An Expectation Model of Referring Expressions

by

John Kræmer

B.A., B.S., University of Texas at Austin (2000)

Submitted to the Department of Brain and Cognitive Science
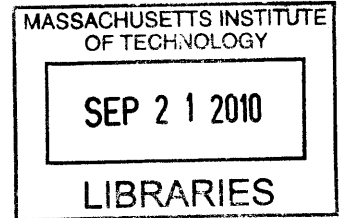in partial fulfillment of the requirements for the degree of
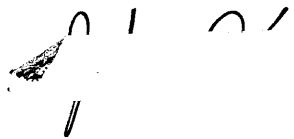
Doctor of Philosophy

at the

MASSACHVSETTS INSTITVTE OF TECHNOLOGY

September 2010
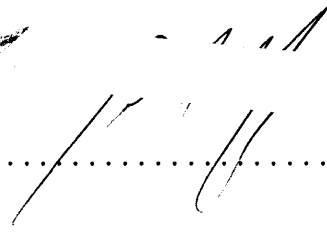
Author ....
Department of Brain and Cognitive Science
Aug 19, 2010

Certified by....................................
Edward Gibson
Professor of Cognitive Science
Thesis Supervisor

Accepted by ....
Earl Miller
Picower Professor of Neuroscience
Director, Brain and Cognitive Science Graduate Program

An Expectation Model of Referring Expressions

by

John Kræmer

Submitted to the Department of Brain and Cognitive Science
on Aug 19, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis introduces EMRE, an expectation-based model of referring expressions. EMRE is proposed as a model of non-syntactic dependencies — in particular, discourse-level semantic dependencies that bridge sentence gaps. These include but are not limited to anaphora (references to noun phrases in previous sentences) and coherence predicates such as causality, temporal ordering and resemblance — two domains that have typically been treated as entirely distinct aspects of language.

EMRE is a computational-level model, and is agnostic about any particular algorithms, cognitive faculties, or neurological substrates that might be applied to the problem of semantic reference. Instead, it describes reference as a computational problem framed in terms of expectation and inference, and describes a solution to the problem based on rational top-down expectations about the likely targets of referring expressions, and on bottom-up feature-based matching that occurs when a referring expression is encountered.

EMRE is used to derive novel empirical predictions about how people will construe particular discourse constructions involving NP anaphora and coherence predicates. These predictions are tested in controlled behavioral experiments, in which participants read and answer questions about short texts.

The results of these experiments are shown to be consistent with a model of reference as an expectation-based computational structure with different underlying rules than those governing syntactic processing.

Thesis Supervisor: Edward Gibson
Title: Professor of Cognitive Science

Acknowledgments

My first thanks go out to my advisor, Ted Gibson. Graduate school is a rough business, but Ted is one of the four people who have taught me the most in my life thus far. Thanks also to my committee and all the other BCS and course 6 professors who have taught and encouraged me. I have also learned a great deal from my labmates, Mara Breen, Ev Fedorenko, Mike Frank, Steve Piantadosi, Melissa Troyer, Melissa Klein and Leon Bergen. Special thanks go out to Steve, for always being willing to help me work through my weird math problems, Liss for being a late-night sounding board on many occasions when she needed to be working, Leon for nerdy conversations that helped me maintain my mood through this final push, and Mara, for being a friend.

Thanks to all the rest of the BCS crew. Michelle, Kate, Barbara, Ethan, Ethan, Chris, Dominique, David, and many others — my time here would have been poorer without you. Thanks also to Vikash Mansinghka and Ed Vul, for being such huge smarty-pants.

My thanks to CSAIL, and MIT as a whole for teaching me so much about things I never suspected, about how to think like a mad thinking thing, and most of all, about humility. Thanks also to Donald Knuth, Richard Stallman, Guido van Rossum, Guy Steele and Gerry Sussman, and everyone else who wrote all the software I used to figure all this stuff out and write it up. And thanks to Noah Goodman, whose Church program I used for the generative modeling work.

Thanks to my local community in Somerville, who always seemed to believe in me, and special thanks to Rebecca Rosenblum, who was always there with a keen observation or a well-needed kick in the ass.

Thanks to my east coast crew. Emily, Carolyn, Sierra, Aatish, Denise, JP, Jesse, Marta, Martin and Martin, and many, many more.

And thanks to my crew from back home. Thanks to Steve, Greg, Brandon, Rachel and John for almost literally always being there to talk to. Shout outs to my bro Mike, and to Andrew. And, Michael, Alex, Christine — we pushed each other into

this, and I'm the last one out.

Many, many thanks to my amanuensis.

A million of thanks to my partner Sarah, whose patience through this process has been unending.

And the three other people who have taught me the most:

Mom, Dad, I don't have the words. Thank you.

And the person who was my greatest single motivation to get this thing done — to my daughter Bridget, thank you.

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

The problem of discourse

Language is often described as having a hierarchical structure. Each of the levels in this hierarchy (Table 1.1) corresponds to a different type of latent structure that is used in language production (*i.e.*, speech) to *generate* the lower-level structures in the layers below, and in language comprehension (*i.e.*, listening and understanding) to *recover* the higher-level generating structures in the levels above.

Syntactic
Lexical
Morphological
Phonological
Acoustic

Table 1.1: Hierarchical structure of language

Advancement in understanding of the hierarchical structure of language has been facilitated by a precise lexicon for discussing these latent structures (*e.g.*, *'formants'* in phonology, *'dependencies'* in syntax) in a way that separates out the structures themeselves from issues of how they are concretely stored and recovered in memory, or how they are algorithmically applied in production and recovered in comprehension. This separation between abstract representation and cognitive behavior was described by Marr (1982) as the difference between *'level 1'*, or *'computational-level'* models, and *'level 2'* or *'algorithmic-level'* models.

Computational-level explanations can be used to investigate the underlying struc-

ture of a problem that the brain is solving, which can be a first step in determining how the problem might be solved in practice (Marr, 1982). While there are areas of contention about how to most effectively represent language's latent hierarchical structure (Level 1), and much remains unknown about how the structures are generated by speakers and recovered by listeners (Level 2), this distinction between computational- and algorithmic-level questions has been useful in language research, as it has in other areas of cognitive science.

Hierarchical language structures, however, do not represent all of the important dependencies in language. The *syntactic* level, in which words are bound together into phrases, clauses, and finally, sentences, is the highest level in the standard hierarchy — above the sentence level, the rules of syntax no longer apply. But sentences don't typically appear in isolation. To the contrary, coherent text often relies heavily on complex semantic dependencies and pragmatic relationships between sentences.

These dependencies include noun phrase (NP) anaphora such as *'she'* and *'them'* (Example 1) which refer to phrasal *antecedents* — phrases that introduce or mention people, objects, events, &c., and without which an NP anaphor cannot be interepreted. Non-syntactic dependencies also include coherence cue phrases such as *'consequently'* and *'before long'* (Example 2), which are similarly uninterpretable without clausal antecedents describing the cause or precedent of whatever is described in in the clause attached to the cue phrase. There is currently no broadly-acccepted simple computational (*i.e.*, Marr level 1) model covering all such non-syntactic dependencies.

(1)   Joe and I went out for the the [1pizzas1] that [2he and Mary2] had ordered. But we had to wait for a parade of [3circus performers3] to pass before we could get across the street to pick *them*1 up.

(2)   [1I was one of the first people to arrive in the movie theater.1] [2I'd been looking forward to the movie2] since [3I saw the trailer two months ago.3] *Before long*1, the show was beginning.

Coherence and NP anaphora have previously been treated as unrelated types of language dependency (Section 1.2). This work argues that similar probabilistic mod-

els can explain aspects of both. It is proposed that these and other types of *referring expressions* can be modeled in terms of the probabilistic *expectation theory* introduced by Levy (2008). This work further shows the results of applying a simple Marr level 1 expectation model in behavioral experiments involving pronoun antecedents and in experiments involving coherence cue phrases.

Chapter 1 provides a brief critical review of the two main research areas that motivate this work: NP anaphora (Section 1.1) and discourse coherence (Sections 1.2 and 1.3), and closes by discussing relevant contemporary work in other areas of language research (Section 1.4). It is primarily ideas from these outside areas of research that will be applied moving forward, both in developing the model introduced in Chapter 2 and to inform the discussion of the experimental results.

Subsequent chapters include:

- Chapter 2, which introduces EMRE, explains how it may be applied to pronouns, describes how it can be empirically tested, and how the results of such experiments can be analyzed.

- Chapter 3, which describes the designs and results of three experiments which tested EMRE-derived hypotheses about about pronoun reference, and discusses hypotheses that are consistent or inconsistent with the data.

- Chapter ?? explains how the model can be more broadly applied, here to discourse-level coherence cue phrases.

- Chapter 5, which describes two experiments testing EMRE-derived hypotheses about temporal predicates, and discusses how their results demonstrate the link between surprisal and text acceptability.

- Chapter 6, which describes two experiments testing hypotheses about causal predicates, and discusses how their results demonstrate that reference depends on both top-down and bottom-up expectations.

- Chapter 7, which closes with a general discussion.

1.1 The problem of noun phrase reference

Referring expressions are interpreted in the context of a few distinct name-spaces shared by the speaker and listener: (1) *encyclopedic* knowledge (*e.g.*, 'Cambridge', 'Johnny Cash', 'Obama'); (2) physical context (*e.g.*, 'this [[thing I'm pointing at]]', 'that guy [[who just walked into the room]]'); and, (3) discourse context (*e.g.*, 'its owner', 'his counteroffer'). This work will propose that the third case, called *anaphora*, can be modeled in terms of *expectation* (Levy, 2008).

Successful anaphora require a reader and a writer. The writer intends to refer to some antecedent from the prior text, and chooses a referring expression that will allow the reader to determine the intended referent. How both parts of this process occur is the subject of ongoing research, which has proposed many correlated factors that might influence listeners' interpretation of NP-referring expressions (Winograd, 1972; Chafe, 1976; Kintsch, 1988; Ariel, 1990; Gordon et al., 1993; Gundel et al., 1993; Garrod and Sanford, 1994; Gordon et al., 1993; Stewart et al., 2000; Gordon et al., 2002; Gundel et al., 1993; Garrod and Sanford, 1994; Gordon and Hendrick, 1998; Arnold, 2001; Kehler, 2002; Wolf et al., 2004; Cowles and Garnham, 2005). This section provides a brief review of some of the past research literature on NP anaphora.

1.1.1 Cue combination theories

Any theory of anaphora must account for the diversity of the referring expressions that can be used to refer to an antecedent. Some referring expressions, such as definite descriptions and proper names (*e.g.*, 'President Obama'), are sufficiently informative about their referents that they can be used in a context where the referent is not salient (*e.g.*, has not been previously mentioned), while other referring expressions are less informative, and require more context (*e.g.*, 'the committee chair'). Still others, *e.g.* pronouns, are even less informative, and could refer to anything that matches their basic semantic constraints (*e.g.*, 'he', 'they', 'it'). Any theory of anaphoric reference must explain why pronouns are clear referring expressions in some contexts but not

others.

Some theories of anaphora describe the use of pronouns as an additional communicative channel that speakers and listeners use to indicate the discourse structure (Ariel, 1990; Gundel et al., 1993). Ariel claimed that speakers sometimes use pronouns and other less informative referring expressions in preference to more complete descriptions because reduced referential form *signals* to the listener that the antecedent is a highly accessible referent.

From a more contemporary perspective (Grosz et al., 1995; Brennan et al., 1987; Asher, 1993; Arnold, 2001; Beaver, 2004), Ariel's account of the forms of referring expressions is counterintuitive — recent work supposes the causal arrow to run in the opposite direction. That is, most contemporary research proposes that the purpose of reduced referring expressions (i.e., pronouns rather than proper names) is to take advantage of a referent's accessibility by using a shorter (and hence less informative) referring expression. Ariel, by contrast, describes reduced referring expressions as a 'signal' of the antecedent's accessibility, suggesting that the reduced expression constitutes an additional channel that speakers can use to make their intentions known. Gordon et al. (1993) continues this line of reasoning, suggesting that the use of pronouns can be a signal of topic continuity, while switching to a more extended form of reference can signal a shift in topic. Although Ariel's account is now rather heterodox, the factors proposed by Ariel to affect accessibility remain standard components of contemporary theories of NP anaphora.

Ariel's factors, which inform the expectation model introduced and tested in this thesis (Section 2.1), are (1) distance from the last mention; (2) competition with other referents; (3) salience of the antecedent; and, (4) whether the antecedent is in the same or prior sentence or paragraph as the reference. Gundel et al. (1993) offer a similar account of reduced referential forms, hypothesizing that they can be used when the antecedent has sufficient 'givenness', a property that can be the result not just of properties of the discourse. Gundel's explanation — which seems more familiar from a contemporary perspective — proposes givenness as a characteristic that depends not only on features of the text but also on the speaker and listener's

mutual knowledge about the shared features of their internal discourse model.

1.1.2 Memory storage and retrieval theories

Most subsequent work on anaphora has focused not on the question of when it is possible to use reduced referring expressions, but on why using them is desirable, and how they are processed. Both Ariel (1990) and Gundel et al. (1993) observed that it is infelicitous to use a more informative form than is allowed by accessibility/givenness (i.e., a proper name where a pronoun would do; cf. Grice (1975)). This claim has been supported by self-paced reading studies (Gordon et al., 1993) and ERP research (Swaab et al., 2004), which suggest an additional processing cost (which Gordon et al. (1993) termed the 'Repeated Name Penalty') associated with using a more informative referring expression where a less informative one would suffice.

This cost has been variously described as due to constructing a cognitive representation of an entity (Gordon and Hendrick, 1998), due to binding two distinct cognitive representations (Almor et al., 1999), and a predictable effect of rational probabilistic expectation (Arnold, 1998; Arnold et al., 2000, 2004). For instance, Arnold et al. (2004) demonstrate that listeners quickly adjust their expectations about what is coming next in the discourse on the basis of speaker disfluencies, which often indicate an upcoming change in topic. The present work builds on this rational expectation approach to anaphora, introducing a novel expectation-based model of anaphor resolution, as explained in the following chapter.

More contemporary theories of reference tend to describe anaphor resolution in terms of processing difficulty due to memory access (Gernsbacher, 1989; Gordon and Hendrick, 1998; Gordon et al., 2000, 2002; Almor, 1999; Arnold et al., 2004; Cowles and Garnham, 2005; Foraker and McElree, 2007). For instance, the Informational Load Hypothesis (ILH) (Almor, 1999) posits that more similar referents will interfere with each other more in memory. Under this theory, two people of the same gender would share more key semantic features than people of different genders, and the one that was otherwise more available in memory (due to factors such as locality and prominence), would have a stronger memory representation, at a cost to the

22

representation of the other. Retrieval of the less favored referent would carry an informational load, which could make the text less readable.

1.2 The problem of coherence

A complete theory of discourse, defined as the organization of language above the sentence level, would have to account not only for NP anaphora between sentences, but for all dependencies and structures in text that span more than a single sentence. These include:

- Discourse-level NP anaphora (referring expressions with noun phrase antecedents in other sentences), as well as NP co-reference.

- Discourse-level coherence predicates (*e.g.*, causal, temporal, and resemblance predicates between sentences).

- Hierarchical topic organization in documents (i.e., outlines).

The dominant computational model of discourse, *Rhetorical Structure Theory* (hereafter *RST*), represents all three of these phenomena as a single tree based on context-free grammars (Section 1.2.1, Mann and Thompson, 1988). However, RST is inadequate to capture all the important discourse-level information in normal text (Section 1.2.2, Wolf and Gibson, 2005; Prasad et al., 2008). An alternate theory of discourse-level coherence — that the dependencies in discourse-level coherence predicates are anaphoric connections similar to NP anaphora (Webber et al., 2003) — is addressed in Section 1.3 along with other contemporary research applied in the present work.

1.2.1 Rhetorical structure theory

Coherence has commonly been modeled as another layer of hierarchical structure above syntax, forming a tree structure over sentences, much as syntax forms a tree structure over words (Grimes, 1975; Hobbs, 1985; Mann and Thompson, 1988; Marcu,

2000a). RST, as the system most commonly used in computational linguistics to represent discourse structure, is the dominant example of such models. An RST representation of a text is a tree structure based on a context-free grammar (CFG), which groups sets of adjacent clauses and sentences into larger discourse constituents, then groups these into still larger constituents, and so on until the entire text is hierarchically structured. This is proposed to be the latent structure that generates the discourse properties of the text (Mann and Thompson, 1988; Marcu, 2000b).

Unlike a traditional outline, in which each section has a name that describes its contents (*e.g.*, the current subsection is named 'Rhetorical structure theory'), an RST annotation resembles a CFG parse in that each constituent is labeled with the relationship that is asserted to bind its constituents. These relationships include coherence predicates (*e.g.*, causality, temporal ordering, resemblance), as well as anaphoric dependencies (*e.g.*, *'elaboration'*) if there is no predicate relationship between neighboring constituents (sentences or paragraphs) in the same section of the text's hierarchical topic organization.

For more than two decades, computational linguists have used RST as the default representation of discourse, producing measurable improvements in applications like summarization and question-answering (Marcu, 1997, 2000a,b; Marcu and Echihabi, 2002; Carlson et al., 2002; Taboada and Mann, 2006b). There is ongoing debate about the specific set of relations needed, with systems ranging from a dozen to hundreds of relation types (cf. (Knott, 1996)). However, all such work takes as a given that the latent discourse structure of text can be modeled as a tree.

Tree structure is a powerful simplifying assumption, but it is inconsistent with more theory-neutral observations of the structure of coherence. As one of the original designers of RST wrote,

> *"...trees are convenient, easy to represent, and easy to understand. There is, however no theoretical reason to assume that trees are the only possible representation of discourse structure..."* (Taboada and Mann, 2006a).

Though trees are a pragmatic approach for applied computational linguistics, they

24
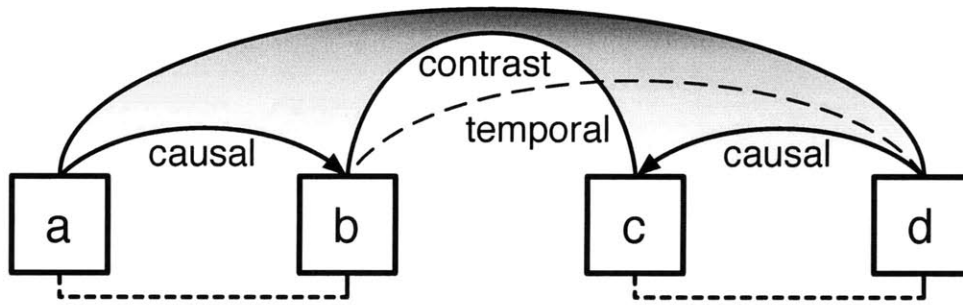
Figure 1-1: Coherence dependencies in Example 3

are not a viable computational description of discourse-level coherence dependencies. Webber et al. (2003) use the following example to illustrate the problem:

(3)   a. John loves Barolo.

        b. So_1 he ordered three cases of the '97.

        c. But_2 he had to cancel the order

        d. $because_3$ he $then_4$ discovered he was broke.

The text is both felicitous and clear. A good discourse model should therefore be able to represent all of its coherence dependencies, but it turns out that a tree-structured model cannot.

As shown in 1-1, a tree model can represent the causal connection signaled by 'So_1' and '$because_3$', (represented by the arrows $a \rightarrow b$, and $c \leftarrow d$) as well as the contrast marked by 'but_2' (represented by the larger shaded connection between $a + b$ and $c + d$). However, the temporal ordering marked by "$then_4$" cannot be modeled without using a more general class of latent structure.

1.2.2 Empirical evidence against RST

Empirical evidence suggests that tree-structured representations of discourse are not adequate to model all of the dependencies in the latent structure underlying coherence. The GraphBank corpus of Wolf et al. (2003) and the Penn Discourse Tree Bank corpus of Prasad et al. (2007) are discourse corpora annotated using novel models that do not

25

require tree structured annotations. Both corpora contain many global dependency structures inconsistent with a tree model of discourse (Wolf and Gibson, 2005; Lee et al., 2008).

The GraphBank corpus (Wolf et al., 2003) applies a novel annotation system to a corpus of 135 texts that was previously used in other discourse research. Instead of requiring a tree structure, the GraphBank system places no explicit global constraints on the graph topology of annotations other than that there be a fully connected graph spanning all the text. Naïve annotators were not given instructions regarding locality of arguments, multiple 'upward' links, crossing dependencies, or other graph-topological properties. Instead of assuming that the relations connecting earlier and later parts of a text must be between large sections, GraphBank annotators were instructed to include only those parts that they believed to be necessary to interpret the connection.

Wolf and Gibson (2005) found many instances of non-treelike structure in Graph-Bank, including nodes with multiple parents (41% of GraphBank's discourse segments) and crossing pairs of coherence relations (13% of dependencies crossing another dependency).

