syntactic island violations considered in the previous chapter, nested dependencies place heavy demands on the processor, making observation of a processing facilitation easier.

The memory facilitation hypothesis predicts that, as in the case of *wh*-islands, a more informative *wh*-phrase should result in faster reading times at the subcategorizing verb (e.g. *break* in (67)). Moreover, prior to the verb, the more informative *wh*-phrase should not cause significantly faster reading in regions between the filler and subcategorizer. The effect, that is, should remain localized to the retrieval and integration site.

In addition, the memory facilitation hypothesis makes no claims about the sorts of phrases subject to the proposed relationship between quantity of information and subsequent retrieval ease. While all the experimental evidence covered so far deals with interrogatives and *wh*-phrases, the same differences in processing at retrieval sites ought to be observable with other kinds of NPs as well.\(^1\) Accordingly, in the second experiment described in this chapter, I explore how differences in the informativity of clefted indefinites impacts processing at the retrieval site.

### 4.1 Experiment IV: Nested Dependencies

#### 4.1.1 Materials

The twenty experimental items in this study were all nested dependency interrogatives of the sort illustrated in (68). The interrogatives varied in terms of whether the fronted, inanimate

\(^1\)Theoretically, other types of phrases such as VPs and PPs that require retrieval, as in cases of ellipsis, should also obey the principle outlined in the memory facilitation hypothesis.
wh-phrase was a bare wh-item (simple), a which-N’ phrase (which), or a which-N’ phrase with an additional adjective (complex). These interrogatives acted as the comprehension questions for preceding sentences, as in the previously described study. Each background sentence mentioned two individuals and at least one inanimate object, and occasionally a second inanimate in the five-word spillover region. Three possible answers were presented after reading the question. Two of these answers shared the same head noun, e.g. ABC political poll & CNN political poll, while the third answer was lexically dissimilar, acting as a distractor.

(68) Scooter hid from the reporter who talked about the recent ABC political poll on a recent evening news segment.

**Simple:** What did the reporter that Scooter avoided discuss during an evening news segment?

**Which:** Which poll did the reporter that Scooter avoided discuss during an evening news segment?

**Complex:** Which political poll did the reporter that Scooter avoided discuss during an evening news segment?

Eighty-eight fillers were included in the item list, including an equal number of subject relatives to prevent expectation of object relatives. Across each experimental list, subjects saw an equal number of questions beginning with who, what, and which, such that there was no experimental bias toward one type of question.

The five most extreme reading times from each condition were removed from the reading time data for each word, affecting less than 2.5% of the overall data. Raw reading times under 150 milliseconds were also taken out of the data, which accounted for less than 1.5% of the remaining data. This method of outlier removal was adopted for this study, instead of setting a cutoff value in standard deviations or milliseconds, because the results clearly revealed sizable asymmetries in outlier distribution across conditions.\(^2\) This outlier removal drastically lowered both the skewness (5.923 vs. .982) and kurtosis values (58.736 vs. 4.562) for the residual reading times at the verb and spillover region. Repeated measures ANOVAs

\(^2\)As it turns out, however, a post-hoc analysis conducted on the data using the standard deviation method for outlier removal produces nearly identical results.
were carried out for each word region to identify an effect of informativity. However, the more interesting tests are the comparisons of the individual treatment or condition levels that are ordered with respect to informativity. Accordingly, planned comparisons were also conducted to evaluate the difference between the three levels. Following the specifications of the Bonferroni correction, pairwise contrasts are thus considered significant if the p-value is under .0167 (.05/3), and marginal if it falls under the .05 level. Hence, in the results below, I report whether a main effect of informativity was found at each region and then specify the individual contrasts.

4.1.2 Participants
Thirty-five native English-speaking Stanford University undergraduates participated in this study to fulfill a course requirement in a linguistics class. The subjects had not participated in the previous experiment on wh-islands. As in the other reading time experiments described here, subjects were informed that they were participating in a reading comprehension experiment. Participants had a mean question-answer accuracy of 89.5% for the entire experiment, including filler items. Participants on average read each word at 337.5 ms ($SD = 61.9$). As in all other experiments described here, I consider responses only from items that were correctly answered. The experimental conditions, however, had no effect on question-answer accuracy (complex: 92.58%, $SE = 1.74$; which: 91.70%, $SE = 1.83$; simple: 92.61%, $SE = 1.73$).

4.1.3 Results
As predicted, the more informative wh-expressions leads to faster reading times at the subcategorizing verb. A main effect of informativity is evidenced at the verb ($F_{1}(2,68) = 4.631$, $p < .05$; $F_{2}(2,38) = 4.197$, $p < .05$). As revealed by individual comparisons, this main effect occurs because the “complex” which-phrase containing two adjectives produces marginally faster reading times at the verb than the bare wh-phrase ($t_{1}(34) = -2.305$, $p < .05$; $t_{2}(19) = -2.576$, $p < .05$). Similarly, the use of the which-phrase without the adjective (which) also results in significantly faster reading times at the critical region of the verb ($t_{1}(34) = 3.078$, $p < .01$; $t_{2}(19) = 3.224$, $p < .01$).

This effect intensifies during the immediately following spillover region ($F_{1}(2,68) = 6.300$, $p < .01$; $F_{2}(2,38) = 5.711$, $p = .012$; complex-simple: $t_{1}(34) = -3.459$, $p = .001$;
t_2(19) = -2.387, p = .028; WHICH-SIMPLE: t_1(1,34) = 2.763, p < .01, t_2(19) = 3.472, p = .003). Analyzing together both the verb and the subsequent word that evidences spillover effects, informativity creates reliably significant differences across conditions (F1(2,68) = 8.960, p = .001; F2(2,38) = 7.880, p < .01; COMPLEX-SIMPLE: t_1(34) = -3.569, p = .001; t_2(19) = -2.616, p = .017; WHICH-SIMPLE: t_1(34) = 3.584, p = .001; t_2(19) = 3.934, p = .001). As is evident from Figures 4.2, however, there is no significant difference between the WHICH and COMPLEX conditions. Presumably, the addition of one extra adjective is insufficient to create a noticeable difference in retrievability.

Prior to the subcategorizing verb, the reading time results initially show a slight advantage for the SIMPLE condition. For instance, at the first word after the wh-phrase, the auxiliary did, the different conditions produce fairly similar results, as shown in Table 4.1. But, at the next word, marginally faster reading times are seen in the SIMPLE condition as compared to the COMPLEX condition (t_1(34) = 2.477, p = .018; t_2(19) = 2.049, p = .055), and non-significantly so in contrast to the WHICH condition. By the following word, which corresponds to the head noun of the subject NP, reading times are again nearly identical across conditions.

At the complementizer, the more informative conditions result in a temporary speed-up in reading times for the first time, creating a main effect of informativity that is only significant by subjects (F1(2,68) = 7.130, p < .01; F2(2,38) = 1.757, p > .1). The raw reading times for both the COMPLEX and WHICH conditions are faster than those in the SIMPLE condition. This contrast, however, only reaches significance by subjects when considering the difference between the WHICH and SIMPLE conditions (t_1(34) = 3.264, p < .01; t_2(19) = 1.891, p = .074).

As in the wh-island investigation, there is a small difference at the embedded subject, as specified in Table 4.1, which does not reach significance by either subjects or items. It is likely that this small effect originates with the differences observed at the complementizer, spilling over onto the subsequent word.

Interestingly, at the verb (avoided in (68) above) immediately preceding the actual subcategorizing verb, a mild effect of informativity is evident: the COMPLEX condition results in marginally faster reading times than the SIMPLE condition, and there is a non-significant trend in the same direction for the comparison between the WHICH and SIMPLE conditions (COMPLEX vs. SIMPLE: t_1(34) = -2.372, p = .023; t_2(19) = -1.974, p = .063; WHICH vs. SIMPLE: t_1(34) = 1.506, p = .141; t_2(19) = 1.743, p = .098). This result,
<table>
<thead>
<tr>
<th>Region + Example</th>
<th>COMPLEX</th>
<th>SIMPLE</th>
<th>WHICH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RWRT</td>
<td>RZRT</td>
<td></td>
</tr>
<tr>
<td>Region 1 (did)</td>
<td>280.26 (5.07)</td>
<td>-40.05 (4.19)</td>
<td>271.24 (4.71)</td>
</tr>
<tr>
<td></td>
<td>272.41 (4.48)</td>
<td>-53.20 (4.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>271.24 (4.71)</td>
<td>-47.47 (4.04)</td>
<td></td>
</tr>
<tr>
<td>Region 2 (the)</td>
<td>265.08 (4.61)</td>
<td>-57.32 (4.23)</td>
<td>257.38 (4.25)</td>
</tr>
<tr>
<td></td>
<td>251.24 (3.75)</td>
<td>-72.36 (3.66)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>257.38 (4.25)</td>
<td>-64.37 (3.69)</td>
<td></td>
</tr>
<tr>
<td>Region 3 (driver)</td>
<td>264.14 (5.63)</td>
<td>-92.11 (6.23)</td>
<td>259.67 (5.33)</td>
</tr>
<tr>
<td></td>
<td>267.27 (6.03)</td>
<td>-94.33 (6.48)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>259.67 (5.33)</td>
<td>-93.45 (5.91)</td>
<td></td>
</tr>
<tr>
<td>Region 4 (that)</td>
<td>259.07 (4.81)</td>
<td>-69.54 (4.78)</td>
<td>246.95 (4.36)</td>
</tr>
<tr>
<td></td>
<td>269.55 (5.75)</td>
<td>-60.91 (5.28)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>246.95 (4.36)</td>
<td>-80.08 (4.18)</td>
<td></td>
</tr>
<tr>
<td>Region 5 (Jill)</td>
<td>262.07 (5.07)</td>
<td>-82.31 (5.53)</td>
<td>262.39 (5.34)</td>
</tr>
<tr>
<td></td>
<td>274.26 (5.48)</td>
<td>-71.66 (5.66)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>262.39 (5.34)</td>
<td>-79.25 (5.38)</td>
<td></td>
</tr>
<tr>
<td>Region 6 (knew)</td>
<td>260.46 (5.04)</td>
<td>-86.48 (5.36)</td>
<td>270.63 (6.29)</td>
</tr>
<tr>
<td></td>
<td>281.38 (5.48)</td>
<td>-67.82 (5.89)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>270.63 (6.29)</td>
<td>-79.69 (6.30)</td>
<td></td>
</tr>
<tr>
<td>Region 7 (win)</td>
<td>267.38 (5.60)</td>
<td>-73.98 (5.83)</td>
<td>270.30 (5.92)</td>
</tr>
<tr>
<td></td>
<td>293.81 (7.69)</td>
<td>-51.67 (7.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>270.30 (5.92)</td>
<td>-73.49 (5.82)</td>
<td></td>
</tr>
<tr>
<td>Region 8 (after)</td>
<td>268.62 (5.41)</td>
<td>-60.77 (5.71)</td>
<td>270.87 (6.11)</td>
</tr>
<tr>
<td></td>
<td>305.33 (6.81)</td>
<td>-32.80 (7.45)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>270.87 (6.11)</td>
<td>-51.67 (3.61)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment IV. Each measure is followed by its mean standard error.
though, is unsurprising given a memory retrieval account, since processing of this verb also requires retrieval of a filler. A retrieval process would then locate two different fillers located in memory, including the relevant *wh*-phrase. The processor then needs to decide which of the available fillers in memory is the correct one.\(^3\) According to the retrieval-based account, retrieval of a more informative *which*-phrase happens faster, allowing the processor to dismiss it as the wrong filler in favor of the closer subject NP.

Figure 4.1 also indicates a notable effect due to the interaction of word region and informativity. As stated above, reading times for the complex and which conditions are initially slower than those of the simple condition; however, at the verb and spillover region, this contrast reverses (see Figure 4.3 as well). The reversal is responsible for an interaction between region and informativity that is highly significant by subjects and items (complex-simple: \(F_1(1,34) = 24.088, p < .0001; F_2(1,19) = 23.107, p = .0001\); which-simple: \(F_2(1,19) = 37.770, p < .0001\)). This interaction of region and informativity corroborates the predictions of the “pay now-pay later” principle, which asserts that retrieval costs can be offset by earlier processing.

\(^3\)Such a decision process would not be necessary if memory for fillers is organized like a stack. If this were the case, though, it would be difficult to understand why informativity effects appear at this verb. Moreover, all retrieval of filler-phrases would presumably be trivially easy and resistant to interference effects if it only involved peeling the topmost element off of a stack.
Figure 4.2: Residual reading times at verb plus next word (e.g. discussed) in experiment IV

Reaction times pattern in the same way as in the investigation of wh-islands. The results exhibit a facilitation for the complex condition (mean RT = 1669 ms) as compared to the simple condition (mean RT = 1897.8 ms) \((t_{1}(34) = -2.539, p = .016, t_{2}(19) = -3.501, p = .002)\). Moreover, there was a numerical trend for faster reactions for the complex condition, in contrast to the which condition \((t_{1}(34) = -1.848, p < .1, t_{2}(19) = -1.535, p > .1)\). As in the experiment on wh-islands, however, the reaction time results are confounded by the fact that the which-N’ phrases contained the same head noun as the correct answer. Similarly, the complex condition explicitly mentions an adjective appearing in the answer that the which condition does not mention. Thus, the explanation for the reaction time results undoubtedly stems (at least, in part) from basic priming or word recognition effects.

4.1.4 Discussion

Paralleling the evidence from the empirical investigations of syntactic islands, the data on nested dependency processing reveal that using more informative filler-phrases improves processing ease at the retrieval site. This result argues for the position that the observed effects in the case of the wh-islands were not a relic of island processing, as they appear in
other types of unbounded dependencies as well. Most importantly, this experiment replicates not only the general effect of information-encoding on subsequent reading times, but also the localization of the facilitation to retrieval sites. Wh-phrase informativity appears to have some marginal effect at the complementizer marking the clause boundary—akin to the findings of the CNPC and wh-island studies—but this effect disappears by the next word. Only at the two retrieval sites, strikingly, does increased syntactic and semantic information create a significant and lasting processing facilitation. This facilitation effect contrasts with the processing slowdown that happens immediately after the more complex wh-phrases—a result that undoubtedly relates to the cost of parsing and integrating the additional syntactic and semantic information. A positional contrast or trade-off, as shown in Figure 4.3, is wholly expected on the resource-driven account outlined in the first chapter. That is, processing resources spent earlier on the encoding of some mental representation effectively offset the cost of retrieving that representation at some later point in sentence processing.

The verification of an effect at the verb prior to the true retrieval site is actually an argument in favor of the facilitation hypothesis. Retrieval of the grammatical subject must
occur at this particular site and it is ultimately reasonable, especially given the growing sentence processing literature on interference effects on dependency processing (Lewis, 1996; Gordon et al., 2001; Gordon, Hendrick, & Johnson, 2004; Warren & Gibson, 2005; Lewis & Vasishth, 2005; Gordon, Hendrick, Johnson, & Lee, 2006; Van Dyke & McElree, 2006), to expect that the nature and accessibility of other similar stored elements should influence the retrieval ease of some target. In fact, if retrieval was not the operative mechanism at play in the effects recorded here, it becomes difficult to comprehend why the effect surfaces on the first verb.

The lack of a significant difference between the two kinds of which-phrases likely stems from an insufficient difference in informativity. The complex condition adds only one word, an adjective, to the description used in the which condition. This extra feature does relatively little in terms of narrowing down the appropriate answer or creating a more distinctive representation. Notice that, while the difference between a which-N’ phrase and who is also only one word, the former specifies much that the latter does not: in addition to specifying number information, the which-phrase further identifies the class of relevant individual or object, e.g. a race. This goes beyond the specification of a mere attribute, as in the difference between race and NASCAR race. Naming the semantic class or type of an individual or thing brings along with it a host of entailed, implied, and predictable features, partly determined by real-world knowledge, e.g. races involve multiple competitors, races typically have winners and losers based on speed, etc. The ability to remember one of these implicit features can aid in the effort of retrieving the true memory target. Thus, conceptually speaking, the two kinds of which-phrases stand distinctly apart from the bare wh-item, but not very far from each other.

This finding raises the point that information discrepancies between two descriptions

4In the right context, however, where multiple entities of the same sort are being discussed (e.g. different wines are being compared), the encoding of a specific attribute (e.g. smoky, unctuous, opaque, earthy, etc.) should prove to be more significant. This goes back to the point, therefore, that the same information will be more or less useful for retrieval purposes and integration, depending upon the context.

5Arguably, the other reason why the two which-N’ phrases do not produce noticeably different reading times concerns the discourse context and the number of possible answers. Looking at the experimental item in (68), there are four possible answers in the simple condition—the two events and the two inanimate objects (the recent evening news segment and the recent ABC political poll), but only one possible answer in both the complex and which conditions. In other words, the extra adjective provides little functional utility over and above what the adjectiveless which-N’ already supplies. I consider this possibility again at the end of the chapter and suggest why this cannot be the complete explanation.
need not necessarily result in processing differences. If two expressions for the same discourse entity differ only in terms of a single, non-crucial (for purposes of identification) piece of information, these expressions are likely to produce identical effects on memory retrieval. Only when differences in information exceed some minimum threshold, therefore, should retrieval times vary significantly. Whether the underlying cause for the role of informativity in retrieval stems simply from the time involved in building the relevant representation or from the distinctiveness of the discourse representation or from some other cause, the evidence here indicates the need for some minimal distance along the relevant dimension to successfully find a contrast in retrieval.

Together with the results from the preceding chapter, the evidence from this experiment advances the idea that the processing advantage for more complex and informative wh-phrases in wh-dependencies extends beyond the context of syntactic islands. The predictions of the memory facilitation hypothesis are verified here in a perfectly grammatical context, signifying that the benefit conferred by the which-N' phrases in the wh-island experiment is not simply a strategic artifact of the island environment.

4.2 Experiment V: Beyond Wh-Phrases

So far, the experimental evidence has been focused on the processing of wh-phrases in filler-gap dependencies, largely because of historical reasons. The linguistic literature concerned with perceived differences in extractability dwells mostly on the case of wh-phrases (Chung & McCloskey, 1983; Chung, 1994; Pesetsky, 1987, 2000), although there are some counterexamples where the issue of specificity or descriptive content is considered more broadly (Cinque, 1990; Deane, 1991; Kluender, 1998).6

6The focus on wh-phrases perhaps stems from the strength of the contrast between extractions of bare wh-phrases and “complex” wh-phrases. As pointed out in the previous section, the informational distance between a bare wh-phrase and a multi-word wh-word is quite substantial, mostly owing to the vagueness and underspecification of the bare wh-phrase. Moreover, direct questions seem much harder to interpret than plain declaratives in decontextualized environments as is standard for theoretical syntactic treatments. Accordingly, the badness or unacceptability of the bare wh-items worsens without contextual support, as listeners or readers have more imaging to do to put the question into an appropriate context. This difference, which basically amounts to how much work the comprehender must do to imagine a scenario for the relevant questions, seems to be part of the underlying explanation for d-linking effects. Notice that, in the experiments covered in this dissertation, the wh-phrases in direct interrogatives are always contextually situated, suggesting that the effects shown here do not stem from differences related to the ease of building
CHAPTER 4. BEYOND ISLANDS

The memory facilitation hypothesis, however, states a general relationship between linguistic form and how this form affects subsequent retrieval. It says nothing specific about the kind of phrase for which this relationship should be relevant.\(^7\) Consequently, the same effects observed for \textit{wh}-phrases should exist for other kinds of extracted elements, given the proper conditions. In the next experimental investigation, therefore, I examine the impact of informativity in a filler-gap dependency involving a different kind of extracted element, namely, clefted indefinites. As with the \textit{wh}-phrases, the MFH predicts that encoding more information in an indefinite description should facilitate retrieval in language comprehension. For filler-gap dependencies, this translates to saying that processing of the subcategorizing head and the subsequent spillover region should be easier when more information is used to introduce the discourse entity.

4.2.1 Materials

Sixteen clefted indefinites constituted the experimental data set, as depicted in (69). Conditions varied in terms of how many adjectives preceded the head noun: zero (\textsc{simple}), one (\textsc{mid}), or two (\textsc{complex}). In all items, the clefted indefinite was followed by a relative pronoun and a five-word subject NP and then a transitive verb with an object gap, requiring the retrieval of the clefted indefinite phrase. In contrast to the other two experiments, these stimuli were followed by comprehension questions, rather than acting as the comprehension questions themselves. The comprehension question for a clefted indefinite like (69) always asked about the identity of the individual denoted by the clefted indefinite, e.g. \textit{Who was prohibited from entering the club by its members?} Three multiple choice answers were provided after displaying the subsequent comprehension question, e.g. \textit{a communist, a socialist, a capitalist}.

\begin{itemize}
  \item \textsc{simple}: It was a communist who the members of the club banned from ever entering the premises.
  \item \textsc{mid}: It was an alleged communist who the members of the club banned from ever entering the premises.
  \item \textsc{complex}: It was an alleged Venezuelan communist who the members of the club
\end{itemize}

\begin{itemize}
\item an appropriate contextual setting.
\end{itemize}

\(^7\)This does not imply, however that the effect of information-encoding on retrieval should be identical across phrase types. For instance, the effect may well be larger for noun phrases that introduce discourse-new entities than discourse-old entities.
These experimental items were included as fillers in the aforementioned nested dependency experiment. Hence, ninety-two other items acted as fillers for this data set.

From the results for each experimental condition, the five highest reading times at each region were excluded from the analysis, which eliminated almost all extreme outliers from the data set. This process affected less than 3% of the data. Furthermore, removing these outliers had the positive effect of lowering the skewness (4.305 vs. 1.631) and kurtosis (25.376 vs. 4.900) values of the data at the critical region of the retrieval site, as reflected by the different histograms in Figure 4.4 and Figure 4.5. As in the previous study, the primary statistical tests of interest are the comparisons among the three levels of informativity.

4.2.2 Participants

The same participants from the study on nested dependencies also took part in this experiment. Overall global reading times for the entire experimental data set thus are identical to those reported in the previous section. As in the other experiments, question-answer
accuracy was unaffected by condition (complex = 94.4%, SE = 3.23; mid = 94.9%, SE = 3.15; simple = 94.4%, SE = 3.32).

4.2.3 Results

Following the results from the previous experiments, the most informative indefinites (complex) induced faster reading times at the verb than the least informative indefinites (simple), as pictured in Figure 4.7. Taking into account the higher family-wise error rate, this effect was only marginally significant by subjects and by items using the standard pairwise t-tests with unadjusted p-values ($t_1(34) = -2.034, p = .05, t_2(15) = -2.373, p = .031$). At the same region, the comparisons between the simple and mid conditions did not reach significance ($t_1(34) = -1.781, p < .1, t_2(15) = -1.690, p = .112$), as was the case for the mid and complex conditions. The difference between these conditions is the presence or absence of one adjective, as was the case with the two kinds of which-phrases in the preceding experiment. Hence, not just any single addition of information necessarily improves retrieval. Note, however, that this does not contradict the predictions of the MFH, as it does not claim or speculate about how much additional information is necessary in
order to significantly improve memory retrieval.

Further inspection of the residual reading times at the verb revealed, however, a significant effect of item sequence—the position in which the item occurred during the experiment. As depicted in the trellis graph in Figure 4.6, most subjects tended to read faster as the experiment progressed. Unfortunately, randomization of items within a list does not control for these effects. Consequently, item sequence should be factored into the overall analysis.\(^8\)

There are multiple techniques for dealing with such effects. It is possible, for instance, to compute for every subject a secondary set of residuals on the basis of regression equations that calculate the effect of item sequence for each subject. In the present case, however, linear mixed effects models provide another fully adequate and reliable method for evaluating the effect of informativity in the presence of other important fixed and random factors (Pinheiro & Bates, 2000; Baayen, 2004).

Hence, a mixed effects model was fitted to the data with residual reading time at the verb dependent upon informativity and the item sequence. Both subjects and items were included as random factors in the model.\(^9\) Mixed linear models themselves do not present \(p\)-values, however, because degrees of freedom can only be approximated. Markov chain Monte Carlo (MCMC) sampling, though, can be used to reliably estimate \(p\)-values for the fixed and random factors.

Using these statistical tools and including item sequence in the model, it is possible to better evaluate the contrast between the reading times for the simple and complex conditions. The mixed effects model and MCMC sampling with 10,000 samples indeed verify that this difference is significant (\(t = 2.530, p = .013\)). The reading time difference between the most informative condition and the least informative condition therefore seems to be a real one, once other experimental factors are controlled for.\(^{10}\)

As in the other experiments, significantly faster reading times for the most informative

\(^8\)This raises the possibility that similar effects of item sequence skewed the data in one direction or the other in the previous experiments. Accordingly, linear mixed effects models that include item sequence as a fixed effect are provided in the appendix for each experiment at the critical retrieval region.

\(^9\)Additional modeling was also conducted to evaluate whether item sequence should be treated separately for each individual (i.e., whether by-subject random slopes are needed for item sequence), given that some subjects appeared to actually slow down over the course of the experiment. However, this failed to create a significant difference in the model, as determined by the likelihood ratio tests on the log likelihoods of the models (Baayen, 2007). Consequently, there is no compelling reason to fit item sequence separately for each subject.

\(^{10}\)Coefficients, standard errors, MCMC means, and other model details are provided in the appendix.
Figure 4.6: Reading time as a function of item sequence for subjects in experiment IV.
condition did not surface prior to the verb where retrieval of the filler occurs. At the word immediately before the verb, there was clearly no effect of informativity (COMPLEX = 328.7 ms, SE = 9.58; MID = 330.4 ms, SE = 9.45; SIMPLE = 321.6 ms, SE = 9.86). If anything, the reading times point to slightly faster processing in the SIMPLE condition, as depicted in Table 4.2. In fact, at the three words prior to the subcategorizing verb, there were no noticeable differences across conditions.

Only at the plural head noun of the subject NP (e.g. members in (69) above), does informativity result in marginally faster reading times, according to pairwise t-tests, but only by items and not by subjects (t₁(34) = -1.220, p = .231; t₂(15) = -2.356, p = .032); however, this effect immediately disappears long before the subcategorizer is reached.¹¹ Moreover, reading times at the first word after the indefinite, the relativizer who, are much slower in the Complex condition than in either the MID or SIMPLE conditions. This result parallels effects seen in previous studies (and to be seen again in other studies), where reading times immediately after the most complex and informative phrases show temporary slowdowns.

Interestingly, spillover effects were not observed in this study, undoubtedly due to the relative processing ease of the sentences. At the first word after the retrieval site, reading times level out across conditions (COMPLEX = -6.52, SE = 5.82; MID = 10.87, SE = 7.26; SIMPLE = 2.63, SE = 3.71). Unlike the other experiments discussed so far, the materials in this study lacked both syntactic island environments and nested dependencies. Thus, these sentences proved relatively easy to comprehend. There was, however, an extremely noticeable rise in reading times after the presence of the gap was confirmed at the first word of the prepositional phrase, which verifies that participants were aware of the gap. Due to the size of the available spillover region and an observable reduction in reading times, the increase in reading times after the verb cannot be ascribed to end of sentence wrap-up effects.

Comparing the reading times of the COMPLEX and SIMPLE conditions in the first word region after the indefinite NP (which was always the relative pronoun who) to reading times at the verb, a significant interaction occurs between word region and informativity (F₁(1,34) = 19.742, p < .0001; F₂(1,15) = 14.413, p < .01).¹² This interaction reflects

¹¹It’s unclear why exactly this effect emerges in the first place, however briefly, but one tentative idea is that the complex condition reduces interference not only on retrieval, but also upon encoding of other discourse entities. See Gordon et al. (2006) for evidence supporting such a proposal.

¹²A similar analysis contrasting the MID condition with the SIMPLE condition failed to produce a significant
<table>
<thead>
<tr>
<th>Region + Example</th>
<th>COMPLEX</th>
<th>MID</th>
<th>SIMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (who)</td>
<td>RWRT 392.54 (12.99)</td>
<td>346.73 (8.82)</td>
<td>340.60 (8.81)</td>
</tr>
<tr>
<td></td>
<td>RZRT 66.4 (11.96)</td>
<td>18.06 (7.56)</td>
<td>17.5 (7.63)</td>
</tr>
<tr>
<td>Region 2 (the)</td>
<td>RWRT 315.2 (7.52)</td>
<td>315.7 (6.71)</td>
<td>309.1 (7.41)</td>
</tr>
<tr>
<td></td>
<td>RZRT -4.0 (6.53)</td>
<td>-4.5 (6.00)</td>
<td>-7.2 (6.12)</td>
</tr>
<tr>
<td>Region 3 (members)</td>
<td>RWRT 319.4 (7.56)</td>
<td>333.9 (9.76)</td>
<td>339.4 (12.79)</td>
</tr>
<tr>
<td></td>
<td>RZRT -40.6 (7.74)</td>
<td>-30.6 (9.61)</td>
<td>-18.6 (10.67)</td>
</tr>
<tr>
<td>Region 4 (of)</td>
<td>RWRT 333.5 (7.07)</td>
<td>348.2 (8.60)</td>
<td>328.5 (7.74)</td>
</tr>
<tr>
<td></td>
<td>RZRT 20.3 (5.93)</td>
<td>33.7 (7.46)</td>
<td>17.6 (6.58)</td>
</tr>
<tr>
<td>Region 5 (the)</td>
<td>RWRT 308.0 (6.86)</td>
<td>312.6 (6.84)</td>
<td>306.5 (6.40)</td>
</tr>
<tr>
<td></td>
<td>RZRT -13.3 (6.17)</td>
<td>-10.4 (6.20)</td>
<td>-12.2 (5.24)</td>
</tr>
<tr>
<td>Region 6 (club)</td>
<td>RWRT 328.7 (9.58)</td>
<td>330.4 (9.45)</td>
<td>321.6 (9.86)</td>
</tr>
<tr>
<td></td>
<td>RZRT -14.7 (8.84)</td>
<td>-12.3 (8.98)</td>
<td>-17.7 (8.76)</td>
</tr>
<tr>
<td>Region 7 (banned)</td>
<td>RWRT 339.5 (10.45)</td>
<td>345.73 (11.37)</td>
<td>365.36 (12.22)</td>
</tr>
<tr>
<td></td>
<td>RZRT -12.35 (9.35)</td>
<td>-8.88 (10.43)</td>
<td>16.84 (11.40)</td>
</tr>
</tbody>
</table>

Table 4.2: Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment V. Each measure is followed by its mean standard error.
the fact that reading times are initially higher in the complex condition than the simple condition, but the converse is true at the retrieval site, as pictured in Figure 4.9. Hence, this positional trade-off echoes the pattern of results found in the nested dependency study and adds further weight to the “pay now-pay later” principle.13

Response times for this study once again reveal the role of lexical and syntactic priming in the speed of response to the comprehension questions. As a reminder, the comprehension questions for this experiment always posed a yes/no question that targeted information about the clefted indefinite NP. To keep the form of comprehension questions and the answers constant across conditions, subjects always saw a simple, unmodified indefinite in the answers, e.g. a poet or a communist. Thus, even if the experimental condition for that item involved one or two adjectives in the description, e.g. a reclusive poet/a reclusive English poet, the possible answers only contained unmodified indefinites. Unsurprisingly, then,

13Experiments I and II also manifest signs of this trade-off, though they are not as significant as in the two studies described in this chapter. For reasons that are unclear, signs of an interaction are not apparent in Experiment III, despite a small (but non-significant) trend for higher reading times immediately after the longer adjunct phrases.
subjects responded fastest in the simple condition where the form of the clefted indefinite matches exactly with the form of the indefinites in the answer set (SIMPLE-COMPLEX: $t_1(34) = 6.543, p < .001, t_2(15) = 4.902, p < .001$; SIMPLE-MID: $t_1(34) = 4.621, p < .001, t_2(15) = 3.601, p < .01$). Moreover, subjects were slowest in the complex condition, where, at the opposite end of the spectrum, the form of the indefinite was most different from the form of the possible answers. This pattern of results therefore confirms the intuition that response times largely depend upon the extent to which elements of the question and answer match.