Like GraphBank, the Penn Discourse Tree Bank (*PDTB* Prasad et al., 2008) is a large annotated discourse corpus that uses a fundamentally different data structure than RST. The authors describe the PTDB model as 'heavily lexicalized', meaning that each coherence predicate they record includes an associated cue phrase. *Cue phrases* signal (or license) coherence predicates, like 'because' in example 4:

(4)   Dave went to the store *because* he needed to buy some milk.

The lexicalized design of the PDTB, in which the cue phrases for coherence predicates are a central component of the discourse model, represents a large theoretical departure from both RST and GraphBank. Despite this difference the PDTB, like GraphBank, contains numerous dependency pairs that are inconsistent with a tree-structured discourse (Prasad et al., 2008; Lee et al., 2008). These results, along with the findings of Wolf and Gibson (2005), suggest that restricting discourse models

26

to tree structures will omit many discourse-level coherence dependencies from the representation.

1.3 A different approach?

This section presents some contemporary research that will be applied throughout the remainder of this thesis. Section 1.3.1 describes the theoretical gap left by evidence that RST is not a viable theory of discourse. Section 1.3.2 describes a proposal introduced by Webber et al. that suggests a way to fill this gap. Section 1.4.2 describes the expectation theory of language processing (Levy, 2008), and Section 1.4.1 describes a theory of top-down and bottom-up cues in ambiguity resolution, both of which are applied in EMRE.

1.3.1 A theoretical gap

Putting aside for a moment the evidence against RST described in the previous section, common arguments in favor of it are that: (1) it is a broad-coverage system that has been successfully applied to many different types of texts (Mann and Thompson, 1988; Marcu, 2000b; Carlson et al., 2002); (2) manual RST annotations of texts can be used to generate summaries and question-answering systems that vastly outperform those generated by unstructured discourse representation systems (Marcu, 2000a; Taboada and Mann, 2006b); and, (3) automatic approaches to RST tree generation can acheive significant agreement with manual annotations (Marcu and Echihabi, 2002). In short, from a computational linguistics point of view, RST annotations have many of the same desirable characteristics as context-free representations of sentence structure.

Framed as a cognitive theory however, RST represents a rather large hypothesis relative to this evidence. It (and other tree-like discourse models) would suggest that readers recover the coherence structure of text by computing a parse structure. The coherence-level significance of clauses and sentences would then be determined through the process of integrating them into the global coherence parse, much as the

27

syntactic interpretation of words is determined in the process of integrating them into a sentence-wide syntactic structure. This would entail either the existence of an additional human parser specialized for dealing with discourse structure, or that the sentence parser somehow does double-duty as a coherence parser.

The evidence that sentences have a mostly tree-like (or *'weakly context-sensitive'* (Kuroda, 1964; Joshi et al., 1975)) latent structure is overwhelming, converging from many fields of research. If the only evidence had been that syntactic trees were useful to computational linguists, it is doubtful that either cognitive science or theoretical linguistics would have put significant effort into the question of how syntactic trees is represented and recovered by the mind. Given the expanding body of evidence against a tree structure for discourse, it seems: (1) unlikely that any progress in a cognitive theory of how people recover coherence dependencies will involve integrating sentences into a global discourse parse; and, (2) likely that if a clear explanation is found for why RST has the useful properties described above, it will involve a latent structure less computationally complex than a context free grammar[1].

However, evidence against a theory is not the same as a new theory. Neither the observation that non-treelike texts such as Example 3 exist, nor empirical evidence that they are relatively common (Wolf and Gibson, 2005; Lee et al., 2008) explain how readers are in fact able to recover the latent structure. Absent any alternate account of how readers recover coherence dependencies, it is difficult for behavioral scientists to frame experimental questions, and most work on coherence in computational linguistics continues to use RST.

---

[1]Such as, perhaps, preferences for: (1) grouping similar topics together (*i.e.*, hierarchical clustering); (2) text structured such that details are closely preceded by explanations of how they relate to the context; and, (3) clearly marked transitions between topics. Taken together, these describe those texts this author has found easiest to annotate using a variety of systems. However, they do not require a parser, as they accommodate globally inconsistent structures without predicting that they will be uninterpretable. They are not a cognitive theory of discourse, but simply a description of well-organised prose.

1.3.2 Discourse-level coherence predicates as anaphora

Contemporary research has demonstrated that a tree-like representation of discourse cannot provide a sufficient description (Wolf and Gibson, 2005; Lee et al., 2008; 1.2.2). Such findings, however, have been largely descriptive and theory-neutral, advancing no specific theory of how the reader locates the (potentially non-local) antecedents of coherence predicates. Webber et al. (2003) suggest a way to fill this theoretical vacuum, proposing that discourse-level coherence predicates are not syntax-like dependencies (cf. Halliday and Hasan, 1976; Grimes, 1975; Hobbs, 1985), but rather a predicate analogue to discourse-level NP anaphora.

This proposal is consistent with both the chain graphs used by Wolf et al. (2003); Wolf and Gibson (2005) and the lexicalized model used by Prasad et al. (2008), but goes a step further by suggesting a cognitive model for coherence predicates about which there is significant prior knowledge. This section places this hypothesis in context. The following section (1.1) describes factors that have been demonstrated or claimed to affect NP anaphora, and how they might also affect the connection between cue phrases and their antecedents in discourse-level coherence predicates.

Coherence predicates are often explicitly marked by a word or phrase. Following Knott (1996), the present work calls these markers *cue phrases*. Cue phrases are often complementizers (also known as 'subordinating conjunctions') — grammatical elements that connect a complement clause to a main clause (Rosenbaum, 1967). Examples of these include, *e.g.*, *'because'*, *'or'*, and *'after'*.

Webber et al. observed that although some of the cue phrases of coherence predicates are indeed complementizers, many are instead syntactically adverbials. Adverbials are words or phrases that modify a clause or sentence and typically affect how the clause should be interpreted. Unlike complementizers, adverbials are syntactically linked to only a single clause. For instance, in Example 5, the adverbial 'unfortunately' modifies the clause 'I'm very hungry,' conveying the speaker's attitude about her hunger.

(5) a. Jack: Do you want to go see a movie?

b. Jill: Unfortunately, I'm very hungry.

The adverbial 'unfortunately' can be interpreted in the context of what Hankamer and Sag (1976) refer to as the 'pragmatic environment', but requires no specific linguistic antecedent. However, some adverbials are only acceptable when they can be interpreted in the linguistic context of a specific prior utterance, as with Example 6, where 'consequently' lacks any reasonable null-context interpretation:

(6)   % Consequently, I'm very hungry.

Adverbial cue phrases that require a specific linguistic antecedent can create discourse-level (i.e., between sentences) language dependencies - as such, Webber et al. (2003) term these phrases *'discourse adverbials'*. The following section (1.3.3) places their hypothesis in context, and explains why it has proved controversial.

1.3.3   Reactions to discourse adverbials hypothesis

The proposal of Webber et al. (2003) that the adverbial cue phrases for discourse-level coherence predicates should be treated not as a type of complementizer, but rather as a type of anaphor is a large departure from earlier theories of coherence predicates between sentences (cf. Grimes, 1975; Halliday and Hasan, 1976; Hobbs, 1985). Four main objections to the central hypothesis of Webber et al. (2003) are that:

1. It flies in the face of essentially all prior accounts of coherence predicates;

2. It shifts a category boundary in a way that could be interpreted as a challenge to basic syntactic theory;

3. It stretches the accepted semantics of anaphora; and,

4. It is lacking in experimental evidence.

This section argues that although these have all been supportable positions with respect to Webber et al.: (1) taken jointly, they have not been sufficient evidence to dismiss the hypothesis; and (2) in the context of the present work, the final point

is no longer true, leaving Webber et al. (2003) with an intriguing and empirically investigable theory.

Coherence assumed to be coordination

Accounts of discourse-level coherence predicates preceding that of Webber et al. (2003) uniformly treat them not as anaphora but as a discourse-level analog to complementizers, which connect clauses (Rosenbaum, 1967). Such accounts described discourse-level predicates (those between sentences) as discourse-level analogues of sentence-level coherence predicates, which involve complementizers, *e.g.*, *'because'*, *'before'*, as in Example 4 from page 26, repeated here:

(4)   Dave went to the store *because* he needed to buy some milk.

The earliest theoretical linguistic work on coherence, by Halliday and Hasan (1976), refers to coherence predicates as *'conjunctions'*, a syntactic category similar to complementizers, referring to words that bind two adjacent clauses (*e.g.*, *'and'*, *'not only'*... *''but also'*) or bind a phrase into a containing clause (*e.g.*, *'despite'*). No modifier was employed by Halliday and Hasan — coherence predicates were simply identified as 'conjunctions', and as such they, as well as their contemporaries, naturally consider only examples in which the predicate's two arguments are adjacent clauses. As with complementizers, which bind two adjacent arguments together into one phrase, some early accounts model discourse-level coherence predicates as having a tree-like structure similar to the latent syntactic structures recovered in sentence parsing (Grimes, 1975; Halliday and Hasan, 1976; Hobbs, 1985).

Mann and Thompson (1988) followed this precedent in their decicison to represent the full structure of coherence using a context-free grammar — in a bare-bones CFG, the arguments are always adjacent (Chomsky, 1957). This convention that coherence predicates are conjunctions is even echoed outside the research community — in a pedagogical monograph written for students, prescriptivist grammarians refer to Webber's 'adverbials' as *'conjuncts'* (Quirk and Greenbaum, 1973).

Later, more methodical work that investigated the semantics of coherence predicates (Asher, 1993; Knott, 1996) made no such specific claims about the structure of coherence predicates, but also considered only examples in which the arguments of the coherence predicate are adjacent clauses or sentences, such as those that might be bound by a complementizer — never examples in which the two arguments are separated by unrelated text[2].

Conjunctions are not anaphoric

Another problem for the anaphoric account of coherence predicates proposed by Webber et al. (2003) is that conjunctions (which, as described above, have been treated as a superset of coherence predicates) have no obvious similarities with NP anaphora. While conjunctions (specifically, complementizers) specify a relation between adjacent clauses (or phrases) that they syntactically bind together, NP anaphora use a referring expression to substitute for an (often syntactically unconnected) NP, typically without specifying how the NP should be interpreted. Anaphora are an extremely different class of language phenomenon from conjunctions, following dissimilar structural and semantic rules.

Working from a presumption that coherence predicates are a type of conjunction, the argument that predicates marked by adverbials are a type of anaphor has been[3] construed as saying that "some conjunctions are anaphoric". The distinction introduced by Webber et al. (2003) appears to have been sufficiently unorthodox that the association between predicate argument structure and hierarchical structure is hard to disentangle.

Stretching the meaning of 'anaphora'

Another reason that the discourse adverbials hypothesis of Webber et al. is surprising is that it significantly stretches the semantics of anaphora. Anaphora are typically

---

[2]This was probably due in part to brevity concerns — as observed in the experimental items in Chapters 3–6, allowing for the possiblity of non-local arguments to coherence predicates makes for rather long-winded examples.

[3](generalizing from this researcher's personal correspondendence)

thought of as simply a way of pointing back at a previously mentioned entity. While the syntactic dependency between an NP and a VP (verb phrase) is usually considered fundamentally different in kind than that between a preposition and an NP, no prior work has proposed different classes of dependency between NP-referring expressions and their antecedents.

Coherence predicates, like those in syntax, are by contrast considered to involve multiple types of dependency. Knott (1996) developed an extensive model describing a directed graph of which cue phrases may always, sometimes, or never be substituted for each other (*e.g.* *'while'* may always be substituted for *'whereas'*, but the reverse is not the case). Both RST (Mann and Thompson, 1988) and the GraphBank system of Wolf et al. (2003) are based on typed dependencies between often-distant sentences in a text, which are modeled as a type-labeled node (in RST) or edge (in GraphBank) in a graph structure joining the two sentences. The typical semantics of anaphora are by comparison much simpler than what coherence researchers are accustomed to.

Lack of experimental evidence

A final reason that the hypothesis advanced by Webber et al. (2003) has not been broadly discussed is a lack of experimental evidence. Their discussion relies primarily on a handful of found and constructed examples and arguments based on linguistic theory. However, the PDTB offers corpus-based data supporting this hypothesis (Prasad et al., 2008; Lee et al., 2008). Furthermore, chapters 5 and 6 of the present work present behavioral data that are also consistent with this hypothesis.

1.4 Probabilistic and cue-based language models

This section considers two recent developments in psycholinguistics: the probabilistic representation of interacting top-down and bottom-up cues introduced by Gibson (2006) (Section 1.4.1), and the expectation theory of language comprehension introduced by Levy (2008) (Section 1.4.2).

1.4.1 Probabilistic top-down and bottom-up cues

Readers use lexical, syntactic, and pragmatic knowledge as *top-down* cues, shaping their rational expectations about what words and phrases are likely to mean. For instance:

- *Priming* readers with words associated with a particular *'frame'* (*e.g.*, *'farmer'*, *'cow'*) decreases the amount of time they need to recognize and name a related word (*e.g.*, *'straw'*), relative to unrelated words (Tulving et al., 1964).

- Readers are sensitive to the likely syntactic categories of upcoming words (Ferreira and Clifton, 1986).

- Readers are able to use verb tense information from recent sentences to disambiguate ambiguous verb tenses (Trueswell and Tanenhaus, 1991).

- Listeners are able to use pragmatic information about objects in the physical environment to distinguish utterances in which a phrase such as *'on the towel'* might be either an NP adjunct or a VP complement from those in which it must pragmatically be an NP adjunct.

There are also situations in which readers rapidly apply *bottom-up* cues that were not available before they needed to be used. For instance, in Example 7 (Trueswell et al., 1994):

(7) a. i. The evidence that was examined by the lawyer turned out to be unreliable.

      ii. The evidence examined by the lawyer turned out to be unreliable.

   b. i. The defendant that was examined by the lawyer turned out to be unreliable.

      ii. The defendant examined by the lawyer turned out to be unreliable.

(8) a. The poster that was drawn by the illustrator was used for a magazine cover.

   b. The poster drawn by the illustrator was used for a magazine cover.

The verb in Example 7, 'examined' can be ambiguous in tense. Unlike verbs like 'drawn' in Example 8, the past tense of which is 'drew', the verb 'examined' has identical past tense and passive participial forms. In the versions of the items in which the relative clause is unreduced ('that was examined by the lawyer'; Examples 7a-i and 7b-i), the verb 'examined' is unambiguously a participial form. However when the relative clause appears in a reduced form ('that was examined by the lawyer'; Examples 7a-ii and 7b-ii), it is ambiguous.

This ambiguity might be expected to produce elevated reading times relative to unambiguous controls (*e.g.*, Example 8), since participial forms are in general less common than past tense forms. This was indeed the case in conditions such as in Example 7b-i, in which the subject of the sentence ('The defendant') is a plausible agent for the verb 'examined'. However, in conditions such as in Example 7a-ii, in which the subject 'the evidence' is an implausible agent for the past tense verb 'examined', but a plausible patient for the passive participial form, no difference in reading time was observed relative to the unreduced conditions.

This result differs from the top-down cues described above in the *time course* of the event. Readers were able to act on their knowledge that evidence does not examine things quickly enough that the ambiguity was undetectable in their reading behavior almost immediately upon encountering the verb form. This cannot be purely explained in terms of inanimate objects being generally more likely to be patients than agents — inanimate objects can certainly be the subject of a verb (*e.g.*, 'the evidence turned out to. . . ') but the main verb reading of 'examined' was sufficiently suppressed by the plausibility information that it was indistinguishable from the unambiguous conditions.

Misleading cues

The study of vision is often more engaging to lay audiences than many other aspects of cognitive scence, not least because of visual illusions. In such illusions, two shapes or regions appear to be very different in properties such as brightness, size, or alignment, and are (almost) always revealed to be identical. These visual illusions can

almost always be explained in terms of well-understood properties of low-level visual processing [4]

In language, as in vision, there are inputs that can lead comprehension systems to go awry:

- Difficulty in integrating long-distance syntactic dependencies like those in object-extracted relative clauses (see Section 2.3.3).

- The incomprehensibility of triply center-embedded structures such as "Oysters oysters oysters eat eat eat."

- Readers' trouble reliably naming words for colors that are printed in a different color (Stroop, 1935).

Most such examples appear to be the result of low-level processes being unable to cope with unorthodox inputs.[5]. Subtler counter-informative bottom-up cues can also be observed using more sensitive methods. For instance, Seidenberg et al. (1982) found that when ambigious words such as *'straw'* appear in an unambiguous context (*e.g.*, *'I drank my soda with a straw.'*, this appears to increase readers' expectation of words related to the other sense of the word (*e.g.*, *'farmer'*). However, this effect appeared to be temporary, and could no longer be observed less than a second later.

Tabor et al. (1997) observed some sentences in which processing difficulty seems to be associated with interference from interpretations that are impossible in that context. This effect was illustrated by Tabor et al. using a determiner/complementizer ambiguity for the word *'this'*; the effect was replicated by Fedorenko and Gibson (2007) for noun/verb ambiguities, which are used to illustrate the idea in Example 10:

(9) *(verb-biased context)*

    a. Rachel had planned to *play*...

---

[4]The standout exception to this rule is the *'checkershadow illusion'* (Adelson, 1995), which can only be explained in terms of higher level reasoning.

[5]One counterexample to this is the phenomenon of *Escher sentences* such as *'More people have been to Russia than I have'*

b. Rachel had planned to *type...*

(10) *(noun-biased context:)*

    a. Rachel thought that a *play...*

    b. Rachel thought that a *type...*

In Example 9, the sentence sets up a *verb-biased context*, in which the next expected head to be integrated into the sentence structure will be a verb. Of the two verbs that appear in this example, *'play'* is more common than *'type'*. Consequently, *'type'* carries a higher *top-down lexical surprisal* than does *'play'* (Tanenhaus et al., 1979; Seidenberg et al., 1982; Tabor et al., 1997, 2004; Gibson, 2006; Fedorenko and Gibson, 2007). As expected, higher reading times were observed for the verbs that were less frequent (and hence, more surprising) in the verb-biased context. This is an example of rational surprisal, in which the differences in reading time are consistent with a purely expectation-based account of processing.

However, Tabor et al. (1997) and Fedorenko and Gibson (2007) also found another type of word-frequency effect on reading times, which is not well-explained in terms of rational expectation. Example 10 sets up a noun-biased context, in which when readers arrive at the emphasized word (*'type'* or *'play'*), they have a very high expectation that next head to be integrated will be an NP, and hence a high expectation that the upcoming word will be a noun and a very low expectation that it will be a verb.

The words *'type'* and *'play'* are equally likely as nouns, and should carry equal top-down lexical surprisal. The verb sense of *'type'* is less frequent than its noun sense, while the verb sense of *'play'* is more likely than its noun sense. Fedorenko and Gibson (2007) replicated the type of effect observed by Tabor et al., finding that in a noun context, verb-biased words like *'play'* are read more slowly than noun-biased words like *'type'*, even when they are matched for their frequency as nouns. It appears that the verb sense of the word somehow interferes with the noun sense, even when there is no rational expectation of a verb.

Probabilistic interpretation

This is an example of a *bottom-up lexical bias* — a type of effect on reading behavior that cannot be explained in terms of rational expectation. The effect, in which the human parser appears to be distracted by irrelevant alternatives, is most easily explained in terms of resource limitations. Because the parser has to respond very quickly to incoming words, it appears to sometimes violate the principle of insensitivity to irrelevant alternatives (Luce, 1959).

Tabor et al. (1997) modeled this behavior using a complicated system based on: (1) training a neural network using a simplified syntax model; (2) interpreting the connection weights of the network as gravitational vectors in a dynamical system; and, (3) dropping sentences into this network and tabulating the frequency with which they fell into the dynamical system's stable periodic attractor states, which were interpreted as representing possible interpretations of the sentence.

In contrast to the somewhat baroque model of Tabor et al. (1997), Gibson (2006) modeled the same data in terms of: (1) *context-dependent syntactic expectation* the expectation $p(c_i \mid E)$ that a particular syntactic category would appear in a given syntactic environment; and, (2) *context-independent lexical category frequency* — the frequency $p(c_i \mid w)$ of a syntactic category given a particular word. These conditional probabilities were estimated using parsed corpora.

The first of these (syntactic expectation) is a top-down cue which is presumably computed and available before a word is encountered. The second (lexical category frequency) is a bottom-up cue, because the relevant values cannot be available before a word is encountered. Gibson (2006) found the normalized product of these estimated probabilities to be predictive of human reading times.

An unusual combination of features exhibited by the model of Gibson (2006) is that: (1) it is a Marr level 1 (computational level) model; and (2) it models top-down and bottom-up cues in probabilistic terms; but, (3) it is not a Bayesian probabilistic model. The model applied in this work follows Gibson's approach in the first of these two ways, and differs in the third.

Gibson's model is a purely computational level model in that it describes a computation that humans could be performing (and compares the results of the computation to human behavioral data) without ever, *e.g.*, suggesting that the observed bottom-up lexical bias is not in fact due to low-level characteristics of the lexical access system. Rather, it simply: (1) describes a quantity that could be computed and used in a particular way; (2) estimates the quantity in the most direct manner possible; and, (3) tests to what extent humans behave as they would be expected to if they were computing and using that quantity. In this sense, it is the direct opposite of the model applied to the same data by Tabor et al. (1997), a purely Marr level 2 model that performs very complicated algorithmic simulations without a clear theory of what is being calculated.

A second characteristic of the Gibson (2006) model is that it breaks new ground by modeling top-down and bottom up cues as probabilities. This is a particularly effective approach for this particular data and model, as the relevant probabilities could be straightforwardly estimated from corpora. However, while this model uses probabilities, it is not a probabilistic model. The reason for this is neither the smoothing applied to the syntactic expectation, which has a Bayesian interpretation (MacKay and L., 1995), nor the use of context-independent lexical category estimates, which is the type of simplifying assumption used in all naive Bayes models, but rather the calculation itself. A Bayesian model would predict:

$$p(c_i \mid w, E) \sim p_{+\alpha}(c_i \mid E) \times p(w \mid c_i)$$
$$= p_{+\alpha}(c_i \mid E) \times \frac{p(w) \times p(c_i \mid w)}{p(c_i)}$$
$$\sim \frac{p_{+\alpha}(c_i \mid E)}{p_{+\alpha}(c_i)} \times p(c_i \mid w)$$

The idea of top-down and bottom-up probabilistic cues being combined will also be applied in EMRE.

1.4.2 Expectation

The model introduced in Chapter 2 models referring expressions in terms of the *expectation* theory of language processing introduced by Levy (2008). Levy introduces this theory in terms of the long-standing observation that some syntactic structures are harder to parse than others, as measured by, *e.g.*, reading times measured by self-paced reading and eye-tracking, changes in the electromagnetic field at the scalp, and fMRI imaging. This has often been explained in terms of some cognitive resource of which the human parser has a limited supply, and which is taxed more by some syntactic structures than by others. Whatever the neural basis of this resource is, readers' behavior demonstrates expectations about what they are likely to observe next. Given these expectations, the optimal way to manage this cognitive resource would be to prepare for the reader's best estimate of the future.

Levy observes that if expectation theory is a good model of readers, the expected behavioral correlate would be what Hale (2001) called *surprisal*, or the negative log probability of a new observation. Levy also proves this to be equivalent to the *KL divergence* from the posterior to the prior expectations, which can be thought of as the Shannon information needed to encode the new expectations in terms of the old. The theory then predicts readers to exhibit signs of processing difficulty when they encounter something that forces them to change their expectations, but not necessarily in the case of ambiguity, which Levy demonstrates can even sometimes be observed to aid in processing.

An underappreciated feature of surprisal theory is that it moves probabilistic expectations from Marr level 2 (the algorithmic level) to level 1 (the computational level). While older theories of language comprehension commonly discussed possible interpretations of ambiguous input in terms of each possible interpretation having a *'weight'* or *'score'*, more recent work has noted that modeling uncertainty in terms of *probabilities* has the dual advantages of: (1) allowing much more powerful mathematical techniques to be brought to bear; and, (2) greatly constraining the space of predictions that could be consistent with their theories (Hale, 2001; Roark, 2001;

Crocker and Brants, 2000; Narayanan and Jurafsky, 1998, 2002; Gibson, 2006; Frank et al., in press).

This application of probability theory has often treated uncertainty as a transient phenomenon. Probability distributions over possible interpretations are intermediate steps in calculations that settle on a single correct interpretation. Levy's expectation theory, however, treats uncertainty not as an algorithmic step, but as a typical outcome of a computation. Under this theory, readers are maximally advantaged by maintaining a state of uncertainty not only about what they would observe in the future, but also about what had been observed, at many levels of interpretation, including acoustic, phonological, lexical, syntactic, and prosodic, as well as uncertainty about past and future inputs.

Levy and collaborators have used surprisal theory to explain experimentally observed behaviors in syntactic circumstances such as high vs. low phrase attachment, incremental speed-up during sentence reading (Levy, 2008), processing of garden path sentences (Bicknell et al., 2009), relative clause processing (Rohde et al., 2010), processing of extraposed structures (Levy et al., 2009b), and processing speed at the end of German verb-final structures (Levy and Keller, 2010).

Though lexical- and syntactic-level analysis comprises the moiety of Levy's work with the surprisal theory, he notes that it could, without basic conceptual changes, be applied to other aspects of language (Levy, 2008), as is done in the present work.

Reflections One possible reason no computational model of NP reference and discourse-level coherence predicates has previously been proposed is that many ideas which usefully inform this research have been separated into different research disciplines. Basic techniques and concepts without which this work could not have usefully proceeded include, at a minumum:

- Behavioral experiments in which sets of items involving minimal changes along carefully controlled dimensions are presented to baseline human participants (Trueswell et al., 1994, 1993; Gibson and Fedorenko, 2010).

- Expectation-based models of surprisal in language comprehension (Hale, 2001; Levy, 2008; Demberg and Keller, 2008).

- Psycholinguistic theories of processing cost in dependency resolution (Gibson, 1998, 2000; Grodner and Gibson, 2005)

- Language models combining distinct top-down and bottom-up effects (Tabor et al., 2004; Gibson, 2006; Demberg and Keller, 2008).

- Fine-grained semantic and structural inspection of coherence predicates (Webber (1988); Knott (1996); Webber et al. (2003).

- Corpus-based discourse analysis (Mann and Thompson, 1988; Wolf et al., 2003; Wolf and Gibson, 2005; Prasad et al., 2007, 2008; Lee et al., 2008).

- Probabilistic language models (Hale, 2001; Roark, 2001; Crocker and Brants, 2000; Narayanan and Jurafsky, 1998, 2002; Frank et al., in press).

- Cognitive theories of anaphoric processing (Halliday and Hasan, 1976; Winograd, 1972; Chafe, 1976; Ariel, 1990; Gordon et al., 1993; Gundel et al., 1993; Gordon et al., 1993; Garrod and Sanford, 1994; Gordon and Hendrick, 1998; Stewart et al., 2000; Arnold, 2001; Gordon et al., 2002; Kehler, 2002; Cowles and Garnham, 2005).

Chapter 2

Expectation and Reference

A reader who encounters an referring expression needs to determine the target antecedent — the phrase to which the writer intends the reader to connect the referring expression. EMRE proposes to model referring expressions in terms of *expectation* (Section 1.4.2; Hale (2001); Levy (2008); Demberg and Keller (2008)).

Unlike models based on explicit hypotheses about processing, probabilistic models (like those applied in expectation theory) are computational-level descriptions of an inference problem that do not distinguish between any representational structures and algorithms that could be applied to this problem (Marr, 1982). Computational-level approaches have the potential to identify aspects of behavior that reflect the underlying latent structure of the problem space, rather than features of the processing mechanism (*i.e.*, the brain). As such, this model describes anaphora not in terms of algorithms or mechanisms the brain could employ when processing them (e.g. 'activation', 'availability', 'interference') but in terms of the computations that must be at least approximately solved (*i.e.*, expectation).

The word 'approximately' is especially relevant in the case of EMRE. Like the model of Gibson (2006), (Section 1.4.1), EMRE does not assume that readers will always use all available information to compute a *maximum a posteriori likelihood* (or *MAP*) reference expectation for a given referring expression. Instead, it is a model of readers who may prepare at leisure, but must decide in haste. The capacity to prepare is modeled as top-down expectation, while the need to decide quickly is

modeled as bottom-up match.

Both of these quantities can depend on an effectively unbounded set of factors, none of which can be said to be outside of the model, so neither top-down expectation nor bottom-up match can be observed or calculated directly. However, as will be explained below EMRE offers a theory of interaction between factors that should effect top-down expectation and bottom-up match — it is these interactions that will be tested in the behavioral experiments described in the following chapters.

Section 2.1 introduces EMRE in terms of NP anaphora, one of the types of referring expression to which it is applied in this work. Section 2.2 describes broadly describes how the model may be tested experimentally. Section 2.3 closes with a description of the behavioral model that links the expectations modeled by EMRE to the dependent measures observed in the experiments in following chapters.

2.1   Expectation Model of Referring Expressions

The model introduced here is intended to be applicable to any type of referring expression that a reader might expect to encounter in a text. This section explains the model in terms of NP anaphora, while the following section (**??**) explains how it can be applied to other types of discourse-level referring expressions. The model uses a prior *top-down expectation* over possible referents (Section 2.1.1) and a feature-based *bottom-up match* likelihood (Section 2.1.2 to compute a posterior *reference expectation* over possible referents (Section 2.1.3).

2.1.1   Top-down expectation

The *top-down expectation* over possible referents models what readers expect of the next word they encounter in a text. This, according to expectation theory (Levy, 2008) is equivalent to what they are prepared for. The expectation of a referent r_i is modeled as being:

- Context-dependent.

- Maintained and updated constantly.

- Formed *before* the reader encounters a referring expression.

- Based on *all available cues* that are statistically informative about what is likely to be mentioned (*i.e.*, *'unbounded rationality'*) (Section 1.4.1).

- Implicitly specifying an *efficient code* for identifying expected referents (Shannon, 1951).

For a given type T of referring expression, EMRE models the reader as having, at any point in the text: (1) an expectation of encountering that referent type; and, (2) an expectation at that point of any given discourse referent r_i being the target of a type T referring expression. In the case where the reference type T is NP anaphora TA:

$$\text{anaphor expectation} \equiv p_{\text{TD}}(A \mid \text{context})$$
$$\text{top-down expectation} \equiv p_{\text{TD}}(r_i, A \mid \text{context})$$

In the case of NP anaphora, the second of these is the expectation that any given NP r_i will be mentioned anaphorically. This expectation is not a proper probability distribution (does not sum to unity), because NP anaphora are only one of the many language constructions that could appear in a given place. These expections will be different for each word that is processed. For instance, in a verb biased context (*e.g.*, *'Whoops, I accidentally...'*), the expectation of an NP anaphor[1] should approach zero. However, the expectations are modeled in terms of probability, because any 'weighting' system not consistent with the basic axioms of probability can be show to predict contradictory results that would be inconsistent with the basic rules of logic (Cox, 1946, 1961; Aczél, 1966; Jaynes, 2003), let alone the more specific framework on which expectation theory depends.

---

[1](or any form of NP)

Taken together, these two values specify the *conditional top-down expectation*:

$$\text{conditional top-down expectation} = \frac{\text{top-down expectation}}{\text{anaphor expectation}}$$

$$\equiv \frac{p_{\text{TD}}(r_i, A \mid \text{context})}{p_{\text{TD}}(A \mid \text{context})}$$

$$= p_{\text{TD}}(r_i \mid A, \text{context}),$$

which is used in calculating posterior reference expectation (Section 2.1.3).

Cues to top-down np expectation

Top-down NP expectation models a reader's use of all available information that the reader could use to rationally predict a reference *before observing it*. This should be affected by cues that have been observed by prior research in NP anaphora to facilitate reference or increase reference likelihood. For instance:

- Locality: how long ago r_i was mentioned (Anderson and Schooler, 1991).

- Syntactic Prominence: whether r_i was mentioned in a main or subordinate clause (Brennan et al., 1987; Gordon et al., 1993; Grosz et al., 1995).

- Repeated use: whether r_i was mentioned only once, or has since been referred to by referring expressions of this or other types.

- Mention of semantically or pragmatically associated referents.

- Changes in the pragmatic environment.

- Topicality or pragmatic importance.

The two top-down cues used in the experiments are locality and syntactic prominence. Locality refers to a potential antecedent's location relative to the referring expression and any intervening clauses. Ariel (1990) claims that the absence of more than one sentence boundary between a referring expression and its antecedent is an key factor in the 'accessibility' of NP antecedents, and that when more than one sentence boundaries (or any paragraph boundaries) occur between a referring expression

and its antecedent, more informative referring expressions are always necessary for the anaphor to be understandable (Ariel, 1990). This assertion is challenged by Miltsakaki et al. (2002), who asserts that the expectation of entities is not always affected by intervening text.

The *target locality* factor used in the experiments assumes that, other things being equal, when separated from a referring expression by two or more clause breaks, an antecedent has lower top-down expectation than when separated by one or zero clause breaks. This extends Ariel's claim that sentence breaks affect the accessibility of antecedents, and formalizes it in terms of expectation theory. If there was a uniform prior expectation over referring expressions such that any referent was equally expected be mentioned again in a given context, then the text could offer no useful intrinsic cues about its anaphoric structure, and all referents would have to be mentioned using fully descriptive NPs. Instead, the likelihood of an entity's repeated mention in natural text appears to drop off roughly according to a power law function of distance (Anderson and Schooler, 1991). The proposed target locality factor is weaker than, but consistent with, the more specific claim that the likelihood follows a power law decay.

The other top-down cue used as a factor in the experiments is *syntactic prominence* (Brennan et al., 1987; Grosz et al., 1995), defined herein as whether the target antecedent is in a main clause or a subordinate clause. It is predicted to be used as an informative cue to top-down expectation, independent of locality.

The final two top-down cues mentioned, pragmatic changes and pragmatic importance, bear further consideration. An argument might be made that, *e.g.*, the pragmatic importance of a referent is surely a feature of the referent itself, and not a feature of the context. While this is true in an intuitive sense, the objection slightly misses the point of top-down expectation. In the Marr level-1 sense of describing an optimal computation to which the mind approximates a solution, the computationally-ideal form of top-down expectation would be one that uses *all available information* as rational cues to expectation, in order to make the best possible estimate of what is most likely to be mentioned based on information available *before* observing a reference.

This can be thought of as the construction of an *optimally efficient code* (Shannon, 1951; Huffman, 1952; Jaynes, 1957; MacKay, 2002) that allows the most expected referents to be mentioned using fewer elements of the code, while mentioning less expected referents requires more, thus achieving *uniform information density* (Shannon, 1951; Huffman, 1952; Reed and Solomon, 1960; Levy and Jaeger, 2007). The elements of the code itself are the features described in the following section.

2.1.2 Bottom-up match

Bottom up match is a model of readers' ability to react to a referring expression and rapidly assess how likely the referring expression is given a each of the possible referents. Where top-down expectation is based on all information available *before* a referring expression is observed, cues to bottom-up match are precisely those that are only available *after* one has been observed.

The bottom-up match computation produces a set of Bayesian likelihoods that can be used to rapidly update the prior top-down expectations, producing a *posterior reference expectation* (Section 2.1.3) The match between a referring expression RE and a referent r_i is modeled as being:

- Context-independent (*i.e.*, orthogonal to top-down expectation).

- Computed after the referring expression is encountered.

- Computed rapidly, using boundedly-rational *feature matching*.

- Estimating the Bayesian likelihood that a referent r_i would have *generated* a referring expression RE.

For a given reference type, the bottom-up match will depend on features that could be used to assess how likely the referent is to be referred to in a given way. Any reference type will require a way to map referents and referring expressions onto

features. This mapping function takes the form:

$$\text{feature mapping} \equiv \omega_A : \begin{cases} r_i \mapsto \{f_{i_1}, f_{i_2}, \dots\} \equiv \text{F}_i \\ \\ \text{RE} \mapsto \{f_{\text{RE}_1}, f_{\text{RE}_2}, \dots\} \equiv \text{F}_{\text{RE}} \end{cases}$$

Bottom-up match is defined as the probability that when a referent r_i with some set of features F_i is referred to using a particular reference type (such as an NP anaphor A), the referring expression RE will have some other set of match features F_{RE}:

$$\text{bottom-up match}(r_i, \text{RE}) \equiv p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, A)$$

This simplification of the Bayes optimal bottom-up match likelihood $(p(\text{RE}) \mid p(r_i, A, \text{context})$ is made because it is cognitively implausible that, *e.g.*, a reader would have no context-independent expectation of how likely it is that a man would be referred to as *'he'* without knowing which man it was and under what circumstances he was being mentioned. Bottom-up match can be thought of as the probability that the features of a possible referent r_i would have 'generated' the features of an observed referring expression RE.

Note the inversion between the definition of bottom-up match used in the present work and the concept of informativeness in prior work (Ariel, 1990; Gundel et al., 1993; Gordon et al., 1993). Informativeness is often described as how generally informative the referring expression is about the antecedent — the context-independent likelihood that a referring expression has a particular antecedent. Bottom-up match is instead the context-independent expectation of the referring expression given the antecedent. This inversion, which is initially unnatural-seeming to most, gives a *'likelihood'*[2]. The inverson allows the desired quantity — the expectation of the antecedent given the referring expression and the context — to be factored into two more easily-estimated expectations. This is key to the tractability of EMRE, as with other Bayesian models.

---

[2](sometimes called, partially in jest, a 'Bayesian backwards probability')

Cues in bottom-up match

As with the bottom-up phenomena described in Section 1.4.1, in which readers sometimes appear to make connections that are not licensed by the top-down context (Stroop, 1935; Seidenberg et al., 1982; Tabor et al., 1997; Gibson, 2006; Fedorenko and Gibson, 2007), EMRE does not assume that readers will use all available information perfectly. In the cases cited above, the non-optimal behavior is consistent with an account in which overlearned connections (*e.g.*, between a color and its name, between a word and its senses) temporarily overwhelm the reader's top-down knowledge.

In the case of anaphoric connections, imperfect reasoning is not likely to be the result of overlearned connections — the word *'he'* will not be strongly associated with any particular person — but rather the result of incomplete use of higher-level knowledge. Rather than modeling this in terms of, *e.g.*, a Marr level 2 processing cost, EMRE represents this imperfect use of knowledge as a Bayesian likelihood computation based on surface features of the referent and referring expression. This decision should not be taken as a committment to a cognitive-realist position that, if readers perform any cognitive process that corresponds to bottom-up match, then they do so by literally computing a set of discrete features and basing a probabilistic calculation on them.[3] EMRE's use of features in bottom-up match is simply a way of representing readers' need to rapidly assess probable reference in a way that might not make full Bayes-optimal use of all available information.

This said, there are some features that do lend themselves well to discrete representation. Pronouns have a very minimal set of such features — the word *'she'*, for instance, has *gender*, *number*, and *animacy* features, but no others, and the word *'they'* carries a solitary[4] feature of *number*. It may be intuitively useful to observe that this likelihood obeys algebraic rules of compositionality — adding features strictly decreases the match, while marginalizing over every possible value would give the same result as not including it[5].

---

[3]This in fact seems extraordinarily *unlikely*.

[4]though plural

[5]For instance, the marginal match for any NP over the words *'he'*, *'she'*, *'it'* and *'they'* will sum to one, since there is no discourse referent that cannot be referred to in one of these ways.

More complex referring expressions will have more features — for instance, *'she'* has fewer features than *'the woman'*, which in turn has fewer features than *'the professor'*, and so on until enough features have been added to identify the referent in a null context. Each added feature will decrease the maximum bottom-up match, but will also decrease the marginal probability of RE's feature set. The ratio of the match likelihood and the marginal likelihood of the features gives a Bayesian *evidence ratio*:

$$\text{bottom up evidence ratio} = \frac{\text{match likelihood}}{\text{marginal feature likelihood}}$$
$$\equiv \frac{p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, A)}{p(\text{F}_{\text{RE}} \mid A)}$$

The features used in bottom-up match are the *code elements* mentioned in the prior section. If the top-down expectation is an optimal code for possible referents, the minimally informative grammatical features like number and gender are the 'short words' of the code, which can indicate the most probable outcomes without expending much entropy (Shannon, 1951; Huffman, 1952; Reed and Solomon, 1960).

The only NP cue used in the experiments described in this work is gender agreement, however more cognitively complex pragmatic cues are possible. See for instance Example 12 (page 56), in which plausibility is used as a bottom-up cue to distinguish between NP referents.

2.1.3   Reference expectation

The *reference expectation* (occasionally 'posterior reference expectation' for emphasis) models the reader's expectation about what a referring expression, in this case a pronoun, refers to. This is the latent parameter used in the behavioral model (Section 2.3) to predict participants' responses.

The reference expectation is computed in two stages: combination and normalization. The combination stage simply computes the pointwise product of product of *conditional top-down expectation* and *bottom-up match* gives the *un-normalized reference expectation* of a referent. This is not a proper distribution, although with just

this much information, a given pair of referents can be directly compared in terms of the ratio of these joint probabilities.