4.2.4 Discussion

Overall, the replication of the memory facilitation effect in dependencies involving indefinites provides evidence that the advantage of using more explicit filler-phrases is not restricted to who-dependencies. A more general factor appears to be operative in language comprehension. The energy and cognitive resources devoted to building a more complex representation pays off inasmuch as subsequent language processing that involves reference to that representation is facilitated. Essentially, this effect reveals a system of investment
Figure 4.9: Residual reading times at first word after the indefinite NP (region 1) contrasted with reading times at the verb (region 7) in experiment V

and reward. As shown in the results of this experiment, as well as the nested dependency and *wh*-island experiments, building a more complicated representation sometimes initially requires a greater degree of resources, but expending these resources early has an observable upside just when these representations need to be re-accessed. A similar point is made in Kluender (1998) in a discussion of referential processing effects:

“Although an NP with an individuated referent will entail a heavier start-up processing cost due to its increased lexical semantic and lower accessibility in discourse representation, this initial cost helps this referent to maintain an elevated level of computational activation throughout the rest of the sentence.” (p.255)

The retrieval effects can thus be thought of as the byproduct of hard work. A relatively onerous task will likely be more memorable (and rewarding) than some trivial task, so recalling the output of the harder task is likely to be easier.

The results from the study of clefted indefinites contribute two critical points to the discussion of memory facilitation: (1) even in relatively simple syntactic constructions, a
significant difference emerges from encoding more or less information in the description of a discourse entity that must be retrieved for purposes of interpreting a filler-gap dependency; (2) this effect is not unique to *wh*-phrases. The first point builds on the result from the nested dependency experiment by illustrating that the nature of the effect found in the examination of *wh*-islands is not unique or dependent upon the atypicality of syntactic island violations. In considering now both the case of nested dependencies and clefted indefinites, the localized effects of processing facilitation restricted to the integration site appears consistently across construction types. That is, this experiment on indefinites adds further weight to the proposal that the degree of encoding impacts sentence processing via memory retrieval in a way that permeates throughout discourse.

With respect to the second point, the results from this experiment deliver some much needed evidence to counter some concerns about the results from the experiments focused on *wh*-phrases. One issue that arises from the consideration of *wh*-dependency processing is how the number of plausible answers in the preceding discourse affects interpretation of the interrogative. It is not unreasonable to speculate that a more informative *wh*-phrase constrains the search for possible answers and that reducing the size of the answer set allows for faster processing at the integration site.\(^{14}\) This experiment shows that facilitated processing occurs with the retrieval of more informative phrases, even when these words or phrases introduce new entities into the discourse.

Lastly, the results from this experiment stress one other conclusion also supported by the nested dependency data: the addition of supplemental information need not result in significantly different processing ease. The evidence at hand suggests that, in order to observe a significant difference in subsequent processing, two different descriptions of the same discourse entity need to be at a minimal informational distance from each other. The version of the clefted indefinite appearing in the MID condition differed from the other conditions by only the presence or absence of one adjective. While this condition resulted in significantly faster reading times at the verb than the SIMPLE condition, the contrast between the MID and COMPLEX conditions was not significant.

\(^{14}\)Note that the experiments in the next chapter address this question again, showing that even when *wh*-phrases are not functionally used to pose direct questions, the effect of informativity still persists.
4.3 Summary

In five different experiments now, the reading-time results consistently point to a processing difference engendered by the quantity of information encoded in the filler-phrase. Each experiment repeated the same processing advantage for the conditions with more informative filler-phrases. This finding surfaces in investigations of the processing of both direct and indirect questions, as well as FGDs with referential fillers; it occurs whether the extracted element is an argument or an adjunct; and it is found whether the construction involves a putative island or not.

Taking stock of the results so far, the question arises as to why memory retrieval and not some other processing factor should be ultimately responsible for the observed processing differences. That is, while the data may unanimously point to clear processing difference in these FGDs, it is worth considering the reasons for identifying memory retrieval as the source of these differences.

The localization of the informativity effects to the retrieval site and the subsequent spillover region poses one of the strongest arguments for this position. The general absence of effects prior to the retrieval site immediately rules out some alternative explanations such as storage-based costs which would be expected to yield much earlier repercussions. Even in the two experiments which showed informativity effects prior to the subcategorizing verb—the CNPC and adjunct extraction studies—the results were nevertheless compatible with the memory facilitation hypothesis, given some reasonable assumptions. Any explanation of these effects, therefore, must ultimately reckon with their strong and consistent emergence at retrieval points.

Given this constraint, however, there is another possible way of interpreting the data so far: integration costs, rather than retrieval, may underlie the informativity effects. On such a view, a more informative filler-phrase improves processing at the subcategorizer for one of the following reasons: (i) the filler-phrase is a more predictable argument of the verb when it encodes more information, i.e. an argument from predictability; (ii) the additional information causes the filler-phrase to be a more likely or plausible argument of the subcategorizer, i.e. an argument from plausibility; (iii) the more an extracted who-phrase restricts the possible answer set, the easier it becomes to compute the answer, i.e. an argument from computation and/or set-size in interrogatives.
The first of these points is perhaps the easiest to confront. An argument from predictability would say that a bare *wh*-phrase like *who* is less likely to be an object of a verb like *evict* than a more informative *wh*-phrase such as *which tenant*. Essentially, such an argument boils down to collocational frequency. The empirical evidence, however, actually aligns against such a position. For one, *which-N’* phrases and other multi-word *wh*-expressions are much less frequent in direct and indirect questions than bare *wh*-words. In the Switchboard corpus of spoken English (Godfrey, Holliman, & McDaniel, 1992), for instance, which contains some 2.7 million words of spoken English from 2340 telephone conversations, dependencies with bare *wh*-words outnumber multi-word *wh*-phrases by a ratio of 10 to 1 (not counting relative pronouns). In the widest sense, therefore, dependencies with complex *wh*-phrases are far less predictable. The other empirical point arguing against a reading of the results based on predictability is that, generally speaking, verbs are not highly predictive of their objects (and clearly, fronted objects are hardly predictive of particular verbs). For instance, the verb *evict* only predicts the presence of a highly plausible argument like *tenants* 4% of the time in the British National Corpus. So while predictability or expectations of upcoming constituents undoubtedly factors into the difficulty of sentence processing, there is little evidence to support an analysis of the experimental data based on predictability.

The second type of integration explanation—the argument from plausibility—is a more thorny issue to tackle. On this view, adding information during the encoding process increases the plausibility of the filler-phrase as an argument of the subcategorizer. So, for instance, *a careless pedestrian* might be a more plausible object of a verb like *hit* or *collided with* than a less informative description like *a pedestrian*, as in a sentence like *The bus hit/collided with a (careless) pedestrian*. Previous reading time studies confirm that implausible arguments in FGDs produce increased reading times at the subcategorizing verb, when compared with a plausible argument (Traxler & Pickering, 1996).

The addition of extra information about some discourse entity can theoretically, under the right circumstances, make an event or proposition easier to imagine and understand, and consequently easier to integrate at a retrieval site (particularly so when the discourse context is limited). Grodner and Gibson (2005) offer such an explanation in a discussion of anti-locality effects: “intervening constituents can make a verb easier to interpret without being dependent on that verb. For instance the sentence in [70b] is more plausible than that in [70a], because the subject modifier provides a basis for the verbal event.”
(70)  a. The fisherman cried.
     b. The fisherman who was cutting onions cried.

Applied to the data documented here, this type of explanation implies that more complex wh-phrases and indefinites present more plausible arguments for their subcategorizing heads. Some of the which-N’ phrases used in the preceding experiments do, in fact, provide information in the head noun that might make the verb-object relation easier to understand, e.g. which prisoner-released, which doctor-consulted. On the other hand, it is unclear for many other stimuli whether they actually increase plausibility or not. Additionally, as one of the last experiments in this dissertation will show, even relatively incongruous extra information (e.g. the adjectives in loveable dictator, dead tourist, etc.) produces faster reading times at retrieval sites than less informative descriptions. So, there is some evidence to conclude that the effects seen at retrieval sites do not strictly depend upon plausibility; nevertheless, the ability to form a meaningful interpretation for the sentence may still be contributing to the overall effects.

The third argument, that the cardinality of the contextually provided answer set influences the speed of integration in questions, obviously only applies to the processing of direct questions, as in the wh-island study. The processing differences noted at the embedded verb in that study relate to the fact that the bare wh-phrase who has three theoretically possible answers (Albert, the managers, the employee with poor sales), while which employee only has one (the employee with poor sales):

(71)  Albert learned that the managers dismissed the employee with poor sales after the annual performance review.

bare: Who did Albert learn whether they dismissed after the annual performance review?
which: Which employee did Albert learn whether they dismissed after the annual performance review?

Thus, at the embedded verb, when computation of the answer can begin in earnest, the task

\(^{15}\text{As mentioned in the preceding chapter, the stimuli for experiment II were divided into high collocational frequency and low collocational frequency pairs. But collocational frequency, which is surely correlated with plausibility, did not have a significant effect on the reading times.}\)
is relatively simpler in the WHICH condition. That said, notice that Albert and the managers are already mentioned before the embedded verb in both conditions. Consequently, these discourse entities are likely to be excluded from the set of answers. By the embedded verb in each condition, then, only one answer is truly plausible. This consideration by itself does not preclude the possibility that the number of theoretically possible answers is influencing processing at the verb; however, it does modulate the difference across conditions.

In FGDs with referential noun phrases, as in the clefted indefinite study (experiment V), the informativity effect persists, despite there being no need to compute an answer. Similarly, in the CNPC study where the stimuli consisted of embedded indirect questions, reading times at the subcategorizing verb also revealed an effect of informativity in the predicted direction. Further experiments in the next chapter that utilize indirect wh-questions also confirm the existence of informativity effects. The fact that informativity effects remain even in contexts where the possible answer set size is irrelevant suggests that this factor cannot be the sole explanation for the facilitated processing at the gap site.

Predictability, plausibility, and computational effort undoubtedly enter into the processing difficulty equation of FGDs at some point. So despite arguing against the interpretation that these factors alone can account for the sentence processing effects documented so far, I still support the notion that these integration-based considerations might influence processing efficiency at a point that is contemporaneous with memory retrieval. The reading time results exhibited so far may thus plausibly involve contributions from one or more of these integration-based factors.

Nevertheless, given the fact that the processing effects still remain visible, even when the integration-based factors are removed from the picture, memory retrieval stands as a more robust candidate for explicating the available data. Furthermore, while there is no extant research which demonstrates or even advances the idea that longer and more informative constituents are easier to integrate, the retrieval hypothesis is bolstered by the existence of other research that directly relates information-linking and retrieval, i.e. the psychology research on elaboration reviewed in the second chapter. Indeed, further evidence from a word recognition study presented in the next chapter, where recall is directly tested, also advances the idea that the informativity effects relate to memory retrieval.
Chapter 5

Exploring the Facilitation Hypothesis

Why does encoding more information have a positive effect on subsequent retrieval in language comprehension? Or, to put it another way, what is the relationship between informativity and retrieval? There are a number of plausible candidate answers to these questions, some based on the preceding discussion. Remember that a similar sort of question plagued the theoretical syntax accounts that observed the acceptability differences related to extracting more or less specific elements (Cinque, 1990; Chung, 1994; Pesetsky, 1987). By far the most outstanding issue for these analyses was the question of why specificity or descriptive content or familiarity should have anything to do with extraction and long-distance dependencies.

The present exegesis attempts to explain this correlation in terms of a very general relationship between information processing and memory retrieval, namely that processing more information about a single entity facilitates subsequent retrieval. Nevertheless, this answer itself leads to the more fine-grained questions raised above, which have been considered previously by psychologists in the study of propositional recall (Anderson & Reder, 1979; Bradshaw & Anderson, 1982; McDaniel et al., 1989). Jacoby and Craik (1979) and Eysenck (1979), for instance, argue that additional propositional information effectively makes a target for recall more distinctive in memory. On this view, representations that are easily distinguishable from other candidate targets in memory are more easily retrieved. Building on this work, McDaniel et al. (1989) find that both elaboration and distinctiveness additively contribute to recall and accuracy, such that each factor individually improves
subsequent recall, but put together, they have an even greater effect.

In the present chapter, I explore several reasonable rationalizations of the link between informativity and retrievability in language comprehension. One possibility is that the longer amount of time spent processing a lengthier and more informative constituent enables more accurate encoding in memory. In addition, I also consider whether both syntactic and semantic information contribute to retrieval facilitation, as the MFH suggests, or if just one type of information matters. The third study in this chapter addresses the question of whether retrieval facilitation can be accounted for solely in terms of encoding cost or the amount of attention elicited by a linguistic description. In other words, this study is used to ask if increasing information improves retrieval because it is just one among numerous possible ways of drawing attention to a particular entity. Lastly, in an evaluation of how general this relationship between informativity and retrievability is (i.e. whether this relationship represents a broader cognitive principle of memory retrieval), I examine how informativity influences longer-term memory in a word recognition task.

5.1 Experiment VI: Time-based Facilitation

Providing more time to study and encode the features of a particular stimulus is usually a good way of ensuring better memory retrieval. If you have to memorize a poem or a telephone number or a combination lock, chances are that if you have more time to encode the relevant information, the better you will be at remembering that information later. Indeed, various works in psychology show this effect of study-time on subsequent retrieval. Bradshaw and Anderson (1982), for instance, argue that time plays a significant role in the improved recall of propositional information.

It is reasonable to speculate, therefore, that stretching out the encoding process of a particular discourse representation over a long period of time will positively benefit retrieval later. In all the evidence considered so far, not only have the experimental conditions differed from one other in terms of information, but they have also differed with respect to the number of words, e.g. who vs. which employee. Plausibly, this extra word may provide additional time for encoding and perhaps even redundantly encoding the necessary information, thereby making more accurate and efficient memory retrieval more probable.

Time-based facilitation in language comprehension has been proposed before to account for so-called anti-locality effects. Konieczny (2000) shows that the German verb hingelegt
gets read faster in structures like (72a) than in structures like (72b). One way of explaining this effect is to say that the intervening material provides extra time for the comprehender to anticipate the upcoming verb, although other plausible explanations exist (which are discussed again in the next chapter (Vasishth & Lewis, 2006)).

(72)  

(a) Er hat das Buch, das Lisa gestern gekauft hatte, hingelegt.
    He has the book that Lisa yesterday bought had laid down.
    'He has laid down the book that Lisa had bought yesterday.'

(b) Er hat das Buch hingelegt, das Lisa gestern gekauft hatte.
    He has the book laid down that Lisa yesterday bought had.
    'He has laid down the book that Lisa had bought yesterday.'

According to Konieczny, the relative clause material interposed between the direct object and final verb in (72a) allows the comprehender to anticipate the clause-final verb and complete processing of all clause-initial constituents, thus facilitating the processing of this lexical item.

A time-based explanation for the effects found in the foregoing studies runs slightly differently. The extra time afforded by extra words in a description concentrates attention on the representation of a single discourse entity for a longer period of time. This reasoning also accords with the theory expressed in Deane (1991), which explains extractability via limits on attentional resources.

In the following experiment, I explore the role of time-based facilitation in the retrieval effects shown previously. By using extracted elements with the same number of words, but which differ with respect to the amount of information encoded, the effect of time can be distinguished from the effect of informativity. Residual reading times remove the effect of word length across conditions, thereby ensuring that even the slight time differences in word length do not translate into a time-based advantage for either condition. If the facilitation effect actually derives only from study-time differences, then two conditions that differ in informativity, but not in the number of words, should lead to similar reading times.

At the same time, holding the number of words constant also eliminates differences of syntactic complexity.1 Retrieval contrasts were discovered previously for phrases that

---

1Phrases of equal length can obviously differ syntactically. It does not seem to be the case, however, that two phrases capable of referring to the same entity could have the same number of words, yet differ in syntactic informativity in the sense of the definition provided in chapter one. That is to say, the syntactic information in one phrase could not be a proper subset of the information in the other phrase. In this sense,
effected differences not only in semantic complexity, but also syntactic complexity.\footnote{Throughout the remainder of this dissertation, I use both the terms “informativity” and “complexity.” The first term is reserved for talking about the differences between linguistic phrases, while the latter term is generally used for distinguishing among the resulting mental representations. That is, since it is awkward to think of mental representations as being more or less informative, I have opted to use the term “complexity.”} The foregoing studies, therefore, do not establish whether both syntactic and semantic complexity are relevant for the retrieval effects, or whether only one of them is responsible for the observed effects. Thus, this experiment also tests whether retrieval effects persist in the absence of differing syntactic complexities. As the MFH states that both syntactic and semantic information factor into the determination of informativity, phrases that differ sufficiently in semantic but not syntactic informativity ought to still result in qualitatively different retrieval processes. Given that the hypothesis invokes both types of information, however, it is fully in line with the predictions of the MFH if eliminating informational differences along one dimension reduces the magnitude or probability of retrieval effects.

### 5.1.1 Materials

Sixteen items constituted the experimental materials for this experiment. The sole manipulation for these items was the amount of information encoded in an embedded `which-N' phrase. In the **PERSON** condition, which expressed less information, the head noun was always “person”. Contrasting with this, in the **TYPE** condition, the head noun describes the type of individual, typically with an occupational title. The `wh'-phrase was separated from its subcategorizing verb by six words, as in (73) below. The six words always consisted of a definite description with a relative clause attached.

\begin{align*}
(73) & \quad \text{a. TYPE: The lieutenant could not remember which soldier the commander who was deeply respected ordered to scout the area ahead.} \\
& \quad \text{b. PERSON: The lieutenant could not remember which person the commander who was deeply respected ordered to scout the area ahead.}
\end{align*}

Each sentence was followed by a yes/no comprehension question on a new screen, such as *Did the captain recall who was doing reconnaissance?*. These questions did not target the content of the `wh'-phrase, since this content necessarily differed across conditions. Half of the comprehension questions had ‘yes’ answers and half had ‘no’ answers.
In contrast to some of the other experiments on *wh*-dependency processing discussed in this thesis, these *wh*-phrases do not require the computation of a set of possible answers. Thus, this experiment removes the potential confound in some of the previous *wh*-dependency studies: a smaller set of possible answers may reduce processing difficulty at the integration site. This potential factor is removed from the list of possible influences, since the *wh*-phrases are not being used to ask a direct question. Additionally, the lack of any preceding context obviates the possibility of looking backward or checking the discourse for possible interpretations of the *wh*-phrase.

Because both *wh*-phrases utilized in this experiment are *which*-N’ phrases, it is also possible to evaluate here whether the differences discovered in the other investigations of *wh*-dependencies were lexically driven by inherent differences between bare *wh*-items like *who* and *what* versus *which*-phrases. Those that argue for (categorical) lexical or phrasal distinctions between specific or non-specific (Chung, 1994; Cinque, 1990), or D-linked versus non-D-linked words or phrases would not predict any difference between these two *wh*-phrases. By contrast, the memory facilitation hypothesis predicts that specifying the type of individual, rather than merely making a statement about some anonymous and featureless individual, should result in faster retrieval.

The materials for this experiment were included in a set of seventy-six total items. Hence, sixty items acted as distractors for the experimental items in this study. All extreme outliers at each region were removed before conducting the statistical analyses. This significantly reduced both the skewness value (3.874 vs. 1.294) and the kurtosis value (20.495 vs. 3.051) for the residual reading times at the verb and subsequent word.

### 5.1.2 Participants

Twenty-seven Stanford undergraduates who were native speakers of English participated in this study for course credit in a linguistics class. None of them had participated in any of the other experiments described here. As in all the other studies in this dissertation, informativity had no significant effect on question-answer accuracy (*person* = 90.74%, *SE* = 1.98; *type* = 87.96%, *SE* = 2.21).

### 5.1.3 Results

The reading-time data from this experiment show a clear conditional reversal of processing effort at the retrieval site. As depicted in Figure 5.1, the *person* condition is read faster
than the type condition immediately after the *wh*-phrases (F1(1,26) = 4.961, *p* < .05; F2(1,15) = 7.520, *p* < .05). After this word region along the filler-gap path, however, reading times across conditions equalize and remain generally equivalent until the retrieval site, as can be seen in Figure 5.2. That is, the reading times for the person condition are never significantly slower than those for the type condition up until the point of retrieval and integration (*ordered* in (73) above).

Here, the roles reverse: reading times for the semantically richer type condition suddenly become faster than those for the person condition (F1(1,30) = 5.201, *p* < .05; F2(1,15) = 4.493, *p* = .05), following the predictions of the MFH. This effect spills over onto the subsequent word as well, creating a marginal effect of informativity (F1(1,26) = 2.757, *p* = .1, F2(1,15) = 5.859, *p* < .05). Looking at both regions simultaneously, the effect of informativity reaches standard levels of significance (F1(1,26) = 5.724, *p* < .025; F2(1,15) = 6.992, *p* < .025).

At the ninth word region, corresponding to the second word after the verb, the person condition momentarily exhibits marginally faster reading times (F1(1,26) = 2.911, *p* = .100; F2(1,15) = 3.045, *p* = .101), but the next word region shows this effect to be fleeting (Fs < 1). The last three word regions in the stimuli once again show equalized reading times.

---

**Figure 5.1:** Residual reading times at first word after *wh*-phrase in experiment VI
CHAPTER 5. EXPLORING THE FACILITATION HYPOTHESIS

Figure 5.2: Residual reading times, word by word in experiment VI, ranging from the first word after the *wh*-phrase to the penultimate word in the sentence.

Figure 5.3: Residual reading times at matrix verb (e.g. ordered) in experiment VI.
<table>
<thead>
<tr>
<th>Region + Example</th>
<th>PERSON</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (the)</td>
<td>RWRT: 370.18 (9.06)</td>
<td>RZRT: -12.56 (7.30)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 393.40 (9.28)</td>
<td>RZRT:  7.93 (7.00)</td>
</tr>
<tr>
<td>Region 2 (commander)</td>
<td>RWRT: 389.74 (10.30)</td>
<td>RZRT: -38.27 (8.03)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 374.71 (9.24)</td>
<td>RZRT: -47.02 (7.21)</td>
</tr>
<tr>
<td>Region 3 (that)</td>
<td>RWRT: 400.91 (11.28)</td>
<td>RZRT: 10.23 (8.97)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 421.52 (12.20)</td>
<td>RZRT: 28.30 (10.19)</td>
</tr>
<tr>
<td>Region 4 (was)</td>
<td>RWRT: 385.71 (9.85)</td>
<td>RZRT: -0.28 (8.31)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 392.20 (9.93)</td>
<td>RZRT:  5.26 (8.62)</td>
</tr>
<tr>
<td>Region 5 (deeply)</td>
<td>RWRT: 364.46 (9.31)</td>
<td>RZRT: -54.50 (8.57)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 361.92 (8.00)</td>
<td>RZRT: -60.10 (7.73)</td>
</tr>
<tr>
<td>Region 6 (respected)</td>
<td>RWRT: 376.37 (9.84)</td>
<td>RZRT: -50.45 (8.02)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 370.69 (9.17)</td>
<td>RZRT: -55.84 (7.72)</td>
</tr>
<tr>
<td>Region 7 (ordered)</td>
<td>RWRT: 468.94 (15.07)</td>
<td>RZRT: 43.42 (12.42)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 422.02 (12.76)</td>
<td>RZRT:  4.34 (9.65)</td>
</tr>
<tr>
<td>Region 8 (to)</td>
<td>RWRT: 409.48 (10.45)</td>
<td>RZRT: 19.07 (8.20)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 388.27 (8.77)</td>
<td>RZRT: -2.46 (6.43)</td>
</tr>
<tr>
<td>Region 9 (scout)</td>
<td>RWRT: 359.45 (8.027)</td>
<td>RZRT: -49.02 (7.12)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 378.26 (8.57)</td>
<td>RZRT: -30.68 (7.58)</td>
</tr>
<tr>
<td>Region 10 (the)</td>
<td>RWRT: 346.32 (7.49)</td>
<td>RZRT: -51.73 (6.85)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 346.56 (7.04)</td>
<td>RZRT: -55.76 (6.33)</td>
</tr>
<tr>
<td>Region 11 (area)</td>
<td>RWRT: 357.42 (7.35)</td>
<td>RZRT: -49.37 (7.00)</td>
</tr>
<tr>
<td></td>
<td>RWRT: 349.72 (8.09)</td>
<td>RZRT: -57.74 (7.41)</td>
</tr>
</tbody>
</table>

Table 5.1: Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment VI. Each measure is followed by its mean standard error.
<table>
<thead>
<tr>
<th>Region</th>
<th>Test</th>
<th>df</th>
<th>$F$</th>
<th>$p - value$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>Subject</td>
<td>(1,26)</td>
<td>5.201</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,15)</td>
<td>4.493</td>
<td>.05</td>
</tr>
<tr>
<td>Next Word</td>
<td>Subject</td>
<td>(1,26)</td>
<td>2.757</td>
<td>.1</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,15)</td>
<td>5.859</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Verb + Next Word</td>
<td>Subject</td>
<td>(1,26)</td>
<td>5.724</td>
<td>&lt; .025</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>(1,15)</td>
<td>6.992</td>
<td>&lt; .025</td>
</tr>
</tbody>
</table>

Table 5.2: Effect of informativity by region inside embedded clause in experiment VI

with no effect of condition.

Overall, there is an initial advantage for the person condition prior to the retrieval site that is manifest at the first word after processing the wh-phrases. But the results also verify that the semantic complexity of the type condition facilitates the retrieval process and sentence processing immediately after retrieval. Informativity consequently has essentially opposite effects on encoding versus retrieval. This is responsible for the significant interaction between region and informativity that is found when the reading times at the first word are contrasted with those at the verb and the immediately subsequent word ($F_1(1,26) = 11.261, p < .01; F_2(1,15) = 18.920, p < .001$).

### 5.1.4 Discussion

According to the evidence, the retrieval effects stemming from differences in informativity do not exclusively depend on differences in encoding time or syntactic complexity. When encoding time and syntactic complexity are held constant, semantically richer filler-phrases in filler-gap dependencies still lead to faster processing times at the subcategorizer. The effect stands out in a particularly striking manner because, in the word region that follows the encoding process, processing is easier in the less informative condition. However, when gap-filling necessitates re-accessing the mental representation encoded by the wh-phrase, the more informative type condition facilitates this process.

These results, however, do not exclude the possibility that time and syntactic complexity contribute to the findings observed in experiments where these factors were not controlled for. Indeed, the retrieval-based informativity effects that appear in this experiment do not
have the same statistical reliability as the effects seen in the previously described experiments. On the basis of this evidence (and additional data from the next experiment), it seems reasonable to conjecture that either syntactic complexity or simply encoding time has some effect on retrieval. Nonetheless, the primary result of this experiment is that the facilitation of memory retrieval in language comprehension can be successfully observed even when time and syntactic complexity are removed from the picture.

In addition to clarifying the role of encoding time (or the lack thereof) in differences in dependency processing, this study also verifies that the effects found in \textit{wh}-dependencies do not stem from inherent lexical or phrasal properties that place syntactic restrictions on FGDs. In earlier experimental studies described here, significant contrasts were observed between informationally disparate \textit{wh}-phrases which also differed with respect to the \textit{wh}-word, e.g. \textit{who} vs. \textit{which}. The evidence from this study confirms, however, that retrieval contrasts can emerge when using \textit{wh}-phrases with the same \textit{wh}-word. Thus, the retrieval-based effects found in experiments I, II, and IV cannot be simply reduced to a dichotomy of \textit{wh}-phrase types. Any processing contrast between bare \textit{wh}-items and \textit{which}-\textit{N'} phrases, based on the cumulative evidence, is not a consequence of categorical, syntactic differences, but an effect of their differing informational complexities.

5.2 Experiment VII: Syntactic Versus Semantic Complexity

Syntactic complexity was held constant in the previous experiment in an attempt to understand the independent contribution of semantic complexity to the retrieval effects. The follow-up experiment detailed in this section explores how syntactic complexity differences affect retrieval, when semantic differences are minimized. Again, since the MFH implicates the role of both syntactic and semantic information, a greater syntactic complexity should still result in faster retrieval, in spite of reduced semantic differences.