There are two equivalent ways to normalize and obtain the posterior reference expectation. One is to sum over all possible referents:

$$\text{reference expectation} \equiv \hat{p}(r_i \mid \text{RE}, T, C)$$
$$= \frac{p_{\text{TD}}(r_i \mid T, C) \times p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, T)}{\sum_i^{R_T} [p_{\text{TD}}(r_i \mid T, C) \times p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, T)]}$$

However, this is unsatisfying as a cognitive model, as it requires summing over a large (potentially unbounded) set of referents. The same posterior may also be more elegantly estimated as:

$$\hat{p}(r_i \mid \text{RE}, A, C) = \frac{p_{\text{TD}}(r_i \mid A, C) \times p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, A)}{p(\text{RE} \mid A, C)}$$

$$\approx \underbrace{\overbrace{\frac{p_{\text{TD}}(r_i, A \mid C)}{p_{\text{TD}}(A \mid C)}}^{\text{top-down prior expectation}} \times \overbrace{\frac{p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, A)}{p(\text{F}_{\text{RE}} \mid A)}}^{\text{bottom-up evidence}}}_{\text{reference expectation}}$$

This is a more cognitively plausible normalization. Rather than directly normalizing over the set of all possible referents, the model simply:

1. Normalizes the prior (top-down) expectation of a type T reference to referent r_i by dividing by the expectation of the type[6], and,

2. Normalizes the (bottom-up) match likelihood by dividing by the expectation of the features of the referring expression[7].

---

[6]This adjusts for the context of the computation, in which the type is assumed.
[7]giving an *evidence* term

2.2 Experiments

2.2.1 Basic Design

The example item presented here co-varies a top-down expectation factor with a bottom-up match factor — a 2×2 factor design producing four conditions. For one of the conditions in each experiment, EMRE assigns a low posterior reference expectation to the target antecedent, which is then disambiguated by world knowledge, leading to surprisal. The experiment used as an example here is from Experiment 1 (Section 3.2), and co-varies (1) the *locality* of the target antecedent, and (2) the *gender agreement* between the pronoun and a distractor:

(11)   a.   *(target local + distractor's gender disagrees:)* Jill takes skiing lessons from James. Jill has been training alone, since James has a sprained wrist. He's expected to make a full recovery soon.

       b.   *(target non-local + distractor's gender disagrees:)* Jill takes skiing lessons from James. Since James has a sprained wrist, Jill has been training alone. He's expected to make a full recovery soon.

       c.   *(target local + distractor's gender agrees:)* Erik takes skiing lessons from James. Erik has been training alone, since James has a sprained wrist. He's expected to make a full recovery soon.

       d.   *(target non-local + distractor's gender agrees:)* Erik takes skiing lessons from James. Since James has a sprained wrist, Erik has been training alone. He's expected to make a full recovery soon.

The top-down expectation of the target (*James*) is manipulated by varying locality — whether the clause containing the target immediately precedes the clause containing the referring expression 'He' (local; higher top-down expectation), or is in a prior clause (non-local; lower top-down expectation). The bottom-up match of the distractor (*Erik / Jill*) with the critical pronoun ('He') is manipulated by varying gender agreement — whether the distractor has the opposite gender as the pronoun (and target) (e.g., *Jill*; low bottom-up match) or the same gender (e.g., *Erik*; high

bottom-up match). In the completion of the sentence the identity of the target an- antecedent (*James*) is clearly disambiguated by world knowledge (*James* is the one who is injured, and could be expected to make a recovery). To the extent that the subject expected the distractor, surprisal will result

Many of the experiments employ the basic 2×2 design described above, in which the two factors in the design vary the relative top-down expectation and bottom-up match of the competing antecedent. In the example item above (11), bottom-up match is varied by manipulating gender agreement, and top-down expectation is varied by manipulating locality.

As explained in Section 2.1.2, *'bottom-up match'* is used in the present work to refer to an expectation — specifically, the expectation that, when mentioned, an antecedent will be referred to using a referring expression with particular features — in this case, the gender of the pronoun. Let the expectation that a reference to a person (e.g., *James*) be in the form of a gender-appropriate pronoun (*'he'*) be denoted as p_{pronoun}. Since bottom-up match is based on a stripped-down feature set, and not on particular context-dependent properties, this should be the same for *Erik* and *James*. Similarly, the expectation that a person would be referred to with a gender-inappropriate pronoun (e.g., *Jill* being referred to as *'he'*) be denoted as ϵ. In this case, p_{pronoun} would be moderate (bottom-up match probably between 0.2 and 0.8, and ϵ is very small (bottom-up match probably less than 10^{-3} — greater than the probability of referring to *James* as *'a full recovery'*, but not by much).

'Top-down expectation' is also an expectation — namely, the expectation that an antecedent will be mentioned (using any expression at all) at a given location. The four top-down expectation values that occur in this item represent probabilities that, at the critical region (the location of the pronoun in the final sentence) there will be a reference (of any form) to the antecedent, or to the distractor. The first top-down value, which will be denoted as $p_{\mathrm{local\ target}}$ is the probability of referring to the target antecedent (*James*) if the (non-prominent) clause in which he is most recently named is local (i.e., the clause immediately preceding the pronoun). The second, $p_{\mathrm{nonlocal\ target}}$ is the probability of referring to *James* if he is most recently mentioned

in a non-prominent, non-local clause. The other two top-down values are $p_{\text{local distractor}}$ and $p_{\text{nonlocal distractor}}$, and are similarly defined, and assumed to be the same for *Jill* and for *Erik*. These values are not known precisely, but are clearly neither zero (in which case the antecedent would be unmentionable, even by its full proper name), nor near unity (in which case it would be unacceptable to refer to anything but the antecedent).

The following argument assumes that $[p_{\text{local target}} > p_{\text{nonlocal target}}]$ and $[p_{\text{local distractor}} > p_{\text{nonlocal distractor}}]$ by some non-trivial margin. Given these definitions, the likelihoods of the target and distractor in each condition are given below. (NB, likelihoods are not probabilities):

Gender	Target	Target Likelihood	Distractor Likelihood
Disagrees	Local	$p_{\text{pronoun}} \times p_{\text{local target}}$	$\epsilon \times p_{\text{nonlocal distractor}}$
Disagrees	Non-local	$p_{\text{pronoun}} \times p_{\text{nonlocal target}}$	$\epsilon \times p_{\text{local distractor}}$
Agrees	Local	$p_{\text{pronoun}} \times p_{\text{local target}}$	$p_{\text{pronoun}} \times p_{\text{nonlocal distractor}}$
Agrees	Non-local	$p_{\text{pronoun}} \times p_{\text{nonlocal target}}$	$p_{\text{pronoun}} \times p_{\text{local distractor}}$

Table 2.1: Target and Distractor Likelihoods

Using these values, it is possible to calculate the competition in each condition, and normalize the target likelihood to give the model's posterior reference expectation (the estimated probability that the pronoun refers to the target). To simplify the representation, this is given with all multiples of ϵ simplified to $k \times \epsilon$ (with $k < 1$):

Gender	Target	Posterior Target Probability
Disagrees	Local	$1 - k \times \epsilon$
Disagrees	Non-local	$1 - k \times \epsilon$
Agrees	Local	$p_{\text{local target}} \times \left(p_{\text{local target}} + p_{\text{nonlocal distractor}}\right)^{-1}$
Agrees	Non-local	$p_{\text{nonlocal target}} \times \left(p_{\text{nonlocal target}} + p_{\text{local distractor}}\right)^{-1}$

Table 2.2: Posterior Reference Expectation

If $[p_{\text{local target}} > p_{\text{nonlocal target}}]$ and $[p_{\text{local distractor}} > p_{\text{nonlocal distractor}}]$, then the posterior target probability must be greater when the target is local than when the target is non-local.

In qualitative terms:

- In conditions where the gender of the pronoun and the distractor disagree,

reference competition from the distractor is very low (near zero), and hence in both cases the posterior probability of the target antecedent is very high (near unity). When the target is disambiguated, this will produce no surprisal.

- In conditions where the gender of the pronoun and the distractor agree, the distractor produces reference competition. In this case the listener's expectations about antecedents depend more strongly on top-down expectation, so the target antecedent will have higher posterior reference expectation in the condition where top-down expectation is higher. When the target is disambiguated, surprisal will be lower in this condition, and participants should report higher acceptability.

As described in Chapter 3, this interaction is indeed reflected in human participant ratings in Experiments 1 and 3. The experimental participants' predicted acceptability judgments in this example are followed up on in Section 2.3, which uses hypothetical values for $p_{\text{nonlocal target}}$ &c. to describe in detail the predictions of the behavioral model used for such predictions.

2.2.2 Factor selection

Many factors should produce measurable effects on both top-down expectation and bottom-up match. The factors used in the experiments were selected on the basis that they were: (1) easily manipulated in a very short text; and (2) suggested by patterns observed in the GraphBank corpus.

The following examples suggest other factors that could be varied in experiments similar to the ones presented in this work.

Example 12 varies the *locality* of two possible referents (a cue to top-down expectation, which is independent of the properties of the referents) with the *plausibility* with which the referents could have performed the actions (a cue to bottom-up match, which could be estimated independently of the context in which the referent is mentioned (*cf.* Trueswell et al., 1994).

(12) ACTOR LOCALITY × ACTION PLAUSIBILITY

a. Mark was hiking through the woods. Up ahead, he could see Susan with her teenage son David. Suddenly, she screamed "Run!". He did so, and climbed up a tall tree.

b. Mark was hiking through the woods. Up ahead, he could see Susan with her teenage son David. Suddenly, she screamed. He did too, and climbed up a tall tree.

c. Mark was hiking through the woods. Up ahead, he could see Susan carrying her infant son David. Suddenly, she screamed "Run!". He did so, and climbed up a tall tree.

d. Mark was hiking through the woods. Up ahead, he could see Susan carrying her infant son David. Suddenly, she screamed. He did too, and climbed up a tall tree.

Here, the presumed reading of the items, in which the final condition is predicted to be less acceptable, is based on the assumptions that: (1) when the action is compatible with the referent (*e.g.*, an infant screaming, a teenager either screaming or running), the reference expectation will favor the local referent (*David*); whereas, (2) in the third condition, where the action/actor pair is implausible (an infant running), the bottom-up plausibility match will lead the reader to favor the interpretation that the pronoun *'he'* refers to *'Mark'*. In the final condition, the reader is predicted to have a higher expectation that the pronoun refers to the infant, which will produce surprisal when the next action is that he climbs a tree.

The next example of another application of this model, Example 13, crosses noun case (or possibly thematic role, as the two are confounded here), with gender agreement. As observed by Kehler (2002) and others, there is a preference for the antecedent of a pronoun to be in the same syntactic position as the pronoun itself. This is a cue to top-down expectation, which does not depend on the particular referent in the favored position. As before, the bottom-up cue is gender agreement.

(13) NOUN CASE × GENDER AGREEMENT

a. Rick spotted Sue. Then Ann greeted her.

b. Sue spotted Rick. Then Ann greeted her.

   c. Betty spotted Sue. Then Ann greeted her.

   d. Sue spotted Betty. Then Ann greeted her.

The dependent measure of interest here would be experimental participants' report of which antecedent the pronoun referred to (*i.e.*, *'Sue'* or *'Rick'/'Betty'*, possibly with *'Ann'* included as a comprehension/compliance test). Also, the target is never meaningfully non-local, so no main effect due to locality would be predicted. The prediction of EMRE is that participants would report that the pronoun *'her'* referred to *'Sue'* in both non-ambiguous conditions. However, in the last two conditions, in which both referents have identical bottom-up match, EMRE predicts higher posterior expectation for the referent favored by the top-down cue of being in the same syntactic position.

Here, as in Experiment 5 (Chapter 5), there is no disambiguation after the referring expression, and consequently, so no surprisal is predicted. The behavioral model therefore predicts that acceptability should be comparable in all conditions

Example 14 crosses locality (a top-down cue) with lexical aspect, which, based on observation of GraphBank (Wolf et al., 2003), is predicted to be a bottom-up match cue in temporal relations:

(14) LOCALITY × LEXICAL ASPECT

   a. Though Bill argued against it, Frank insisted that 0 was an even number. Next, he tried using the chalkboard.

   b. Though Bill didn't believe it, Frank insisted that 0 was an even number. Next, he tried using the chalkboard.

   c. Though Bill insisted that 0 was an even number, Frank argued against it. Next, he tried using the chalkboard.

   d. Though Bill insisted that 0 was an even number, Frank didn't believe it. Next, he tried using the chalkboard.

Here again, the dependent measure of interest would be participants' reports of which referent *'he'* refers to. EMRE predicts that in the first two conditions there

would be a strong preference for the local referent. However, in the second pair of conditions, EMRE predicts there to be a stronger preference for the local referent in the third condition than in the final condition. This is because the lexical aspect of the verb *'believe'* is non-eventive, making it a worse antecedent to *'Next'* than *'argued'*.

2.3 Behavioral model

This work uses a two-stage behavioral model to predict partipant responses (Jaeger, 2006; Jaeger and Levy, 2006; Levy and Jaeger, 2007). The first stage, EMRE, models participants' reference expectations about the targets of referring expressions. This is then used in a linking function, which predicts the pattern of responses based on EMRE's modeled reference expectations.

Where the experimental items involve material that *disambiguates* the target of a referring expression, the model's reference expectation is also used to estimate *surprisal*, which following Levy (2008) is used as a predictor of processing difficulty (Section 2.3.1). When participants are asked to *report the target* of a referring expression, the linking function models their responses as *sampling* from the posterior reference expectation (Section 2.3.2). Additionally, the Dependency Locality Theory of Gibson (2000) predicts additional processing difficulty when there are referring expressions with non-local antecedents (Section 2.3.3). The full behavioral model is presented in Section 2.3.4. Section 2.3.6 closes by giving a concrete example of overview of the behavioral model using hypothetical parameters, with emphasis on what subprocesses of the tasks given to participants are modeled, and how various factors either affect or do not affect the predictions.

2.3.1 Surprisal as processing cost

As noted in Section 1.4.2, the measure of difficulty associated with Levy's expectation theory (2008) is *surprisal*. The surprisal associated with an observation is the KL divergence from the reader's prior expectations before the observation to the posterior

expectations after the observation[8]

In the experiments in which a target referring expression is disambiguated by the continuation of the sentence past the referring expression, the relevant predicted effect on the dependent measure was the participants' ratings of the acceptability of the text. The predictions were made on the basis of the *surprisal of the continuation* (or *SoC*). These experiments were designed such that EMRE predicted a clear interaction in on the SoC. In one baseline pair of conditions (*e.g.*, an unambigious pronoun, an unmarked causal relation), the other factor (*e.g.*, locality, syntactic prominence) was predicted to have a small effect (though see Section 2.3.3). In the other pair of conditions, some effect was introduced that was predicted to increase the SoC in one condition, but not another. (For a fully worked-out example of this, see Section 2.2.1).

2.3.2   Reference resolution as sampling

In Experiment 5 (Chapter 5), participants are asked to report what they think a temporal cue phrase refers to. Following the theory of Vul (2010), which asserts that the mind approximates inference by sampling, the linking function models the process of reporting the target of a referring expression as sampling from the reference expectation. This predicts that for conditions across which the model predicts significantly different reference expectations, participants should report significantly different referents.

Vul's sampling theory models sampling as a low-cost operation that does not collapse a distribution. Since sampling from a distribution is not predicted to change it, there should be relatively little surprisal associated with participants' responses, compared to experiments that involve disambiguation.

---

[8]... or equivalently, the negative log probability of the observation under the priors (Hale, 2001).

2.3.3 The problem of locality

Difficulty with long-distance syntactic dependencies is a well-established and often-observed property of human syntactic processing (Gibson, 1998, 2000; Grodner and Gibson, 2005). This work presents evidence suggesting that readers also have difficulty with long-distance *non-syntactic* dependencies, above and beyond what can be explained in terms of rational probabilistic expectation.

The effect of syntactic non-locality can be observed in the difference in word-by-word reading times between Examples 15a and 15b.

(15)   a.   The reporter$_i$ who $\_i$ attacked the senator disliked the editor.

       b.   The reporter$_i$ who the senator attacked $\_i$ disliked the editor.

Both sentences include a relative clause that contains a *gap*, a syntactic construct representing the relationship between the element extracted from a latent non-relativized clause (*e.g.*, the 'reporter' in in 15), and the position of the missing verb complement, where in the extracted element would have been in an unextracted version of the clause. (*e.g.*. "the reporter attacked the senator" in 15a).

Subject-extracted relative clauses (RCs) like 15a are generally read more quickly than object-extracted RCs like 15b, in which readers slow down in the vicinity of the gap (Gibson, 2000; Grodner and Gibson, 2005). The dominant account of this effect is the Dependency Locality Theory of Gibson (2000) (hereafter, DLT). DLT explains this difference, and many similar slowdown effects across many types of structure in multiple languages , in terms of a processing difficulty that increases with *non-locality*: the distance between: (1) the new element h_2 that is being integrated into a partial syntactic structure; and, (2) h_1, the head of the phrase to which it is being attached.

DLT provides a theoretical account of this processing difficulty, describing it as the result of: (1) structural integration costs due to the difficulty of construction a dependency that spans intervening referents; and, (2) retrieval costs due to the difficulty of accessing a distant referent in memory. According to the DLT, the structural integration cost of adding a new head h_2 (*e.g.*, a word, a gap) is a function of the complexity of the syntactic computation that took place since the reader encountered h_1.

61

DLT includes a straightforward model for estimating the structural integration cost: the cost of integration can be estimated as an increasing linear function of the number of new discourse referents between h_1 and h_2. This straightforward model of integration cost has been empirically successful at predicting reading times (Grodner and Gibson, 2005; Demberg and Keller, 2008).

Unlike the many aspects of reading behavior that have been shown to be well-predicted by expectation (see Section 1.4.2) locality has remained stubbornly hard to explain away as the result of rational expectation (Levy, 2008; Demberg and Keller, 2008, 2009; Levy and Keller, 2010).

Demberg and Keller (2008), used a large regression model to investigate to what extent reading times in the Dundee eye-tracking corpus could be predicted using a syntax model, as well as other factors including word length, word frequency, word position in the sentence, frequency of the previous word, forward and backwards transition probabilities, and DLT dependency length. They found that all of these were predictive of reading time.

Some of these factors are most sensibly construed as expectation effects. Word frequency is trivially linked to expectation — the negative log frequency of a word is its unigram surprisal. Surprisal decreases with word position, because a reader progressing through a sentence is accumulating evidence about what is likely to follow. The syntax model, an unlexicalized probabilistic context free grammar, was used to calculate the surprisal of a word w_k in terms of its *prefix probability*. The prefix probability was defined as the marginal probability of the words $w_1 \ldots w_k$ of the sentence as summed over all possible latent syntactic structures that could have produced them. This was estimated by the summed probabilities of the incremental parses given by the Roark parser Roark (2001).

The three factors that could not be given a probabilistic interpretation were word length, frequency of the previous word, and dependency length. Word length, independent of frequency, will affect reading time for obvious psychophysical reasons. Frequency of the previous work is a simple proxy for spillover effects — a reader who just saw something surprising is likely to slow down for a moment while processing

62

catches up.

Dependency length, however, is an anomaly. Naively, it might seem trivial to assign a probabilistic interpretation to the effect of dependency lengths — long-distance dependencies are, after all, less frequent in text than local dependencies (Demberg and Keller, 2009). But the specific formulation of the DLT given by Gibson (2000) was found by Demberg and Keller (2008) to be a strong predictor of reading times for nouns, verbs, and auxillaries[9], and this formulation has nothing to do with probability or expectation.

This work finds a similar non-probabilistic effect of locality. When the results of Experiments 1 – 3 are analyzed, both in the regression analyses (Chapter 3) and the generative probabilistic model (Section ??), the effect of non-local dependencies is very large. Nor is this the result of long-distance dependencies being surprising — as explained in Section 3.5.1, the data are inconsistent with a surprisal based account of locality. It appears to have a large, independent, and non-probabilistic effect on how experimental participants rate the acceptability of stimuli containing referring expressions with non-local antecedents.

2.3.4   Full behavioral model

This model predicts participants' ratings of the texts on the basis of the independent measures, and some free parameters. It is used both for qualitative predictions of acceptability ratings that were tested using linear mixed effects regression, and quantitatively in the generative Bayesian models used to analyze the results of Experiments 1 and 3. The form of the full behavioral model for predicting behavioral responses in the experiments is:

$$\overline{rating} = X_0 + X_1 \times [local?] + X_2 \times F(SoC, ToA)$$

The terms are defined as follows:

---

[9]the effect from auxillaries was a second-order interaction, in which auxillaries decreased reading time for verbs with long-distance dependencies.

- $\overline{rating}$ = Estimated participant acceptability rating.

- X_0 = the intercept: the rating that an entirely acceptable text resembling the item could be expected to get given the pool of participants.

- [*local?*] = Whether the target is local or not. This is the first and simpler of the two formulations of locality cost presented by Gibson (2000) — rather than counting the number of referents between the local and non-local elements, it simply indicates whether there are zero or more intervening elements[10]. (Section 2.3.3).

- $F(\text{SoC} - \text{ToA})$:

  - SoC = The *surprisal of the continuation*. This is the KL divergence from the subject's reference expectation of the pronoun and the final knowledge state in which the pronoun has been disambiguated. (Section 2.3.1).