\footnote{It is not clear how the effects of these factors could be disentangled from one another. When I refer to syntactic complexity in the remainder of this chapter, therefore, this term should be considered shorthand for syntactic complexity as well as string length.}
5.2.1 Materials

This experiment expands upon the previous investigation by including a third experimental condition that tests the processing of wh-dependencies with a bare wh-phrase. Two additional items were added to the stimuli set from the previous experiment to make a total of eighteen experimental items, each of which has three conditions, as depicted below:

(74) a. bare: The lieutenant could not remember who the commander who was deeply respected ordered to scout the area ahead.

b. TYPE: The lieutenant could not remember which soldier the commander who was deeply respected ordered to scout the area ahead.

c. PERSON: The lieutenant could not remember which person the commander who was deeply respected ordered to scout the area ahead.

In all other respects, the stimuli for this experiment are identical to those in the previous study. These experimental items were paired with 54 other filler items that differed from the fillers used in the previous study.

Data points that lay more than 2.5 SDs from a subject’s mean were excluded from further analysis. This affected less than 5% of the total data. Looking at the four-word spillover region following the critical verb (e.g. ordered), the removal of these outliers improved the skewness statistic (6.203 vs. 1.273) and the kurtosis statistic (61.823 vs. 5.873). As in the two experiments from the last chapter, three sets of planned comparisons are of interest here. A modified level for statistical significance is therefore necessary. As before, I employ the Bonferonni correction here to determine significance, which means that a p-value must be under .0167 to be deemed statistically significant. The results of the repeated measures ANOVAs that test for informativity effects are also reported below, although these results have limited interpretive value.4

5.2.2 Participants

Thirty-seven Stanford undergraduates who were all native speakers of English took part in this study for course credit. The results of two of these subjects were removed because each

---

4In fact, it is possible to have a statistically significant individual comparison, but no overall effect of condition, depending upon the number of conditions and other factors. One might reasonably ask, then, why these ANOVAs are reported at all, as they say very little and can actually be confusing. Unfortunately, the answer seems to be that the reporting of these statistics is just tradition.
of these subjects scored less than 60% on question-answer accuracy. After excluding these subjects, the question-answer accuracy results showed no effect of informativity (bare = 81.08%, SE = 2.64; person = 84.23%, SE = 2.45; type = 79.73%, SE = 2.70). Similarly, reaction times to the comprehension questions were unaffected by the informativity of the wh-expressions (bare = 2742.34, SE = 115.49; person = 2756.40, SE = 99.90; type = 2738.33, SE = 112.89).

5.2.3 Results

The reading time results for this experiment roughly parallel those of the previous study. Immediately after the wh-phrases, higher reading times surface for the type condition, as compared to the bare condition (bare-type: t1(34) = 2.884, p < .01; t2(17) = -2.271, p = .036). The comparison between the bare condition and the person condition at the first word after the wh-phrase is also marginally significant by subjects, and highly significant by items (t1(34) = 1.627, p = .113, t2(17) = -3.225, p < .01). There is not a significant difference, however, between the two which-N' phrases at this site (ts < 1.3). The ANOVA results further show an overall effect of informativity (F1(2,68) = 4.142, p < .05; F2(2,34) = 4.237, p < .05). As in all the other reading time studies reported on in this dissertation, therefore, syntactically and semantically simpler phrases create less immediate comprehension difficulty, as demonstrated by the reduced spillover effects in the bare condition.

Between this first word after the wh-expressions and the retrieval site, the reading times do not evidence any significant effect of condition. At the word region immediately prior to the retrieval site, however, there is a numerical trend for slower reading times in the bare condition (bare-person: F1(1,34) = 3.690, p = .063; F2(1,17) = 2.863, p = .109; bare-type: F1(1,34) = 3.234, p = .081; F2(1,17) = 3.817, p = .067). This marginal effect potentially relates to the fact that, in most of the items, a verb appeared in this position, thus making it a theoretically possible retrieval site. The absence of a contrast between the person and type conditions at this sentence position mirrors the fact that no such contrast appeared in the previous experiment either.

At the true retrieval site (e.g. the verb ordered), reading times dramatically increase, but no effects of informativity emerge at this sentence position. The subsequent spillover regions, however, show strong effects of informativity.\(^5\) One word after the verb, however, there is

\(^5\)The lack of informativity effects at the verb itself undoubtedly relates to the extremely elevated reading
### Table 5.3: Effect of informativity by region inside embedded clause in experiment VII.

<table>
<thead>
<tr>
<th>Region</th>
<th>contrast</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BARE-Person</td>
<td>(1,34)</td>
<td>2.673</td>
<td>&lt; .01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>2.433</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>BARE-Type</td>
<td>(1,34)</td>
<td>1.233</td>
<td>.226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>1.675</td>
<td>.247</td>
</tr>
<tr>
<td></td>
<td>PERSON-Type</td>
<td>(1,34)</td>
<td>-1.341</td>
<td>.189</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>-1.198</td>
<td>.247</td>
</tr>
<tr>
<td>Region 9</td>
<td>BARE-Person</td>
<td>(1,34)</td>
<td>2.661</td>
<td>&lt; .012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>2.025</td>
<td>.059</td>
</tr>
<tr>
<td></td>
<td>BARE-Type</td>
<td>(1,34)</td>
<td>4.314</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>3.35</td>
<td>&lt; .01</td>
</tr>
<tr>
<td></td>
<td>PERSON-Type</td>
<td>(1,34)</td>
<td>1.960</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>1.087</td>
<td>.292</td>
</tr>
<tr>
<td>Region 10</td>
<td>BARE-Person</td>
<td>(1,34)</td>
<td>1.727</td>
<td>&lt; .1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>2.323</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>BARE-Type</td>
<td>(1,34)</td>
<td>2.622</td>
<td>&lt; .025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>2.503</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>PERSON-Type</td>
<td>(1,34)</td>
<td>&lt; 1</td>
<td>.230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>&lt; 1</td>
<td>.333</td>
</tr>
<tr>
<td>Region 11</td>
<td>BARE-Person</td>
<td>(1,34)</td>
<td>1.627</td>
<td>.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>2.040</td>
<td>.057</td>
</tr>
<tr>
<td></td>
<td>BARE-Type</td>
<td>(1,34)</td>
<td>2.884</td>
<td>&lt; .01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>1.912</td>
<td>&lt; .1</td>
</tr>
<tr>
<td></td>
<td>PERSON-Type</td>
<td>(1,34)</td>
<td>&lt; 1</td>
<td>.213</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,17)</td>
<td>&lt; 1</td>
<td>.289</td>
</tr>
</tbody>
</table>
### Table 5.4: Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment VII. Each measure is followed by its mean standard error.

<table>
<thead>
<tr>
<th>Region + Example</th>
<th>BARE</th>
<th>PERSON</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region 1 (the)</strong></td>
<td>RWRT: 345.54 (8.96)</td>
<td>RWRT: 377.06 (12.53)</td>
<td>RWRT: 371.76 (9.43)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -41.57 (7.29)</td>
<td>RZRT: -10.90 (11.12)</td>
<td>RZRT: -15.41 (8.31)</td>
</tr>
<tr>
<td><strong>Region 2 (commander)</strong></td>
<td>RWRT: 402.85 (15.48)</td>
<td>RWRT: 399.39 (15.72)</td>
<td>RWRT: 403.14 (15.70)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -43.99 (13.01)</td>
<td>RZRT: -51.77 (14.36)</td>
<td>RZRT: -51.65 (13.95)</td>
</tr>
<tr>
<td><strong>Region 3 (that)</strong></td>
<td>RWRT: 394.26 (14.03)</td>
<td>RWRT: 399.12 (13.23)</td>
<td>RWRT: 418.84 (15.34)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -6.67 (12.90)</td>
<td>RZRT: -2.901 (11.46)</td>
<td>RZRT: 15.20 (14.22)</td>
</tr>
<tr>
<td><strong>Region 4 (was)</strong></td>
<td>RWRT: 388.50 (12.70)</td>
<td>RWRT: 383.43 (9.99)</td>
<td>RWRT: 375.50 (10.20)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -.96 (11.47)</td>
<td>RZRT: -4.38 (9.09)</td>
<td>RZRT: -16.12 (9.91)</td>
</tr>
<tr>
<td><strong>Region 5 (deeply)</strong></td>
<td>RWRT: 376.39 (12.01)</td>
<td>RWRT: 382.26 (11.22)</td>
<td>RWRT: 376.27 (10.47)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -69.20 (11.00)</td>
<td>RZRT: -62.50 (11.52)</td>
<td>RZRT: -64.09 (11.54)</td>
</tr>
<tr>
<td><strong>Region 6 (respected)</strong></td>
<td>RWRT: 434.80 (18.19)</td>
<td>RWRT: 412.93 (13.86)</td>
<td>RWRT: 410.34 (15.17)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -9.57 (15.93)</td>
<td>RZRT: -43.14 (12.85)</td>
<td>RZRT: -43.03 (12.67)</td>
</tr>
<tr>
<td><strong>Region 7 (ordered)</strong></td>
<td>RWRT: 546.78 (29.95)</td>
<td>RWRT: 537.13 (27.83)</td>
<td>RWRT: 536.01 (16.38)</td>
</tr>
<tr>
<td></td>
<td>RZRT: 101.90 (26.31)</td>
<td>RZRT: 81.77 (24.63)</td>
<td>RZRT: 84.01 (24.02)</td>
</tr>
<tr>
<td><strong>Region 8 (to)</strong></td>
<td>RWRT: 454.75 (16.43)</td>
<td>RWRT: 400.97 (12.30)</td>
<td>RWRT: 429.56 (13.87)</td>
</tr>
<tr>
<td></td>
<td>RZRT: 58.96 (15.60)</td>
<td>RZRT: 10.86 (11.00)</td>
<td>RZRT: 32.68 (12.22)</td>
</tr>
<tr>
<td><strong>Region 9 (scout)</strong></td>
<td>RWRT: 415.33 (12.81)</td>
<td>RWRT: 380.86 (10.93)</td>
<td>RWRT: 357.84 (8.76)</td>
</tr>
<tr>
<td></td>
<td>RZRT:.307 (12.82)</td>
<td>RZRT: -46.01 (9.30)</td>
<td>RZRT: -65.81 (10.04)</td>
</tr>
<tr>
<td><strong>Region 10 (the)</strong></td>
<td>RWRT: 378.76 (10.28)</td>
<td>RWRT: 355.04 (7.87)</td>
<td>RWRT: 350.35 (9.01)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -31.95 (8.34)</td>
<td>RZRT: -49.58 (7.08)</td>
<td>RZRT: -67.80 (9.79)</td>
</tr>
<tr>
<td><strong>Region 11 (area)</strong></td>
<td>RWRT: 378.74 (9.86)</td>
<td>RWRT: 359.49 (8.49)</td>
<td>RWRT: 352.99 (7.67)</td>
</tr>
<tr>
<td></td>
<td>RZRT: -38.56 (10.22)</td>
<td>RZRT: -59.29 (8.84)</td>
<td>RZRT: -72.64 (8.95)</td>
</tr>
</tbody>
</table>
an overall effect of informativity ($F_1(1,34) = 3.373, \ p < .05$; $F_2(2,34) = 3.785, \ p < .05$): the PERSON condition leads to significantly faster processing than the BARE condition ($t_1(34) = 2.673, \ p = .011; \ t_2(17) = 2.433, < .05$). The contrast, however, between BARE and TYPE was not significant at this region ($t_1(34)= 1.233, \ p = .226; \ t_2(17) = 1.675, \ p = .247$), nor was the difference between the PERSON and TYPE conditions ($t_2(34) = -1.341, \ p = .189; \ t_2(17) = -1.198, \ p = .247$).

In regions 9 through 11, the TYPE condition consistently generates faster reading than the other two conditions, as Table 5.4 states. At region 9—the second word after the verb—reading times for the TYPE condition are more than 60 ms faster on average than those for the BARE condition ($t_1(34) = 4.314, \ p < .001; \ t_1(17) = 3.335, \ p < .01$). Moreover, the PERSON condition is marginally slower at this region by subjects, compared to the TYPE condition ($t_1(34) = 1.960, \ p = .058; \ t_2(17) = 1.087, \ p = .292$). The PERSON condition nevertheless remains faster than the BARE condition ($t_1(34) = 2.661, \ p = .012; \ t_2(17) = -1.198, \ p = .247$).

As evidenced by the large standard errors in Table 5.4, reading times at this word region reflect a high degree of variance. The absence of an effect at the verb, however, verifies that the informativity effects found in the subsequent spillover region are not based on processing differences emanating from the word immediately before the verb.

Figure 5.4: Mean residual reading times for regions 9 to 11 in experiment VII
2.025, \( p = .059 \)). These differences account for the overall effect of informativity at this word (\( F(1,26) = 10.697, p < .01 \); \( F(2,34) = 5.683, p < .05 \)).

At the next word, the in (74), the informativity effects persist (\( F(1,26) = 3.773, p < .05 \); \( F(2,17) = 2.503, p = .023 \)); BARE-PERSON: \( t(34) = 1.727, p < .1 \); \( t(17) = 2.323, p < .05 \)); PERSON-TYPE: \( ts < 1.3 \)). Lastly, at the penultimate word in the sentence, the same trends continue though slightly reduced in magnitude as would be expected under the assumption that these effects reflect spillover costs (BARE-TYPE: \( t(34) = 2.884, p < .01 \); \( t(17) = 1.912, p < .1 \)); BARE-PERSON: \( t(34) = 1.627, p = .113 \); \( t(17) = 2.040, p = .057 \)).

As depicted in Figure 5.4, the last three words of the spillover region cumulatively point to a post-retrieval processing advantage for the most syntactically and semantically complex wh-phrase, but also a clear benefit for the syntactically complex but semantically simpler PERSON condition (BARE-TYPE: \( t(34) = 5.300, p < .0001 \); \( t(17) = 3.385, p < .01 \)); BARE-PERSON: \( t(34) = 3.133, p < .01 \); \( t(17) = 2.883, p = .01 \)); PERSON-TYPE: \( t(34) = 2.026, p = .051 \); \( t(17) = 1.348, p < .195 \)).

Looking more broadly at the entire post-gap spillover region, the reading times cumulatively indicate a processing advantage for the most syntactically and semantically complex condition (TYPE) (BARE-PERSON: \( t(34) = 4.288, p < .001 \); \( t(17) = 3.146, p < .01 \)); BARE-TYPE: \( t(34) = 4.636, p < .001 \); \( t(17) = 3.348, p < .01 \)); PERSON-TYPE: \( ts < 1 \)). At no point in this spillover region does any other condition elicit significantly faster reading times. Only at the word immediately after the verb does the PERSON condition produce faster reading times, but as stated above, this numerical difference is not significant. This meets the expectations generated by the MFH, as processing difficulty correlates with the quantity of information encoded in the wh-expression.

An interaction between word position and condition is also evident. Comparing the reading times at the first word after the wh-phrases with the reading times in the entire spillover region (which corresponds to word regions 8 through 11 in Table 5.4), the contrast between the BARE condition and the other two conditions yields highly significant interactions of region and condition (BARE-PERSON: \( F(1,34) = 15.331, p < .001 \); \( F(1,17) = 32.391, p < .0001 \)); BARE-TYPE: \( F(1,34) = 26.841, p < .0001 \); \( F(1,17) = 25.050, p = .0001 \)). The two syntactically complex conditions (PERSON and TYPE) are read slower at
the beginning of the filler-gap dependency, as compared to the bare condition. As in many of the other experimental studies, this contrast flip-flops at the retrieval site, such that the syntactically complex conditions lead to quicker reading in the spillover region, as seen in Figure 5.6.

5.2.4 Discussion

The purpose of this experiment was to further examine how informativity effects depend upon syntactic complexity versus semantic complexity. The experimental investigations from the previous chapters conflate syntactic and semantic informativity, obscuring the extent to which the retrieval effects stem from one or the other. Accordingly, this study contrasts a semantically “light” phrase like which person with the syntactically simpler who, as well as semantically richer which-N’ phrases that encode the kind or type of relevant individuals. This does not mean, however, that I interpret phrases like which person as being semantically identical to who or that I assume that both of these wh-expressions are at an equal semantic distance from a which-N’ phrase like which soldier. Minimally, which
CHAPTER 5.  EXPLORING THE FACILITATION HYPOTHESIS

Figure 5.6: Mean residual reading times at the first word after the *wh*-phrase (region 1) compared with aggregated residual reading times for post-retrieval spillover region (region 2) in experiment VII.

Figure 5.6: Mean residual reading times at the first word after the *wh*-phrase (region 1) compared with aggregated residual reading times for post-retrieval spillover region (region 2) in experiment VII.
person encodes presuppositional and number information that the bare wh-word does not. Nevertheless, the semantic distance to a bare wh-item is unquestionably minimized by the use of non-specific wh-phrase like which person as compared to a phrase like which soldier.

As in the previous study, the data from this follow-up experiment establish a mild facilitation for the retrieval of semantically richer representations, holding syntactic complexity constant. This dataset adds to the previous one, however, by showing a facilitation effect even when the semantic or conceptual distance between two representations is minimized, but a clear distinction in terms of syntactic complexity is retained. Combined, these processing contrasts thus argue in favor of the view that both syntactic and semantic information contribute to the retrieval-based effects. The semantically and syntactically richest representations lead to the most efficient retrieval process, according to the results. In contrast, representations that differ only in syntactic complexity or else in semantic complexity create less differentiation at retrieval.

Once again, the trade-off between early or late processing costs linked to retrieval effort recurs in this experiment, supporting the notion of a “pay now-pay later” principle of processing and retrieval. Increasing the complexity of linguistic representations along both the syntactic and semantic dimension has an immediate cognitive cost, but both types of complexity ultimately facilitate the retrieval process. Interestingly, the immediate processing costs for the two which-N’ phrases are roughly equal in this experiment, which corresponds with the fact that the processing times at the retrieval site and spillover region are not strongly differentiated, as evidenced in Figure 5.6. The evidence from this dissertation thus points over and over again to a strong relationship between the quantity of resources expended during the encoding phase and the resources necessary for successful retrieval.

5.3 Experiment VIII: Distinctiveness

Intuitively, events and individuals with distinctive properties stand out in memory. It can be almost hard to forget a very unpredictable or rare event or occurrence. Research in psychology has confirmed that concepts that express uncommon relationships or unexpected objects are easier to remember (in cued recall tasks, at least) (Stein, Morris, & Bransford, 1978; Eysenck, 1979; McDaniel et al., 1989). Distinctiveness has been found, for instance,
to affect propositional recall, such that more distinctive encodings improved recall. Within these studies, distinctiveness is defined as “the uniqueness of the relationships between the target concepts relative to the learner’s previous experience” (McDaniel et al., 1989, p. 358). Hence, it is not the degree to which some concept stands out from other competitors in the context (i.e. isolation or interference effects) that is relevant here. Instead, it is the distinctiveness of the relationship between sub-constituents. Another way to think about this is in terms of the parts of a mental representation and how well they naturally go together, based upon prior experience. McDaniel et al. (1989) discovered that people were better at remembering sentences that expressed unique relationships between the component concepts (e.g. The boy found a huge diamond in his applesauce). The authors conclude that “only certain types of elaborations will facilitate recall of target information” which echoes the statement in Stein et al. (1978) that “effective elaboration seems to depend on the quality rather than the quantity of the information expressed.”

Retrieval from working memory during language comprehension may consequently be subject to a similar principle. In particular, phrases that express atypical or aberrant combinations of features may be easier to retrieve just because they command a high degree of attention or cognitive resources during the initial encoding. A phrase like “the five year-old smoking a cigar” expresses an uncanny situation, which means that it likely captures a listener or reader’s attention. This increased attention at encoding potentially translates to facilitated retrieval.

Throughout the preceding experiments, I have repeatedly pointed to the interaction between encoding costs and retrieval costs. In many of the experiments, a strong interaction was observed between the processing effort expended at encoding and retrieval. What I have left unclear, however, is whether retrieval facilitation can be accounted for strictly on the basis of encoding cost. In other words, more informative expressions may benefit

---

7This is not to suggest that these effects are not interesting or relevant for the discussion of informativity effects. A large part of the next chapter, in fact, is spent reviewing the relevance and import of similarity-based interference effects.

8This is not quite the interpretation that McDaniel et al. (1989) assume. Instead, they advise that there is “better memory for those items that are presented in such a way as to create relations that are unique and non-overlapping with prior knowledge.” Why exactly this should help memory retrieval is left undisussed, but it as at least theoretically compatible with the notion that these items receive greater attention during the encoding process.
retrieval simply because they demand more attention and cognitive resources during encoding. Consequently, this attention-focusing may be at the heart of the informativity effects. In this sense, increasing informativity constitutes one of many theoretically possible ways of compelling comprehenders to focus their attention on one particular representation.

This experiment tests the hypothesis that the relationship between informativity and memory retrieval basically hinges upon how distinctive the representation is and thus how much attention the encoding process elicits. On the hypothesis that faster retrieval ultimately depends upon how distinctive the information encoded is, stereotypical information should help retrieval less than unexpected or atypical information. Alternatively, faster or equivalent retrieval times when a representation with commonly associated information is targeted would signal that distinctiveness is not the key component to retrieval facilitation.

5.3.1 Materials

The materials for this experiment consisted of twenty-four items that each contained three levels or conditions. The experimental conditions differed with respect to the amount of information provided about the object NP. In the bare condition, participants saw only a definite NP with no additional modifiers, but in the other two conditions, the definite object NP contained two additional words. The second word was held identical in both conditions, but the first word was manipulated for distinctiveness or typicality. In the typ(ical) condition, the first word expressed a common or highly predictable feature of the head noun, e.g. ruthlessness is a predictable characteristic of dictators. As a contrast, the first word in the atyp(ical) condition encoded an unpredictable or uncommon feature, e.g. dictators are not typically associated with being lovable:

(75)  
\begin{enumerate}
\item The diplomat contacted the dictator who the activist looking for more contributions encouraged to preserve natural habitats and resources.
\item The diplomat contacted the ruthless military dictator who the activist looking for more contributions encouraged to preserve natural habitats and resources.
\end{enumerate}

\footnote{Based on the evidence from previous experiments, it seemed clear that one extra word of information was likely to be incapable of producing statistically reliable informativity differences. For this reason, a second additional word was included in both the TYP and ATYP conditions.}
CHAPTER 5. EXPLORING THE FACILITATION HYPOTHESIS

148

Mean Frequency of Adjective  \( P(w|w+1) \)  \( P(w|w-1) \)
---
TYP  127613.17  .00377  .00359
ATYP  86544.87  .00012  .00063
Ratio  1.475:1  31.176:1  5.697:1

Table 5.5: Corpus results from Gigaword for TYP & ATYP conditions from experiment VIII: mean frequencies of adjectives, the mean conditional probability of the adjective preceding the head noun (\( P(w|w+1) \)) and the mean conditional probability of the head noun following the adjective (\( P(w|w-1) \)).

c. The diplomat contacted the lovable military dictator who the activist looking for more contributions encouraged to preserve natural habitats and resources.

Stimuli were followed by yes/no comprehension questions, half of which had ‘yes’ answers and half ‘no’ answers. An equal number of comprehension questions (4) targeted information in the matrix subject, matrix verb, matrix object, embedded subject, embedded verb, and post-verbal spillover region. These items appeared along with forty-eight filler items, twelve of which contained similarly complex sentences but with informationally richer subject phrases, instead of object phrases. An equal number of ‘yes’ and ‘no’ questions appeared in the entire experiment.

Residual reading times greater than 2.5 SDs were removed from the dataset before the statistical analysis was conducted. This process affected less than 2% of the data. At the critical retrieval site, this altered the skewness value from 2.607 to .989 and the kurtosis value from 13.881 to 2.420. Once again, three planned comparisons are relevant for the dataset here. Significance is therefore set at the level of .0167. As usual, corrected p-values are provided in the appendix.

Distinctiveness was determined by two methods. First of all, Gigaword, a 1.2 billion word corpus of written English, was used to evaluate how often the critical adjective appeared with the head noun. The results of these corpus searches are shown in Table 5.5. While the overall conditional probabilities for the adjective-noun combinations proved to be quite low for both conditions, the ratios of the probabilities show that the adjective-noun combinations in TYP condition are much more likely. Even taking into account the difference in mean frequencies of the adjectives, the conditional probabilities for the combinations in the TYP condition are still much higher.
Secondly, a norming study was conducted with eleven separate participants to verify the categorizations. All of these participants were native speakers of English at Stanford University and were compensated with course credit for their participation. These subjects were asked to rate on a scale of 0-10 how likely an individual of the sort described by the object head noun (e.g. dictator) is to have the characteristic described by the adjective (e.g. ruthless). Each subject saw each of the critical twenty-four head nouns used in the reading experiment with either the typical or atypical adjective, along with twenty-six filler items. Z-scores were computed for each data point by subtracting the original score from the subject’s mean.

As depicted in Figure 5.7, items categorized as typical (TYP: 2.00, SE = .127) were judged to be more likely characteristics without exception than items categorized as atypical (ATYP: -2.00, SE = .180) (F1(1,10) = 83.276, p < .0001; F2(1,23) = 200.174, p < .00001, minF’(1,19) = 58.81, p = .0001).
5.3.2 Participants

Thirty-seven students from Stanford University and Foothill Community College participated in this experiment. Subjects either received course credit or were payed $12 for their participation. Question-answer accuracies did not differ across experimental conditions (atyp: 75.68%, SE = 2.498; bare: 75.00%, SE = 2.521; typ: 77.03%, SE = 2.449). Likewise, there were no significant differences in response times to the comprehension questions (atyp: 2805.36 ms, SE = 95.395; bare: 2943.88 ms, SE = 111.734; typ: 2841.19 ms, SE = 97.524).

5.3.3 Results

Residual reading times at the most deeply embedded verb (e.g. encouraged in (75)) reveal facilitated processing for both the typ and atyp conditions compared to the less informative bare condition. Figure 5.9 illustrates that the typ condition produces the fastest reading times at the verb, followed by the atyp condition. A repeated measures ANOVA at this region thus verifies a cumulative effect of informativity (F1(2,72) = 4.231, p < .05; F2(2,46) = 4.504, p = .016). Planned comparisons among the conditions substantiate that the slightly faster reading times for the typ condition, as compared to atyp condition, are not significant (t < 1.4). However, the typ condition leads to unequivocally faster reading times at the verb than the bare condition (t1(36) = -2.752, p < .01; t2(23) = -3.055, p = .006). The atyp condition, in contrast, only yields a numerical trend for faster reading times, according to pairwise t-tests (t1(36) = -1.812, p = .078; t2(23) = -2.191, p = .039).

Further exploration of the data with mixed effects modeling, however, suggests that item sequence once again has a significant effect on reading times. In contrast to the situation in experiment V, separate intercepts and slopes for the effect of item sequence are needed for each subject. Looking at the trellis graph in Figure 5.10, we see a non-uniform effect of item sequence on subjects. While most subjects demonstrate quicker reading times throughout the course of the experiment, some subjects exhibit a more profound effect of the item position, i.e. these subjects have steeper regression slopes. Additionally, while most subjects speed up throughout the experiment, a few subjects appear to slow down as the experiment progresses. This subject variability is the reason for needing to treat the effect of item sequence separately for each subject.

Once item sequence is included in the linear regression model, the contrast between the
Figure 5.8: Residual reading times, word by word in experiment VIII, ranging from the first word after the object NP to the word after the embedded verb.
The ATYP condition and the BARE condition is judged significant (t = -2.422, p = .0157). The difference between the TYP and BARE conditions, of course, remains highly significant after item sequence is taken into account. In fact, the contrast becomes even more significant (t = -3.433, p = .0006). Even after controlling for item sequence, though, there is still no significant difference the TYP and ATYP conditions. Overall, what the results show is faster processing at the retrieval site when the targeted material is more informative. These results add to the findings already covered in this dissertation. More importantly, these results indicate that the informativity effects do not depend upon distinctiveness or some other type of oddball effect. Representations with increased complexity that have less common or atypical combinations of features show no processing advantage over representations with an equivalent complexity but more commonly associated features.

Reading times after the verb do not evidence any noteworthy spillover effects. Immediately after the verb, mildly slower reading times are evident for the ATYP condition. No significant effects, however, are observed at this region. Controlling for item sequence with linear regression, the reading times in the ATYP condition are marginally slower than those of the other two conditions, as reported in appendix C. Combining the reading times for the verb and the immediately subsequent word, linear mixed-modeling points to significantly faster reading times for the TYP condition as compared to the ATYP condition (t = 2.188, p = .029). If anything, then, the results indicate a mild advantage for common feature combinations over uncommon ones.

As in the other experiments, the slower reading times at the verb for the least informative condition (BARE) are counterbalanced by faster reading times immediately after processing the bare definite. At the first word after the definite (which was always the relative pronoun who), subjects read significantly faster in the BARE condition, as compared to the ATYP condition (t1(36) = 3.100, p = .004; t2(23) = 2.727, p <= .012). Furthermore, the TYP condition produces marginally faster reading times than the ATYP condition (t1(36) = 2.384, p = .023; t2(23) = 2.369, p = .027). These contrasts explain the main effect of informativity found at the first word region (F1(2,72) = 5.898, p < .01; F2(2,46) = 5.892, p < .025).

This configuration of reading times continues into the second word after the definite NPs, although the differences between conditions are reduced. Taking the two regions after the definite NPs together, therefore, there is a strong main effect of informativity (F1(2,72) = 6.315, p = .003; F2(2,46) = 4.769, p < .02). Planned comparisons suggest that this overall effect stems from the significance between ATYP and BARE (t1(36) = 3.291, p =
Figure 5.9: Mean residual reading times at verb in experiment VIII

.002; \( t_2(23) = 2.929, p = .008 \), as well as the marginally faster reading times for the TYP condition compared with the ATYP condition (\( t_1(36) = 2.219, p = .033; t_2(23) = 1.577, p = .128 \)). There is also a numerical trend for faster reading in the BARE condition, compared to the TYP condition (particularly at the second word region), but this does not reach significance by subjects or items. In short, the definites NPs with the atypical modification cause participants to slow down the most, as expected. Definites with more predictable or typical modifiers lead to marginally faster reading times; however, the bare definites with no modification result in the fastest referential processing times, as reflected by spillover costs.

Interestingly, at the fourth word following the definite NP, the constellation of reading times looks similar to the pattern of results seen around the retrieval site. The fourth word was always a participial modifier that constituted the first element in a reduced subject relative, e.g. *looking for more contributions*. As at the correct retrieval site that occurs five words later, the TYP condition produces the fastest reading times. In this case, though, the ATYP and BARE conditions are not significantly different from one another, but the TYP condition is marginally faster than the ATYP condition (\( t_1(36) = 2.361, p = .025; t_2(23) = 1.989, p = .059 \)). Given the existence of informativity effects in previous experiments
Figure 5.10: Residual reading times as a function of item sequence in experiment VIII
at verbs that were not the true subcategorizer, it seems reasonable to assume that the conditional effects seen at this verb form also stem from premature retrievals.

5.3.4 Discussion

The main finding of this experiment is that the strength of retrieval facilitation is not dependent upon the distinctiveness of the encoded information. Nor, for that matter, does initial referential processing cost appear to be the sole predictor of retrieval cost. The complex definite with atypical information elicited more initial processing than either of the other two conditions, yet it was not processed fastest at the retrieval site.

This fact is unexpected under the assumption that there is a strict correlation between the amount of resources expended during the encoding process and the amount of resources needed for successful retrieval. While the evidence from this experiment and the other experiments in this dissertation generally support the idea that encoding costs predict retrieval costs, the current results suggest that this relationship is not absolute. Put slightly differently, other discourse considerations and facts about retrieval may override this otherwise general processing principle.

Since the TYP condition was responsible for marginally faster reading at both the early premature retrieval site and at the true retrieval site (especially when the spillover region is considered), retrievability appears to be sensitive to not only the complexity of the targeted representation, but also the “cohesion” of the information encoded in that representation. Features or attributes that constitute a highly plausible set of shared features appear to be highly effective in facilitating future retrieval. In contrast, adding information that creates combinations of features with unlikely relationships or that are otherwise incongruous seems less beneficial for purposes of retrieval, although even this kind of information seems to be better than none at all.

A viable explanation for these facts is that inferential mechanisms or high degrees of associative strength underlie the memory retrieval benefits for complex representations that contain cohesive information. Reder (1982) and Anderson et al. (2001) point to the importance of inferential processes in the recognition of propositional information. If situational or contextual information can be recalled about some event, then that information can lead

---

10The marginally slower reading times for the ATYP condition at the word after the verb may either reflect spillover from the retrieval process or they may indicate renewed interpretive difficulties that occur following retrieval. It is unfortunately impossible to tell if just one or both of these interpretations are correct.
## Table 5.6: Raw reading times (RWRT) and residual reading times (RZRT) by region in experiment VIII. Each measure is followed by its mean standard error.

<table>
<thead>
<tr>
<th>Region + Example</th>
<th>RWRT</th>
<th>RZRT</th>
<th>RWRT</th>
<th>RZRT</th>
<th>RWRT</th>
<th>RZRT</th>
<th>RWRT</th>
<th>RZRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (who)</td>
<td>440.93 (15.75)</td>
<td>60.72 (14.12)</td>
<td>380.28 (9.84)</td>
<td>4.00 (7.91)</td>
<td>394.01 (9.61)</td>
<td>19.52 (7.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 2 (the)</td>
<td>385.55 (10.65)</td>
<td>5.24 (8.42)</td>
<td>354.51 (8.64)</td>
<td>-21.94 (7.21)</td>
<td>377.65 (10.23)</td>
<td>.533 (9.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 3 (activist)</td>
<td>413.02 (14.10)</td>
<td>-34.51 (10.60)</td>
<td>404.31 (14.27)</td>
<td>-34.39 (11.35)</td>
<td>396.49 (12.48)</td>
<td>-45.70 (10.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 4 (looking)</td>
<td>449.04 (14.19)</td>
<td>1.04 (11.39)</td>
<td>428.75 (14.55)</td>
<td>-14.15 (11.12)</td>
<td>410.79 (12.81)</td>
<td>-35.22 (9.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 5 (for)</td>
<td>391.19 (10.06)</td>
<td>17.53 (9.16)</td>
<td>393.77 (10.45)</td>
<td>23.91 (8.93)</td>
<td>388.53 (9.48)</td>
<td>17.63 (7.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 6 (more)</td>
<td>359.09 (9.87)</td>
<td>-39.04 (6.81)</td>
<td>361.58 (8.74)</td>
<td>-33.71 (6.97)</td>
<td>358.76 (8.50)</td>
<td>-34.17 (7.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 7 (contributions)</td>
<td>382.43 (10.59)</td>
<td>-49.41 (8.15)</td>
<td>374.96 (10.49)</td>
<td>-58.82 (8.41)</td>
<td>376.58 (10.91)</td>
<td>-48.36 (7.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 8 (encouraged)</td>
<td>410.37 (12.24)</td>
<td>-31.24 (9.44)</td>
<td>441.60 (15.19)</td>
<td>5.04 (11.33)</td>
<td>396.09 (11.26)</td>
<td>-39.94 (8.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region 9 (to)</td>
<td>407.09 (10.34)</td>
<td>4.06 (8.78)</td>
<td>387.57 (10.11)</td>
<td>-7.93 (8.34)</td>
<td>384.65 (9.21)</td>
<td>-9.11 (7.13)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
to the (potentially misleading) inference of other commonly associated events. For instance, Anderson et al. (2001) note that “even if we cannot directly recall that Bob ate the meal, if he went to a restaurant, ordered a meal, and paid the bill we might be willing to infer that the meal was consumed.” Thus, the retrieval of information that is highly associated with some targeted chunk or representation in memory (based upon prior linguistic and non-linguistic experience) is a means for indirectly re-accessing that target. This inferential link, however, would be unavailable if the two information features rarely if ever appear together.

Applied to the data set at hand, the possibility of differing associative strengths between memory chunks justifies the unexpectedly fast retrieval times for the TYP condition. Given a noun phrase like the ruthless dictator or the wealthy celebrity, the successful recall of the adjectival information provides a means for recovering the information corresponding to the head noun: dictators are often thought of as being ruthless and celebrities are usually assumed to be wealthy.\textsuperscript{11} Basically, if you can recall the adjective, you can get to the corresponding noun quickly, due to their strong association. An adjective like lovable, however, would not provide such efficient means for recovering a noun like dictator, since they have a very low associative strength. That said, such an atypical characteristic would still have some associative strength with the head noun at retrieval, because of their recent co-occurrence at encoding. In other words, the atypical feature does present a possible retrieval path, but just not a particularly efficient one.

In sum, the slowest retrieval times occur when the targeted representation both contains the least amount of information and elicits the least amount of processing effort at encoding. Faster retrieval times, in contrast, accompany more complex representations. Several aspects of these complex representations ultimately appear to be relevant for retrieval facilitation: they require more referential processing at encoding and they present more retrieval paths. The complex, distinctive representations elicit the highest degree of referential processing, but offer only weak associative strengths between points on the retrieval path. On the other hand, the complex, but non-distinctive representations consume fewer resources at encoding, but present an information network with strong associative strengths. That is, each kind of complex representation has a retrieval advantage to it that the other only

\textsuperscript{11}This reasoning works in the other direction as well. If retrieval at gap sites (or any overt anaphor, for that matter) involves recovering all the syntactic and semantic information in the filler, then retrieving the head noun can facilitate the recall of modifiers like adjectives.
possesses in a weaker form. In the end, therefore, both of the more informative definite expressions benefit the retrieval process, but for slightly different reasons.

As a final point, this experiment further ties the informativity effects to an unequivocally referential noun phrase. These effects have now been found in studies of dependencies involving argument and adjunct wh-phrases, indefinite NPs, and definite NPs. The effect of informativity upon retrieval consequently appears to be a rather general principle of language processing that occurs independently of phrasal type.

5.4 Experiment IX: Word Recognition

So far, every experiment in this dissertation has dealt with the effects of informativity on subsequent sentence processing. These experiments, therefore, produce measures of the impact of informativity on working memory resources used during the encoding and retrieval processes. What they do not do, however, is indicate whether the informativity of linguistic expressions affects retrieval from long-term memory. Although previous research in psychology has documented similar effects for the recall of propositional information, it has not been previously confirmed that the quantity of information encoded in a linguistic expression predicts the long-term recall of the associated mental representation or lexical information contained within the expression. Furthermore, these experiments on propositional recall have asked participants to memorize the contents of sentences as they read them. This is obviously not how people normally read and so the ensuing results may simply characterize memorization strategies rather than natural and automatic aspects of language comprehension.

In the following experiment, I depart from the self-paced reading methodology utilized up to this point and probe long-term recall of lexical information in a word recognition task. There are two basic points to this study: (1) to accumulate additional evidence that differences in memory retrieval underlie informativity effects, as opposed to integration or some other processing-based consideration, and (2) to test the generality of the relationship between information-encoding and retrievability. To address these points, I examine in this study how long-term recall accuracy varies with linguistic informativity when subjects do not intentionally plan for memorization.
5.4.1 Materials

This experiment consisted of two phases: an incidental encoding phase and a recall phase. During the encoding phase, participants read a total of forty-eight simple transitive sentences of English without knowing that their memory of those sentences would later be tested. Half of these items consisted of the experimental stimuli for this recognition experiment. The main experimental manipulation was the amount of information used to describe the patient or grammatical object. Informativity varied with the presence of zero (short), one (post and adj), or two (long) modifiers, as illustrated below.

\[(76)\] short: The general cornered the dictator in the secret room at the palace.
post: The general cornered the dictator of East Timor in the secret room at the palace.
adj: The general cornered the paranoid dictator in the secret room at the palace.
long: The general cornered the paranoid dictator of East Timor in the secret room at the palace.

Both the post and adj conditions were included in order to evaluate the positional effect of additional information.

Unlike the self-paced reading tasks described previously, the rate of word presentation in this study was not controlled by participants. Instead, following the appearance of a center-oriented cross-hair at the beginning of every item, each word in the sentence was presented for 400 milliseconds in the center of the screen. This method of stimuli presentation avoids the threatening confound of recall differences due to different study-times. That is, someone reading a given sentence more slowly might recall the contents better simply because of the extra time spent studying that information. A comprehension question followed each sentence to once again ensure that subjects actually read the sentences for understanding. The comprehension questions for the twenty-four experimental stimuli never repeated any information about the thematic patient. Half of these questions had negative answers and the other half had positive answers. Filler items for the encoding phase contained inanimate NPs in one or both positions and half of the filler items contained complex subject NPs with one or more modifiers.

Immediately after this encoding phase of the experiment, subjects were informed that their memory of the sentences seen in the encoding phase would be tested. In particular,
they received instructions stating that they would see a series of verb-noun combinations, such as “cornered-dictator,” with one word appearing on the screen at a time. Subjects were asked to indicate whether they saw both words in the same sentence, such that the verb (= the first word) described what happened to the individual referred to by the noun (= the second word). This task thus tested recognition of verb-noun pairs that appeared during the previous reading task. Subjects were also explicitly provided with an example showing that they should respond affirmatively even if multiple words intervened between the verb and noun.

Each word was presented on screen for 300 ms plus 16.667 ms for each character in the word, e.g. “dictator” would stay on the screen for 433.33 ms (300 + 16.667 x 8). Subjects were given five seconds to respond from the appearance of the first word on the screen. Failure to respond within that time limit was counted as an inaccurate response.

In total, subjects saw forty-eight verb-noun pairs, twenty-four of which appeared in the reading task, while the other twenty-four pairs did not. In order to prevent dependence on recognition of the verb, twelve of the twenty-four distractors presented a verb which did appear in the preceding task. These distractors were followed by a noun that did not occur at any point during the encoding phase. These verb-noun pairs consequently required a negative response.

5.4.2 Participants

In total, thirty-four Stanford University undergraduates (who had not completed any other experiment discussed here) participated in this two-part study for course credit; however, only the results of the nineteen subjects who recognized at least 67% of the verb-noun pairs were included in the final analysis. Recognition accuracy on this task was not surprisingly quite low, given that subjects were unaware that their memory would later be tested, the relatively short study-time allowed for each sentence, and the number of items appearing in the study.

\[\text{\footnotesize 12} \text{In this way, subjects were not expected to be familiar with the concepts of verb and noun.}\]
\[\text{\footnotesize 13} \text{16.667 ms is the amount of time it takes for the screen to refresh at 60 Hz—otherwise known as frame duration.}\]
\[\text{\footnotesize 14} \text{To maintain decent chances of finding statistically significant results, it was necessary to use this low number of items (48). Previous pilot testing confirmed that a larger number of items results in near-to-chance performance at the task.}\]
5.4.3 Results

Two dependent measures are available from the recognition task: response time and recognition accuracy. With respect to the first of these, the amount of information encoded about the thematic patient failed to influence response times at all. All four experimental conditions had nearly identical response times for correctly recognized stimuli (adj: 1382.90 ms, SE = 66.12; long: 1418.33, SE = 64.67; post: 1473.04, SE = 70.57; short: 1374.16, SE = 78.93). Even considering incorrect responses, response times remain consistent across conditions. Thus, informativity had little effect on the response speed for this task.

Contrasting with this, recognition accuracy does demonstrate some effects of the encoding process. Specifically, the LONG condition, which contains both an adjectival modifier and a prepositional phrase, yielded the highest overall recognition accuracy, as pictured in Figure 5.11. Statistical analysis shows that the recognition accuracy for this condition was significantly better by subjects than the accuracies for the conditions with only one modifier, but only marginally so by items (LONG-ADJ: $t_1(18) = 2.348$, $p < .05$, $t_2(23) = -1.819$, $p = .082$; LONG-POST: $t_1(18) = 2.233$, $p < .05$, $t_2(23) = 1.887$, $p = .072$). Similarly, the LONG condition facilitated marginally better recognition than the SHORT condition ($t_1(18)$

![Figure 5.11: Mean recognition accuracy of verb-noun pair in experiment IX](image)
= 1.701, \( p = .106 \), \( t_2(23) = 1.224, \ p = .233 \). All other comparisons of recognition accuracy prove to be non-significant.

5.4.4 Discussion

While the results from this word recognition experiment are not definitive, they do offer some tantalizing suggestions. Unlike the other experimental studies detailed within this dissertation that examine memory retrieval during on-line language comprehension, this experiment looks at recall after comprehension processes have been completed. In other words, the recall task targets information stored in long-term memory, as opposed to working memory, since the interval between encoding and recognition ranged anywhere from two to fifteen minutes. The recognition data nevertheless pattern after the findings of the reading time studies: the most informative encoding conditions lead to better performance. In the recognition study, of course, this facilitated performance was manifested in terms of recognition accuracy, rather than retrieval speed. However, in a series of articles (McElree, 1993; McElree & Nordlie, 1999; McElree & Griffith, 1995, 1998; McElree, 2000; McElree et al., 2003), McElree and colleagues argue that the likelihood of successful recall affects response rates in behavioral tasks like self-paced reading, so slower reading times at retrieval points are not necessarily indicative of slower retrievals so much as unsuccessful retrievals. A wealth of evidence from speed-accuracy tradeoff experiments supports their claim that while the accuracy of recall decreases over time, the speed of retrieval does not. In this sense, the results from the word recognition task and the self-paced reading studies are quite compatible, since they both potentially depend upon the accuracy of recall.

The absence of a statistically significant difference between the most complex condition and the simplest condition may relate to the proximity or contiguity of the verb and head noun in the SIMPLE condition. Since participants were tested on their recognition of verb-noun pairs, recognition accuracy may be partly affected by the distance between the noun and verb during the encoding process. In the SIMPLE condition, only the determiner the separates the verb from the target noun, but in the LONG and ADJ conditions, an extra word separates the two lexical items. The POST condition was included among the stimuli in order to control for the contiguity effects. However, as the lexical heads in this condition were not recognized with a higher accuracy, it is difficult to judge whether the recognition percentage for the SIMPLE condition reflects a contiguity effect or not.\(^{15}\)

\(^{15}\)The extra prepositional phrase in the POST condition may simply not add enough information to raise
The results further imply that a single modifier, whether adjectival or prepositional, is insufficient to effect a significant advantage for recognition. Items in POST and ADJ conditions were not recognized with any better accuracy than the SHORT condition. In contrast, the LONG form with two modifiers apparently pushes past some threshold of low informativity, such that the recognition process is facilitated. These accuracy results thus parallel the reading time data in Experiment V where one adjective failed to significantly facilitate retrieval compared to the unmodified indefinite NP, but two adjectives managed to produce a statistically reliable difference in processing.

To the extent that these results represent a real recognition contrast, they suggest that the relationship between informativity and retrievability is, in fact, determined by a general cognitive principle that links encoding and retrieval. That is, the informativity effects established for working memory processes appear to have some impact (albeit weaker) on long-term memory, as well. In this sense, there is perhaps nothing unique to working memory about this relationship. Indeed, it is even possible that this informativity-retrievability connection is independent of language (although the current results do not provide any support for this claim).

Moreover, these suggestive results add further weight to the claim that the reading time differences which occur at subcategorizing verbs stem from differences in retrievability. The data in this experiment hints that informativity has an effect on recognition memory in a way that is consistent with how informativity affects processing at retrieval sites in comprehension, i.e. more informative linguistic forms elicit better subsequent performance. Given this parallelism, it would be unintuitive and unparsimonious to claim that the informativity effects stem from memory retrieval differences in one task, but that they derive from other processing considerations in another task.

Of course, further experimental investigation is necessary to solidify these conclusions. It is first of all necessary to consider how recall accuracy varies in the absence of contiguity effects. Secondly, more research and discussion is needed to determine the extent to which the results of long-term memory tasks can inform us about the nature of working memory processes. This experiment should consequently be interpreted as a launching point for the informativity to a level where it benefits long-term memory. If this is the case, then it would be incapable of establishing the existence of a contiguity effect.

This conclusion is particularly sensible if some of the same architectural properties of of working memory also constrain long-term memory, such as decay and interference. See Anderson (1974) for some evidence of interference on longer-term recall of propositional information.
further inquiries into the parallels between working memory and longer-term memory biases.

5.5 Summary

While the previous chapters have largely been concerned with amassing evidence for the central claim of this dissertation—that informativity affects retrieval—this chapter was spent addressing the reasons for the relationship between the informational content of a linguistic expression and the cognitive effort required to retrieve the corresponding mental representation. A greater quantity of information in a linguistic expression is often accompanied by a greater amount of morphophonemic material, a greater amount of syntactic complexity, a greater expenditure of cognitive resources to process the additional information, and an increasing probability of atypical or uncommon feature combinations. Any one of these facets of informativity could theoretically be responsible for the connection between informativity and retrieval costs.

The accumulated evidence from this chapter advances the idea that some of these factors truly matter for purposes of retrieval, while others do not. Experiments VI and VII provided evidence that both syntactic and semantic complexity contribute to the memory facilitation effects. Holding complexity constant in experiment VI, significant differences were observed between \textit{wh}-phrases that differed in semantic complexity.\footnote{It is perhaps worth saying a word or two about why the results from experiments VI and VII are not identical, insofar as experiment VI revealed a significant effect of semantic complexity, but this effect was only marginal in experiment VII. First of all, the stimuli for experiment VII contained two extra items and, of course, contained an extra condition and different filler items. Secondly, the results from the two experiments are ultimately quite compatible, since in both cases, the semantically complex condition leads to faster reading: the difference between the experiments is merely one of magnitude or effect size.} And in experiment VII, it was found that varying syntactic complexity while minimizing semantic differences between phrases was sufficient to create informativity effects. In this sense, the \textit{kind} of information that is relevant for retrievability is not specific or pre-determined. While the MFH mentions only syntactic and semantic information, other types of information, such as prosodic (in the case of listening) or visual features (in the case of reading), potentially also influence the retrieval process.\footnote{In this sense, the wording of the MFH is perhaps conservative, in order to reflect the current state of knowledge about the types of information that are relevant for retrieval in language comprehension.}

Experiment VIII, in contrast, does not find any advantage for information that is unexpected or incongruous. The theoretical upside to processing such information is that
it focuses attention on the representation at hand because of its unnaturalness. The results did, in fact, confirm that such references yielded markedly slower reading times, as expected. It was not the case, however, that this led to a substantially improved retrieval process as compared to phrases that exhibited the same amount of informational complexity, but which contained stereotypical or predictable information. Actually, the data hints that the representations with commonly associated features are retrieved slightly faster. Both at the intermediary (but false) retrieval site and the correct retrieval site (plus spillover), definite NPs with stereotypical adjective-noun combinations led to slightly faster reading times.

In short, even though the distinctive condition in experiment VIII generated the most initial referential processing, this early expenditure of cognitive resources failed to produce the fastest processing at retrieval sites. Note that this finding does not rule out a relationship between encoding effort and retrieval effort, i.e. the “pay now-pay later” principle. In most of the experiments considered in the present work, there has been a repeated and consistent interaction of position (encoding vs. retrieval) and condition. In fact, no experiment has shown a case where the fastest encoding times are matched with the fastest retrieval times. Even in experiment VIII, the bare definites that were processed the fastest still resulted in the slowest retrieval processes. So, there does seem to be a quite ubiquitous and general relationship between the cost of encoding and the cost of retrieval. But, as with many general processing principles, there can be other competing or interacting cognitive factors that obscure or hide the influence of some otherwise quite regularly observed cognitive process. Given a correlation between the quality of two cognitive processes, nothing prevents some additional property of working memory from influencing the outcome of one or both of these processes.

The anomaly or unexpected result here (from the perspective of the PNPL principle) is the absence of a stronger retrieval facilitation for the most distinctive encoding. There are at least two plausible reasons for why the informativity effects were not stronger. The simplest explanation is that retrieval could not be facilitated any more—that is, optimal

---

19As one example, the anti-locality effects discussed in the next chapter counter-exemplify the distance or locality-based principle of retrieval and integration. Nevertheless, no one—not even the researchers responsible for identifying the anti-locality effects—interprets the anti-locality effects as a refutation or denial of a general distance-based processing principle.
efficiency was reached. However, the faster reading times for the non-distinctive, but syntactically and semantically complex, condition argues against such an interpretation. Another reasonable explanation discussed in this chapter appeals to properties of memory retrieval and associative strength. Specifically, retrieving a whole mental representation from working memory may reasonably involve first accessing part of that representation. From this memory chunk, the rest of the mental representation that it is linked to can be accessed. Bradshaw and Anderson (1982) talk about this kind of retrieval process wherein the retrieval of some memory chunk leads to the retrieval of other connected chunks in terms of the availability of retrieval paths. In their view, the retrieval of one chunk in memory can provide a conduit to the retrieval of a related chunk either through redundancy or inferential processes. The improved recall accuracies they observed for sentences that involved causally related continuations is thus captured by the postulation of information networks in memory that permit recall transitions from one node to the next.

With respect to the retrieval of filler-phrases, the ability to transition from one node or chunk to the next was described in terms of associative strength, rather than pure inferential processes. Inference implies a certain amount of guessing, but it is actually rather unlikely that one could guess the right noun on the basis of an adjective (or vice versa). Speaking in probabilistic terms, an adjective like ruthless or wealthy does not present a good predictor of an upcoming head noun, despite the intuition that dictators are commonly ruthless and celebrities are stereotypically wealthy. For instance, the conditional probability that dictator shows up after ruthless is just .014, based on the 1.2 billion word English Gigaword corpus. Thus, any attempt at guessing would probably run afoul. Accordingly, I have suggested that the way to describe the capacity to access one representational element via another element is in terms of associative strength, which can be independently measured by elicitation tasks or by the amount of priming that happens in a word recognition task (Postman & Keppel, 1970; Neely, 1990, inter alia).

If two chunks of information that were encoded as part of the same larger representation have a high degree of association, then successful retrieval of one chunk should facilitate retrieval of the other chunk (where this is actually necessary). In contrast, two chunks that exhibit a low degree of associative strength will demonstrate a correspondingly diminished faculty for facilitating each other’s retrieval. Nevertheless, any complex network of linked information nodes appears to be better for purposes of retrieval in language comprehension than just a single node or a less complex subset network.
Finally, in experiment IX, I sought to determine whether evidence for the MFH could be found in tests of longer-term memory. This experiment was not intended to address the functional motivation for the MFH, but to test the generality of the informativity effects. While the ensuing evidence was only suggestive due to the minimal effect size, the results patterned in a way that conformed to the predictions of the MFH. The most informative descriptions elicited the highest rate of question-answer accuracy, while question-answer accuracies were essentially equal in all the other conditions. Thus, there is some reason to suspect that the relationship between informativity and retrieval is a rather general one that has ramifications beyond the sphere of what is traditionally accepted as working memory.

This chapter has thus laid out some possible explanations for the informativity effects detailed in the previous chapters and explored some alternative hypotheses. On the basis of all the evidence collected so far, two primary explanations have emerged for the relationship between informativity and retrievability: (1) the greater expenditure of cognitive resources in the processing of more informative expressions leads to a facilitated retrieval process and (2) more complex representations present more retrieval paths, which offsets the probability of failing to retrieve any single chunk from memory. The possibilities that I have considered here for the underlying functional relationship between informativity and retrieval, though, are hardly exhaustive. This discussion, in fact, carries over to the next chapter where the findings of other recent experimental work in linguistics are compared to this dissertation research. We will see that the explanations and theories developed to account for several different types of sentence processing effects are analogous and compatible with the present conclusions. Looking at these other analyses, in fact, will cast some further light on the nature of informativity effects and help further define the relationship between informativity and retrieval.
Chapter 6

Memory and Reference in Discourse

There are two purposes for this chapter. The first part of this chapter addresses how the Memory Facilitation Hypothesis complements other general theories of discourse processing and representation. Theories of referential form choice typically are restricted to events and properties of discourse that lead up to that choice, but are less concerned (if at all) with the consequences of a linguistic form choice. The MFH contributes to the understanding of how the activation and retrievability of discourse entities varies throughout the course of a conversation or a text by specifying the retrieval-related byproducts of these choices. Combining the MFH with previous theories of discourse and reference, therefore, we can develop a more comprehensive view of how activation fluctuates throughout discourse and how that affects comprehension processes.

The second part of this chapter considers how the experimental results documented here pair with other findings in the sentence processing literature. In particular, I discuss how other sentence processing findings on anti-locality and similarity-based interference are in line with the predictions of the MFH. Indeed, the same cognitive limitations that define these two sets of effects are argued to be responsible for the character of informativity effects. More precisely, I argue that increased informativity facilitates future retrieval because the process of building a complex representation boosts activation—similar to anti-locality effects—and produces a mental representation that reduces the likelihood of interference. Lastly, to explicitly illustrate how these intuitions can be precisely modeled, I provide an
overview of a cue-based cognitive model of retrieval that presents all the formal tools necessary to capture these hypothesized relationships.

6.1 Reflecting and Attenuating Retrievability

Memory has been at center stage for numerous theories of reference (Chafe, 1976; Gernsbacher, 1990; Ariel, 1990, 2001; Almor, 1999, 2004; Almor & Nair, 2007). The accessibility theory of Ariel (1990), for instance, states that speakers indicate the difficulty of a given retrieval from memory with the choice of linguistic form:

“... retrievability is crucially dependent on degree of activation, or Accessibility. It is the specific degree of Accessibility of mental entities attributed by the speaker to the addressee which is the crucial criterion determining the forms of retrieval marking. An addressee is instructed to retrieve a mental representation which may be characterized by reference to the individual features (‘wise’, ‘short’), but always also with a feature establishing its current Accessibility to him. In other words, a speaker signals to her addressee how easy/automatic the retrieval is. The various types of referring expression, then, each represent different sets of instructions for the search process” (Ariel, 1990, p. 16).

Longer and more informative forms (e.g. low accessibility markers like definite descriptions) accordingly express the speaker’s belief that the appropriate mental representation will be hard for the addressee to retrieve. This echoes collaborative theories of referring that claim that speakers and addressees work together to agree upon the conceptualization of a discourse entity (Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989; Brennan & Clark, 1996), since speakers respond to the processing demands of their addressees.\(^1\) Clearly, though, a linguistic form choice does more than simply imply what the speaker thinks about the addressee’s discourse representation. It also compensates for retrieval difficulty by providing the necessary information to retrieve or re-access the correct representation, as determined by the activation level of the representation (see Figure 6.1). According to

\(^1\)Nevertheless, these theories lay out different tasks and responsibilities for speakers and addressees. The formulation of accessibility theory above suggests that speakers form a mental model of their addressee’s cognitive state. Such a move, however, is not necessary on the collaborative approach where a conceptualization for a given entity is mutually agreed upon.
accessibility theory, the speaker’s belief is therefore embodied in the amount of information provided for retrieval. In this sense, linguistic form choices reflect and aid retrievability.

The MFH and the evidence presented in the preceding chapters combine with such referring theories to suggest that linguistic form not only reflects the retrievability of some mental representation for a discourse entity at the time of utterance, but it also attenuates the future retrievability of that representation. The empirical evidence in this dissertation indicates that longer and more informative forms make re-accessing the corresponding representations easier in a variety of linguistic contexts. Accordingly, at subsequent references, shorter and less informative forms will be necessary to achieve successful communication. In turn, these shorter and less informative forms contribute relatively little to the next retrieval process, such that the repeated use of reduced forms is likely to eventually cause retrieval problems.