  - ToA = The *threshold of acceptability*. This term is discussed below.

  - $F(SoC, ToA)$ ≡ The linking function. See Section 2.3.5.

- $X_{(1|2)}$ = the fixed effects of locality and surprisal.

The X_n terms are those used in any linear effects model. They are implicitly fit to the data in the LMER analysis used in most of the analyses, and sampled over in the generative models.

The other term that has not yet been introduced is the ToA, or threshold of acceptability. This pararameter is included in the behavioral model because while language must always entail surprisal[11], surprisal does not not always lead to unacceptability. Surprisal is predicted to be associated with processing difficulty because the reader has prepared for something other than what is observed. In situations such as, *e.g.*, the beginning of a sentence, when expectations about what will come next are extremely flat, the relatively high surprisal involved in narrowing in on much

---

[10][x] is the Iverson bracket — it is 0 if the value of the expression x is 0 or False or the empty set, and 1 otherwise (Iverson, 1962; Knuth, 1992).

[11]...at least to the degree that it is informative ...

more specific expectations obviously does not lead readers to find the beginning of every sentence unacceptable. However, as readers progress through a sentence — as they learn more about what the sentence says — their expectations about what will come next become progressively sharper, and their reading speed steadily increases.

As suggested by Levy and Jaeger (2007), this model only includes situations with *low entropy and high surprisal* — that is, when readers have relatively sharp expectations that must then be adjusted. Readers can manage some degree of surprisal and still retain what Oppenheimer and Frank (2007) call *'fluency'* — a "meta-cognitive experience of ease of processing" that will not lead to a percept of difficulty. This model aims to model only what could be termed *'unexpected surprisal'* — a construction that only appears contradictory until you note that any person that cannot perfectly predict the future will expect to see things that he or she could not have expected.

The threshold of acceptability was included for two reasons. The first was in order to capture this idea of expected vs. unexpected surprisal. The only effect it had on the linear regression analysis of the data was that, since the ToA is unknown, the behavioral model was able to make no predictions about the fixed main effect of ambiguity. If it were zero, such that no SoC was acceptable, then the model would predict a main affect of gender agreement. If it were very high, such that any degree of surprisal was acceptable, then the model would predict no effects other than the main effect of locality. The second reason for including the ToA for use in the generative modeling, for which see Section 2.3.5.

As the experimental items were designed with the intention that they be unsurprising[12], the only things in the text that should be unusually surprising are: (1) the continuation of the sentence past the pronoun, and (2) the nonlocal pronoun itself, even in unambiguous conditions. The second of these is addressed in detail in Section 3.5.1; the effect of the first is discussed further in the following section.

---

[12] ... if not to say boring ...

2.3.5 Linking function

Three main linking functions between the SoC and participant ratings were considered, and two tested using the generative model. The first was $F(SoC, ToA) \equiv R(SoC - ToA)$, where R is the ramp function $R(x) = x \times [x]$. This was motivated by Levy's demonstration that surprisal is the natural ranking of processing difficulty under the surprisal model, and so, adjusting for some ToA, it should have a linear effect on participant ratings.

The other two linking functions considered were: (1) $\log(SoC - ToA)$, as suggested by Stevens' power law[13]; and, (2) $R(e^{-SoC} - e^{-ToA})$, which simply converts the surprisals into natural probabilities, and thresholds in that space. However, since the log of surprisal is roughly linear across most of the range $[0, 1]$, these turned out to be substantially equivalent.

The second of the two (converting back to probabilities) was chosen on the basis that: (1) the log of surprisal, though it is mostly linear, has highly eccentric behavior at the extremes, which the generative model sometimes drifted into; and (2) using the raw probabilities is conceptually simpler.

2.3.6 Example of behavioral model

This section will further consider the example introduced in Section 2.2.1. The experimental materials for each condition were given in Example 11 (page 53), repeated here:

(11)   a.   *(target local + distractor's gender disagrees:)* Jill takes skiing lessons from James. Jill has been training alone, since James has a sprained wrist. He's expected to make a full recovery soon.

     b.   *(target non-local + distractor's gender disagrees:)* Jill takes skiing lessons from James. Since James has a sprained wrist, Jill has been training alone. He's expected to make a full recovery soon.

---

[13]Roughly, that the percept of a stimulus is proportional to the log of its objective magnitude.

c. *(target local + distractor's gender agrees:)* Erik takes skiing lessons from James. Erik has been training alone, since James has a sprained wrist. He's expected to make a full recovery soon.

d. *(target non-local + distractor's gender agrees:)* Erik takes skiing lessons from James. Since James has a sprained wrist, Erik has been training alone. He's expected to make a full recovery soon.

Also, posit that:

- Taken together, the marginal prior expectation $p(\text{target} \vee \text{distractor} \mid A)$ of an anaphoric reference to either the target or the distractor is $\frac{1}{5}$[14].

- The marginal likelihood of a reference to the local referent r_L, whether it is the target or the distractor, is twice that of the non-local referent r_{nL}. In this hypothetical example, this is a *fixed effect*, which is the same across experimental items.

- The target referent r_T (*'James'*), who broke his wrist, is slightly more interesting, and has $\frac{5}{4}$ marginal odds favoring an anaphoric reference to r_T over a reference to the distractor r_D.

- These two factors have independent effects on top-down prior expectation of a given referent r_i.

- The probability of using a gender-mismatched pronoun to refer to a person is $\epsilon \approx 1\text{E}^-3$.

---

[14]There are many other things that could be mentioned, *e.g.*, the skiing lessons, the broken wrist, a third person.

Conditioned on a referent to either the target or the distractor, this is:

$$p(A, \{r_T \lor r_D\}) = p_{A_x} = \text{ }^1\!/_5$$

$$p(\text{local} \mid A, \{r_T \lor r_D\}) = p_{\text{local}} = p_{x_1} = \text{ }^2\!/_3$$

$$p(\text{nonlocal} \mid A, \{r_T \lor r_D\}) = p_{\text{nonlocal}} = p_{x_2} = \text{ }^1\!/_3$$

$$p(\text{target} \mid A, \{r_T \lor r_D\}) = p_{\text{target}} = p_{y_1} = \text{ }^5\!/_9$$

$$p(\text{local} \mid A, \{r_T \lor r_D\}) = p_{\text{distractor}} = p_{y_2} = \text{ }^4\!/_9$$

Since the two factors are presumed to act independently on expectation, the joint probability of any combination of factors p_{x_a} and p_{y_a} is simply:[15]

$$
\begin{aligned}
p(x_1, y_1, A) &= p_{A_x} \times \frac{p_{x_a} \times p_{y_a}}{p_{x_a} \times p_{y_a} + (1 - p_{x_a}) \times (1 - p_{y_a})} \\
&= p_{A_x} \times \frac{p_{x_a} \times p_{y_a}}{p_{x_a} \times p_{y_a} + p_{x_b} \times p_{y_b}}
\end{aligned}
$$

giving joint top-down prior expectations of:

$$p(\text{local target}, A) = \text{ }^1\!/_7 \approx 0.143$$

$$p(\text{nonlocal target}, A) = \text{ }^2\!/_{35} \approx 0.123$$

$$p(\text{local distractor}, A) = \text{ }^8\!/_{65} \approx 0.077$$

$$p(\text{nonlocal distractor}, A) = \text{ }^1\!/_{13} \approx 0.057$$

In this case, the predicted posterior target probabilities for each condition are given in Table 2.3 (from the formulæ given in Table 2.2, reproduced below): At the end of sentence 2, before the reader has begun to observe sentence 3, the reader is modeled by EMRE as having a state of expectation about what might come next that includes, among many possible observations, the expectation that the next word will be an anaphoric reference the to target (*e.g.*, *'James'*, *'he'*), and an expectation that the

---

[15]x_a and x_b in the following expression simply mean "one marginal condition" and "the other one". (If this will not suffice, take $b = 3 - a$ with the definitions given above.)

Condition		Posterior Target Expectation	
Gender	Target	Algebraic	Numeric
Disagrees	Local	$1 - k \times \epsilon$	0.999
Disagrees	Non-local	$1 - k \times \epsilon$	0.999
Agrees	Local	$\dfrac{p_{\text{local target}}}{p_{\text{local target}} + p_{\text{nonlocal distractor}}}$	$^{13}/_{20} = 0.65$
Agrees	Non-local	$\dfrac{p_{\text{nonlocal target}}}{p_{\text{nonlocal target}} + p_{\text{local distractor}}}$	$^{13}/_{41} \approx 0.317$

Table 2.3: Example Posterior Reference Expectation

next word will be an anaphoric reference to the distractor (*e.g.*, *'Jill / Erik'* or an an appropriate pronoun).

The reader is also modeled by the behavioral model as having a certain level of acceptance of the text thus far. The texts are designed to read like normal, if slightly dull, English up to this point, so before the target pronoun has been observed the behavioral model predicts no differences in acceptability between conditions[16]. Using the data from Experiment 1, posit that the predicted rating of the text is the behavioral model's intercept $X_0 = 4.34$. Comparing this to the actual condition means observed in Experiment 1:

	Acceptability		
Condition	Actual	Estimate$_0$	δ
Local + Unambiguous	4.34	4.34	(0)
Local + Ambiguous	4.10	4.34	0.24
Non-local + Unambiguous	3.65	4.34	0.69
Non-local + Ambiguous	2.83	4.34	1.51

Table 2.4: Prediction: Intercept Only

From this point on, the behavioral model's predictions for the four conditions diverge, based on two independent factors: target non-locality, and the surprisal of the continuation.

The first factor, non-locality, is modeled as being equivalent in both conditions in which the target is non-local. This implicitly assumes that the memory integration

---

[16]This assumption could be behaviorally tested by collecting acceptability judgments on texts that include only the first two sentences of the item conditions, but this experiment has not been performed.

cost of assigning a non-local dependency is the same whether it occurs all at once, or after a period of uncertainty. In either condition, the cost of the non-local integration is modeled as $X_1 \times [local?]$ — a fixed cost associated with a fully-disambiguated non-local dependency. As explained in Section 3.5.1, this is a non-probabilistic memory integration cost, which is applied equally in both 'target non-local' conditions. A posteriori fitting from the results of Experiment 1 suggests this memory integration cost might hypothetically correspond to a difference of 0.69 points in the Likert-scaled acceptability judgements: Even after the linear effect of non-locality is residualized

| | Acceptability | | |
Condition	Actual	Estimate$_0$	δ
Local + Unambiguous	4.34	4.34	(0)
Local + Ambiguous	4.10	4.34	0.24
Non-local + Unambiguous	3.65	3.65	(0)
Non-local + Ambiguous	2.83	3.65	0.82

Table 2.5: Prediction: Intercept+Locality

away (as above), EMRE predicts a further difference between the two 'ambiguous' conditions based on the difference in posterior target expectation shown in Table 2.3. Assuming that the linking function between target expectation and acceptability judgements is the thresholded linear ramp function, in which any gap between target expectation and some threshold of acceptability ToAscales linearly by a factor X_2 with a decrease in target acceptability $(R(e^{-SoC} - e^{-ToA}))$; Section 2.3.5), the MAP estimates of the parameters (which fit the data perfectly, since the number of free parameters[17] is the same as the number of conditions) would be:

$$\text{acceptability intercept} \equiv X_0 = 4.34$$

$$\text{non-locality cost} \equiv X_1 = 0.69$$

$$\text{surprisal scaling factor} X_2 = 1.73$$

$$\text{threshold of acceptability} = \text{ToA} = 0.79$$

---

[17](discounting the expectation ϵ of referring to a person with a gender-mismatched pronoun)

Factors and time course of ratings

Many of the factors critical to predicting the ratings are not modeled using EMRE, but are instead treated by the behavioral model as fixed effects independent of pronoun reference expectation (*i.e.*, item and condition). These include:

- The acceptability X_0 of the text prior to the target pronoun.

- The world knowledge-based process through which a reader's reference expectation for a pronoun is narrowed to the target (which is the only possible referent that makes sense given the continuation of the sentence).

- The difficulty of integrating non-local dependencies.

The only sub-process involved in the task of rating the item conditions' acceptability that is directly modeled by EMRE is the reader's computation of the pronoun reference expectation. The only effect this expectation has on the predicted acceptability is based on the surprisal that results from theworld knowledge-based disambiguation of the pronoun that occurs as the continuation of the sentence is observed. This disambiguation process is far outside the scope of this of this model.

Given that EMRE is based on an unbounded-rationality model of top-down expectation, which uses all available world knowledge to calibrate reference priors, it might not be immediately clear why the world knowledge used in disambiguation is not part of the model domain. The simplest reason for this is that the pronoun reference disambiguation in the experimental materials cannot be explained by mere reference expectation, but instead depends on world knowledge like the fact that, *e.g.*, a sprained wrist is something that a person *'recovers'* from, while solo skiing practice will hopefully not require it. This simple explanation leads into a thicket of questions about the nature of computational models of cognition which, during the general discussion in Section 7.1, are shown to be even more perplexing than they might immediately appear.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Expectation of pronoun reference

As described in Section 2.2.1, EMRE can be used to derive empirical hypotheses about about how people will interpret referring expressions such as pronouns. The three experiments described in this chapter tested several of these predictions.

The chapter opens with an explanation of the motivation for the experiments (3.1). The next three sections (3.2–3.4) describe the experiments themselves, describing their design, execution and results. As the designs of these experiments are similar, their results are most clearly considered jointly, so discussion in sections 3.2–3.4 is kept to a minimum. Section 3.5 closes with a comprehensive discussion of the results of all three experiments, and what they might be interpreted to mean about about what referring expressions are.

3.1   Background

Pronoun antecedent resolution is a quintessential example of non-structural dependency resolution. A reader encountering a word like 'he', or 'it' in a sentence must determine what the pronoun is meant to refer to, else the sentence cannot be assigned a coherent interpretation. This section describes how EMRE models pronoun interpretation in texts like those used in Experiments 1 – 3[1]

---

[1]The numbers assigned to the experiments in this thesis are arbitrary labels — the order of presentation was chosen to improve clarity of explanation. The order in which the experiments were performed was: 1, 2, 7, 6, 3, 4+5.

As described in Chapter 2, this work models pronoun resolution as an expectation-based rational inference process. EMRE models the reader as having a *top-down expectation* that describes the probability that any referent (*e.g.*, a person, an event, an utterance) will be mentioned at a given point in the text (2.1.1). When the reader encounters a pronoun, the *bottom-up match* between the pronoun and the possible referents provides a certain degree of *evidence* for or against each of the possible referents included in the top-down expectation (2.1.2). The result of using the evidence to update the top-down expectation is a *reference expectation*, describing the reader's expectation that the pronoun refers to each of the possible referents (2.1.3).

The items in these experiments present readers with sentence-initial pronouns such as the word '*she*' in Example 16. In this example, the pronoun is potentially ambiguous, because there are two females named in the text: '*Julia*' (called the *target antecedent*, because she is the person to whom the pronoun will refer) and '*Kim*' (the *distractor*). EMRE models the reader as computing a reference expectation that specifies how probable it is that the pronoun refers to each of the two referents.

(16)   Kim is babysitting Julia on Saturday afternoon. Kim put a soccer ball and net in her car's trunk, because Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

As the reader continues with the sentence, it becomes clear that the antecedent must be '*Julia*',[2] the reference expectation becomes a delta distribution, with all the probability mass concentrated on the target antecedent. As described in Section 2.3, the difference between the initial and final expectations is associated with a *surprisal* value[3]. The greater the initial expectation of the target (or '*target expectation*') is, the smaller the surprisal associated with the continuation of the sentence (the '*surprisal of the continuation*', or '*SoC*') is. The proposed behavioral model (Section 2.3) predicts that texts in which the SoC is higher will be rated as less acceptable.

---

[2]Car-owning babysitters are not typically enrolled in kindergarten.

[3]The surprisal is equal to the KL divergence between the final and initial expectations (Levy, 2008) (or perhaps the continuously integrated equivalent (Jaynes, 1957)).

These experiments co-vary: (1) the gender agreement between the distractor referent and the pronoun, which affects bottom-up match; and, (2) the locality and/or syntactic prominence of the clause containing the target antecedent, which are modeled as informative cues to top-down expectation.

The conditions in which the distractor disagrees in gender are the baseline cases — they estimate how acceptable the text is in when the antecedent of the pronoun is immediately clear, and the the SoC is zero[4]. These baseline *'gender disagrees'* conditions are compared to corresponding conditions in which only change to the text is that the gender of the distractor referent is the same as that of the target.

If locality and prominence are *informative cues* that a reader can use to calibrate his or her top-down expectation of possible referents (in particular, the target and the distractor), then the SoC should be greater in conditions in which these cues are incorrectly informative about the antecedent than in those where the cues are correctly informative. If this is the case, then the behavioral model (Section 2.3.4) predicts that experimental participants will assign lower ratings to items presented in those conditions. In particular, there should be a statistical interaction in which the conditions predicted by EMRE to produce higher SoC should be less acceptable than would be predicted by the main effects estimated from the *'gender disagrees'* conditions, in which SoC is zero.

3.2 Experiment 1: Locality × Gender agreement

EMRE predicts that (1) the locality of a target referent, and (2) the gender agreement of a distractor referent will be used as cues to (1′) top-down expectation, and (2′) bottom-up match. This entails a pattern of target expectation that is predicted by the behavioral model to produce an interaction effect between these factors on acceptability. Experiment 1 tested for this interaction, and it was observed in the data.

---

[4] ... *i.e.*, the continuation of the target sentence past the pronoun generates no surprisal.

3.2.1 Design

Experiment 1 used a 2 × 2 factor design that co-varied: (1) the locality of the clause containing a target antecedent; with, (2) the gender agreement of a distractor with the target antecedent and its referring pronoun. Both the target antecedent and the distractor were people who were introduced and referred to using common gender-unambigous U.S. American first names.

The top-down expectations of the target and distractor were manipulated by placing one in a local clause (immediately preceding the key pronoun at the beginning of the following sentence), and the other in a non-local clause (preceding the local clause). The bottom-up match between the distractor and the pronoun was manipulated by varying whether the distractor had the opposite gender as the target antecedent and pronoun (low bottom-up match) or the same gender (high bottom-up match).

3.2.2 Materials

All 16 experimental items had the same format. The first sentence had one clause, which introduced the target and distractor by name. The second sentence had two clauses, with the target in the subordinate clause. The third sentence began with a pronoun, which was then disambiguated by the continuation of the sentence, which clarified that it referred to the target rather than the distractor. As explained above, the two factors varied were the locality of the clause containing the target, and the gender agreement of the distractor. Also included were 32 filler items which were similar to the experimental items but with slight structural variations. An example experimental and filler item from this experiment appear below[5]:

(17) SAMPLE EXPERIMENT 1 ITEM:

    a. *(target local + distractor's gender disagrees:)* Josh and Karen are in the same tennis club. Josh often skips practice to sleep in, while Karen arrives early every Saturday morning. She's always eager for a game.

---

[5]Full materials for all experiments are provided in Appendix A.

b. *(target non-local + distractor's gender disagrees:)* Josh and Karen are in the same tennis club. While Karen arrives early every Saturday morning, Josh often skips practice to sleep in. She's always eager for a game.

c. *(target local + distractor's gender agrees:)* Jenny and Karen are in the same tennis club. Jenny often skips practice to sleep in, while Karen arrives early every Saturday morning. She's always eager for a game.

d. *(target non-local + distractor's gender agrees:)* Jenny and Karen are in the same tennis club. While Karen arrives early every Saturday morning, Jenny often skips practice to sleep in. She's always eager for a game.

(18) SAMPLE EXPERIMENT 1 FILLER: Deborah and Walter are on the school board together. Walter cares only about athletics, while Deborah doesn't care about athletics at all. They often fight over the allocation of money.

Four Latin-square sets of item conditions were generated such that each participant saw exactly one condition for each experimental item. Each of these sets, along with the 32 filler items, was shuffled into two randomly ordered lists, for a total of eight presented lists of 48 stimuli. Each stimulus was paired with two 'Yes/No' comprehension questions, to check for participant compliance.

3.2.3   Participants

Each of the eight lists was presented to 10 human participants in the form of an online survey, giving 80 survey responses. Participants were selected and paid via Amazon's Mechanical Turk crowd-sourcing system. Participants were screened to have internet addresses within the United States, and to be self-reported native English speakers. When a participant left more than 10% of the items blank, answered more than 25% of the comprehension questions incorrectly, or responded to more than one list in the experiment (confounding the results due to exposure to multiple versions of the same item), all data from that participant were excluded from analysis. 10 participants were excluded in this manner, leaving 70 usable survey responses.

3.2.4 Predictions

The behavioral model predicts an interaction between target locality and gender agreement. Specifically, the model prediction is that that participants should rate items in the *'target non-local + distractor gender agrees'* condition as significantly less acceptable than would be predicted by the main effects observed for locality and gender.

The specific assumptions entailing this model prediction are:

1. Before encountering the key pronoun (*e.g.*, *'She's'* in Example 17), reader has a higher top-down expectation of a reference to the local antecedent (whether it is the target or the distractor) than of a reference to the non-local antecedent (Section 2.1.1).

2. If the two antecedents have the same gender, the reader has roughly the same null context expectation that either would be referred to with a (gender-agreeing) pronoun (Section 2.1.2).

3. The reader has a very low expectation that a person will be referred to with a gender-disagreeing pronoun ($p_{\text{mismatch}} < 1\text{E}^-3$).

4. Item conditions in which EMRE predicts a greater SoC for the final sentence will be rated as lower in acceptability (Section 2.1.3).

5. Additionally, DLT predicts a main effect of dependency locality across conditions (Section 2.3.3; (Gibson, 1998, 2000; Grodner and Gibson, 2005)).

If these are all the case, and if EMRE is a good model of how the reader will interpret the pronoun, then readers should find the item conditions with a non-local target and a gender-agreeing distractor less acceptable than would be predicted by the observed main effects.

3.2.5 Analysis

Data were analyzed using the R statistical language (R Development Core Team, 2010), including the lme4 package (Bates et al., 2008) for linear mixed effect regression (LMER) modeling, and the languageR package for calculating statistical significance (p) values using Highest Posterior Density (HPD) parameter estimation and Markov Chain Monte Carlo (MCMC) sampling from the resulting mixed effects model (Baayen, 2008).

The data were also separately analyzed using a structured generative Bayesian model implemented in the probabilistic programming language Church (Goodman et al., 2008). This analysis is described in Section ??.

3.2.6 Results

The interaction predicted by EMRE was clearly observed in the data ($p < 1\mathrm{E}^{-}4$). This interaction is visible from inspection of the mean acceptability rating for each condition (Table 3.1).

Condition	Acceptability
Local + Unambiguous	4.34
Local + Ambiguous	4.11
Non-local + Unambiguous	3.65
Non-local + Ambiguous	2.83
(Residual Std. Error)	0.12

Table 3.1: Experiment 1: Mean Acceptability / Condition

In the LMER estimates of the fixed effects, the interaction has a large effect — approximately $^4/_5$ the magnitude of the the main effect of locality (Gibson, 2000), which is the largest fixed effect, (Table 3.2).

The predicted interaction appears clearly in a linear effects plot of the acceptability means (Figure 3-1). This plot clearly illustrates that when the target referent is in the local clause, the difference between ambiguous and unambiguous conditions is quite small (though significant; $p = 2\mathrm{E}^{-}3$). There is a large main effect of locality, with both gender conditions rated significantly worse in the non-local condition, and

Fixed Effects	Estimate	StdErr	p-value
(Target Local + Distractor Disagrees)	4.34	0.113	0.0000
Target not Local	-0.693	0.0782	0.0000
Distractor Agrees	-0.240	0.0782	0.0022
Non-local : Agreement	-0.571	0.111	0.0000

Table 3.2: Experiment 1: Linear Fixed Effects (LMER)

gender/locality interaction
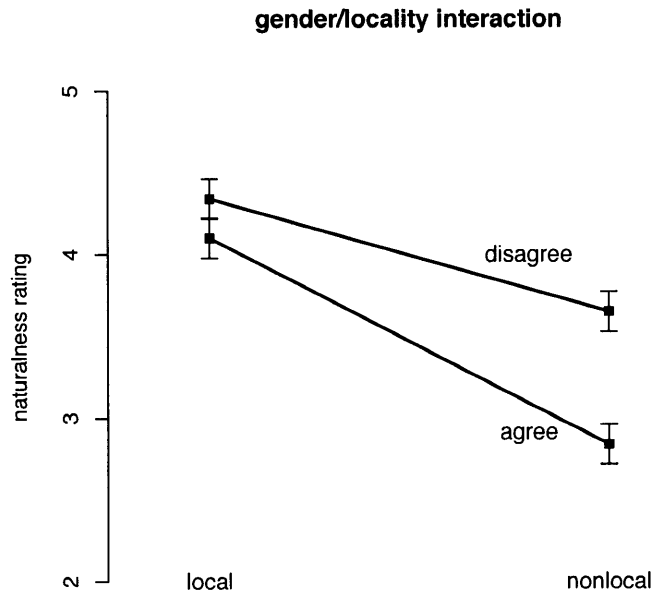
Figure 3-1: Experiment 1: Locality/Gender Interaction

a comparably large interaction, with the gender-ambiguous non-local condition worse than the unambiguous condition (both $p < 1\mathrm{E}^{-}4$).

3.2.7 Discussion

Observation of the predicted interaction gives evidence in favor of both: (1) the specific assumptions enumerated above; and, (2) EMRE being a useful computational model of how pronoun antecedents can be resolved. The interaction, in which the non-local target is rated as less acceptable when the distractor has the same gender as the target, can be explained by readers also using locality as an *informative cue*.

Readers know that non-local references are less likely (Anderson and Schooler, 1991), so they have lower expectation that an ambiguous pronoun will refer to them. If they discover that the pronoun did refer to the nonlocal entity, SoC will be greater than if they discover that it did not. As described in Section 2.3, greater surprisal is predicted to lead to lower ratings, which is what was observed.

The expected main effect of target locality was also observed. The overall pattern of fixed effects cannot be explained by modeling this main effect as due to surprisal, so this is best explained as a pure processing cost effect.[6] For further discussion, including data from this experiment and Experiments 2 and 3, see Section 3.5.

3.3 Experiment 2: Syntactic prominence × Gender agreement

EMRE predicts that (1) the prominence of a target referent, and (2) the gender agreement of a distractor referent will be used as cues to (1') top-down expectation, and (2') bottom-up match. This entails a pattern of target expectation that is predicted by the behavioral model to produce an interaction effect between these factors on acceptability. Experiment 2 tested for this interaction, but it was not observed in the data (though see also Experiment 3 in Section 3.4 below).

3.3.1 Design

Experiment 2 used a 2 × 2 factor design that co-varied: (1) the syntactic prominence of the clause containing a target antecedent; with, (2) the gender agreement of a distractor with the target antecedent and its referring pronoun. Both the target antecedent and the distractor were humans, introduced and referred to using common gender-unambigous U.S. American first names.

The relative top-down expectation of the target and distractor was manipulated by placing one in the main clause of a sentence, and one in a subordinate clause of the

---

[6]See Section 3.5 for explanation of this point.

same sentence. The bottom-up match of the distractor was manipulated by varying whether the distractor had the opposite gender as the target antecedent and pronoun (low bottom-up match) or the same gender (high bottom-up match).

3.3.2 Materials

All 16 experimental items had the same format. The first sentence had one clause, and introduced the target and distractor by name. The second sentence had two clauses, with the target in the non-local clause. The third sentence began with a pronoun, which was then disambiguated by the continuation of the sentence as referring to the target rather than the antecedent. As explained above, the two factors varied were the syntactic prominence of the clause containing the target, and the gender agreement of the distractor. Also included were 32 filler items, similar to the experimental items but with slight variations in structure. An example experimental item appears below:

(19) SAMPLE EXPERIMENT 2 ITEM:

   a. *(target main clause + distractor's gender disagrees:)* Josh is Karen's next door neighbor. Karen is overseas doing factory safety inspections, so Josh is mowing both lawns. She won't be back for a month.

   b. *(target subordinate clause + distractor's gender disagrees:)* Josh is Karen's next door neighbor. Because Karen is overseas doing factory safety inspections, Josh is mowing both lawns. She won't be back for a month.

   c. *(target main clause + distractor's gender agrees:)* Jenny is Karen's next door neighbor. Karen is overseas doing factory safety inspections, so Jenny is mowing both lawns. She won't be back for a month.

   d. *(target subordinate clause + distractor's gender agrees:)* Jenny is Karen's next door neighbor. Because Karen is overseas doing factory safety inspections, Jenny is mowing both lawns. She won't be back for a month.

Four Latin-square sets of item conditions were generated such that each participant saw exactly one condition for each experimental item. Each of these sets, along

82

with the 32 filler items, was shuffled into two randomly ordered lists, for a total of eight presented lists of 48 stimuli. Each stimulus was paired with two 'Yes/No' comprehension questions, to check for participant compliance.

3.3.3 Participants

Each of the eight lists was presented to 10 human participants in the form of an online survey, giving 80 survey responses. Participants were selected and paid via Amazon's Mechanical Turk crowd-sourcing system. Participants were screened to have internet addresses within the United States, and to be self-reported native English speakers. When a participant left more than 10% of the items blank, answered more than 25% of the comprehension questions incorrectly, or responded to more than one list in the experiment (invalidating the results due to viewing multiple versions of the same item), all data from that participant were excluded from analysis. 8 participants were excluded in this manner, leaving 72 usable survey responses.

3.3.4 Predictions

The behavioral model predicts an interaction between target prominence and gender agreement. Specifically, the model prediction is that that participants should rate items in the *(target subordinate + distractor gender agrees)* condition as significantly less acceptable than would be predicted by the main effects observed for locality and gender.

The specific assumptions that produce this model prediction are:

1. Before the key pronoun, the reader has a higher top-down expectation of the prominent antecedent (whether the target or the distractor) than of the non-prominent antecedent.

2. If the two antecedents have the same gender, the reader has roughly the same null context expectation that either would be referred to with a (gender-agreeing) pronoun.

3. The reader has a very low expectation that any entity would be referred to with a gender-disagreeing pronoun ($p_{\mathrm{mismatch}} < 1\mathrm{E}^-3$).

4. Item conditions in which EMRE predicts a greater surprisal of the continuation of the final sentence will be rated as lower in acceptability (2.3).

If these are all the case, and if EMRE is a good model of how the reader will interpret the pronoun, then under the behavioral model described in Section 2.3, a reader should find the item conditions with a non-prominent target and a gender-agreeing distractor less acceptable than would be predicted given only main effects.

3.3.5 Analysis

Data were analyzed using the R statistical language (R Development Core Team, 2010), including the lme4 package (Bates et al., 2008) for linear mixed effect regression (LMER) modeling, and the languageR package for calculating statistical significance (p values) using Highest Posterior Density (HPD) parameter estimation and Markov Chain Monte Carlo (MCMC) sampling from the resulting mixed effects model (Baayen, 2008).

3.3.6 Results

The interaction predicted by EMRE was not present in the data — neither statistically ($p > 0.5$), nor in any observable trends.

Condition	Acceptability
Main Clause + Unambiguous	3.94
Main Clause + Ambiguous	3.27
Subordinate + Unambiguous	3.67
Subordinate + Ambiguous	3.06
(Residual Std. Error)	0.12

Table 3.3: Experiment: 2 Mean Acceptability / Condition

LMER estimates of the fixed effects, show small but significant main effects of both prominence ($p = 1\mathrm{E}{-}3$) and gender agreement ($p < 1\mathrm{E}^-4$), with the 'gender-
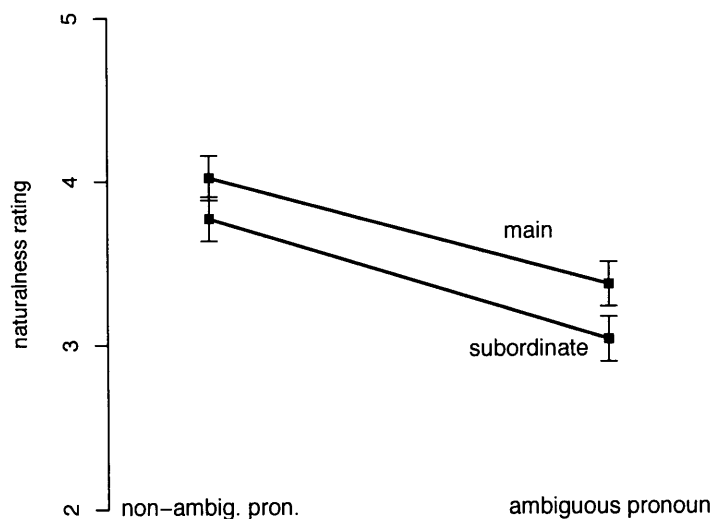
Figure 3-2: Experiment 2: Prominence/Gender (no interaction)

ambiguous' and 'subordinate' conditions rated as less acceptable, but counter to the
model predictions, these effects appear to be entirely independent. (Table 3.4).

Fixed Effects	Estimate	StdErr	p-value
(Main Clause + Gender Disagrees)	3.94	0.122	0.0000
Subordinate	-0.267	0.0780	0.0007
Gender Agreement	-0.665	0.0780	0.0000
Subordinate : Agreement	0.0531	0.110	0.631

Table 3.4: Experiment 2: Linear Fixed Effects (LMER)

The lack of an interaction between the two factors is shown by the virtually
identical slopes for prominent and non-prominent conditions in a linear effects plot
(Figure 3-2).

3.3.7 Discussion

The lack of an observed interaction between the two factors is evidence against the
account of participant responses in which pronoun reference is well modeled by EMRE

(2.1), and acceptability judgements by the behavioral model (2.3).

The divergence in results between Experiment 1 and Experiment 2 was difficult to interpret. Although the key conditions were the same between the two experiments (target antecedent in a non-local, subordinate clause) acceptability ratings of the key condition were quite different between the two experiments — 0.23 points lower in Experiment 1 than Experiment 2, which is comparable in magnitude to the fixed effects in these experiments[7]. The materials for Experiments 1 and 2 had been deemed similar enough to exclude subjects who had participated in Experiment 1 from participating in Experiment 2, but they were different enough that between-subjects, between-items comparisons were effectively uninterpretable. This confusion motivated Experiment 3, to clarify the divergence in results between the two experiments.

The results of Experiment 2 are discussed further in Section 3.5 below.

3.4 Experiment 3: Locality × Prominence × Agreement

EMRE predicts that (1) the locality and prominence of a target referent will be used as cues to top-down target reference expectation, and (2) the gender agreement of a distractor referent will be used as a cue to the distractor's bottom-up match. This entails a pattern of target expectation that is predicted by the behavioral model to produce two interaction effects on acceptability: one between locality and agreement, the other between prominence and agreement. Experiment 3 tested for these interaction, and they were observed in the data.

3.4.1 Design

Experiment 3 used a 2 × 2 × 2 factor design that co-varied all three of the factors varied in Experiments 1 and 2: (1) locality of the target antecedent; (2) syntactic

---

[7]The null hypothesis is of course that this type of between-subjects comparison is simply a random effect.

prominence of the target; and, (3) gender agreement of the distractor. As before, both target and distractor had common gender-unambigous American names.

As before, locality and prominence were both varied to manipulate the top-down expectation of the target and distractor, and gender agreement was varied to manipulate the bottom-up match of the distractor.

An additional element was added to the design, in order to test for a possible degenerate cause of the interaction observed in Experiment 1. This interaction could have been driven by a random effect due to responses in which participants had interpreted the distractor to be the intended antecedent of the pronoun in the key condition, or due to a subset of participants who did so systematically. Because the final disambiguating sentence was designed to be nonsensical if applied to the distractor, such responses would be expected to rate the text low in acceptability. EMRE predicts an interaction in the responses in which the participant construed the pronoun as referring to the target antecedent. However, responses in which the distractor was construed as the antecedent might be better explained by some other account, unrelated to expectation.

To test whether the interaction observed in Experiment 1 was independent of such responses, subjects were asked which of the two antecedents the pronoun referred to in each stimulus. As described below, analysis was limited to responses in which the subject reported that the pronoun referred to the target.

3.4.2 Materials

The format was largely the same as in Experiments 1 and 2, with 16 experimental items and 32 filler items. The most obvious difference was in the design, with each item having eight, rather than four, conditions.

Also, items in which there was a causal coherence relation between the two clauses in the second sentence (*e.g.*, Example 20, in which the clauses are joined by 'so' in the two *'nonlocal + main'* conditions, and 'because' in the four *'subordinate'* conditions), the clause boundary in the second sentence was replaced by a full stop in the two *'local + main'* conditions. This was done because there is no English-language

complementizer that marks the abstract referent of a subordinate clause as having been caused by the abstract event in a subsequent main clause.

(20) SAMPLE EXPERIMENT 3 ITEM:

    a. *(local + main + disagree):* John is taking Beth along on a cross-country trip. John has to do all the driving. Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

    b. *(nonlocal + main + disagree):* John is taking Beth along on a cross-country trip. Beth can't work the clutch without stalling, so John has to do all the driving. She has never driven a car with a stick shift.

    c. *(local + subordinate + disagree):* John is taking Beth along on a cross-country trip. John has to do all the driving because Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

    d. *(nonlocal + subordinate + disagree):* John is taking Beth along on a cross-country trip. Because Beth can't work the clutch without stalling, John has to do all the driving. She has never driven a car with a stick shift.

    e. *(local + main + agree):* Susan is taking Beth along on a cross-country trip. Susan has to do all the driving. Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

    f. *(nonlocal + main + agree):* Susan is taking Beth along on a cross-country trip. Beth can't work the clutch without stalling, so Susan has to do all the driving. She has never driven a car with a stick shift.

    g. *(local + subordinate + agree):* Susan is taking Beth along on a cross-country trip. Susan has to do all the driving because Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

    h. *(nonlocal + subordinate + agree):* Susan is taking Beth along on a cross-country trip. Because Beth can't work the clutch without stalling, Susan has to do all the driving. She has never driven a car with a stick shift.

Eight Latin-square sets of item conditions were generated such that each participant saw exactly one condition for each experimental item. Each of these sets,

along with the 32 filler items, was shuffled into two randomly ordered lists, for a total of eight presented lists of 48 stimuli. Each stimulus was paired with two 'Yes/No' comprehension questions, to check for participant compliance, and a question about which referent was the antecedent of the target pronoun.

3.4.3 Participants

Each of the eight lists was presented to 10 human participants in the form of an online survey, giving 80 survey responses. Participants were selected and paid via Amazon's Mechanical Turk crowd-sourcing system. Participants were screened to have internet addresses within the United States, and to be self-reported native English speakers. When a participant left more than 10% of the items blank, answered more than 25% of the comprehension questions incorrectly, or responded to more than one list in the experiment (invalidating the results due to viewing multiple versions of the same item), all data from that participant were excluded from analysis. 6 participants were excluded in this manner, and 8 more on the basis of not having English as a primary language, leaving 66 usable survey responses.

3.4.4 Predictions

The behavioral model predicts that readers' reference expectation for the pronoun will produce: (1) an interaction between target locality and gender agreement, and (2) an interaction between target prominence and gender agreement. Additionally, if locality and prominence are used as independently informative cues, EMRE predicts no three-way interaction between these factors and gender agreement (see Section 3.5 for proof).

Locality, which is known to contribute to processing difficulty in syntactic dependencies (Gibson, 1998, 2000), was also presumed likely to have effects on the ratings. Locality produced a strong main effect in Experiment 1, and was expected to do so here as well.

3.4.5 Analysis

Data were analyzed using the R statistical language (R Development Core Team, 2010), including the lme4 package (Bates et al., 2008) for linear mixed effect regression (LMER) modeling, and the languageR package for calculating statistical significance (*p* values) using Highest Posterior Density (HPD) parameter estimation and Markov Chain Monte Carlo (MCMC) sampling from the resulting mixed effects model (Baayen, 2008).

3.4.6 Results

Both of the interactions predicted by EMRE were clearly present in the data, as was the main effect of locality. There was no three-way interaction — the two-way interaction effects combined linearly. A large main effect was also observed for non-locality. No other significant effects of any type were detected in the data. Restricting analysis to only responses that indicated that the pronoun referred to the target antecedent had a negligible effect, as these constituted 0.99 of the non-empty responses.

As can be seen by inspection of the first three rows of Table 3.5, there is no main effect for either syntactic prominence or gender agreement. These rows show the means of conditions that involve none of the three effects (non-locality and the two interactions), and the values are indistinguishable. This table also clearly shows the large main effect of non-locality (0.66 rating points) — all local conditions are rated more acceptable than all non-local conditions.

Condition	Acceptability
local + main + disagree	4.42
local + main + agree	4.34
local + subord. + disagree	4.48
local + subord. + agree	4.06
nonlocal + main + disagree	3.76
nonlocal + main + agree	3.12
nonlocal + subord. + disagree	3.69
nonlocal + subord. + agree	2.97
(Residual Std. Error)	0.12

Table 3.5: Experiment 3: Mean Acceptability / Condition

gender/locality interaction
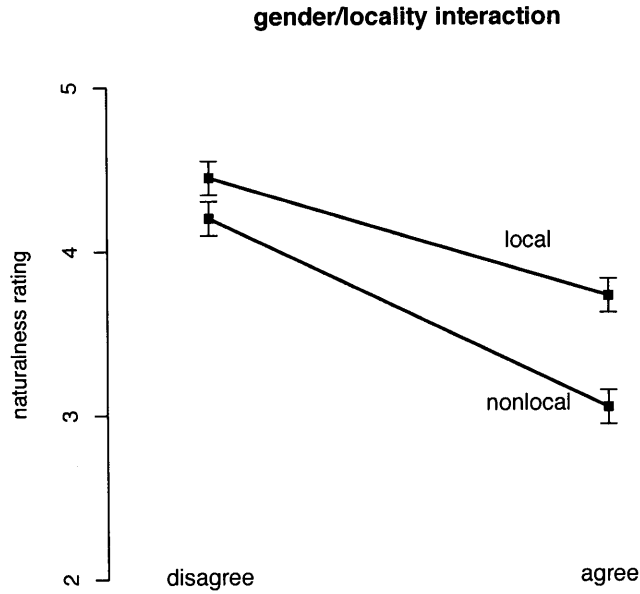
Figure 3-3: Experiment 3: Gender / Locality Interaction

As the table of fixed effects (Table 3.6) shows, there is an interaction effect between non-locality and gender agreement that is only slightly smaller than the non-locality effect (0.56 points). Gender ambiguity has no main effect on *'local'* conditions, but the ratings effect of locality in gender-ambiguous conditions is roughly twice that in unambiguous conditions (Figure 3-3).