Put together with principles of working memory, these theories—the MFH and theories of reference choice—thus present a picture of fluctuating activation levels modulated by linguistic form (in addition to other factors). The claim that reference to a particular discourse entity raises the corresponding activation level has been made before (Gernsbacher, 1989; Anderson et al., 2001; Lewis et al., 2006; Vasishth & Lewis, 2006, inter alia). What I am advocating here, however, is the notion that the magnitude of these “boosts” depends upon the informativity of the linguistic expression, akin to the explicitness principle proposed by Gernsbacher (1989). On this view, an initial reference, typically long and specific (e.g. the 1969 winner of the Tour de France, Eddy Merckx), introduces and highly activates a mental representation for a discourse entity. This high activation level licenses the subsequent use of a shorter and less informative form for this discourse referent (e.g. he, Merckx, the Tour winner, etc.), until decay and interference effects eventually weaken the activation level. Repeated references, even with these high accessibility markers, temporarily increases the activation; however, these boosts are assumed to be substantially weaker than those cumulatively created by long and informationally rich descriptions. Over time, then, the activation level associated with a discourse entity can fall to such a low level that an uninformative referring expression would induce retrieval difficulty. A long and informative reference, in contrast, raises the probability of a successful reference. At the same time, such an explicit reference would (re)-strengthen the activation signal, such that informationally minimal forms would again be licensed. This cycle then repeats throughout the course of a conversation or a text until this discourse entity loses its topical status or
Figure 6.1: Relationship between anaphoric form choices and three hypothetical activation profiles
CHAPTER 6. MEMORY AND REFERENCE IN DISCOURSE

Many texts, in fact, evidence this cycle of informativity where complex forms lead to simpler and simpler forms that are, in turn, followed by one or two complex forms and a return to simple forms, and so on. For instance, a Washington Post article delving into the political life of Dick Cheney refers to him in the first few paragraphs with the following referring forms in the order shown:\(^2\) Vice President Cheney, Cheney, he, him, the vice president, Cheney, him, he, he, he, he, Cheney. This is just one simple example for illustrative purposes, and further corpus-based evidence is necessary to judge the accuracy of the claims

\(^2\)http://blog.washingtonpost.com/cheney/chapters/chapter1/
made here about the distribution of referring forms throughout discourse. The point nevertheless remains that speakers do not simply settle on the most economical form and then never shift from the use of this uninformative expression. Particularly after repeated uses of pronominal forms over a relatively long period, the use of semantically and syntactically complex forms is justified to raise the chances of successful and efficient interpretation. In the terminology of Brennan and Clark (1996), the re-use of such complex forms can be thought of as a means for reconfirming a conceptual pact, i.e. an agreement made between interlocutors about how to conceptualize a discourse referent. That is, interlocutors may choose to revisit how they frame or conceptualize a particular object in the discourse, especially after a significant amount of time has passed. This process may even involve adding new information to that conceptualization or modifying previously mentioned information. This pragmatic explanation, however, does not elucidate what happens cognitively in order to solidify this pact. The present account does just that, by showing that increased informativity makes the corresponding mental representation easier to access.

While I have focused primarily on the effects of linguistic form on subsequent comprehension, speakers thus often seem to demonstrate a cooperative spirit in dialogue, in order to maximize the chances of successful communication. Speakers appear to know just when to embellish a description with additional information and when to cut back on information. Based on previous research that suggests that speakers at least occasionally choose linguistic variants that also facilitate comprehension (e.g. Jaeger, 2006), production patterns may reveal some trace of the cognitive principles behind the MFH. Long and informationally rich descriptions are, as discussed at the outset of this dissertation, reserved for particular discourse functions. The resource costs associated with informationally rich descriptions make them unjustifiable in most discourse contexts.

Given the predictions of the MFH and the accumulated reading time evidence presented here, longer filler-gap dependencies are more likely to be processed quickly and accurately with more informative filler-phrases. Conversely, long and challenging FGDs beginning with syntactically and semantically simple filler-phrases have been seen to induce significant comprehension difficulties. If speakers are sensitive to this probability at some level, then dependencies of greater length ought to be licensed after more complex fillers. To acquire some evidence for this position, a small scale corpus study was conducted to see if the informativity of the filler-phrase had any significant impact on the corresponding dependency length. The question is, in other words, does increasing the informativity of
the filler-phrase make a longer FGD more likely?

To answer this question, all instances of direct and indirect questions with bare wh-words and complex “multi-word” wh-phrases (e.g. [What]/[What kind of dress] are you looking for?) were extracted from the parsed portion of the Switchboard corpus of spoken American English.³ This corpus contains some 800,000 words from 650 telephone conversations (Godfrey et al., 1992). For each class of wh-phrase, the mean number of words from the right-edge of the wh-phrase to its gap site was calculated. For instance, in the bare wh-question in (77a), the distance between the right edge and the gap site is three words. Similarly, the distance between the complex wh-phrase in (77b) and its gap is three words (here the gap is represented with the notation ‘*T*’, as it appears in the tagged version of the corpus):

(77) a. [What] did you wear *T*-1 to work today?

b. [What kind of food] do you like *T*-1?

Paired sample t-tests indicate that dependencies after complex wh-phrases (mean length = 3.11 words, \( n = 199 \)) are significantly longer than dependencies after bare wh-words (mean length = 2.62 words, \( n = 2944 \)) (\( p (T < t) = .0002 \), assuming equal variances). If greater informativity promotes higher activation (= faster memory retrieval), then longer wh-phrases, with more encoded features, can be farther from their gap sites. This doesn’t imply that speakers somehow plan the length of the dependencies they utter on the basis of the informativity of the filler-phrases that they started the dependency with. Instead, speakers may simply use longer dependencies after more informative wh-phrases because the use of the longer phrase creates a larger window for successful communication. If bare wh-dependencies make comprehension more difficult at the retrieval site, there is good reason from the speaker’s perspective to constrain the length of the dependency. To be clear, however, this discussion and small-scale corpus investigation is only intended to be suggestive and to keep open the possibility that production data may reveal further indications of the MFH.

Regardless of the ultimate role of speakers in catering to addressee needs, knowledge of how linguistic form choices bear on subsequent comprehension processes provides a more comprehensive view of discourse, especially with respect to the cognitive factors that drive

³Due to the limited number of multi-word wh-phrase tokens in the corpus, more fine-grained evaluations of the effect of each extra word in a wh-phrase was impossible, due to statistical limitations.
linguistic choices throughout discourse. The present hypothesis makes important predictions about how activation, which prior theories of reference have identified as critical to the choice of referring form, varies as a function of informativity. In this sense, this research on informativity feeds back directly into theories of reference and discourse and can potentially help make these become more precise and penetrating theories of linguistic variation.

6.2 Anti-Locality Effects

Contrary to the predictions of locality-based notions of sentence processing (Gibson, 1998, 2000; Grodner & Gibson, 2005), Vasishth and Lewis (2006) report on anti-locality effects found in a variety of linguistic constructions in Hindi, including center embedding structures, subject relatives, and object relatives (see also Konieczny, 2000). For example, the dependency length in object relatives between the verb and noun (e.g. kaagaz below) can be increased by adding a phrase that modifies the noun, as shown in the following example:

(78) a. Vo kaagaz jisko us lar.ke-ne dekhaa bahut puraanaa thaa.
    that paper which that boy-ERG saw very old
    ‘That paper which that boy saw was very old.’

b. Vo kaagaz jisko us lar.ke-ne mez-ke piiche gir.e.hue dekhaa bahut
    that paper which that boy-ERG table-GEN behind fallen saw very
    puraanaa thaa.
    old
    ‘That paper which that boy saw fallen behind a/the table was very old.’

According to a strict locality-based view of retrieval and integration difficulty, interposing the extra material between the two linguistic dependents should increase the processing difficulty; however, Vasishth & Lewis find that the interposed material actually facilitates processing at the subcategorizing verb. That is, reading times are shorter at the verb dekhaa in (78a) than in (78b).

Vasishth & Lewis consider a number of possible explanations for these anti-locality effects. Already alluded to in the last chapter, one possibility is that the interposed material provides time to anticipate or pre-activate the upcoming verb. Grodner and Gibson (2005) provide another candidate explanation: information intervening between two dependents can help to contextualize a verb, making it easier to understand the relationship between the verb and its dependents (e.g. The fisherman cried—The fisherman who was cutting
onions cried). In the end, Vasishth and Lewis conclude that these effects can best be captured by appealing to the cognitive effect of repeatedly retrieving some representation from memory. They argue that processing a word or phrase whose interpretation depends upon or modifies some representation already in memory requires the re-activation of that memory representation. This reactivation process supplies a boost in activation:

“If some event causes the retrieval of an NP or VP node before its retrieval at a verb . . . the activation of that node will increase and its subsequent retrieval at the verb will be faster. One such situation occurs when a relative clause intervenes between the verb and the NP—the NP must be retrieved to attach the modifying RC” (Vasishth & Lewis, 2006, p. 784).

According to this logic, the verb in (78b) is read faster than the one in (78a) because the interposed phrase mezke piiche gîrehue triggers an additional retrieval of kaagaz before it has to be retrieved again at the verb. Any successful retrieval process, therefore, increases the activation of the memory target. Hence, the more some particular representation is retrieved, the higher its activation level will be, thereby making future memory retrievals easier. In essence, retrieval makes retrieval easier.

The underlying relationship between informativity and memory retrieval can consequently be partly justified by a similar rationale. For instance, the results of the clefted indefinite study (experiment V) state that the presence of multiple adjectival modifiers lead to faster processing at the retrieval site. On the view espoused by Vasishth and Lewis, each of these modifiers pre-activates the head noun (these pre-activations of predicted nodes count as a type of retrieval), and the head noun itself naturally boosts the activation level. The retrieval differences in many of the experiments documented here can thus be reasonably interpreted as a derivative of the syntactic complexity of the NPs and the process of repeatedly retrieving some syntactic head. This pairs well with the results from Experiment VII that reveal significant differences in retrieval for phrases that generate representations of differing syntactic complexity but minimally different semantic complexity (who vs. which person). That is, even after minimizing the semantic distance between the two NPs, significant retrieval differences still emerged, suggesting that syntactic complexity has a major role to play in the informativity effects.

A (re)-activation or boosting account encounters some trouble, however, when it comes to the observed differences between phrases containing the same syntactic information that
nonetheless differ semantically. Experiments VI and VII verify that, holding syntactic complexity constant, informativity effects are still observable, although reduced in magnitude. Additionally, the results of Cowles and Garnham (1995) show that antecedents which differ in their specificity, but have the same syntactic form, have noticeably different effects on the processing of anaphors. In accord with the predictions of the MFH, Cowles and Garnham (1995) found that an antecedent like *the hatchback* improves the processing of an anaphor like *the car*, as compared to a less specific antecedent like *the vehicle*. These results, too, cannot be explained by a re-activation process alone. Which is to say that something else must be responsible for the semantic complexity effects. This raises the intriguing possibility that semantic complexity and syntactic complexity improve retrieval for a slightly different set of reasons. The next two sections, in fact, provide additional grounds for thinking this might be so.

In sum, re-activation processes offer a powerful means for understanding the relationship between informativity and retrievability, insofar as informational complexity is a function of syntactic complexity. On the other hand, the rationale provided by Vasishth & Lewis for anti-locality effects has no means for explaining semantically-based effects. Ultimately, the anti-locality results and informativity results seem to naturally complement one another. They both point to the conclusion that retrieval costs can be offset by modifying the to-be-retrieved representation. Furthermore, the present approach and the cognitive model employed by Vasishth & Lewis link retrievability to the level of activation and assume that the natural process of activation decay is counteracted by modification. Thus, the effect of representational complexity on retrieval appears to be intrinsically tied up with the process of building the complex representation, but, as discussed in the previous chapter and as we shall see again, it is also the resulting complexity itself that contributes to the retrieval facilitation.

### 6.3 Similarity-Based Interference

According to theories that claim that language processing is constrained by similarity-based interference (Gordon et al., 2001; Gordon, Hendrick, & Levine, 2002; Gordon et al., 2004, 2006; Warren & Gibson, 2005; Vasishth, Suckow, Lewis, & Kern, in prep.; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006), the retrievability of a mental representation depends upon the presence of other representations in working memory that overlap the
feature specifications of the retrieval target. In particular, demands on retrieval are thought to increase when a target shares features with similar representations. This conclusion is based on repeated findings which show that when two referential NPs are matched in type (e.g. the banker-the barber), reading times and comprehension accuracies suffer.\footnote{Ongoing research is attempting to identify what sorts of features are relevant for similarity-based evidence. Gordon et al. (2004) report that overlapping number and definiteness features do not create interference. In the opinion of Gordon and his collaborators, “the critical dimension of similarity is referential, with the predication of the common noun achieving reference in a manner that is psychologically distinct and more complex than the more direct reference given by names and pronouns” (Gordon et al., 2006, p. 1305). Other research, however, identifies effects of similarity driven by overlapping syntactic cues and semantic components (Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006), so it still remains an open question as to what kinds of overlapping features can induce similarity-based interference.}

Gordon et al. (2006), for instance, discovered a significant interaction between relative clause type (object vs. subject relative clauses) and whether the embedded NP matched the type of the relative clause head. While subject relatives are generally processed more easily than object relatives, Gordon and colleagues found that this difference increases when the embedded noun phrase matches the type of the relative clause head. In fact, the classic subject vs. object relative clause difference essentially disappears when the second noun phrase is a pronoun or a proper name (which forces the embedded NP and the relative clause noun to be different types, since names and pronouns cannot be relativized). Moreover, within relative clause types, two matching, unintegrated NPs drastically slowed processing both in the RC itself, at the matrix verb, as well as in the subsequent spillover region:

(79) a. The banker that praised the barber/Sophie climbed the mountain just outside of town.

b. The banker that the barber/Sophie praised climbed the mountain just outside of town.

Additional evidence verifies that these similarity effects do not simply stem from the difficulty of processing definite descriptions as compared to names. In cleft constructions, like the one in (80), the matched conditions led to significantly slower reading times than the unmatched conditions:

(80) It was the barber/John that the lawyer/Bill saw in the parking lot.

Similarity-based interference has also been implicated in the processing complexity of the standard center-embedding examples where multiple definite descriptions occur in subject
position (Lewis, 1996).

The emerging evidence, therefore, argues that interference plays an important role in sentence comprehension. Most researchers agree that both proactive interference (when interfering material occurs before encoding the future target) and retroactive interference (when interfering material occurs after encoding the future target) hinder the retrieval process, although it may influence the encoding phrase, too, as suggested by Gordon et al. (2006). Memory representations that overlap in feature specifications with other representations in memory will be harder to retrieve.\(^5\)

This conceptualization of retrieval during comprehension being subject to interference effects suggests another reason why greater syntactic and semantic complexity would facilitate subsequent retrieval. A greater number of features in a mental representation will likely (although not necessarily) raise the probability that some of those features are not shared by other stored memory items. Thus, to the extent that these unique features can be targeted, they should increase the probability of successfully retrieving a particular memory representation, due to reduced interference. Speaking metaphorically, the larger and more conspicuous a target is among a field of false targets, the easier it will be to hit.

Conversely, a low number of very general and ubiquitous features increases the probability of feature overlap with other memory items. A referring form that only expresses animacy information like \textit{something} should be subject to a high degree of interference in any discourse context in which other inanimate entities are discussed. Thus, less informative linguistic forms introduce a greater risk of interference.

As Van Dyke (2007) suggests, this interference may occur along both a syntactic and semantic dimension. Thus, the facilitating effects of syntactic and semantic complexity uncovered in Experiments V and VI can theoretically both be linked to interference. That is, not only can additional semantic features cause a representation to be more unique, but additional syntactic features can also distinguish a representation. Hence, while the activation-boosting account of Vasishth and Lewis is only compatible with the syntactic complexity effects, an interference-based view of the present data could theoretically predict the impact of both syntactic and semantic informativity. That being said, it is worth noting

\(^5\)On the basis of eye-tracking evidence, Gordon et al. (2006) claim that similarity-based interference effects are restricted to unintegrated linguistic constituents that must be held together in working memory, as in complex center embeddings or object relative clauses. This view, however, appears to be too strong, given evidence of both syntactic and semantic interference of already integrated constituents in Van Dyke and Lewis (2003) and Van Dyke (2007).
that no current account of interference effects (to the best of my knowledge) addresses the role of syntactic or semantic complexity. In other words, while prior research has established that certain kinds of syntactic and semantic representations can interfere with one another, it has not been previously shown whether complexity strengthens or dampens these interference effects. To the extent, therefore, that the current results embody the outcome of interference effects, they present a novel means of modulating these effects.

Similarity-based interference thus plausibly constitutes part of the underlying motivation for the retrieval benefits associated with more complex representations. As with the anti-locality effects described in the previous section, the interference effects demonstrate a high degree of compatibility with the informativity effects found here. Conceivably, therefore, informativity effects may be largely reducible to a combination of anti-locality and similarity-based interference effects. On this view, informativity partly determines the number of times a syntactic head category is re-activated, but it also adjusts the distinctiveness of the representation in memory.

The main point of this section and the last has been to show that informativity effects are quite compatible with the results and experimental findings of other recent research in sentence processing. I assume, in fact, that the re-activation process that accounts for anti-locality effects in Hindi and German also causes syntactic complexity to result in faster retrieval. Reduced interference, too, has also been identified as a viable contributor to the informativity effects, although this explanation is limited by definition to contexts where there is a sufficient number of memory competitors to create noticeable interference. Overall, the review of these two sets of psycholinguistic effects further substantiates the present hypothesis and conclusions by demonstrating that the current results are very much related to other recent experimental findings and may, in fact, have partly overlapping explanations with these other psycholinguistic phenomena.

6.4 Cognitive Models of Cue-Based Retrieval

The precise series of steps that are involved in retrieval from memory during language comprehension have not been explicitly laid out in this text. This section, therefore, aims to specify the stages of the retrieval process and to clarify some features about this cognitive operation that I have assumed throughout this text, based upon previous research and cognitive modeling. In particular, the bulk of this section reviews the basic elements of
the Adaptive Control of Thought-Rational (ACT-R) model of central cognition, which is defined by the interaction of a number of independent cognitive modules (Anderson & Lebiere, 1998; Anderson et al., 2001, 2004; Anderson, 2005). ACT-R theory is furthermore not specifically tailored to the shape of language processes; instead, sentence parsing is seen as a specialized task whose particular characteristics stem from general cognitive principles. It has been applied to the modeling of numerous, well-known sentence processing effects, as detailed extensively in Lewis and Vasishth (2005), from which this section draws heavily. The overview of this theory supplied here, however, is not meant to be comprehensive. The main purpose is to ground the claims made about memory retrieval during language comprehension in a highly specified and computationally complete model of cognition and to indicate how the model would need to be updated to reflect the current set of discoveries.

The present focus on the ACT-R model of retrieval, however, does not exclude the applicability of other models of cognition and sentence processing. Other cognitive models may be able to capture the same array of facts equally well, if not better. Nevertheless, there a number of principled reasons for exclusively considering the relevant informativity effects in terms of the ACT-R model. Perhaps the most important reason among these is that the ACT-R theory addresses the severe time constraints that any account of sentence processing must grapple with. As Anderson et al. (2001, p. 388) put it, “there is just not enough time, at normal reading or listening rates, to do more than a minimal number of inferences.” This fundamental fact about language processing is met in ACT-R by a commitment to a very limited set of processes and a deep concern about the time it takes for these processes to unfold. That is, the theory makes fine-grained predictions about the time course of cognitive processes in a wide range of tasks, among which language processing is just one applicable domain.

The main architectural features of the model are a number of independent, but interacting modules. Most pertinent for the current discussion are the declarative memory and procedural memory systems. Declarative memory contains so-called information chunks that are comprised of a number of feature-value pairs, as illustrated in Figure 6.3, adapted from similar representations in Lewis and Vasishth (2005). Values consist of basic symbols or pointers to another stored item in memory. Hence, the value dp3 of the SPEC feature in

---

6In this sense, these chunks look much like the attribute-value matrices employed in lexicalist syntactic theories like LFG and HPSG. There are, however, strict limitations on the size of chunks. As expressed by Lewis and Vasishth (2005), “[t]he feature contents of two items and the novel relation between them cannot
the example chunk simply points to another chunk in memory, but it does not provide any other information about this chunk (the DP label is purely for convenience). The procedural module, in contrast, is the home of production rules that determine how the content of cognitive buffers is evaluated and acted upon. Effectively, all action and behavior is controlled or mediated by this particular module. Lewis and Vasishth (2005) assume that “much grammatical knowledge is encoded procedurally in a large set of quite specific production rules that embody the skill of parsing.” This still leaves room for some syntactic knowledge to be stored in declarative memory, but it is assumed that skilled behavior largely reflects procedural processing. A third module, referred to alternatively as the control or goal module, tracks the current goal that is being attended to by the system. For instance, the current goal could be to construct an IP. Once satisfied, this goal is removed from the current goal buffer, but all action by the system is guided by the contents of this buffer until it is emptied.

Within these modules, multiple processes are carried out in parallel: multiple searches of declarative memory can happen simultaneously, multiple searches of procedural memory for the appropriate or best production rule can operate in parallel, etc. Each of these modules, however, has an extremely limited buffer that can only contain the result of the most recent process, i.e. a single chunk of information. The buffer associated with the declarative module therefore has only a single chunk in it at a time. Processing bottlenecks or limitations arise, therefore, in the transmission of information from one module to another. And since only a single production rule can fire at a time, the procedural module presents the main processing bottleneck for cognition. Consequently, while module-internal processes occur in parallel,
there is a serial communication between modules.

During sentence processing, declarative memory is repeatedly accessed to retrieve lexical information, as well as recently encoded discourse references and (potentially incomplete) syntactic representations. Hence, there is no explicit distinction made between long-term versus short-term or working memory; however, the limited buffer size of the modules is compatible with the notion of strict limitations on working memory. An incoming word triggers a series of cognitive processes: an initial production rule is fired (and sent to the perceptual-motor system) to encode the word, followed by a second production rule that initiates a retrieval from declarative memory to access the lexical entry and syntactic information associated with the word, while a third firing within the procedural module utilizes this retrieved lexical information to update or set new predictions for syntactic and semantic representations. For instance, this lexical information, in combination with the contents of other buffers that track syntactic expectations, gets passed back to the procedural module to determine a set of retrieval cues for previously constructed syntactic and semantic representations. These retrieval cues have been thought of as information chunks themselves with feature-value pairs that are matched against suitable targets in memory. This second retrieval process that targets recently processed constituents has been the focus of this dissertation. The time it takes for this retrieval process to complete, therefore, is hypothesized to be contingent on the representational complexity of the target chunk. Once this material is retrieved, yet another production takes place that integrates the most recently seen word with the retrieved material. This entire process subsequently repeats at each new incoming word.

As a simplified example, consider a situation where a subject NP like *The dragon* has already been processed, and the word *dozed* has just been uttered. After encoding this word, the procedural module orders a retrieval of the lexical and syntactic information for this word from the declarative module. Given successful retrieval, this retrieved information temporarily resides in the declarative buffer and can be used to guide the retrieval of the recently seen subject NP. A procedural rule might say something like, “If the new lexical entry is an intransitive verb, and the current goal is to build a sentence, then retrieve the

---

7Lewis and Vasishth (2005) argue for a non-standard module and buffer dedicated to storing lexical information. In contrast, the theory of sentence memory provided by Anderson et al. (2001) does not include such a buffer, nor do other standard versions of ACT-R, which situate all lexical information in declarative memory. Since the current section is intended only as an overview for illustrative purposes and to clarify certain representational assumptions, I will proceed with a discussion of only the standard version.
subject NP.” All procedural rules in ACT-R have this basic form of evaluating the contents of the other buffers, and then programming a set of actions based on those conditions, i.e. if-then statements. At this point, therefore, a set of retrieval cues are developed and used to retrieve the dragon. In this case, of course, the retrieval is likely to be quite trivial, barring any simultaneous processing demands. Following the retrieval, a final production rule in this cycle attaches the verb and subject and updates the goal buffer as necessary.

Beyond the detailed specification of the individual steps involved in comprehension, this theory models two important aspects of sentence processing that I have identified as significant determinants of the nature of informativity effects: decay and interference. The theory depends upon the assumption that the activation of stored memory items fluctuates over time, as a function of time and usage, as well as associative interference. These ideas are realized in the formula for activation of a memory chunk expressed below, which says that the activation level of a chunk \( A_i \) is a function of its base level activation \( B_i \) plus a boost from the retrieval cue(s), whose value depends upon the associative strength \( S \) between a retrieval cue \( j \) and the target \( i \). \( W_j \) is set to \( 1/n \), where \( n \) is the number of elements currently in focus, i.e. elements contained in the goal buffer:

\[
A_i = B_i + \sum_j W_j S_{ji} \tag{6.1}
\]

The base-level activation, in turn, reflects the history of past usage. Its value depends upon how many times an item has been retrieved, as well as the time between retrievals, i.e. how long ago the item was last retrieved. Here, \( t_j \) represents the time that has elapsed since the \( j \)th retrieval of the item \( i \). The exponent \(-d\), however, denotes a fixed value that accounts for the effect of activation decay, which previous modeling in ACT-R has estimated at .5 (Anderson & Schooler, 1991; Anderson & Lebiere, 1998).

\[
B_i = \ln \left( \sum_{j=1}^{n} t_j^{-d} \right) \tag{6.2}
\]

According to this equation, the base level activation incorporates the effect of all previous retrievals. Each retrieval effectively increases the base-level activation value and counters the effect of decay over time. Thus, this formula captures the intuition that memory retrievals that occurred far in the past will have little impact on how accessible an item in memory
is. In contrast, multiple retrievals over a relatively short period of time will effect a high level of activation.

The last component of the equation in (6.1) accounts for the role of associative interference. The associative strength between a retrieval cue \((j)\) and a memory item is limited by how many items the retrieval cue is associated with—referred to as the \(\text{fan of } j\).

\[
S_{ji} = S - \ln(\text{fan}_j) \tag{6.3}
\]

The underlying intuition here is that the more items a particular retrieval cue is associated with, the less useful it will be for retrieving any particular item from memory. A retrieval cue that matches with multiple items in memory will consequently have its associative strength distributed amongst these multiple targets, thereby lowering its associative strength to any one target. An appropriate analogy would be to what happens if a waterway is diverted into many channels: the flow of water in each channel is weaker than it would have been if fewer channels existed.

As demonstrated by simulations in Anderson et al. (2001), Lewis and Vasishth (2005), and elsewhere, these formulae can be marshaled to successfully predict retrieval and reaction times with the addition of a final equation that relates the retrieval latencies \((T_i)\) to activation \((A_i)\). In this equation, the term \(F\) denotes a scaling constant that depends upon the cognitive task.

\[
T_i = Fe^{-A_i} \tag{6.4}
\]

In essence, these mathematical formulae characterize a cognitive system where retrieval from memory is complicated by both retrieval and interference. This aligns with empirical evidence that demonstrates both factors are at work in behavioral responses (Anderson & Lebiere, 1998; Altmann & Schunn, 2002). Moreover, the base-level activation equation means that the activation for items in memory fluctuates throughout discourse as a function of their use.

Applied to the current reading time data sets, the ACT-R theory proves to be quite compatible with many of the findings, owing to the theoretical commitments regarding decay, interference, and fluctuating activation levels. Take, for instance, the retrieval time differences observed between indefinite NPs like “an alleged Venezuelan communist” versus “a communist” in experiment V. According to the ACT-R theory, each of the adjectives in the more informative description launches a retrieval for a predicted noun category to
attach to, based on a procedural firing along the lines of the following:

(81) \(\text{IF } \) goal category is NP
and lexical entry is an ADJ
\(\text{THEN set retrieval cues to N expectation}\)

At the first adjective, therefore, a predicted noun category is created, which is subsequently retrieved at the next adjective, and then retrieved once more when the head noun is finally encountered. Each of these retrievals boosts the activation level of the corresponding chunk in memory, according to the base-level equation above. This reasoning is precisely the same as that applied to anti-locality effects in Hindi by Vasishth and Lewis (2006), as described earlier in this chapter. When the gap site is eventually reached, the syntactically more complex expression is more likely to have a higher activation level, making a fourth retrieval easier, so long as not so much time has elapsed that the levels of activation have decayed away to nothing.

ACT-R also presents the necessary machinery for modeling the effects of associative strength or multiple retrieval paths. In an investigation of word literality and lexical priming, Budiu (2001) and Budiu and Anderson (2004) offer a formal means for calculating associative strength on the basis of semantic relatedness. They offer an alternative means for calculating the associative strength \(S_{ji}\) between concepts on the basis of semantic similarity or relatedness:

\[
S_{ji} = C + M \times \sigma(j, i)
\]  

\(C\) here is a negative constant, reflecting the idea that positive associative strength requires relatedness above a certain threshold. Similarly, \(M\) is also just a scaling constant, while \(\sigma(j, i)\) represents the semantic relatedness of chunks \(i\) and \(j\) with possible values between 0 and 1. In essence, activation spreads to sufficiently similar or related chunks in declarative memory, making them easier to retrieve within some future window of time, if necessary.

Activation presumably also spreads when one chunk in a network of linked chunks is retrieved, since chunks contain pointers to other chunks, although ACT-R does not presently

---

\(^8\)Budiu and Anderson employ a pre-existing theory to calculate semantic relatedness—the Latent Semantic Analysis of Landauer and Dumais (1997)—which I do not have the space (or the mathematical know-how, for that matter) to adequately describe here. As the authors note, the theory of relatedness or similarity that they employ could be “replaced by more reliable definitions of similarity (e.g. participant ratings) if these were available.”
specify how this is realized in mathematical terms.\footnote{This characterization is more faithful to the sentence processing model in ACT-R that appears in Lewis and Vasishth (2005). In Anderson et al. (2001), syntactic nodes are connected with links that are themselves chunks that contain hierarchical information about the constituents they conjoin, while the nodes do contain any structural information.} Anderson and his colleagues indeed have presumed such a process where the retrieval of one chunk leads to the retrieval of other connected chunks in order to explain certain aspects of sentence memory. It is also reasonable to assume that connecting two chunks of information would increase their associative strength, at least temporarily. The availability of recall transitions from one node to another in a network of stored chunks would consequently end up benefiting more informative linguistic expressions, because failure to retrieve one chunk could be avoided by the successful retrieval of a related chunk, which would provide a route to the target.

Besides having a clear justification for the relationship between retrieval and syntactic complexity, the ACT-R model rationalizes the effect of semantic complexity evident in experiments VI & VII. Syntactic complexity involves the combination of multiple memory chunks and repeated retrievals, but increased semantic informativity without an accompanying increase in syntactic information is not realized in terms of larger networks of memory chunks or repeated access. Representationally speaking, the only difference is in terms of the number of slots or feature-value pairs within the chunks. From the perspective of ACT-R, therefore, the term “complexity” as used here means something different for semantic representations, rather than syntactic representations. For instance, the semantic representation for a noun like soldier can theoretically contain a slot/feature-value specification for occupation that a less informative noun like person does not have. Now if a set of semantic retrieval cues has a slot or specification for occupation, this information can be matched up against the information stored in the memory chunk for soldier, but not person. In other words, a greater number of semantic specifications provides more targets for cue-target matching in retrieval. And to the extent that those features are unique in the context, the effect of associative interference is reduced. ACT-R therefore can link semantic informativity effects to better targets for retrieval.

Additionally, equation (6.5), which models activation spread, can at least partly explain some semantic informativity effects. Essentially, this equation along with (6.1) is meant to express the idea that memory items in focus (chunks currently in a buffer) spread activation to other items they are associated with, which facilitates the retrieval of related items.
Thus, the word *Cheshire* may have a high degree of associative strength with the word *cat*, which means, practically speaking, that the word *cat* will be relatively easy to retrieve from declarative memory after reading or hearing the word *Cheshire*. Along similar lines, the retrieval of filler-phrases at subcategorizing verbs can be affected by the degree of association between the verb and filler. A verb like *ordered* (when its meaning has to do with giving a command), for instance, may ultimately have a greater associative strength with *soldier*, rather than *person*. In the experiments where semantic informativity effects appeared (experiments VI & VII), in fact, most of the subcategorizing verbs show a bias toward the semantically more informative noun phrase (e.g. *rejected-applicant*, *hired-janitor*, *bribed-representative*, etc.). Hence, semantic informativity effects may be limited to contexts where retrieval cues support the retrieval of a more semantically complex representation due to differences in associative strength and interference. This means that in retrieval contexts where the retrieval cues lack an associative strength bias and where interference is at a minimum, semantic complexity effects may well be absent.

For both syntactic and semantic representations, ACT-R is also compatible with the claim that additional processing that occurs during the encoding phase benefits the retrieval process. Recall of surface details, for instance, can be significantly improved by instructions to pay attention to them in the utterance or text (Murphy & Shapiro, 1994). As stated in Anderson et al. (2001), “ACT-R predicts that memory for any chunk, syntactic or semantic, will be enhanced by greater processing.” This prediction is satisfied in the current set of experiments where we repeatedly saw facilitated memory retrieval following greater processing during the encoding phase. And while experiment VIII demonstrated that this correlation was unable to explain the full range of retrieval time differences, the aggregated evidence suggests a powerful effect of encoding costs on retrieval.

Most of the necessary components for successfully modeling the informativity effects documented within this text therefore already exist in the ACT-R theory. Syntactic complexity forces repeated retrievals of some head category or projection. Plus, syntactic complexity translates to more developed networks of syntactic information that pose multiple access points for retrieval, i.e. the successful retrieval of one node can lead to the retrieval of other connected nodes. Likewise, semantic informativity effects can be linked to the process of cue-based retrieval: elements in the set of retrieval cues can be matched against feature-value specifications contained in more semantically informative representations. Additionally, retrieval cue elements can also spread activation to stored memory chunks, raising
the activation level of future targets. Thus, when the retrieval target has more semantic information, there is a greater chance that the chunk will receive an activation boost. Finally, ACT-R assumes that more syntactic and semantic processing leads to facilitated retrieval, so either linking more chunks to a particular chunk or adding more feature-value specifications to that chunk should improve the ability to re-access that chunk.

The hypotheses that I have advanced within this dissertation are thus further substantiated by their compatibility with a thoroughly documented and tested model of cognition. Certainly, some of the proposals regarding the retrieval process need to be further tested, refined, and formalized; however, this section has demonstrated that most of the requisite machinery for explaining the effects of representational complexity is available and in line with the predictions of an established cognitive model. Even more significantly, this model supports the conclusion that the underlying reason for informativity effects arises from (1) the process of creating a complex representation and (2) the resulting complexity in terms of larger memory networks and more feature-value specifications. Cumulatively, these elements serve as devices that offset the retrieval difficulties introduced by the forces of decay and interference which constitute key features of the cognitive architecture.

6.5 Summary

The theme of this chapter has been that representations in memory are associated with fluctuating activation levels. In discourse, the activation of a particular discourse entity varies throughout the conversation or text, as a function of a number of independent factors. Chief among these factors that reduce the activation of a discourse entity are decay and interference from other competing entities. Of course, previous discourse theories have acknowledged this fact about varying degrees of activation, and have indeed used this knowledge to determine why certain choices among possible linguistic alternatives are made. Some theories have even credited repeated references to a particular entity with producing activation boosts (or somehow making the discourse entity more salient). What they have not acknowledged, however, is that these choices themselves have consequences for the activations of the corresponding mental representations.

Supplementing these previous theories with the present hypotheses on representational complexity results in a more complete picture of discourse and the factors influencing reference comprehension. It can explain how future form choices are processed and why some
form choices engender future processing difficulty, but others do not. These explanations are simply unavailable under theories which do not give credit to the informativity of linguistic expressions on subsequent comprehension. The reasons speakers choose to use highly informative expressions, I have suggested, may be just to (re-)solidify a conceptual pact about a discourse object. In psycholinguistic terms, this is realized in terms of choices that increase activation.

These ideas were formally embedded within the ACT-R theory of cognition that models sentence processing in the same fundamental way as any other specialized cognitive task. The ideas of decay and interference are central to this model, and hence pose a natural context for situating the current findings. The theory also possesses the necessary components to understand why syntactic and semantic complexity benefit retrieval. Interestingly, this cognitive model points to distinct reasons for these effects: syntactic complexity effects largely stem from the process of building a complex representation, while semantic complexity effects emerge from features of the ensuing representation, including the resulting associative strengths and extra feature-value specifications. Hence, comparing the current set of results to the predictions and assumptions of a recognized model of cognition has served to not only clarify certain assumptions regarding the retrieval process, but also to elucidate the potential cognitive reasons for the effects of syntactic and semantic complexity on retrievability.
Chapter 7

Conclusion

7.1 Overview of Findings

Given the possibility of characterizing the same entity in countless different ways, the research in this dissertation proceeds from the basic assumption that these choices of form ought to impact how these entities are processed or thought about later on. Some descriptions effectively background certain entities, making them relatively unimportant for future discourse, while other descriptions increase the relevance of their corresponding discourse referents. I have attempted to cash out these intuitions in cognitive terms by suggesting that the information content of a linguistic expression used to build or update a mental representation has repercussions for how easily that representation can be re-accessed. That suggestion by itself should hardly be surprising or controversial. Plenty of other research has documented that qualitative aspects of a memory target affect its retrievability, e.g. the distinctiveness of the relationship between propositions (Stein et al., 1978; Bradshaw & Anderson, 1982; McDaniel et al., 1989). What is novel about the current proposal is the claim that the complexity of a memory representation is an important predictor of how difficult it is to retrieve that representation from memory. In particular, as complexity increases, retrieval becomes easier. As discussed in the first chapter and in the preceding one, this relationship between complexity and retrieval is limited in practice by (1) the expectation that information has a discourse purpose and (2) the cost of processing information. Increasing syntactic and semantic information therefore will eventually produce more processing difficulty than the facilitated retrieval process can compensate for.

This dissertation set out to determine whether such a relationship could be observed
CHAPTER 7. CONCLUSION

in sentence comprehension using filler-gap dependencies as the primary context for testing and evaluation. These syntactic constructions with dislocated constituents impose numerous processing difficulties and can strain working memory resources, as demonstrated via multiple experimental methodologies. Furthermore, since there is no overt constituent marking at the position of the retrieval site, the question of how much the antecedent and the anaphor match is laid aside. Due to these facts, FGDs present a natural testing ground for a proposal about memory facilitation.

Over the course of multiple comprehension-based experiments, a variety of phrasal types in the filler position were seen to participate in informativity effects—where the informativity of a linguistic expressions influences the subsequent retrieval of the corresponding mental representation. The empirical investigations in this dissertation began with a concentration on wh-phrases, and even more particularly the contrast between bare wh-words like who and the more informative which-N’ phrases in syntactic island settings. The retrieval contrasts observed between FGDs with these two types of fillers were subsequently replicated with adjunct phrases in another island context. Hence, these studies cumulatively argued for the significance of informativity on retrieval from working memory, at least within syntactic islands. Follow-up experiments, however, verified that the relationship between informativity and retrievability is visible beyond the insular realm. In other complex, but unquestionably grammatical contexts like nested dependencies, more informative wh-phrases again led to faster retrieval times. It was also shown that the retrieval differences observed in wh-dependencies cannot be ascribed to a simple binary contrast between bare wh-words and which-N’ phrases, as advocated in previous syntax research: differences in retrievability persist even when comparing which-N’ phrases that differ in terms of semantic complexity. These differences (especially when taken together with the observed contrasts with other types of phrases) are therefore incompatible with theories that make a dichotomous distinction among wh-phrases, in order to explain behavioral responses in comprehension and acceptability tasks. Additional reading time studies further showed that informativity effects were not unique to wh-dependencies: the manipulation of the informativity of both indefinite and definite NPs resulted in significant retrieval differences. Added together, these results convincingly affirm that the relationship between informativity and retrievability is based on a general cognitive principle that indiscriminately applies to phrases of all sorts.

This leaves open the possibility that the magnitude of informativity effects varies across phrasal types. In actual discourse, phrases like wh-expressions that introduce new discourse
entities (i.e. phrases that are not anaphoric) may demonstrate a greater sensitivity to manipulations of informativity because the corresponding mental representations lack any prior activation level. Undoubtedly, other contextual factors influence the magnitude of these effects as well. For instance, the contrast in experiment VIII between complex definite NPs and “bare” definite NPs may have partly depended upon the presence of multiple other definite NPs in the sentential context. That is, the informativity effects potentially owe their weight to the presence of interfering NPs of the same phrasal type.

Given the cumulative evidence that supports the MFH, the secondary goal of this dissertation has been to address the question of what underlies this relationship. It is one thing to establish that one cognitive process affects the outcome of another process, but it is theoretically more interesting to determine why such a relationship exists in the first place. In the course of confronting this issue, several theoretically possible interpretations of the informativity effects were discounted.

One plausible source of these effects that I considered was the study time allocated to forms of differing informativity. Since greater syntactic complexity is essentially tantamount to more morphophonemic material that requires a greater amount of time to process, the retrieval facilitation for more informative expressions potentially originates with longer amounts of time building a single representation. The first two experiments in the fifth chapter, however, counter this hypothesis. It was demonstrated that when the number of words are held constant, but the semantic complexity of the filler-phrase is varied, the retrieval effects still remain (although in a weakened form). These same experiments also clarified that it was neither syntactic information or semantic information alone that created these effects. Minimizing semantic differences between two forms, but varying syntactic structure, again produced statistically reliable differences. Combined, these experiments backed the original formulation of the MFH that states that both syntactic and semantic information contribute to the retrieval effects.

Another alternative hypothesis that I addressed was focused on whether increasing information constituted one of innumerable ways of drawing attention to a particular phrase. On this hypothesis, there is nothing particularly special about informativity—it just represents one means for modulating the amount of attention comprehenders give to sentence constituents. In fact, I have recommended some version of this hypothesis throughout the pages of this dissertation. The critical question to ask, therefore, is whether all the retrieval
effects could be accounted for via a simple correspondence between encoding costs and retrieval costs (as one goes up, the other goes down). To test this hypothesis, I contrasted forms of equal syntactic complexity that nonetheless differed in terms of the typicality or distinctiveness of the encoded feature combinations. An analysis that draws a strict correlation between attention at encoding and retrievability would predict that the oddness of the distinctive feature combinations would have the consequence of faster retrieval times. Indeed, the distinctive condition did attract the most attention as measured by increased reading times around the reference. Counter the expectations of this alternative hypothesis, however, the most distinctive or atypical description did not produce significantly faster processing times at the retrieval site. Nevertheless, this condition still produced faster reading times at the retrieval site than the less informative definite NPs. These results therefore justify the conclusion that the informativity effects are not exclusively due to the amount of reference processing or the amount of attentional demands that are imposed during the initial encoding phase. This still allows encoding costs to have a significant impact on the retrieval process, but it suggests that other factors are relevant, as well.

In particular, I made the claim that the number of retrieval paths and the associative strength between points on that path is another critical component of the retrieval effects. To explain why the distinctive feature combinations did not cause the fastest retrieval times, I suggested that these features had low associative strength, which weakened the ability to transition from the recall of one feature to the next. Thus, I have advocated the view that there are at least two fundamental reasons for the relationship between informativity and retrieval. One of these has to do with the fact that it takes more cognitive resources to construct more complex representations. At heart, this correspondence reflects the fact that constructing complex representations requires activating the same representation on multiple occasions in order to add to it or otherwise modify that representation. In other words, following the ideas expressed in recent models of language comprehension (Anderson et al., 2001; Vasishth & Lewis, 2006; Lewis & Vasishth, 2005; Lewis et al., 2006), repeated access of some memory representation increases the corresponding activation level. This accessing process and integration of new information is necessarily costly from a short-term perspective, but it actually facilitates processing if and when the relevant representation needs to be accessed once again. Hence, the “pay now or pay later” principle depends upon the idea that reference processing raises the activation level of the corresponding representation. Since retrievability is a function of activation, the effort required to retrieve
some linguistic entity from memory will partly depend upon the amount of processing that went on during the encoding phase.

The other factor that I have concluded is driving this relationship is the availability of multiple retrieval paths that come with complex representations. Given the assumption that it is possible to transition from one node in a memory network to another, complex representations increase the chances of successful retrieval. Recovery of one node in an information network can lead to the recovery of other information in that network. Again, this hypothesis draws on a large body of memory research that identifies inferential processes and other indirect methods of imputation at work in recall. The unique aspect to the current proposal is the suggestion that such methods of recovery are operative in on-line language comprehension. As indicated in the previous chapter, this aspect of my analysis is complemented by research on similarity-based interference. Multiple retrieval paths can ameliorate some of the difficulty that accompanies the simultaneous presence of representations with overlapping features in working memory. This amelioration is by virtue of the fact that increasing representational complexity increases the probability that some features will be unique and therefore help distinguish a representation from other competitors in memory.

These two aspects of informativity—activation boosting and multiple retrieval targets—act as compensatory mechanisms for the natural processes of decay and interference that are omnipresent factors in language comprehension. Representations in memory are constantly becoming weaker, and as other representations get introduced into the discourse, they are likely to lose their uniqueness. Representation complexity partially offsets these hinderances to sentence processing, because the process of building the complex representation creates a highly activated memory network and because the ensuing cognitive structure allows for multiple retrieval strategies. Put slightly differently, it is both the process and the result of building a complex representation that facilitates the retrieval process. If there were no other cognitive pressures but decay and interference and no other task but to communicate successfully, then expansive descriptions that overlap only with their own linguistic antecedents would be universally preferred in comprehension (which would undoubtedly influence production patterns). As it stands, however, there are other relevant and critical factors that constrain the utility of information. In particular, parsing, categorizing, and integrating information about a particular discourse entity consumes valuable resources that are limited in quantity by the architecture of working memory. Accordingly,
CHAPTER 7. CONCLUSION

196
discourse reflects a balancing act between too much information and too little information. Yet within the confines of that balancing act, there is still a certain amount of latitude when it comes to choices of linguistic form. To thus understand how the way we talk about things now affects how we talk and think about them later, it is necessary to have a theory of the consequences that the informational complexity of a linguistic description has for subsequent language comprehension—which is what I have attempted to provide here.

7.2 Implications

Aside from their relevance for theories of sentence processing and retrieval from working memory, these empirical findings have clear and significant repercussions for numerous areas of linguistic research, ranging from work on anaphor resolution to discourse structure to syntactic theory. Practically speaking, any linguistic task that requires retrieving a stored mental representation is affected by the issues raised here.

This implies, for instance, that the search for a referent when an anaphor is encountered partly depends upon the informativity of the antecedent. A syntactically and semantically more complex antecedent should make this search easier, provided that the antecedent is not “too accessible.” The underlying intuition here is quite simple: the ability to access an antecedent at an anaphoric reference is conditioned, in part, by how that antecedent was previously presented and framed in the discourse. This idea has surfaced to a limited extent in previous work (Gernsbacher, 1989; Gordon et al., 1993; Grosz et al., 1995), but the predominant focus there has been on the extent to which an anaphor and antecedent overlap in features—how similar they are to each other.

Distinct from any principle of antecedent-anaphor harmony, the current research program predicts a systematic relationship between antecedent form and the ability to find that same antecedent at a subsequent anaphoric reference.\footnote{Harmonization between antecedent and anaphoric form is not a principle that follows from anything I have said here, so, to the extent that such a principle holds (and it is by no means a settled question that it does), it originates from some source independent of the MFH.} This relationship was cast in terms of activation levels: the form of an antecedent phrase modulates the activation of the corresponding representation, which influences the ease of identifying the correct referent at subsequent anaphoric references. Specifically, increased informativity boosts the activation level, meaning that retrieval will be easier when the same representation needs to be
accessed again. This scenario permits successful communication even when the anaphoric form itself contains relatively little information, as in the case of pronouns or deictic references. Essentially, then, the present analysis offers a means for understanding how the availability or accessibility of discourse referents varies throughout discourse. This understanding supplements the insights expressed within previous literature that confronts the question of what factors determine the choice of a particular form by clarifying what the consequences of those choices are for comprehension. Linguistic form choices consequently do not represent the culmination or final stop in a series of cognitive processes; instead, each form choice contributes to and modulates the nature of a new cycle of processes, until the conversation ends or the topic is abandoned.

The pairing of reading time results with the judgments from acceptability tasks also demonstrates the applicability of this research to syntactic theory. Multiple experiments exemplified that acceptability judgments varied in accordance with processing difficulty: where processing difficulty increased, acceptability judgments decreased and where where difficulty decreased, acceptability increased. This negative correlation was specifically identified in the context of syntactic island constructions, which have been historically labeled as instances of universal constraint violations.

Noting, however, the variety of exceptions and gradience associated with these constructions, I have repeated the assertion made numerous times over the past several decades that the degraded quality of these constructions likely stems from the processing difficulty they impose. The reading time evidence here repeatedly confirmed that FGDs into islands could be processed as easily as non-syntactic island tokens when a more informative filler-phrase begins the long-distance dependency. Reading times, response times, and question-answer accuracies all pointed to this conclusion. Moreover, this improved processing corresponded to higher acceptability ratings. At a bare minimum, especially when the plethora of other processing difficulties associated with island constructions are taken into account, these results present the possibility that processing factors are behind the perceived unacceptability of many syntactic island constructions.

This line of thinking results in a more parsimonious and spare grammar and appeals to the existence of empirically verified results, rather than the intuitions of linguists. At a functional level, too, there is no obvious reason why grammars should contain seemingly arbitrary rules on dependencies that require a daunting degree of complexity to express even in the language of grammarians. This is not to suggest that languages never exhibit
a certain amount of arbitrariness, but when that arbitrariness purportedly applies to all the languages of the world, there are good grounds for being skeptical. In contrast, a processing-based account of island violations appeals to independently motivated cognitive principles and possesses an eminently transparent functional explanations: islands sound bad because they’re hard. This reasoning is no different than the reasoning applied to constructions like center-embeddings and garden path sentences that were never treated in terms of grammatical constraints.

More broadly, the assertion that processing difficulty may account for a certain amount of variance in acceptability judgments and apparent grammaticality distinctions has ramifications beyond just the domain of syntactic islands. Of course, a number of other researchers have demonstrated that acceptability judgments are sensitive to a variety of factors beyond the purview of grammatical theory (see, for instance, Schütze (1996) and Cowart (1997)). This research is unique, however, in that it shows that manipulation of the same non-structural factor (i.e. the informativity of the filler-phrase) affects both the processing difficulty and corresponding acceptability associated with a variety of island constructions. Variation and gradience in other types of dependencies may therefore demonstrate a connection to this same factor, although a discussion of such issues goes beyond the scope of this document.

### 7.3 Future Research

Numerous questions about the relationship between informativity and retrievability have been left unanswered in this dissertation. Given that the role of informativity in language comprehension has essentially been unexplored prior to this dissertation work, I have attempted first and foremost to present as much evidence as possible to establish the existence of the informativity effects in an array of syntactic constructions and with a diversity of phrasal types. A secondary goal was to find evidence that could explain the relationship between informativity and retrievability. Along the way, quite a few ancillary questions have arisen that deserve future consideration.

The studies focusing on direct interrogatives, for instance, led to the following question: to what extent does the possible answer set size influence reading times at retrieval and integration sites in *wh*-questions? One theoretically possible reason for the emergence of faster reading times in questions that began with more informative *wh*-phrases has to do
with the limitations that such phrases place on the possible answer set size. Upon this interpretation of the data, the greater quantity of information in the filler-phrase spares computational resources used in the task of developing an answer to the question.

One simple way of testing this hypothesis is to develop contexts in which there is only one possible answer for direct questions beginning with either a complex wh-form or an informationally simpler wh-form. For instance, in the example below, the initial context mentions only a single animate discourse referent.

(82) The front-page story suggested that rising oil prices concerned the German prime minister because of the repercussions on the whole economy.

Who\ Which prime minister did the report suggest that the price of oil concerned due to its effect on the economy?

Consequently, the construction of the answer set and the subsequent evaluation of the possible answers should require the same amount of cognitive effort. The MFH predicts that, even in such circumstances, the syntactically and semantically more complex wh-phrase should lead to faster retrieval, as long as processing performance is not operating at optimal efficiency.

Additionally, the experimental research in this dissertation has focused almost entirely on the processing of filler-gap dependencies. As mentioned on numerous occasions, though, the implications of the MFH extend far beyond the realm of FGDs. Thus, other comprehension processes that necessitate the retrieval of previously mentioned constituents should reflect a similar bias that affects the processing of FGDs. Perhaps the most obvious candidate for future consideration that comes to mind, therefore, is anaphoric interpretation.\footnote{Note that gaps can be reasonably considered a type of null anaphor. Although there are reasons for differentiating filler-gap dependencies from other types of linguistic dependencies, there is no a priori reason for supposing that filler-gap dependencies must conclude with a missing constituent. In fact, there is a range of cross-linguistic evidence that suggests that FGDs may end with overt constituents, which are normally labeled resumptives. Even in English, there are attested examples of FGDs ending with not just pronominals but also epithets, as in \textit{There was one prisoner that we couldn't figure out why the guard beat the poor guy.}}

Suggestive evidence for such an effect already appears in Cowles and Garnham (1995), as mentioned previously. Also, Frazier and Clifton (2002) offer experimental evidence that
which-N’ phrases are deemed better antecedents to subsequent pronouns than bare wh-words. In essence, the syntactically and semantically more informative phrases provide better antecedents. One way of interpreting these results is to say, in the spirit of this dissertation, that the informationally richer which-N’ phrases activate the corresponding mental representation more than a bare wh-words. This means that the activation level for the representation associated with the which-N’ phrase is likely to be higher at a subsequent pronominal reference. The predictions of the MFH can be further tested in this regard by examining the reading times of sentences containing pronominal anaphors in passages like those in Chapter 2.2, where the only manipulation in the entire text is the form of some second reference to a discourse entity.

This raises a separate issue or area for further research: how does informativity impact retrieval in longer texts and discourse, i.e. when dependencies cross sentence boundaries? The experimental stimuli employed in the reading time studies all contained single sentences with subsequent comprehension questions. In this sense, they do not represent paradigmatic instances of texts that people read on a daily basis. Accordingly, further insights into the nature of the retrieval effects can be gained by considering how comprehension processes are affected by manipulations of informativity in longer texts, where discourse entities are tracked and managed over a longer period of time. Indeed, considering that sentence and paragraph boundaries mark points when new predictions of upcoming constituents are made and attention to surface details of a sentence ends, informativity effects may prove to be even more profound when the targeted material for retrieval occurs in a previous sentence or paragraph.

The production side of the MFH has also gone more or less unexplored, although some suggestive evidence was presented that accords with the predictions of the MFH. Experimentally speaking, there are more challenges associated with testing for informativity effects in production, rather than comprehension. This requires finding evidence that speakers purposefully choose linguistic forms that a more informative than some available alternative. It may be possible, however, to evaluate this proposal in production by testing how speakers alter their descriptions when they are led to believe that their addressee suffers from a type of memory impairment. Or, speakers could be deceptively convinced that their addresses will have to recall their statements at a later point that would stress accurate retrieval. Either of these designs would evaluate how speakers modulate their form choices on the basis of predictions about the needs of addressees. In addition, further corpus work can also help
CHAPTER 7. CONCLUSION

establish how the retrieval benefits for more informative phrases affects production.

It has also been pointed out on numerous occasions that there is clearly some point at which more information no longer helps retrieval, and may, in fact, pose more of a processing hinderance. The reasons for this limitation concern the cost of processing and integrating new information about a discourse representation, as well as the expectation that information has some discourse purpose. What is unclear, however, is when these considerations of processing cost and discourse function outweigh any rewards from shorter retrieval times. For the purpose of constructing cognitive models of informativity and retrieval, therefore, it will ultimately be illuminating to establish whether there is an identifiable threshold at which additional information no longer improves the retrieval process.

Perhaps one of the biggest unanswered question lurking behind this research is the following: what exactly gets retrieved during language comprehension? Clearly, this is not what I have set out to answer, but it is ultimately of great importance for the theory developed here. At one extreme, everything about a linguistic phrase gets retrieved at retrieval sites, including phonetic, syntactic, and semantic information. Some researchers, in fact, have explicitly claimed that retrieval from working memory during language comprehension (in the case of gap-filling, at least) involves total recall, i.e. retrieval of all the linguistic information that was input at the time of encoding (Tanenhaus et al., 1985). But it is also reasonable to suspect that only a very limited amount of information gets retrieved, particularly so in light of the time and resource constraints that comprehenders face. The preceding analysis has assumed that at least some syntactic and some semantic information is retrieved at gap sites. At a bare minimum, my theory requires that lexical information about the syntactic head is retrieved during language comprehension. The present findings are also compatible with the hypothesis that other information in mental representations corresponding to filler-phrases is also being retrieved. But it is wholly possible that information that is basically irrelevant at the point of retrieval does not get retrieved. For instance, the lexical content of a specifier may be entirely unimportant for interpreting a gap and therefore need not be retrieved. From the standpoint of economy, then, it makes the most sense to believe that comprehension involves the retrieval of the minimal amount necessary for successful comprehension.

In this light, the present body of results advocates the idea that what comprehenders retrieve from memory at gap sites (and presumably other retrieval points, as well) is not a linearly ordered list of words, but pre-packaged chunks of information. If comprehenders
simply recover each and every word in a filler-phrase at the corresponding gap site, then we should have found slower reading times when the filler-phrase contained more information. But, in fact, the exact opposite was seen. This data thus stands in direct conflict with any proposal that suggests linguistic retrieval necessarily consists of revisiting linguistic objects, i.e. words or phrases, as implied by working memory research on the articulatory loop. According to the data amassed here, mental representations in memory that are built during language comprehension look like any other type of representation in memory and hence do not have to be accessed via their acoustic signatures.

Other remaining questions include the following: how much do the informativity effects depend upon the presentation of new information, as opposed to previously mentioned information? How do informativity effects vary across phrase types? Can the definition of informativity be extended or modified to reflect more than just subset relations? As presently formulated, the MFH and the definition for informativity only makes predictions about pairs of phrases where one phrase contains a subset of the information contained in the other phrase. Because of this definitional constraint, the MFH does not discriminate between phrases like "the Secretary of Defense during the Vietnam War who later became president of the World Bank" and "Robert McNamara."

Theoretically, informativity could be redefined on the basis of the quantity of information features that a description encodes. This would not alter any of the conclusions made here, but there are several problems associated with such a reformulation: (1) it is not clear exactly what features should be counted in order to determine some overall informativity statistic and (2) different features may have different “informational weights.” That is, contextual factors may make some kinds of information particularly important for retrieval. For this reason, the concept of informativity is only defined in terms of subset relations, which controls for the possibility of different informational weights and obviates the necessity for calculating an information statistic. If future research, however, provides techniques for coping with these issues, then the scope of informativity effects can be expanded to consider a wider range of contrasts.

Finally, the experiments in this dissertation have almost exclusively depended upon self-paced reading to amass comprehension data. The self-paced reading methodology, however, does have its limitations and drawbacks. In many experiments using this methodology, subjects see scores of unique proper names and other referential descriptions in a short period of time. This preponderance of references that do not have to be tracked further than a
single sentence may incite participants to develop abnormal referential processing strategies, i.e. processing references on a very shallow level. Moreover, comprehension is assessed mainly on the basis of yes/no or multiple choice questions that permit a high quantity of false positives.\(^3\) That is, it is difficult, if not impossible, to evaluate how accurately a subject understood a given sentence if the comprehension question was answered correctly. These considerations and others like them do not invalidate the body of experimental results presented here (nearly every experimental methodology in linguistics research demonstrate \textit{some} shortcomings), but they do recommend the idea that the reading time findings should be verified with other methodologies. Hence, it is ultimately necessary to corroborate the findings reported here with other experimental methodologies such as ERPs, fMRIs, speed-accuracy trade-offs, eye-tracking, as well as other direct tests of memory like free recall and recognition of entire sentences where the informativity of certain constituent phrases is systematically manipulated.

7.4 Final Thoughts

At the most basic level, the theme of this dissertation is that memory retrieval is sensitive to how stored representations were formed, as well as what the resulting form of that representation is. This elementary proposition constitutes a seemingly very general fact about cognition. Considering the arguments that the precise form of this sensitivity in comprehension ensues from architectural facets of memory, specifically decay and interference, any other cognitive task that is subject to the same constraints ought to exhibit a similar type of sensitivity. Retrieval of visually processed information, for instance, is potentially also influenced by the complexity of the visual stimulus. In this sense, the arguments and conclusions developed here are not necessarily specific to language.

For the study of language and language processing, though, this text began with the question of whether the information contained in a linguistic reference affects the way the corresponding mental representation is accessed later. The variety of possible alternative linguistic formulations should compel us to understand not only how one particular form is settled on, but also what consequences those decisions have on later language understanding.