Fixed Effects	Estimate	StdErr	p-value
(local + main + disagree)	4.42	0.117	0.0000
nonlocal	-0.662	0.117	0.0000
subordinate	0.0578	0.117	0.620
agree	-0.0768	0.117	0.512
nonloc : subord.	-0.125	0.165	0.448
nonloc : agree	-0.558	0.167	0.0009
subord. : agree	-0.339	0.166	0.0411
nonloc : subord. : agree	0.253	0.235	0.282

Table 3.6: Experiment 3: Fixed Effects (LMER)

The final effect observed was the predicted interaction between syntactic promi-

prominence/gender interaction
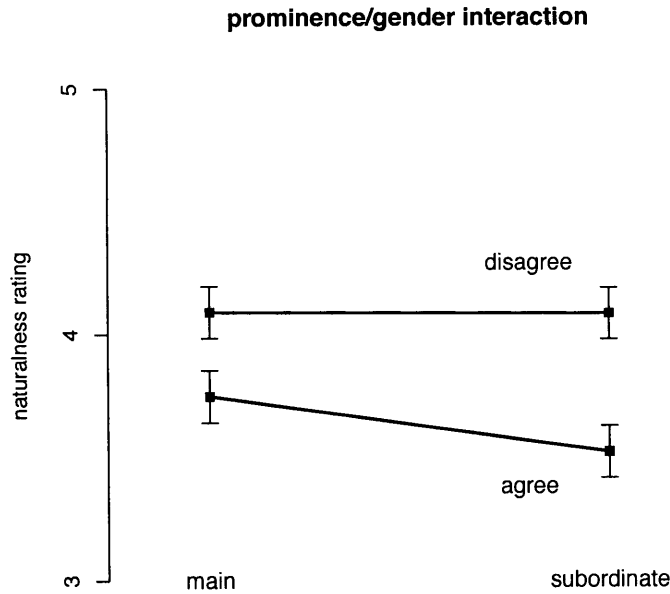
Figure 3-4: Experiment 3: Prominence / Gender Interaction

nence. This is a smaller effect (0.33 rating points — roughly half the the locality effect). But unlike the gender/locality interaction, which involves a factor with a large main effect, this interaction is produced by two factors neither of which is associated with a main effect. Note for instance that in the linear effects plot illustrating the prominence:gender interaction (Figure 3-4), the line for the gender-disagreeing conditions is essentially flat, reflecting the null effect of prominence in those conditions.

The points generating this flat effects line combine both 'disagree + local' conditions and much lower-rated 'agree + local' conditions, but this is hidden by the minimal interaction between these effects.

The lack of main effects for prominence and agreement is more clearly illustrated by considering only the ''local' conditions, as an independent data set. As shown in the table of fixed effects for this restricted data set (Table 3.7), the interaction between these factors is fairly large (0.36 rating points — about half the size of the locality main effect) while the fixed effects for the two factors both have an observed
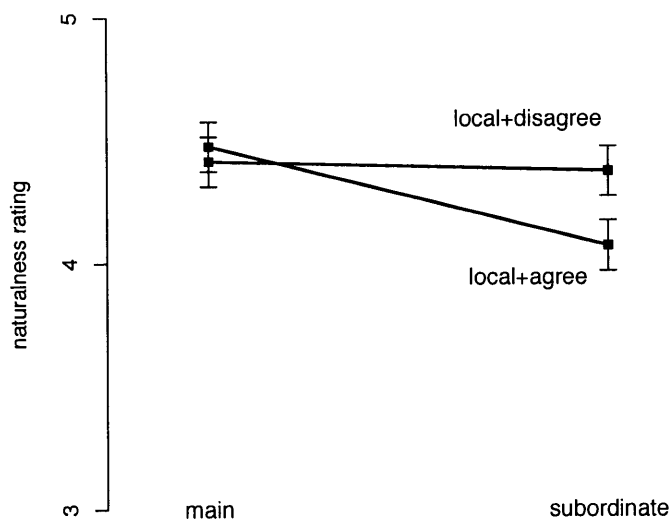
Figure 3-5: Experiment 3: Prominence / Gender Interaction (Local Conditions Only)

magnitude around 0.05 points, and $p > 0.5$.

Fixed Effects	Estimate	StdErr	p-value
(main + disagree)	4.41	0.102	0.0000
subordinate	0.0618	0.0878	0.546
gender agreement	-0.0407	0.0882	0.692
subordinate : agreement	-0.364	0.125	0.0126

Table 3.7: Experiment 3: Fixed Effects — 'Local' conditions only

The lack of main effects for these factors can also be observed in a linear effects plot of the local data (Figure 3-5), in which the interaction provides the only deviation from the intercept.

3.5 Discussion

This discussion will first address the predicted significant fixed effects observed in Experiments 1 and 3 (3.5.1). This is followed by a cautious discussion of possible

interepretations of the fixed effects which were neither predicted by the model nor observed to be significant in the data (3.5.2). Lastly, the results of Experiment 2 are considered, with an eye to those fixed effects that are inconsistent with the interpretation of the data described in the prior sections (3.5.3).

3.5.1 Significant effects: expectation and locality

Two different types of significant effects were observed in Experiments 1 and 3, and will be discussed separately. There were: (1) the interactions predicted by EMRE and the behavioral model; and, (2) the main effect of locality.

Predicted interactions

Of the significant fixed effects detected in Experiments 1 and 3, all but the main effect of locality are predicted by EMRE, and can be explained in terms of expectation. Also in these experiments, all the effects predicted by EMRE were observed. This is consistent with an account in which: (1) participants formed posterior reference expectations for the pronouns consistent with those predicted by EMRE; and, (2) these reference expectations led to surprisal in the *'gender agrees'* conditions that influenced their acceptability ratings. In this account, when a reader's rational expectations about the target of an ambiguous pronoun turned out to be incorrectly informative, the reader became confused and annoyed.[8]

This annoyance is parametrized in the behavioral model (Section 2.3.4) as the 'surprisal of the continuation' (SoC). The SoC predicted by the behavioral model is the KL divergence from the final disambiguated knowledge state[9] to the reference expectation generated by EMRE. The modeled surprisal predicts the observed interactions between top-down and bottom-up factors. As discussed further in Section 3.5.2, these data are consistent with an account in which readers treat factors like

---

[8]Some might object that when framed in these terms, the ratings of the readers seem logically obvious, and wonder whether EMRE is an overcomplicated solution to the problem. The precedent of the past several decades of research in cognitive science and artificial intelligence, however, suggests that describing a cognitive task as 'logically obvious' is seldom theoretically fruitful. (See Chapter 7 for further discussion of this point.)

[9](given the continuation of the sentence past the pronoun)

locality and prominence as *informative cues*, and respond to these cues with rational belief updates.

Spurious main effects of gender   While Experiment 3 detected no main effect of gender, such effects were observed in Experiments 1 and 2. The small main effect of gender agreement in Experiment 1 (-0.24 rating points; ($p < 1\text{E}^-2$), and the outsized effect in Experiment 2 (-0.66; ($p < 1\text{E}^-4$) do not not even appear as trends in Experiment 3. However, these two main effects are explained away by the interactions observed in Experiment 3.

In all of the experimental items in Experiment 2, the target is in a non-local clause; the *'locality : gender'* interaction observed in Experiments 1 and 3 predicts a large difference between both pairs of *'disagree'* and *'agree'* conditions in Experiment 2. This was indeed the case — as mentioned above, gender agreement produced the largest fixed effect observed in Experiment 2. Similarly, the items in Experiment 1 all had the target in non-prominent clauses, such that the *'prominence : gender'* interaction observed in Experiment 3 predicts the main effect of gender agreement observed in Experiment 1. In Experiment 3, in conditions where both prominence and locality are correctly informative about the target antecedent, no main effect of gender was observed.

Non-probabilistic locality effects

One question that could be asked about the behavioral model presented in Section 2.3.4 is why locality must be treated as a separate, non-expectation based effect. Three reasons for doing so are presented below. First, this approach respects current findings that locality-based processing effects are not well-explained by surprisal, but appear to be a separate processing cost. Second, this would entail including a latent parameter representing sentence-initial surprisal, which is not cognitively plausible. Third, this would produce a model that does not predict any of the observed interactions, but rather predicts the main effect of locality to be the only significant fixed effect, which would be both counterproductive and inconsistent with the observed

data.

Findings of other researchers Demberg and Keller (2008) found that a regression model which predicted reading times in the Dundee eye-tracking corpus found a significant effect of dependency length that could not be explained in terms of surprisal, but was well-accounted for by DLT (Gibson, 1998, 2000; Grodner and Gibson, 2005). This research is described in some detail in Section 2.3.3.

Sentence-initial surprisal The expectation model (Hale, 2001; Levy, 2008) predicts that when a reader encounters the sentence-initial pronoun in the experimental items, this will produce some surprisal. When the pronoun is unambiguous, the pronoun will generate more surprisal when the target is non-local than when the target is local, because the reader has more expectation that a local referent will be mentioned, and less expected events generate more surprisal than more-expected events.

What might be less obvious is that both of these conditions will have less surprisal than the gender-ambiguous conditions. When the two referents are of the same gender (let them both be female, to aid clarity of explanation), a mention of either of them could generate the pronoun 'she'. Expectation marginalizes over these possibilities: the expectation of 'she' is the sum of the expectations of all the different latent causes that could have led to it being the word observed. In this case, these possiblities would be a mention of the target and a mention of the antecedent, either of which could generate the word 'she'. As Levy puts it, the two referents *conspire* to produce a higher marginal expectation than in unambiguous cases. As discussed in Section 2.3.1, this can make ambiguous constructions less surprising than their unambiguous counterparts (Levy, 2008).

The surprisal of the sentence-initial pronoun is not included as a component of the behavioral model. This is because surprisal is high at the beginning of all sentences, and people appear to be insensitive to this. Before the start of a sentence there are typically many things that could sensibly be mentioned — expectation is much flatter than at later points in the sentence. Since the beginnings of all sentences are high

in surprisal, this surprisal is unlikely to be predictive of acceptability. This approach follows a suggestion of Levy and Jaeger (2007) that processing should be more difficult when there is low entropy and high surprisal than when both are high.

Model Accuracy Lemma 1 shows that the most straightforward approach towards a probabilistic account of non-locality integration costs produces a model that makes inaccurate predictions. Including the sentence-initial surprisal so as to give a surprisal-based (rather than non-probabilistic, integration cost-based) account of the main effect of non-locality produces a model that does not predict any of the interactions observed, but rather one that predicts no differences between conditions with gender ambiguity and those without:

Lemma 1. *The total surprisal from the start of sentence 3 in the experimental items of Experiments 1 and 3 (Chapter 3) is not affected by the gender of the distractor. As such, the pattern of fixed effects observed in the data is not consistent with interpreting the main effect of locality as due to surprisal.*

Proof. Let E_0 be the reader's top-down expectation over referents before encountering the pronoun at the start of sentence 3 in the experimental items.

$$E_0 = \left\{ \text{local ref.} : p_L, \text{non-local ref.} : p_N, x_0 : p_0, x_1 : p_1 \dots \right\}$$

The local and non-local referents both have a certain expectation, as do an uncertain number of other things that could possibly be mentioned instead.

In the *'gender disagrees'* conditions, the reference expectation after having read the pronoun and after having read the sentence continuation are identical — all probability mass has shifted to either the local or the non-local referent:

$$E_{\text{pronoun}} = E_F = \begin{cases} \left\{ \text{ref}_{\text{local}} : 1, \text{ref}_{\text{non-local}} : 0, x_0 : 0, x_1 : 0 \dots \right\} & \text{if target local} \\ \left\{ \text{ref}_{\text{local}} : 0, \text{ref}_{\text{non-local}} : 1, x_0 : 0, x_1 : 0 \dots \right\} & \text{if target non-local} \end{cases}$$

In such cases, the KL divergence from the posterior to the prior is simply the surprisal

97

of the outcome in question (or, framed differently, the information required to specify that hypothesis): its negative log probability. In the nonlocal case:

$$D_{KL}(E_F \parallel E_0) = 1 \log \frac{1}{p_N}$$
$$= -\log(p_N),$$

...and similarly for the *'local'* condition. In the *'gender agrees'* conditions, surprisal occurs in two stages. First, reference expectation for the pronoun is computed:

$$D_{KL}(E_{\text{pronoun}} \parallel E_0) = \frac{p_L}{p_N + p_L} \log \frac{\frac{p_L}{p_N+p_L}}{p_L} + \frac{p_N}{p_N + p_L} \log \frac{\frac{p_N}{p_N+p_L}}{p_L}$$
$$= \frac{1}{p_N + p_L} [P_L \log \frac{1}{p_N + p_L} + P_N \log \frac{1}{p_N + p_L}]$$
$$= -\log(p_N + p_L).$$

Since the pronoun does not provide more evidence for either referent over the other, their relative likelihood does not change. Consequently, the surprisal is simply the surprisal of the joint event *'either* ref$_{\text{local}}$ *or* ref$_{\text{non-local}}$*'* The second surprisal is associated with discovering the true referent of the pronoun. In the non-local case:

$$D_{KL}(E_F \parallel E_{\text{pronoun}}) = 1 \times \log \frac{1}{\frac{p_N}{p_N+p_L}}$$
$$= \log \frac{p_N + p_L}{p_N}$$
$$= \log(p_N + p_L) - \log p_N,$$

...and again, similarly for the *'local'* condition.

So the total surprisal across both events is $[-\log(p_N + p_L)] + [\log(p_N + p_L) - \log p_N]$, or simply $-\log(p_N)$, which is the same as the surprisal in the non-ambiguous condition. So regardless of the gender of the distractor, the total surprisal of the sentence is constant. $\square$

Including sentence-initial surprisal in the model would produce a model that predicted gender agreement to have no effect on acceptability. Since this is clearly not

the case, it is difficult to argue that this model would be superior to the behavioral model used in the work.

3.5.2 Non-significant effects in Experiment 3

The data observed in Experiments 1 and 3 suggest that the top-down factors manipulated in these experiments (locality and syntactic prominence) were used as informative, *non-interacting* cues. As predicted, the top down cues *do not interact* with each other, but do *independently interact* with the bottom-up cue. Modeling reference in terms of probabilistic expectation makes it possible[10] to predict *interactions between factors which independently generate no main effects.*

Out of all the possible fixed effects that could have been observed in Experiment 3, the only significant effects observed in the data were the main effect of locality, and the two interactions predicted by EMRE. This section will consider the possible fixed effects that were not detected, and explain why they are not called for by the behavioral model.

Null results are necessarily hard to interpret, and doing so calls for great caution. However, under an information-theoretic interpretation of evidence[11], absence of evidence is in fact (weak) evidence of absence. Every day that the sun fails to rise in the west provides one bit of evidence against the theory that the sun is equally likely to rise in the east or the west, and has merely happened to rise in the east thus far, and $3\mathrm{E}^-8$ bits against the theory that it will rise in the west with probability $1\mathrm{E}^-9$.[12] As such, null results do provide weak evidence for theories that predict no effect over theories that strongly predict an effect.[13]

Bearing in mind the risks of over-interpretation, this section discusses the fixed effects not observed to be significant in Experiment 3, in the interest of explaining

---

[10] (and, when possible, required)

[11] (the same intepretation as the one applied in EMRE)

[12] $1\mathrm{E}^-9 \times \log_2(1\mathrm{E}^-9) = -2.98973\mathrm{E}^-8$

[13] An analogy could be made between null effects and syntactic gaps. In a place where something might have been expected, one instead observes nothing. The missing element, though, can provide information as to the how the global context might be construed.

why are not predicted by the model[14]. First, an explanation is provided for why there is no reason to predict main effects of syntactic prominence or gender agreement, or for an interaction betwen prominence and locality. This is followed by an explanation of why the behavioral model also does not predict a three-way interaction between prominence, locality and agreement.

Prominence + Agreement

Syntactic prominence and gender agreement were not were not observed to have significant main effects in Experiment 3. Further, the 95% confidence intervals for these effects ($[-0.1638, 0.2768]$ for prominence, and $[-0.2996, 0.1415]$ for gender agreement) provide modest evidence that if these effects do exist, they are likely to be smaller than the effects that were observed to be significant. The behavioral model similarly does not predict main effects for prominence or agreement, or for their interaction.

Syntactic Prominence   Unlike locality, which has been shown to have measurable effects across a broad range of structures and languages (Gibson, 2000), syntactic prominence has no such known effect. This work models prominence not as a processing cost, but as a rational cue to reference. As such, there is no reason to expect that it would have a large effect in unambiguous conditions.

Gender Agreement   The case for not including a gender agreement parameter is slightly more complicated. While the model contains no terms for ideas like 'competition' or 'interference', it is certainly the case that EMRE predicts a greater SoC (surprisal of the continuation) for even a local, prominent antecedent when the pronoun is ambiguous than when it is not. Since neither locality nor prominence is a perfectly reliable cue, the the model will never assign zero expectation to the distractor, so the SoC will always be nonzero. However, when there is no ambiguity, the SoC is strictly zero.

---

[14] ... such that their failure to appear in the data could be construed as evidence against an expectation-based account of reference.

However, some level of surprisal is normal and even necessary in language comprehension — to provide any new information, an utterance must be somewhat unpredictable. At any given point in a sentence, there is a certain amount of surprisal that the reader is willing to accept as natural. For instance, as explained in Section 3.5.1 above, readers will take in stride a great deal of surprisal at the beginning of a sentence. Similarly, there may be some level of distractor reference expectation that a reader will accept as within the bounds of normal levels of surprisal.

This acceptable level of surprisal is represented in the behavioral model as the *'threshold of acceptability'* (ToA) parameter described in Section 2.3.4. This parameter is included in the behavioral model because to the extent that language is informative, it is not entirely predictable or expected, and will always involve some surprisal. If this threshold was very high, and readers were willing to accept very little SoC, there would be a main effect of gender across all pairs of conditions. If the ToA was very low, then participants might be willing to accept the SoC produced in even the conditions where it was predicted to be highest, in which case the only observed fixed effect would be the non-surprisal based main effect of locality (Section 3.5.1).

Since it is *a priori* unclear what ToA should be expected, there is no principled way to make a prediction. As it happened, the data are consistent with the ToA being approxmately equal to the SoC in the ambiguous condition with a local, prominent antecedent. It cannot be determined from these data whether this is due to random chance, or a replicable effect (perhaps due to some form of adaptation).

Locality : Prominence   As explained above (3.5.1), the main effect of locality is included in the model as a pure processing cost that is not due to surprisal, and which has been widely observed in empirical research (Gibson, 2000; Grodner and Gibson, 2005; Demberg and Keller, 2008). Just there is no similar effect known for syntactic prominence, neither have there been prior findings that this processing cost is affected by syntactic prominence. So these two factors are not predicted to interact in the non-ambiguous *'gender disagrees'* conditions.

Locality : Prominence : Gender

This section argues that the lack of a significant three-way interaction is consistent with the joint proposition that: (1) readers treat locality and prominence as independently informative cues (2.1.1); and, (2) there is a quasi-linear linking function between participants' target expectation and their acceptability ratings (see Section 2.3.5).

Independently informative cues   There was no significant three-way interaction between locality, prominence and gender, but as shown in 2, a null result is exactly what should be expected if locality and prominence are being treated as *independently informative cues* to top-down expectation.