\(^3\)This is not an inherent problem of self-paced reading. Comprehension can also be checked by more rigorous methods such as requiring participants to write out the answer to a \textit{wh}-question (a non-yes/no question), rather than selecting the correct answer from a range of displayed answers. Nevertheless, most self-paced reading experiments reported in the literature do not employ this type of comprehension assessment.
Answering such questions helps to identify the dynamic forces that operate behind the scenes of language use. In the present case, the inquiry into the consequences of linguistic form choice has uncovered the operation of a system that balances current and future resource needs. If retrieval is relatively difficult now, then the linguistic forms necessary to guide successful recall and comprehension will make subsequent retrieval easier. Conversely, when retrieval is easiest and accompanied by minimal forms, subsequent re-accessing is likely to grow more difficult. In a sense, then, memory retrieval in language comprehension behaves like a self-correcting system. Indeed, if production processes are logically added to the picture, then the resulting image is akin to a negative feedback system. When the activation level of a discourse representation gets too low, more informative forms are necessary to access that representation, which boosts the activation. But when the activation rises above some upper threshold, the system prefers less informative forms to preserve resources. Eventually, though, these less informative forms will be insufficient to keep activation above the threshold, and the system will need to correct again. These processes emerge, I have claimed, in response to the simultaneous pressures of resource limitations and the need to successfully communicate. This investigation of how the information in our utterances affects how we remember has therefore let us catch a glimpse of some of the dynamic processes that underlie language comprehension and the ways in which they interact.
Appendix A

Experimental Stimuli

A.1 Experiment I Stimuli

1. He knew which country\who Emily heard (a/the rumor(s)) that we had invaded due to increased political instability.
   Was there word of an invasion? Y

2. She discovered which passage\what Jacob read (a/the allegation(s)) that they had copied into the final written report.
   Did Jacob have proof of what they had done? N

3. I saw which convict\who Emma doubted (a/the report(s)) that we had captured in the nationwide FBI manhunt.
   Was Emma skeptical that we had captured someone? Y

4. She remembered which article\what Michael denied (a/the suggestion(s)) that they had plagiarized in order to sound intelligent.
   Were there suggestions that they had cited an article? N

5. He forgot which song\what Jessica reiterated (a/the contention(s)) that we had stolen from the original German composer.
   Were we accused of stealing the work of a German? Y
6. I verified which patient who Chris held (a/the conviction(s)) that they had cured with the new experimental treatment.
Did the new treatment cure Chris? N

7. He realized which prisoner who Ashley countered (a/the belief(s)) that we had interrogated without regard to international law.
Was the legality of our interrogation practices being discussed? Y

8. She wondered which company who Matthew confirmed (a/the suspicion(s)) that they had sued for its unethical accounting practices.
Did she confirm who had been sued? N

9. I researched which student who Amanda conveyed (a/the threat(s)) that we would reject despite an outstanding academic record.
Did Amanda say who had such a good record? Y

10. She guessed which client who Joshua disputed (a/the notion(s)) that they had defended before the federal appeals court.
Were there notions that a client had been refused by them? N

11. He insinuated which actor who Jennifer overheard (a/the comments(s)) that we had arrested for drunk driving last night.
Did Jennifer find out who was driving while intoxicated? Y

12. I acknowledged which novel what David expressed (a/the worry(ies)) that they would ban due to its racy content.
Was there worry that they would ban a biography? N

13. He indicated which concert what Sarah answered (a/the objection(s)) that we had canceled unnecessarily because of a disagreement.
Was the cancellation on account of a difference of opinion? Y

14. She understood which intern who Daniel envisioned (a/the prospect(s)) that they
would hire for the emergency room position.
Was the emergency room position filled? N

15. I surmised which agency\who Erin echoed (a/the complaint(s)) that we had overcharged for a routine financial report.
Did Erin agree that the report had cost too much? Y

16. She testified which supervisor\who James established (a/the expectation(s)) that they would fire because of his lewd behavior.
Were there expectations that they would promote a supervisor? N

17. He determined which project\what Nicole shared (a/the intuition(s)) that we would complete with extra time to spare.
Did he figure out what Nicole thought would be finished ahead of schedule? Y

18. I learned which route\what Andrew proposed (a/the hypothes(is)/(es)) that they had used to cross the mountain range.
Were there hypotheses that they had used a plane? N

19. He published which structure\what Brittany announced (a/the plan(s)) that we would build to replace the condemned building.
Was there an announcement about building a replacement? Y

20. She perceived which river\what Robert addressed (a/the fear(s)) that they had polluted with dangerous levels of toxins.
Were there fears that they had drained a river? N

21. I recorded which dictator\who Heather contested (an/the assertion(s)) that we had supported in his rise to power.
Was our role in his rise to power contested by Heather? Y

22. She uncovered which quality\what John debated (a/the perception(s)) that they had lacked in the home loan application.
Were there perceptions that she lacked a particular quality? N

23. He confirmed which student\who Elizabeth believed (a/the charge(s)) that we had suspended due to their poor grades.
Did Elizabeth believe that someone had been suspended? Y

24. I specified which car\what Ryan considered (a/the demand(s)) that they should recall because of failed safety tests.
Were there demands that they should recall a tire? N

25. He divulged which base\what Megan recommended (a/the proposal(s)) that we should abandon in order to minimize casualties.
Did Megan think we should leave to prevent further losses? Y

26. She recalled which leak\what Joseph appreciated (a/the comment(s)) that they would repair within a week from now.
Did they comment that we caused a leak? N

27. I proved which mineral\what Melissa relayed (a/the message(s)) that we had identified on the surface of Mars.
Did Melissa relay what had been located? Y

28. She admitted which community\who Brandon printed (a/the warning(s)) that they would forget in the disaster relief effort.
Did Brandon print that they saved a community? N

29. He confided which species\what Amber verified (a/the theor(y)/(ies)) that we would discover with enough time and energy.
Did Amber affirm that we could find something? Y

30. I asked which project\what Justin recognized (a/the concern(s)) that they should coordinate after a series of disasters.
Were there concerns that they should coordinate a fundraiser? N
31. He investigated which election what Lauren repeated (a/the claim(s)) that we had rigged in favor of the Democrats.
Did Lauren say that an election had been rigged? Y

32. She decided which company who William concealed (a/the sign(s)) that they had ruined with numerous illegal takeover attempts.
Were there signs that they assisted a company? N

33. I resolved which spacecraft what Rachel noted (a/the signal(s)) that we had lost due to an insulation problem. Was Rachel aware of the loss? Y

34. She perceived which tax what John protested (a/the request(s)) that they should pay for the next ten years. Were there requests to lower taxes? N

35. He studied which resource what Danielle confessed (a/the feeling(s)) that we had depleted over many years of mismanagement.
Did Danielle feel that we had used up something? Y

36. I noticed which player who Nick disregarded (a/the comment(s)) that they would lose because of a serious injury.
Were there comments that they would lose a contract? N

A.2 Experiment II Stimuli

1. Kathy wondered if her friends consulted the doctor from New Madrid at the hospital last night.

Who Which doctor did Kathy wonder whether they consulted at the hospital last night?
“the doctor from New Madrid” “the doctor from Los Angeles” “the pediatrician that practiced nearby”

2. Nathan mentioned that the generals promoted the sergeant from western Tennessee
during the war with Iraq.

Who\Which sergeant did Nathan mention whether they promoted during the war with Iraq?
“the lieutenant from West Virginia” “the sergeant from western Tennessee” “the sergeant from West Virginia”

3. Winston stated that the prosecutors trained the assistant who needed practice shortly before the trial began.

Who\Which assistant did Winston state whether they trained shortly before the trial began?
“the assistant who needed money” “the juror who needed money” “the assistant who needed practice”

4. Sarah wondered if the voters elected the senator from Portland, Oregon in spite of the scandal.

Who\Which senator did Sarah wonder whether they elected in spite of the scandal?
“the senator from Portland, Oregon” “the senator from Seattle, Washington” “the representative from Portland, Oregon”

5. Stephen confirmed that the radicals released the prisoner from New Zealand just after the demands were met.

Who\Which prisoner did Stephen confirm whether they released after the demands were met?
“the soldier from New Zealand” “the prisoner from New Zealand” “the soldier from London, England”

6. Brandy pondered if the investigators identified the suspect in Wednesday’s robbery at the station near Scranton.
Who\Which suspect did Brandy ponder whether they identified at the station near Scranton?
“the salesman from San Francisco” “the suspect in Thursday’s robbery” “the suspect in Wednesday’s robbery”

7. Jason learned that the doctors examined the patient with back pain very quickly at the clinic.

Who\Which patient did Jason learn whether they examined very quickly at the clinic?
“the patient with back pain” “the nurse with flat arches” “the patient with chronic arthritis”

8. Charlotte pondered if the delegates nominated the candidate who was liberal at the recent Republican convention.

Who\Which candidate did Charlotte ponder whether they nominated at the recent Republican convention?
“the candidate who was liberal” “the candidate who was conservative” “the governor who was famous”

9. Marvin mentioned that the violinists accompanied the pianist who played Beethoven at the concert on Sunday.

Who\Which pianist did Marvin mention whether they accompanied at the concert on Sunday?
“the pianist who played Mozart” “the pianist who played Beethoven” “the classical guitarist from Spain”

10. Maureen speculated that the TAs tutored the student who was failing before the final physics exam.

Who\Which student did Maureen speculate whether they tutored before the final physics exam?
“the undergraduate who was failing” “the musician who loved Chopin” “the student who was failing”

11. Albert learned that the managers dismissed the employee with poor sales after the annual performance evaluations.

Who/Which employee did Albert learn whether they dismissed after the annual performance evaluations?
“the employee with poor sales” “the employee with poor attendance” “the cashier who stole money”

12. Anna said that the members rejected the applicant from Yale University at the meeting about admissions.

Who/Which applicant did Anna say whether they rejected at the meeting about admissions?
“the researcher from St. Petersburg” “the applicant from Yale University” “the applicant from Harvard University”

13. Oscar stated that the detectives dispatched the officer with little experience after the tragic murder yesterday.

Who/Which officer did Oscar state whether they dispatched after the tragic murder yesterday?
“the officer with violent tendencies” “the mechanic with nine fingers” “the officer with little experience”

14. Vera indicated that the Yankees retired the batter with two homers after the previous hitter doubled.

Who/Which batter did Vera indicate whether they retired after the previous hitter doubled?
“the batter with two homers” “the batter with two triples” “the fan who was screaming”
15. Victor announced that the defendants intimidated the witness who was twenty with numerous threats of violence.

Who\Which witness did Victor announce whether they intimidated with numerous threats of violence?
“the witness who was thirty” “the witness who was twenty” “the expert who testified twice”

16. Matthew verified that the landlords evicted the tenant who wasn’t quiet to satisfy the other residents.

Who\Which tenant did Matthew verify whether they evicted to satisfy the other residents?
“the janitor who wouldn’t work” “the squatter who wasn’t quiet” “the tenant who wasn’t quiet”

17. Susan disclosed that the parents adopted the child who was sick after making several failed attempts.

Who\Which child did Susan disclose whether they adopted after making several failed attempts?
“the child who was sick” “the infant from San Jose” “the child who was violent”

18. Eric announced that the agents arrested the criminal who escaped yesterday at a motel in Ohio.

Who\Which criminal did Eric announce whether they arrested at a motel in Ohio?
“the thief who stole jewelry” “the criminal who escaped yesterday” “the criminal who sold drugs”

19. Erin disclosed that the terrorists killed the hostage from the U.S. in a moment of panic.
Who, Which hostage did Erin disclose whether they killed in a moment of panic? “the hostage from Sri Lanka” “the laborer helping with reconstruction” “the hostage from the U.S.”

20. Thomas indicated that the professors taught the graduate who studied psychology for at least two years.

Who, Which graduate did Thomas indicate whether they taught for at least two years? “the graduate who studied psychology” “the graduate who studied mathematics” “the brilliant girl from Missouri”

21. Thelma speculated that the jurors acquitted the defendant charged with arson after discussing all the evidence.

Who, Which defendant did Thelma speculate whether they acquitted after discussing all the evidence? “the defendant charged with kidnapping” “the defendant charged with arson” “the accountant who committed fraud”

22. Ryan verified that the zealots followed the leader of the cult to the islands without forethought.

Who, Which leader did Ryan verify whether they followed to the islands without forethought? “the lunatic who worshipped slugs” “the leader of the pack” “the leader of the cult”

23. Rachel said that the cornerbacks sacked the quarterback for the Seahawks multiple times during the game.

Who, Which quarterback did Rachel say whether they sacked multiple times during the game? “the quarterback for the Seahawks” “the athlete who kept fumbling” “the quarterback for the Giants”
24. Crystal confirmed that the teams rescued the survivor with severe injuries from the wreckage without difficulty.

Who\Which survivor did Crystal confirm whether they rescued from the wreckage without difficulty?
“the passenger from San Diego” “the survivor with severe injuries” “the survivor with minimal injuries”

A.3 Experiment III Stimuli

1. Katherine was informed that the salesman often only got four hours of rest per night because of his insomnia.

(How long\How many hours per night) did Katherine find out whether the salesman had slept on account of his schedule?
Is it possible to answer the question? N

2. Hilary appreciated the fact that the foreman was going to keep talking about his personal life until the work day was over.

(How long\For how much of the day) did Hilary understand whether the foreman would work before finally leaving for home?
Is it possible to answer the question? N

3. Jack was aware that, twice a day, the sheriff had stopped by to gather additional information about the burglary.

(How often\How many times a day) did Jack know whether the sheriff had come to inquire about the burglary?
Is it possible to answer the question? Y

4. Julie discerned that the survivor had managed to stay alive for eight days after the
crash in the harsh conditions.

(How long\For what period of time after the crash) did Julie observe whether the passenger had survived in the unbelievably harsh conditions?
Is it possible to answer the question? Y

5. Stephen was informed that the representative had gone missing for almost three hours prior to the committee meeting.

(How long\How many hours) did Stephen learn whether the representative had talked at the executive committee meeting?
Is it possible to answer the question? N

6. Meghan received information that the patient was working out five days a week without losing any weight.

(How long\How many days per week) did Meghan discover whether the patient had exercised without seeing any weight loss?
Is it possible to answer the question? Y

7. In her graduate research, Lily looked into the unusual phenomenon of a group of Democrats voting Republican for 3 or more consecutive elections.

(How long\For how many consecutive elections) did Lily ascertain whether the Democrats had voted Republican while writing her dissertation?
Is it possible to answer the question? Y

8. Mark claimed that his students had spent half of the quarter editing haikus instead of discussing modern poetry.

(How long\What part of the academic quarter in total) did Mark state whether his students had been reading through the novels he assigned?
Is it possible to answer the question? N
9. Mark claimed that his students had spent half of the quarter editing haikus instead of discussing modern poetry.

(How long\For how many months) did Hilary doubt whether some people would fish during the official fishing season?
Is it possible to answer the question? Y

10. Erica’s article determined that her school’s athletes had consumed the right amount of protein only once a week.

(How often\How many days per week) did Erica establish whether the athletes had eaten the recommended amount of protein?
Is it possible to answer the question? Y

11. Ethan concluded that one official had information about the modified bus routes for at least a few weeks.

(How long\For what length of time) did Ethan assess whether the official had known about the recent controversial firings?
Is it possible to answer the question? N

12. Zack informed the others that the guard had not been coming to work punctually on Sunday mornings during the last 6 months.

(How long\For how many months) did Zack report whether the guard had noticed some suspicious activity taking place?
Is it possible to answer the question? N

13. Kevin allowed one of his employees to spend two and a half weeks in the Bahamas for her outstanding performance.
(How long\For how many weeks) did Kevin decide whether the employee could go to vacation in the Bahamas?
Is it possible to answer the question? Y

14. Morgan mentioned that the customer had lingered for five hours over the decision to buy the computer.

(How long\For how many hours) did Morgan indicate whether the customer had got into an argument with her?
Is it possible to answer the question? N

15. Dana left her car with the mechanic for a week in July, before she left for a vacation on Cape Cod.

(How long\For how many days in the summer) did Dana speculate whether the mechanic had worked on fixing the broken engine?
Is it possible to answer the question? N

16. Leslie noted that it was only for the last three hours before daylight that the lioness prowled without making a sound.

(How long\For how many hours before daylight) did Leslie perceive whether the lioness had moved without making a single sound?
Is it possible to answer the question? Y

17. Renee did not believe that the reporter would be interested throughout the entire show, since it went on for so long.

(How long\For how much of the event) did Renee confirm whether the reporter would stay to speak with the celebrities?
Is it possible to answer the question? N

18. Alex confessed that, for at least the first week of the school year, the tenant was
happy with the living arrangement.

(How long\For how many weeks in the school year) did Alex say whether the tenant had felt happy with the living arrangement?
Is it possible to answer the question? Y

19. Andrew overheard the daycare staff discussing how they wanted to get away from the children for a few hours.

(How long\How many hours) did Andrew hear whether the children had played during the daycare’s afternoon recess?
Is it possible to answer the question? N

20. Bill couldn’t help but see that his officemate was in a meeting with the boss for several hours after the news of the merger.

(How long\How many hours) did Bill notice whether his colleague had argued about the wisdom of merging?
Is it possible to answer the question? N

21. Peter found in his old notes that the garden had blossomed for two weeks less than their normal three months because of a drought last year.

(How long\How many weeks less than normal) did Peter note whether the flowers had bloomed during last spring’s dry spell?
Is it possible to answer the question? Y

22. Jennifer’s father told her that in 1990 he took five years off to learn some new skills before starting a different career.

(How long\How many years) did Jennifer verify whether her father had traveled before returning to new work?
Is it possible to answer the question? N
23. One day, Mark was informed that his favorite painting at his great uncle’s house had been there for a hundred years.

(How long/How many years) did Mark determine whether the painting had hung at his great uncle’s house?
Is it possible to answer the question? Y

24. Jane knew that her nephew had been crashing at her parents for a couple of weeks, judging from the mess in the living room.

(How long/How many weeks) did Jane realize whether her nephew had stayed at her parents’ now-messy house?
Is it possible to answer the question? Y

A.4 Experiment IV Stimuli

1. The driver who was friends with Jill came in first in the 2005 NASCAR race but finished last the previous year.

Which NASCAR race did the driver that Jill knew win after losing the previous year?
“2005 NASCAR race” “2004 NASCAR race” “2003 NASCAR race”

2. The mathematician who was the husband of Carmen showed that, unlike design theory, matrix theory was flawed which pleased almost everybody.

Which math theory did the mathematician that Carmen married discredit to the delight of everyone?
“matrix theory” “design theory” “chaos theory”

3. Joshua helped construct the Sears Tower in Chicago and was a big fan of the architect of the DeWitt Tower in Chicago that a large accounting firm commissioned.
Which Chicago skyscraper did the architect that Joshua admired design for a large accounting firm?

“the Dewitt Tower” “the Sears Tower” “the Chrysler Building”

4. The actress who Sandra met in Venice traveled to the Mediterranean city of Palermo for some well-deserved rest and relaxation.

Which Mediterranean city did the actress that Sandra met visit for some rest and relaxation?

“Palermo” “Sicily” “Athens”

5. Samuel purchased a copy of the newspaper that printed the Dutch cartoon about Mohammad that instigated all the violent protests.

Which Dutch cartoon did the paper that Samuel bought publish thereby causing numerous violent protests?

“the cartoon about Mohammad” “the cartoon about Jesus” “the cartoon about Hitler”

6. The Spanish explorer that was brought up by Fernando on the island of New Guinea was the first European to land on the island of Tonga in the Pacific in the early spring of 1623.

Which Pacific island did the Spaniard that Fernando raised discover in the spring of 1623?

“Tonga” “Togo” “Tahiti”

7. Scott was in love with a celebrity that brought a lawsuit against the National Enquirer for printing unflattering photos.

Which tabloid magazine did the celebrity that Scott loved sue for printing some unflattering photos?

“the National Enquirer” “the National Empire” “the Daily Globe” “the National Enquirer”
8. Leah questioned the producer that came up with the NBC sitcom called The Golden Girls after first seeing the Broadway version.

Which NBC sitcom did the producer that Leah interviewed create after seeing the Broadway show?


9. Jack held the chemist who determined the natural plasticity of amber in high esteem for the discovery.

Which physical property did the chemist that Jack respected find to be inherent in amber?

“plasticity” “conductivity” “transparency”

10. The musician who Robert had a chance to see made the rock album Live on 22nd Street with two other popular blues guitarists.

Which rock album did the musician that Robert saw record with two popular blues guitarists?

Live on 22nd Street “22nd Street Blues” “Sabbath Bloody Sabbath”

11. The pope who followed Pius X II advocated the third Christian crusade fervently in the early 12th century.

Which Christian crusade did the pope that Pius X II succeeded support in the early 12th century?

“the third Crusade” “the second Crusade” “the fourth Crusade”

12. Erica was afraid of the judge who reversed the decision in the Lecter murder trial when new evidence was brought forth.

Which criminal case did the judge that Erica feared overturn because of the new evidence?

“the Lecter murder trial” “the Schachter murder trial” “the Lecter arson trial”
13. Scooter hid from the reporter who talked about the recent ABC political poll on a recent evening news segment.

Which political poll did the reporter that Scooter avoided discuss during an evening news segment?
“an ABC political poll” “a CNN political poll” “a NY Times political poll”

14. Natalie believed the critic who said the foreign film by Ingmar Bergman was boring and not worth seeing.

Which foreign film did the critic that trusted review and say it was horrible?
“the Ingmar Bergman film” “the Ingrid Bergman film” “the Werner Herzog film”

15. Logan had a contract to build a mind reader with the scientist that designed the polygraph machine to find out when people lie.

Which electronic device did the scientist that Logan employed invent to detect when people lie?
“the polygraph machine” “the time machine” “the Erasmus machine”

16. Nicole did not believe the plaintiff that denied knowing the killer after the prosecutor mentioned it yesterday.

Which incriminating fact did the plaintiff that Nicole doubted dispute after the prosecutor mentioned it?
“knowing the killer” “being the killer” “being at the crime scene”

17. Ethan frisked the terrorist transporting a weaponized bio-agent and discovered the small disguised canister.

Which dangerous weapon did the terrorist that Ethan searched hide unsuccessfully inside a small disguised canister?
“a weaponized bio-agent” “a pipe bomb” “a box cutter”

18. Brenda really upset the governor who shut down the Folsom state prison because of some critical safety issues.

Which state prison did the governor that Billy angered close due to critical safety issues?
“Folsom state prison” “Nevada state prison” “Joliet state prison “

19. Stephanie paid no attention to the geologist who foresaw the massive earthquake in Kobe several weeks before it struck Japan.

Which massive earthquake did the geologist that Stephanie ignored predict several weeks before its occurrence?
“the Kobe earthquake” “the Mumbai earthquake” “the New Madrid earthquake “

20. Emma helped the vandal who stole one Renaissance painting and ruined another Renaissance painting by Raphael for reasons that no one understands.

Which Renaissance painting did the vandal that Emma abetted deface for reasons that are unknown?
“the Raphael painting” “the Leonardo painting” “the Donatello sculpture”

A.5 Experiment V Stimuli

1. It was a (reclusive English) poet that the dignitaries at the ceremony awarded with a medal in Stockholm.

Who was given a medal at the ceremony in Stockholm?
“a poet” “a dignitary” “a novelist”

2. It was a (famous deaf) sculptor that the aristocrats at the gallery ridiculed during the exclusive art show.
Who was being laughed at during the gallery exhibition?
“a sculptor” “an aristocrat” “a peasant”

3. It was a(n) (unsuspecting, nervous) sophomore that the pranksters at the fair frightened by setting off loud firecrackers.

Who did some loud firecrackers scare at the fair?
“a sophomore” “a junior” “a softball”

4. It was a (dangerous Russian) mobster that the jurors in the trial imprisoned for thirty years without parole.

Who was put in jail for thirty years without parole?
“a mobster” “a banker” “a thug”

5. It was a (clueless hospice) nurse that the surgeons in the hospital accused of poisoning the elderly man.

Who did they blame for the old man being poisoned?
“a nurse” “a janitor” “a surgeon”

6. It was a (heartless professional) mercenary that the commanders at the base hired for the mission in Guatemala.

Who was enlisted for the mission in Guatemala by the commanders?
“a mercenary” “a marine” “a major”

7. It was a(n) (incompetent prison) guard that the warden of the prison blamed for the escape attempt yesterday.

Who did the escape attempt get blamed on?
“a guard” “a warden” “an inmate”
8. It was a (notorious, rich) Frenchman that the authorities in San Francisco identified as the smuggler of diamonds.

Who was smuggling diamonds, according to the authorities?
“a Frenchman” “a Dutchman” “an Irishman”

9. It was a (poor local) fisherman that the villagers on the beach saw on the stormy ocean seas.

Who did the villagers see out in the stormy seas?
“a fisherman” “a sailor” “a swimmer”

10. It was a (Texas cattle) rancher that the officials for the state subsidized throughout the worst drought periods.

Who was given assistance during the droughts?
“a rancher” “a farmer” “a cowboy”

11. It was a (crooked government) bureaucrat that the citizens of the county despised for always accepting illegal bribes.

Who did the citizens hate because of the acceptance of illegal bribes?
“a bureaucrat” “a Democrat” “a technocrat”

12. It was a(n) (alleged Venezuelan) communist that the members of the club banned from ever entering the premises.

Who was prohibited from entering the club by its members?
“a communist” “a socialist” “a capitalist”

13. It was a(n) (successful marketing) entrepreneur that the investors in the company invited to the banquet on Thursday.

Who was it that was invited to the Thursday banquet?
“an entrepreneur” “an executive” “a board member”

14. It was a(n) (fearless German) environmentalist that the activists for energy conservation congratulated for all the crucial accomplishments.

Who did they applaud for all that had been accomplished?
“an environmentalist” “an atheist” “a mentalist”

15. It was a(n) (generous oil) billionaire that the organizers of the campaign thanked for the largest ever contribution.

Who made the largest contribution ever?
“a billionaire” “a millionaire” “a debutante”

16. It was a (helpless, crying) child that the neighbors from next door pulled from the burning apartment building.

Who did they manage to retrieve from the building?
“a child” “a cat” “a carpet”

**A.6 Experiment VI & VII Stimuli**

1. The lieutenant could not remember which soldier the commander that was deeply respected ordered to scout the area ahead.
   Did the captain recall who was doing reconnaissance? N

2. Stephen could not identify which volunteer the terrorist that was widely feared captured during a humanitarian aid mission.
   Did Stephen negotiate with the terrorist? N

3. Harold figured out which student the author that was frequently quoted encouraged to try writing a novel.
   Did Harold teach the student to write? N
4. The foreman sensed which novice the architect that had recently died inspired to attempt a daring design.
Did the architect try for one last bold design before dying? N

5. The manager forgot which customer the salesman that hardly ever laughed conned into accepting the extended warranty.
Did the buyer forget which extended warranty they were sold? N

6. Naomi indicated which applicant the committee that was annoyingly rushed rejected without reading the personal statement.
Did Natalie say who was rejected? N

7. She recorded which trustee the chairman that was just replaced phoned to vehemently express his anger.
Was the chairman bothered about leaving the position? Y

8. He learned which toddler the pediatrician that was usually insightful diagnosed with a slight stomach flu.
Did the doctor think anything was wrong with the patient? Y

9. The activist determined which representative the tycoon that was without morals bribed to log the ancient rainforest.
Was the activist bribed so the forest could be logged? N

10. Heather admitted which scientist the dean that was recently instated reprimanded for plagiarizing some rare texts.
Did Heather acknowledge who behaved unethically? Y

11. He theorized which friend the host that was getting angry ignored for reasons nobody really knows.
Was the host refusing to pay attention to someone? Y
12. She kept secret which model the photographer that had been drinking molested to avoid a serious scandal.
Did she want to hide what the photographer did? Y

13. A jury decided which accomplice the kidnapper that was already sentenced corrupted too much to let go.
Was the kidnapper going to go to prison? Y

14. A journalist investigated which athlete the trainer that was quickly indicted injected with some illegal steroid substance.
Was the trainer providing someone with steroids? Y

15. A health inspector asked which janitor the chef that had two restaurants hired to clean on the weekends.
Did the chef own two different restaurants? Y

16. A bystander noticed which steelworker the girder that was dangerously dangling struck without doing any serious harm.
Could the bystander tell who got hit by the beam? Y

17. An insider realized which client the bank that was going bankrupt cheated out of eighty thousand dollars.
Was it the bank that robbed thousands of dollars from someone else? Y

18. A local official knew which survivor the rescuer that was severely injured saved from the burning plane wreckage.
Did the rescuer save the local official from the wreckage? N

A.7 Experiment VIII Stimuli

1. The reviewer criticized the (famous\blind young) actor who the director who was making the art film ignored during the opening night festivities.
Did the director complain about the actor? N
2. The official congratulated the (sensitive\childish English) poet who the professor who was trying to get tenure admired since reading the book reviews.
Did the official condemn the poet? N

3. The diplomat contacted the (ruthless\lovable military) dictator who the activist who was looking for more contributions encouraged to preserve natural habitats and resources.
Did the diplomat communicate with the activist? N

4. The reporter interviewed the (wealthy\bankrupt teen) celebrity who the photographer who was struggling to find work embarrassed last week at a charity dinner.
Did the photographer struggle to find the celebrity? N

5. The student defended the (brilliant\secretive schizophrenic) mathematician who the dean who was attempting to avoid controversy removed from the university ethics committee.
Did the dean take the student off the ethics committee? N

6. The defendant accused the (corrupt\scholarly narcotics) cop who the judge who was presiding in the case silenced after a disturbing courtroom outburst.
Was the cop silenced before the case went to court? N

7. The guard aided the (dangerous\responsible exiled) criminal who the agent who was overseeing the dramatic pursuit apprehended following a long and tiring chase.
Did the guard help the criminal? Y

8. The employee obeyed the (professional\immature plant) supervisor who the inspector who was reviewing the safety measures cautioned about the poor work conditions.
Did the employee follow the orders of the supervisor? Y

9. The captain evaluated the (brave\lonely volunteer) fireman who the veteran who was planning to retire soon trained over the course of six months.
Was it the performance of a fireman that was being assessed? Y
10. The investigator summoned the (injured\sickly rookie) athlete who the coach who was running out of ideas invited to try out for a spot on the team.  
Was the coach running out of ideas? Y

11. The pundit ridiculed the (Republican\demented senate) candidate who the leader who was making the final decision supported despite the misgivings of other members.  
Did the leader decide to support the candidate? Y

12. The interrogator questioned the (heartless\courteous former) mercenary who the commander who was organizing the armed rebellion abandoned without any explanation or warning.  
Did the commander explain the decision to leave the mercenary behind? N

13. The fugitive robbed the (lost\dead American) tourist who the guide who was showing the group around warned about straying too from the group.  
Did the guide steal from the tourist? N

14. The industrialist threatened the (poor\gifted Russian) peasant who the investor who was searching for new opportunities protected without the least bit of hesitation.  
Did the industrialist protect the peasant? N

15. The executive infuriated the (liberal\morbid socialist) politician who the lobbyist who was representing some oil companies bribed to vote for the upcoming bill.  
Did the politician infuriate the executive? N

16. The ranger followed the (fearless\girlish crocodile) hunter who the landowner who was building a new house despised for trespassing on his land.  
Was the landowner constructing a new dock? N

17. The advisor lectured the (handsome\insane blonde) prince who the dignitary who was learning to speak English thanked at the end of the ceremony.  
Did the dignitary snub the prince after the ceremony? N
18. The waitress married the (reclusive\vindictive musical) genius who the psychiatrist who was known for being insightful treated for a mild case of depression. Was the genius being treated for depression? Y

19. The pedestrian dodged the angry\sullen taxi driver who the bystander who was waiting for a bus identified later on at the police station. Did the pedestrian avoid being hit? Y

20. The customer offended the (helpful\senile new) assistant who the manager who was hoping for a raise called into his office after the incident. Was the new assistant hurt by what the customer said? Y

21. The nurse consoled the (dying\hostile elderly) patient who the doctor who was working a double shift forgot due to a lack of sleep. Did the nurse try and make the elderly patient feel better? Y

22. The representative consulted the (conservative\conservative) legal strategist who the prosecutor who was looking into corporate fraud knew since they went to school together. Was the prosecuting involved in a fraud investigation? Y

23. The scientist avoided the (impartial\enraged academic) observer who the technician who was assisting the busy researchers escorted around the recently built facilities. Did the technician show the observer around? Y

24. The secretary aggravated the (young\old female) intern who the partner who was negotiating a huge settlement hired less than three weeks ago. Was the intern hired in the last month? Y

A.8 Experiment IX Stimuli

1. The advisor cornered the paranoid dictator of East Timor at the royal palace. Was an advisor at the enemy palace? N
2. The demonstrator assaulted the annoying reporter from Washington DC after the anti-war rally.
Was there a riot before the rally? N

3. The rebel obeyed the violent commander of the Colombians without any hesitation.
Did a rebel lose his nerve? N

4. The spy strangled the dishonest ambassador from South Africa without being seen.
Did the spy manage to stay hidden? Y

5. The traitor begged the masked assassin wielding an axe for one more chance.
Did a traitor want to die? N

6. The priest reprimanded the immoral follower of the cult before the baptismal ceremony.
Was there a baptism about to happen? Y

7. The representative congratulated the ambitious executive in the oil industry after the promotion.
Did the representative get promoted? N

8. The model slapped the insensitive photographer for the tabloid due to constant harrassment.
Was the model bothered by all the attention? Y

9. The playwright coached the stammering actress with stage fright on how to deliver the lines.
Was the director making disparaging comments? N

10. The saxophonist kissed the lonely drummer from Def Leppard following the long interview.
Did the interview last for quite a while? Y
11. The historian angered the defensive legislator from New Orleans by questioning the report on the evacuation. Did the historian take issue with what was disclosed about the evacuation? Y

12. The auditor called the scrupulous attorney for the government and complained about the missing documents. Was the auditor unhappy with the situation? Y

13. The inventor welcomed the famous researcher from Harvard University into the genetics laboratory. Did the inventor gain access to the propulsion laboratory? N

14. The mechanic criticized the inexperienced designer of the vehicle for putting the gearbox near the fan blades. Did the mechanic put the gearbox next to the fan blades? N

15. The shopkeeper insulted the affluent customer with outrageous demands without considering the consequences. Was the shopkeeper acting without thinking? Y

16. The bartender ignored the drunken manager of La Bohême until the bachelor party finished. Was the bartender getting married? N

17. The waitress tripped the indiscreet busboy with no manners after seeing the note. Did the waitress have a chance to learn the contents of the note? Y

18. The coroner studied the unidentified victim with neck wounds in the county morgue. Was the coroner at the crime scene? N

19. The driver struck the blind pedestrian crossing the street near the busy intersection. Was the immediate area heavily trafficked? Y
20. The guard consoled the innocent inmate with severe depression after the parole hearing. Was the parole hearing going to happen soon? N

21. The mediator quieted the greedy plaintiff with numerous complaints during the meeting. Was there a mediator present at the meeting? Y

22. The writer contacted the lone survivor from the crash after two months elapsed. Did the writer wait until 60 days had gone by? Y

23. The aviator radioed the inattentive operator without enough sleep at the base. Was the co-pilot at the base? N

24. The rancher warned the trembling trespasser without any shoes to leave the property immediately. Did the rancher need to escape as soon as possible? N
Appendix B

MinF’ Values & Bonferroni Corrections

<table>
<thead>
<tr>
<th>Region</th>
<th>df</th>
<th>$F'$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complementizer</td>
<td>(1,53)</td>
<td>7.7978</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Embedded Subject</td>
<td>(1,43)</td>
<td>3.541</td>
<td>.067</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>(1,58)</td>
<td>5.473</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Verb</td>
<td>(1,59)</td>
<td>3.002</td>
<td>.088</td>
</tr>
<tr>
<td>Two-Word Spillover Region</td>
<td>(1,58)</td>
<td>5.137</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

Table B.1: MinF’ values by region in experiment I

<table>
<thead>
<tr>
<th>Region</th>
<th>df</th>
<th>$F'$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preceding Word</td>
<td>(1,38)</td>
<td>1.14</td>
<td>.292</td>
</tr>
<tr>
<td>Verb plus next word</td>
<td>(1,39)</td>
<td>2.722</td>
<td>.107</td>
</tr>
<tr>
<td>Verb + 3-word spillover region</td>
<td>(1,41)</td>
<td>6.603</td>
<td>.014</td>
</tr>
</tbody>
</table>

Table B.2: MinF’ values by region inside embedded clause in experiment II for contrast between WHICH and BARE conditions
APPENDIX B. MINF’ VALUES & BONFERRONI CORRECTIONS

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>$df_1$</th>
<th>$t_1$</th>
<th>$p$</th>
<th>$df_2$</th>
<th>$t_2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preceding Word</td>
<td>BARE-BASELINE</td>
<td>19</td>
<td>2.491</td>
<td>.051</td>
<td>23</td>
<td>2.969</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>BARE-WHICH</td>
<td>19</td>
<td>1.366</td>
<td>.387</td>
<td>23</td>
<td>1.716</td>
<td>.117</td>
</tr>
<tr>
<td></td>
<td>BASELINE-WHICH</td>
<td>19</td>
<td>-1.140</td>
<td>.806</td>
<td>23</td>
<td>-.940</td>
<td>1.00</td>
</tr>
<tr>
<td>Verb + Next</td>
<td>BARE-BASELINE</td>
<td>19</td>
<td>2.212</td>
<td>.118</td>
<td>23</td>
<td>2.127</td>
<td>.133</td>
</tr>
<tr>
<td></td>
<td>BARE-WHICH</td>
<td>19</td>
<td>2.127</td>
<td>.140</td>
<td>23</td>
<td>2.615</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>BASELINE-WHICH</td>
<td>19</td>
<td>-.587</td>
<td>1.00</td>
<td>23</td>
<td>-.378</td>
<td>1.00</td>
</tr>
<tr>
<td>Verb + 3-word Spillover</td>
<td>BARE-BASELINE</td>
<td>19</td>
<td>2.367</td>
<td>.086</td>
<td>23</td>
<td>2.356</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td>BARE-WHICH</td>
<td>19</td>
<td>3.600</td>
<td>.006</td>
<td>23</td>
<td>3.669</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>BASELINE-WHICH</td>
<td>19</td>
<td>.190</td>
<td>1.00</td>
<td>23</td>
<td>.231</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table B.3: Bonferroni corrected values by region in experiment II

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>$df_1$</th>
<th>$t_1$</th>
<th>$p$</th>
<th>$df_2$</th>
<th>$t_2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj. Determiner</td>
<td>BARE-BASELINE</td>
<td>27</td>
<td>5.017</td>
<td>&lt;.001</td>
<td>23</td>
<td>4.198</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>BARE-LONG</td>
<td>27</td>
<td>3.484</td>
<td>.005</td>
<td>23</td>
<td>3.513</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>BASELINE-LONG</td>
<td>27</td>
<td>-.733</td>
<td>1.00</td>
<td>23</td>
<td>-.719</td>
<td>1.00</td>
</tr>
<tr>
<td>Entire Clause</td>
<td>BARE-BASELINE</td>
<td>27</td>
<td>3.218</td>
<td>.01</td>
<td>23</td>
<td>2.518</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>BARE-LONG</td>
<td>27</td>
<td>2.356</td>
<td>.078</td>
<td>23</td>
<td>1.648</td>
<td>.339</td>
</tr>
<tr>
<td></td>
<td>BASELINE-LONG</td>
<td>27</td>
<td>-.667</td>
<td>1.00</td>
<td>23</td>
<td>-.550</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table B.4: Bonferroni corrected values by region inside embedded clause in experiment III

<table>
<thead>
<tr>
<th>Region</th>
<th>$df$</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Determiner</td>
<td>(1,50)</td>
<td>6.118</td>
<td>.0169</td>
</tr>
<tr>
<td>Entire Clause</td>
<td>(1,42)</td>
<td>1.824</td>
<td>.184</td>
</tr>
</tbody>
</table>

Table B.5: MinF’ values by region inside embedded clause in experiment III
### Table B.6: Bonferroni corrected values by region in experiment IV

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>( t_1 )</th>
<th>( p )</th>
<th>df</th>
<th>( t_2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>COMPLEX-SIMPLE</td>
<td>34</td>
<td>-2.372</td>
<td>.070</td>
<td>19</td>
<td>-1.974</td>
<td>.189</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-WHICH</td>
<td>34</td>
<td>-0.616</td>
<td>1.00</td>
<td>19</td>
<td>-0.709</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SIMPLE-WHICH</td>
<td>34</td>
<td>1.506</td>
<td>0.424</td>
<td>19</td>
<td>1.743</td>
<td>0.293</td>
</tr>
<tr>
<td>V2</td>
<td>COMPLEX-SIMPLE</td>
<td>34</td>
<td>-2.305</td>
<td>0.082</td>
<td>19</td>
<td>-2.576</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-WHICH</td>
<td>34</td>
<td>-0.085</td>
<td>1.00</td>
<td>19</td>
<td>0.113</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SIMPLE-WHICH</td>
<td>34</td>
<td>3.078</td>
<td>0.012</td>
<td>19</td>
<td>3.224</td>
<td>0.013</td>
</tr>
<tr>
<td>Spillover</td>
<td>COMPLEX-SIMPLE</td>
<td>34</td>
<td>-3.459</td>
<td>0.004</td>
<td>19</td>
<td>-2.387</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-WHICH</td>
<td>34</td>
<td>-0.276</td>
<td>1.00</td>
<td>19</td>
<td>-0.264</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SIMPLE-WHICH</td>
<td>34</td>
<td>2.763</td>
<td>0.028</td>
<td>19</td>
<td>3.472</td>
<td>0.008</td>
</tr>
<tr>
<td>Verb + Spillover</td>
<td>COMPLEX-SIMPLE</td>
<td>34</td>
<td>-3.569</td>
<td>0.003</td>
<td>19</td>
<td>-2.616</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-WHICH</td>
<td>34</td>
<td>-0.253</td>
<td>1.00</td>
<td>19</td>
<td>-0.085</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SIMPLE-WHICH</td>
<td>34</td>
<td>3.584</td>
<td>0.003</td>
<td>19</td>
<td>3.934</td>
<td>0.003</td>
</tr>
</tbody>
</table>

### Table B.7: MinF’ values by region in experiment IV

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>( F )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Verb</td>
<td>WHICH-SIMPLE</td>
<td>(1,52)</td>
<td>1.298</td>
<td>.260</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-SIMPLE</td>
<td>(1,43)</td>
<td>2.303</td>
<td>.137</td>
</tr>
<tr>
<td>Subcategorizing Verb</td>
<td>WHICH-SIMPLE</td>
<td>(1,51)</td>
<td>2.951</td>
<td>.092</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-SIMPLE</td>
<td>(1,50)</td>
<td>4.956</td>
<td>.031</td>
</tr>
<tr>
<td>Next Word</td>
<td>WHICH-SIMPLE</td>
<td>(1,53)</td>
<td>4.674</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-SIMPLE</td>
<td>(1,37)</td>
<td>3.860</td>
<td>.057</td>
</tr>
<tr>
<td>Verb + Next Word</td>
<td>WHICH-SIMPLE</td>
<td>(1,51)</td>
<td>7.02</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-SIMPLE</td>
<td>(1,39)</td>
<td>4.45</td>
<td>.041</td>
</tr>
</tbody>
</table>
### Table B.8: MinF' values at retrieval site in experiment V

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>COMPLEX-SIMPLE</td>
<td>(1,46)</td>
<td>2.385</td>
<td>.130</td>
</tr>
</tbody>
</table>

### Table B.9: Bonferroni corrected values at retrieval site in experiment V

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>t1</th>
<th>p</th>
<th>df2</th>
<th>t2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>COMPLEX-MID</td>
<td>34</td>
<td>-.546</td>
<td>1.00</td>
<td>15</td>
<td>-.081</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>COMPLEX-SIMPLE</td>
<td>34</td>
<td>-.2034</td>
<td>.150</td>
<td>15</td>
<td>-2.373</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td>MID-SIMPLE</td>
<td>34</td>
<td>-1.781</td>
<td>.252</td>
<td>15</td>
<td>-1.690</td>
<td>.335</td>
</tr>
</tbody>
</table>

### Table B.10: MinF' values by region inside embedded clause in experiment VI

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>PERSON-TYPE</td>
<td>(1,36)</td>
<td>2.410</td>
<td>.130</td>
</tr>
<tr>
<td>Next</td>
<td>PERSON-TYPE</td>
<td>(1,41)</td>
<td>1.875</td>
<td>.178</td>
</tr>
<tr>
<td>Verb + Next</td>
<td>PERSON-TYPE</td>
<td>(1,40)</td>
<td>3.147</td>
<td>.084</td>
</tr>
</tbody>
</table>
### Table B.11: Bonferroni corrected values by region inside embedded clause in experiment VII

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df₁</th>
<th>t₁</th>
<th>p</th>
<th>df₂</th>
<th>t₂</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reg. 8</td>
<td>BARE-PERSON</td>
<td>34</td>
<td>2.673</td>
<td>.034</td>
<td>17</td>
<td>2.433</td>
<td>.079</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>34</td>
<td>1.233</td>
<td>.678</td>
<td>17</td>
<td>1.675</td>
<td>.337</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>34</td>
<td>-1.341</td>
<td>.566</td>
<td>17</td>
<td>-1.198</td>
<td>.742</td>
</tr>
<tr>
<td>Reg. 9</td>
<td>BARE-PERSON</td>
<td>34</td>
<td>2.661</td>
<td>.035</td>
<td>17</td>
<td>2.025</td>
<td>.177</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>34</td>
<td>4.314</td>
<td>&lt; .001</td>
<td>17</td>
<td>3.350</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>34</td>
<td>1.960</td>
<td>.175</td>
<td>17</td>
<td>1.087</td>
<td>.877</td>
</tr>
<tr>
<td>Reg. 10</td>
<td>BARE-PERSON</td>
<td>34</td>
<td>1.727</td>
<td>.280</td>
<td>17</td>
<td>2.323</td>
<td>.099</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>34</td>
<td>2.622</td>
<td>.039</td>
<td>17</td>
<td>2.503</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>34</td>
<td>1.22</td>
<td>.690</td>
<td>17</td>
<td>.997</td>
<td>.998</td>
</tr>
<tr>
<td>Reg. 11</td>
<td>BARE-PERSON</td>
<td>34</td>
<td>1.627</td>
<td>.339</td>
<td>17</td>
<td>2.040</td>
<td>.172</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>34</td>
<td>2.884</td>
<td>.020</td>
<td>17</td>
<td>1.912</td>
<td>.219</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>34</td>
<td>1.269</td>
<td>.639</td>
<td>17</td>
<td>1.094</td>
<td>.868</td>
</tr>
<tr>
<td>Regs. 8-11</td>
<td>BARE-PERSON</td>
<td>34</td>
<td>4.288</td>
<td>&lt; .001</td>
<td>17</td>
<td>3.146</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>34</td>
<td>4.636</td>
<td>&lt; .001</td>
<td>17</td>
<td>3.348</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>34</td>
<td>1.005</td>
<td>.967</td>
<td>17</td>
<td>.867</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Table B.12: MinF’ values by region inside embedded clause in experiment VII

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reg. 8</td>
<td>BARE-PERSON</td>
<td>(1,42)</td>
<td>3.237</td>
<td>.079</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>(1,51)</td>
<td>.986</td>
<td>.325</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>(1,42)</td>
<td>.798</td>
<td>.377</td>
</tr>
<tr>
<td>Reg. 9</td>
<td>BARE-PERSON</td>
<td>(1,36)</td>
<td>2.595</td>
<td>.116</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>(1,37)</td>
<td>6.961</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>(1,28)</td>
<td>.903</td>
<td>.350</td>
</tr>
<tr>
<td>Reg. 10</td>
<td>BARE-PERSON</td>
<td>(1,51)</td>
<td>1.920</td>
<td>.172</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>(1,44)</td>
<td>3.278</td>
<td>.077</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>(1,39)</td>
<td>.596</td>
<td>.445</td>
</tr>
<tr>
<td>Reg. 11</td>
<td>BARE-PERSON</td>
<td>(1,50)</td>
<td>1.617</td>
<td>.209</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>(1,32)</td>
<td>2.538</td>
<td>.121</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>(1,40)</td>
<td>.686</td>
<td>.412</td>
</tr>
<tr>
<td>Regs 8-11</td>
<td>BARE-PERSON</td>
<td>(1,35)</td>
<td>6.434</td>
<td>.016</td>
</tr>
<tr>
<td></td>
<td>BARE-TYPE</td>
<td>(1,35)</td>
<td>7.367</td>
<td>.010</td>
</tr>
<tr>
<td></td>
<td>PERSON-TYPE</td>
<td>(1,41)</td>
<td>.431</td>
<td>.515</td>
</tr>
</tbody>
</table>

Table B.12: MinF’ values by region inside embedded clause in experiment VII

### Table B.13: MinF’ values by region inside embedded clause in experiment VIII

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Verb</td>
<td>ATYP-BARE</td>
<td>(1,59)</td>
<td>1.95</td>
<td>.168</td>
</tr>
<tr>
<td></td>
<td>TYP-BARE</td>
<td>(1,58)</td>
<td>4.18</td>
<td>.046</td>
</tr>
</tbody>
</table>

Table B.13: MinF’ values by region inside embedded clause in experiment VIII
**APPENDIX B. MINF’ VALUES & BONFERRONI CORRECTIONS**

<table>
<thead>
<tr>
<th>Region</th>
<th>Contrast</th>
<th>df₁</th>
<th>t₁</th>
<th>p</th>
<th>df₂</th>
<th>t₂</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regs 1 + 2</td>
<td>ATYP-BARE</td>
<td>36</td>
<td>3.291</td>
<td>.007</td>
<td>23</td>
<td>2.929</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>ATYP-TYP</td>
<td>36</td>
<td>2.219</td>
<td>.099</td>
<td>23</td>
<td>1.577</td>
<td>.385</td>
</tr>
<tr>
<td></td>
<td>BARE-TYP</td>
<td>36</td>
<td>-1.490</td>
<td>.435</td>
<td>23</td>
<td>-1.643</td>
<td>.342</td>
</tr>
<tr>
<td>Verb</td>
<td>ATYP-BARE</td>
<td>36</td>
<td>-1.812</td>
<td>.235</td>
<td>23</td>
<td>-2.191</td>
<td>.116</td>
</tr>
<tr>
<td></td>
<td>ATYP-TYP</td>
<td>36</td>
<td>.767</td>
<td>1.00</td>
<td>23</td>
<td>.455</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>BARE-TYP</td>
<td>36</td>
<td>2.752</td>
<td>.028</td>
<td>23</td>
<td>3.055</td>
<td>.017</td>
</tr>
<tr>
<td>Next</td>
<td>ATYP-BARE</td>
<td>36</td>
<td>1.186</td>
<td>.730</td>
<td>23</td>
<td>.974</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>ATYP-TYP</td>
<td>36</td>
<td>1.132</td>
<td>.796</td>
<td>23</td>
<td>1.267</td>
<td>.653</td>
</tr>
<tr>
<td></td>
<td>BARE-TYP</td>
<td>36</td>
<td>-.168</td>
<td>1.00</td>
<td>23</td>
<td>.352</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table B.14: Bonferroni corrected values by region inside embedded clause in experiment VIII
Appendix C

Linear mixed-effects models

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>52714</td>
<td>-26351</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>52697</td>
<td>-26341</td>
<td>21.021</td>
<td>2</td>
<td>2.725e-05</td>
</tr>
</tbody>
</table>

Table C.1: Model comparison of reading times at first six words of embedded clause in experiment I. Model A: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + \beta_3\text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + (\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$

| Comparison          | Estimate | Std.Error | t-score | Pr(>|t|) |
|---------------------|----------|-----------|---------|----------|
| BARE vs. BASE       | -18.0780 | 6.908     | -2.617  | .0089    |
| BARE vs. WHICH      | -29.2699 | 4.895     | -5.980  | < 0.0001 |

Table C.2: LME analysis of reading times for first six words of embedded clause in experiment I

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>18265.8</td>
<td>-9126.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>18199.6</td>
<td>-9091.8</td>
<td>70.192</td>
<td>2</td>
<td>5.728e-16</td>
</tr>
</tbody>
</table>

Table C.3: Model comparison of reading times at verb + three-word spillover region in experiment II. Model A: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + \beta_3\text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1\text{BASELINE}_i + \beta_2\text{WHICH}_i + (\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$
Comparison | Estimate | Std.Error | t – score | Pr(>|t|)
---|---|---|---|---
BARE vs. BASE | -14.0696 | 3.5899 | -3.919 | 0.0001
BARE vs. WHICH | -13.0188 | 2.9612 | -4.397 | < 0.0001

Table C.4: LME analysis of reading times at verb + spillover region in experiment II

Comparison | Estimate | Std.Error | t – score | Pr(>|t|)
---|---|---|---|---
BARE vs. BASE | -50.583 | 9.096 | -5.561 | 0.0001
BARE vs. LONG | -38.807 | 9.127 | -4.252 | 0.0001

Table C.6: LME analysis of reading times at first word in embedded clause in experiment III

Model | Df | AIC | logLik | Chisq | Df | Pr(> Chisq)
---|---|---|---|---|---|---
A | 6 | 6759.7 | -3373.8 | | | |
B | 8 | 6731.0 | -3357.5 | 32.736 | 2 | 7.79e-08 |

Table C.5: Model comparison of reading times at first word in embedded clause in experiment III. Model A: \( y_{ij} = \mu + b_i + \beta_1 \text{BASELINE}_i + \beta_2 \text{LONG}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij} \); Model B: \( y_{ij} = \mu + b_i + \beta_1 \text{BASELINE}_i + \beta_2 \text{LONG}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij} \)

Comparison | Estimate | Std.Error | t – score | Pr(>|t|)
---|---|---|---|---
BARE vs. BASE | -28.964 | 5.756 | -5.03 | < 0.0001
WHICH vs. SIMPLE | -28.052 | 5.739 | -4.89 | < 0.0001

Table C.8: LME analysis of reading times at verb + next word in experiment IV
### APPENDIX C. LINEAR MIXED-EFFECTS MODELS

<table>
<thead>
<tr>
<th>Model</th>
<th>$D_f$</th>
<th>AIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>$D_f$</th>
<th>$Pr(&gt; \text{Chisq})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>6457.8</td>
<td>-3222.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>6461.3</td>
<td>-3222.7</td>
<td>4.642</td>
<td>2</td>
<td>0.7929</td>
</tr>
</tbody>
</table>

Table C.9: Model comparison of reading times at retrieval site in experiment V. Model A: $y_{ij} = \mu + b_i + \beta_1 \text{mid}_i + \beta_2 \text{simple}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1 \text{mid}_i + \beta_2 \text{simple}_i(\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$

| Comparison       | Estimate | Std.Error | $t$ - score | $Pr(>|t|)$ |
|------------------|----------|-----------|-------------|-----------|
| complex vs. simple | -33.569 | 13.52 | -2.483 | 0.0134 |
| complex vs. mid   | -5.459   | 13.54 | .0403  | 0.6870 |

Table C.10: LME analysis of reading times at verb in experiment V

<table>
<thead>
<tr>
<th>Model</th>
<th>$D_f$</th>
<th>AIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>$D_f$</th>
<th>$Pr(&gt; \text{Chisq})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>9120.8</td>
<td>-4555.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>7</td>
<td>9124.8</td>
<td>-4555.4</td>
<td>9e-04</td>
<td>2</td>
<td>.996</td>
</tr>
</tbody>
</table>

Table C.11: Model comparison of reading times at retrieval site in experiment VI. Model A: $y_{ij} = \mu + b_i + \beta_1 \text{type}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1 \text{type}_i + (\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$

| Comparison     | Estimate | Std.Error | $t$ - score | $Pr(>|t|)$ |
|----------------|----------|-----------|-------------|-----------|
| person vs. type | -31.500  | 9.057 | -3.478 | 0.0005 |

Table C.12: LME analysis of reading times at verb + next word in experiment VI

<table>
<thead>
<tr>
<th>Model</th>
<th>$D_f$</th>
<th>AIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>$D_f$</th>
<th>$Pr(&gt; \text{Chisq})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>25033</td>
<td>-12511</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>25030</td>
<td>-12507</td>
<td>6.8553</td>
<td>2</td>
<td>.0325</td>
</tr>
</tbody>
</table>

Table C.13: Model comparison of reading times in regions 8 through 11 in experiment VII. Model A: $y_{ij} = \mu + b_i + \beta_1 \text{type}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1 \text{type}_i + (\beta_3 + \zeta_i)\text{Sequence}_i + \epsilon_{ij}$
### Table C.14: LME model of reading times at regions 8-11 in experiment VII

| Comparison         | Estimate | Std.Error | t - score | Pr(>|t|) |
|--------------------|----------|-----------|-----------|----------|
| BARE vs. PERSON    | -30.0512 | 7.255     | -4.142    | .0001    |
| BARE vs. TYPE      | -37.982  | 7.332     | -5.180    | .0001    |

### Table C.15: LME analysis of reading times at regions 9-11 in experiment VII

| Comparison         | Estimate | Std.Error | t - score | Pr(>|t|) |
|--------------------|----------|-----------|-----------|----------|
| BARE vs. TYPE      | 40.4207  | 6.9055    | 5.853     | .0001    |
| PERSON vs. TYPE    | 17.8913  | 6.8683    | 2.605     | .0106    |

### Table C.16: Model comparison of reading times at retrieval site in experiment VIII. Model A: $y_{ij} = \mu + b_i + \beta_1 \text{ATYP}_i + \beta_2 \text{TYPO}_i + \beta_3 \text{Sequence}_{ij} + \epsilon_{ij}$; Model B: $y_{ij} = \mu + b_i + \beta_1 \text{ATYP}_i + \beta_2 \text{TYPO}_i + (\beta_3 + \zeta_i) \text{Sequence}_i + \epsilon_{ij}$

| Model | Df | AIC  | logLik | Chisq  | Df | Pr(>|Chisq|) |
|-------|----|------|--------|--------|----|-----------|
| A     | 6  | 8458.3 | -4223.1 |        |    |           |
| B     | 8  | 8439.6 | -4211.8 | 22.641 | 2  | 1.212e-05 |

### Table C.17: LME model of reading times at verb in Experiment VIII

| Comparison         | Estimate | Std.Error | t - score | Pr(>|t|) |
|--------------------|----------|-----------|-----------|----------|
| ATYP vs. BARE      | -30.8787 | 12.7489   | -2.422    | 0.0157   |
| TYP vs. BARE       | -43.6261 | 12.7067   | -3.433    | 0.0006   |

### Table C.18: LME analysis of reading times at first word after verb in experiment VIII

| Comparison         | Estimate | Std.Error | t - score | Pr(>|t|) |
|--------------------|----------|-----------|-----------|----------|
| BARE vs. ATYP      | -19.349  | 10.859    | -1.782    | 0.075    |
| TYP vs. ATYP       | -17.487  | 10.769    | -1.624    | 0.105    |
| Comparison    | Estimate | Std.Error | t - score | Pr(>|t|) |
|---------------|----------|-----------|-----------|---------|
| ATYP vs. TYP  | 19.317   | 8.827     | 2.188     | 0.0288  |
| BARE vs. TYP  | 22.895   | 8.857     | 2.585     | 0.0098  |

Table C.19: LME analysis of reading times at verb + next word in experiment VIII
References


Allwood, J. (1976). The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. In J. Stillings (Ed.), University of Massachusetts Occasional Papers in Linguistics, Volume II. Amherst, MA: University of Massachusetts at Amherst.


References

*International Congress of Linguists, 914–978.*


Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some


References


Kluender, R., & Kutas, M. (1993a). Bridging the gap: evidence from erps on the processing
References


References

computing. New York: Springer.
Rizzi, L. (2000). Reconstruction, Weak Island Sensitivity, and Agreement. (Ms. Università di Siena)


References