Lemma 2. *Assuming that locality and prominence are independently informative about the expectation that a given referent is the antecedent of a pronoun, there will be no interaction between their effects on posterior expectation about a pronoun's antecedent.*

Proof: The proof follows directly from the definition of independence. Let p_{local} be the likelihood that the antecedent of a pronoun is local, and p_{prom} be the likelihood that it is syntactically prominent (in a main clause):

$$p_{local} = p(local(x) \mid antecedent(x, pronoun))$$
$$p_{prom} = p(prominent(x) \mid antecedent(x, pronoun))$$

The four conditions in Experiment 3 (Section 3.4) relevant to this lemma are the *'gender agrees'* conditions, in which there is a pronoun that could refer either to a target that is in one clause, or a distractor that is in another. If the local clause (adjacent to the pronoun is also the main clause of the sentence, the posterior expectation over

102

the two referents is:

$$\Big(p(local, main \mid pronoun), p(nonlocal, subordinate \mid pronoun)\Big)$$

$$=\left(\frac{p_{local} * p_{prom}}{p_{local} * p_{prom} + (1 - p_{local}) * (1 - p_{prom})}, \frac{(1 - p_{local}) * (1 - p_{prom})}{p_{local} * p_{prom} + (1 - p_{local}) * (1 - p_{prom})}\right)$$

These of course sum to one — there are no other plausible antecedents to the pronoun in the experimental items. If the local clause is subordinate, the posterior expectations are,

$$\Big(p(local, subordinate \mid pronoun), p(nonlocal, main \mid pronoun)\Big)$$

$$=\left(\frac{p_{local} * (1 - p_{prom})}{p_{local} * (1 - p_{prom}) + (1 - p_{local}) * p_{prom}}, \frac{p_{local} * (1 - p_{prom})}{p_{local} * (1 - p_{prom}) + (1 - p_{local}) * p_{prom}}\right),$$

which also sum to one. Interpreted as linear effects, the effect of locality is $[p(local, main \mid pronoun) - p(nonlocal, main \mid pronoun)]$, and the effect of non-prominence is $[p(local, main \mid pronoun) - p(local, subordinate \mid pronoun)]$. The value expected from linearly combining these two effects would be:

$$p(loc, main \mid pron) - [p(loc, main \mid pron) - p(\neg loc, main \mid pron)] -$$

$$[p(loc, main \mid pron) - p(loc, \neg main \mid pron)]$$

$$= \quad - p(loc, main \mid pron) - p(loc, \neg main \mid pron) + p(\neg loc, main \mid pron))$$

$$= \quad 1 - p(loc, main \mid pron),$$

which is of course equal to $p(nonlocal, subordinate \mid pronoun)$. $\square$

It is important to note that the null result is not even weak evidence that locality and prominence are in fact independently informative about the probability that a discourse object will be mentioned. It could be the case, for instance that prominence is much stronger evidence in non-local clauses than in local clauses, but that people are not sensitive to this correlation and treat them as independently informative.[15]

---

[15]This assumption that treating weakly-interacting cues as cognitively independent will adequately model human behavior is the foundation of naive Bayes behavioral models.

Or of course it could be the case that people do make use of this second order term, and that with more subjects or different materials, a significant three-way interaction would appear.

However, the lemma above shows why a three-way interaction in target reference expectation is not predicted by EMRE.

Behavioral linking function Since reference expectation is a latent variable in the behavioral model, how this translates into what is predicted about interactions on the dependent measure (experimental participants' acceptability ratings) depends on the linking function in the behavioral model.

As explained in Section 2.3.5, the linking function between SoC, a latent (unobserved) variable, and the participants' observed acceptability ratings of the items is unknown. If the effect of target reference expectation on ratings is linear or quasi-linear, then independent effects of locality and prominence on reference expectation would translate into linear independence of the factors' effects on acceptability, and no three-way interaction. If instead it is the SoC that has a linear effect on ratings, then independent cues would produce a three-way interaction on acceptability, because as discussed in Section 2.3.5), for lower probabilities, surprisal increases superlinearly with linear decreases in probability.

The observed data are consistent with an account in which the the two factors are treated as independently informative cues, and the linking function is linear between acceptability and target reference expectation (as opposed to SoC).

3.5.3 Results of Experiment 2

As discussed in Section 3.5.1, the results of Experiments 1 and 3 can be explained in terms of: (1) The SoC, as estimated by EMRE, and (2) a processing cost of locality. However, this account is inconsistent with the results of Experiment 2.

The significant fixed effects observed in Experiment 2 were: (1) a large main effect of gender; and, (2) a smaller fixed effect of prominence. There was no interaction between factors, nor a trend towards one. The main effect of gender is, as described

in Section 3.5.1, explained by the interaction between locality and agreement observed in the other experiments. Since all conditions in Experiment 2 had non-local targets while gender agreement was manipulated, this interaction would appear as a main effect of gender.

However, given the results of Experiment 3, what would have been expected for Experiment 2 was not a main effect of promence, but an interaction between prominence and gender. The data of Experiment 2 would be better explained by considering prominence as a processing cost rather than as a rational cue to expectation. This inconsistency cannot be accounted for using the data presently available.

3.6 Other theories of reference processing

Memory-based theories of NP anaphor reference resolution are not incompatible with EMRE. Although EMRE is agnostic about concepts like *'memory'*, a human reader who is resolving the antecedent of a referring expression must perforce rely on some basic cognitive capacity for storing abstract structures as a pattern of neural activation and later retrieving usable representations of these structures.

The point of modeling referring expression resolution as an expectation-based rational inference is not to claim that the actual cognitive process involves constructing cognitively primitive representations of model elements like 'top-down expectation' or 'reference expectation'. As stressed in Section 2.1, EMRE is a model of reference — not a theory of reference processing. Memory-based accounts of anaphora are *algorithmic-level* theories ('level 2' in Marr (1982)), which seek to explain how anaphora are processed. EMRE is a *computational-level* model ('level 1' in Marr (1982)) of referring expressions, designed to investigate to what extent readers' responses to texts containing referring expressions are consistent with interpreting reference as a problem of rational expectation.

These two levels of explanation do not necessarily entail strong mutual constraints. It could, for instance, be the case that the underlying cognitive processing of reference resolution is strongly dependent on the kind of memory interference described by

Almor (1999), and that due to the particular details of how entity representations are stored and retrieved, the results of this process would be hard to distinguish from those that would be expected if readers were treating properties like locality and prominence as informative cues.

Of particular relevance here is the notion of a special cognitive status for certain privileged antecedents (Foraker and McElree, 2007). It is manifestly the case that after enough time has passed[16] while an antecedent has gone unmentioned, it can only be referred to using a referring expression that would suffice to identify it in a 'null' context. This is almost certainly associated with some form of short-term memory ceasing to have an active representation of the referent. However, this is entirely consistent with the 'efficient code' account mentioned in Section 2.1.1. Indeed, this is precisely the behavior that would be expected of a system of working language memory the behavior of which was well-explained by expectation theory.

The experiments presented here demonstrate that some aspects of how people respond to referring expressions can be explained by a model of referring expressions based on top-down expectation and bottom-up match. Which, if any, of the many other effects demonstrated by researchers in this cognitive tradition can also be explained in terms of expectation remains to be seen.

---

[16](or enough intervening text processed)

Chapter 4

Expectation modeling of coherence

EMRE was also designed to be applied to coherence predicates. This chapter introduces the definition of a *reference type* (Section 4.1), discusses the implications of patterns observed in the GraphBank corpora (Section 4.2), explains how top-down expectation and bottom-up match apply to coherence predicates in general (Section 4.3), and then gives examples how they can be applied to particular types of coherence cues (Section 4.4).

4.1 Reference types

The model supposes that for any type of non-syntactic connection T a reader might expect to encounter in a text, there is a corresponding set R^T of possible referents. Each referent $r_i \in R^T$ is associated with: (1) a *top-down expectation* that r_i will be involved in a connection $t_x \in T$; (2) a set of *bottom-up features*, a limited set of context-independent features[1] of r_i that can be determined locally; and (3) a link back to a full concrete cognitive representation of r_i[2]. When a lexical head (*i.e.*, a word, phrase or gap; (Bloomfield, 1914; Pollard and Sag, 1994)) is encountered that cannot be interpreted without outside linguistic context, its features are compared to

---

[1]'Context-independent' in the sense that they can be computed locally, without reference to rich representations of either the linguistic environment, or semantic knowledge.

[2]This third feature is not used by EMRE, and will not be mentioned again. It is merely included because in order for the expectations computed from this bare-bones representation to be useful, they would have to have some way to be connected to the proper representation.

those of the possible referents, to compute a posterior reference expectation.

Reference types as defined include, *e.g.*, NP anaphora, coherence predicates, *'do so'* references, clausal ellipsis (*e.g.*, *'I did _.'*), and sluicing (*e.g.*, *'Someone took in our mail while we were away.'*; *'Do you know who _?'*).

For a given coherence cue type C, both top-down expectation and bottom-up match will depend on the particular semantic properties of the cue phrase (*e.g.*, causal, temporal). Top-down expectation of a cue phrase target r_i models the reader's prior expectation that a clause will be the target of a given type of coherence cue phrase, based on a context-dependent, *deductive* model of the clause's meaning:

$$\text{coherence class expectation} \equiv p_{\text{TD}}(C \mid \text{context})$$
$$\text{top-down coherence expectation} \equiv p_{\text{TD}}(r_i, C \mid \text{context})$$

Bottom-up match between a a target clause and a cue phrase with its attached clause models the reader's estimated *inductive* likelihood that a coherence dependency from a target r_i with a given set of surface features F_i would attach to a cue phrase and attached clause RE with another set of surface features F_{RE}, independent of the global context:

$$\text{bottom-up match}(r_i, \text{RE}) \equiv p_{\text{BU}}(\text{F}_{\text{RE}} \mid \text{F}_i, C)$$

4.2 Corpus evidence for top-down and bottom-up cues

Analysis of the Penn Discourse Treebank (Prasad et al., 2008) and the discourse coherence GraphBank (Wolf et al., 2003) shows many instances of coherence cue phrases with non-local arguments. For instance, of the 225 temporal relations in GraphBank, 148 were observed by this researcher and another expert annotator to have cue phrases, and 20 of these had antecedents separated from the cue phrase (and

its attached clause) by one or more intervening clauses. This strongly suggests that upon encountering a cue phrase, readers must perform some non-trivial computation to recover the intended antecedent.

Also, in 18 of these 20 temporal cue phrases with long distance dependencies, the initial word of the cue phrase was either the first or the second word of the sentence. This means that when the cue phrase itself was read, the reader did not know the content of the attached clause, suggesting that the reader had to delay computing the intended referent until the attached clause could be at least partially processed. This hypothesis of delayed assignment is further supported by three resemblance cue phrases (*e.g.*, *'similarly'*, *'also'*, *'however'*) which were both the first or second word in the sentences and had long-distance dependencies that attached to clauses in separate paragraphs.

This is evidence that readers can use *bottom-up* information about a cue phrase and its attached phrase (hereafter a *coherence anaphor*) — information that is only available *after* the coherence anaphor has been observed — to select an appropriate coherence antecedent from the preceding text. This could in principle offer a complete theory of how readers process coherence anaphora[3]: upon observing a coherence anaphor, select out of all prior clauses the clause that is the most sensible target.

However, other patterns observed in the GraphBank suggest that this is not the case. For instance, not one of the roughly 400 causal cue phrases in GraphBank has a long distance dependency spanning the main clause of another sentence. This suggests a hypothesis (tested in Experiments 6 and 7) that causal coherence cues are subject to strong top-down locality expectations, and that clauses that are otherwise good targets for causal cue phrases will not be considered if a main clause is between the target and the coherence anaphor. The 20 long-distance temporal dependencies observed also show evidence of a top-down locality preference — the clauses that intervene between the temporal cue phrases and their targets are all inappropriate targets for the dependency because they are, *e.g.*, in an inappropriate tense, or do

---

[3](analogous to the annotation procedure specified in the instruction manual for the GraphBank annotation system)

not describe an event.

The patterns for different coherence cue types that are suggested by GraphBank are:

- Causal cues are subject to sharp top-down expectations that prevent long distance dependencies spanning a main clause.

- Temporal cues can have less sharp top-down expectations allowing longer distance dependencies, and are sensitive to bottom-up verb tense cues.

- Resemblance cues have very broad top-down expectations, and readers can use bottom-up cues to locate appropriate targets even if they are very distant in the text.

Framed in this way, these can be framed in terms of the referring expressing model described in Section ?? above and experimentally tested. The first hypothesis, about causal cue phrases, is tested in Experiments 6 and 7, while the second, about temporal cue phrases, is tested in Experiment 5. Taken on their own, however, these hypotheses seem somewhat arbitrary. However, as explained in the next section, they are all consequences of a relatively simple theory.

4.3   Predicability and inductive traits

The top-down expectation of coherence predicates is defined similarly to that of NP anaphora. It represents, at any point in the text, the expectation *before* a coherence cue phrase has been observed that a given clausal referent will be the target of a type of discourse-level coherence cue phrase. [4] The patterns of top-down expectation for the coherence cue types described above are those that would be expected if they had been designed to be combined with a reference computation based on *predicability* and *local features*.

---

[4]The point that could be most easily missed here is that EMRE presumes a distinct set of expectations for each type of reference type.

Predicability *Predicates* are properties that can be true or false about a particular object. As an example of a single-argument predicate, one could say of a banana that has the property of being `ripe`, or that it does not have this property. An analogous two-argument predicate would be `riper` — one can say of two bananas that one is `riper` than the other, or that it is not.

Some predicates have zero expectation of being true for a member of some class of objects. For instance, while if all that was known about an object was that it was a car, there would be some expectation of it being `blue`. However, a banana has zero expectation of having the `blue` predicate.[5] Nonetheless, it is sensible to consider the possibility of a blue banana. Since bananas are physical objects, and colors are predicates that apply to physical objects, there is an underlying logic to the question of whether a particular banana is blue.

By constrast, not only does a banana have zero expectation of having the property `one hour long`, the question fails to make sense in a way that the question of whether it was blue does not. This distinction between predicate/object compositions that are merely false and ones that are nonsensical is called *predicability*. Although predicability is typically considered in terms of single argument predicates, the concept can be applied to two-argument predicates as well: the question of whether one person is riper than another or one banana is the father of another fail to make sense in much the same way as the example of the hour-long banana.

Coherence cues signal two-argument predicates, and the concept of predicability can be applied to these just as it can to other predicates. For instance, in the context of the reference type consisting of phrases like *'later on'*, *'after a while'*, and *'eventually'*, which are cue phrases for similar types of temporal antecedence predicates (Knott, 1996), the top-down expectation of a sentence like Example 21 being the target should be zero, because this is a non-eventive sentence, stating a proposition that cannot come before or after anything.

(21) The number seven is prime.

---

[5]All examples involving bananas are taken directly from Schmidt et al. (2006).

This might at first appear to be a bottom-up cue and not a top-down cue, as it depends on the particular properties of the sentence. However, this is a top-down cue because it can be computed *before* any temporal cue phrase is observed. EMRE models top-down expectation as an optimal expectation over referents using all available sources of information. In purely computational terms (*i.e.*, Marr Level 1 (1982)), this could be modeled as the reader marginalizing in advance over all possible uses of all referring expressions of this type, noting that the marginal probability of referring to Example 21 is zero, and updating expectations accordingly.

Local features Local features are properties of the arguments to predicates that allow the predicates' truth to be estimated in a purely *bottom-up* manner, by inspecting the properties of the arguments themselves (Knott, 1996). For instance, in order to assess whether it is sensible to that *'Cod and haddock are similar in flavor'*, one need only consider the particular qualities of their flavor, without reference to any global context.

EMRE models bottom-up match in coherence predicates as being based on local features — judgments that can be made about the plausibility of a dependency between a coherence anaphor and a target based on locally-computable features of the attached clause (*e.g.*, tense, basic semantic category, named entities) but without reference to the global context. Experiments 6 and 7 (Chapter 6) will show evidence that readers rely on some bottom-up information about the attached arguments of causal cue phrases, but that bottom-up match does not suffice as a complete explanation.

4.4   Cue phrase types

The present work proposes similar predicability constraints on coherence predicates. Specifically, it proposes that: (1) resemblance predicates have predicability constraints based on the semantic categories of their arguments; (2) temporal predicates have predicability constraints based on tense and lexical aspect; and, (3) causal coherence predicates have few inherent predicability constraints. These hypotheses are

tested in Experiments 4, 6, and 7 (Chapters 5 and 6). These patterns can be explained in terms of predicability and local features.

4.4.1 Resemblance predicates

Resemblance predicates require strong similarities between their arguments (Knott (1996); Section ??). Example 22 famously illustrates this point (Carroll, 1865):

(22)   ? Why is a raven like a writing desk?

This example demonstrates that when two things that are not in the same basic-level category it is difficult to provide a clear description of how they are similar or dissimilar that is not either unsatisfying[6] or more clever than informative[7].

These similarities are strictly local features — they can be computed without reference to the global discourse environment. A reader searching for the non-attached argument of a resemblance predicate might filter the search using bottom-up details such as the basic semantic category of the local argument, and only further consider referents that shared enough matching properties to make a comparison meaningful. (This contrasts with the causal predicate in Example 24 below, in which the cause and effect share no superficial details.)

If local features can be used to decisively eliminate many possible targets of resemblance relations as not merely improbable, but categorically non-predicable, such a sharp bottom-up match would make very broad top-down prior expectations reasonable.

4.4.2 Temporal predicates

Temporal predicates are somewhat informative about their antecedents , due to constraints based on verb tenses and lexical aspect (Reichenbach, 1966; Hornstein, 1977; Webber, 1988). For instance, an event in the future perfect tense may be 'before' an event in the present tense (23a), while an event in the present perfect may not (23b):

---

[6] . . . such as by comparing the superordinate categories, e.g., how animals differ from artifacts . . .

[7] "they produce notes that are very flat, and are 'nevar' backwards."

(23)   a. Rachel will have won an Emmy before she retires.

        b. # Rachel has won an Emmy before she retires.

These tense constraints mean that, as with resemblance predicates, some predica-bility constraints on temporal ordering can be determined from bottom-up properties of the sentence, without considering top-down deductive factors. In observation of the GraphBank corpus it was noted that temporal relations marked by cue phrases commonly had distant arguments, but in all such cases, the clauses intervening be-tween the cue phrase with its attached clause and the non-local argument had tenses, lexical aspects, or grammatical moods that made it implausible to interpret them as the non-local argument of the relation.

However, unlike resemblance predicates, other than consistency verb tense, tempo-ral predicates do not require their arguments to have any other consistency in terms of local features. Consequently, there are likely to be many targets in a text that would be consistent with the temporal anaphor. A hypothetical reader with unlimited cog-nitive resources might approach this computation by giving equal consideration to all possible targets, and choosing the one that made most sense based on all avail-able information. However, a top-down expectation favoring more recent clauses over less-recent ones would simplify matters for a reader with limited cognitive resources — the reader would simply select the most recent clause with a consistent tense, and the writer, knowing this, would attempt to ensure that this was the intended target.

4.4.3   Causal predicates

Causal predicates commonly exist between domains with no obvious similarity (Cheng, 1997). For instance, in Example 24, although there is no obvious overlap in content or topic between the two sentences, the causal predicate is quite plausible:

(24)   My wife had a bad fall on some ice last night. Consequently, I couldn't concentrate at work today.

The famous 'butterfly effect' proposes that almost any kind of event could be oc-casioned by almost any other kind of event (*e.g.*, "the flap of a butterfly's wings in

114

Brazil might set off a tornado in Texas"; (Lorenz, 1963)). While it is not difficult to find exceptions (*e.g.*, normal events cannot change whether a given number is prime), cue phrases for causal predicates (*e.g.*, *'because'*, *'therefore'*) appear to have relatively weak constraints on predicability.

As such, almost any clause could be a superficially plausible target for a discourse-level causal cue phrase. Given a large number of possible causal targets based only on local features might allow a reader to determine the intended cause at an above-chance frequency, but this would make trying to communicate clearly using causal cues risky at best. However, given very sharp top-down expectations limiting the target to (for instance) either the most recent clause or the most recent main clause, the lack of strongly-informative local features would not be problematic.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Expectation of Temporal Relations

EMRE can be used as a model of discourse-level coherence cue phrases as well as a model of pronouns. As described in Section ??, these cue phrases are modeled as depending on the top-down expectation that a given sentence or clause will be the target of a temporal cue phrase, and the bottom-up match between the features of the possible target clause and those of the temporal cue phrase with its attached argument.

5.1 Background

Temporal predicates have been observed to demonstrate constraints on the tenses of their about their arguements based on the tenses and lexical aspects of the verbs involved (Reichenbach, 1966; Hornstein, 1977; Webber, 1988). For instance, in Example 23 (from page 114, repeated below) an event in the future perfect tense may be *'before'* an event in the present tense (23a), while an event in the present perfect may not (23b):

(23) a. Rachel will have won an Emmy before she retires.

   b. # Rachel has won an Emmy before she retires.

Inspection of long-distance temporal relations in GraphBank (Wolf et al., 2003) revealed several relations like Example 25. The temporal cue phrase *'then'* (25c)

attaches to the closest plausible target, but that target (*'carries'*) is several sentences back.

(25) a. In her lab, a conveyer belt at waist level *carries*$_i$ parts from storage shelves to a robot that dominates the room like a silent, 4-foot metal sentry.

b. The robot is a V-shaped, jointed arm on a pivoting base. The arm can be fitted to allow it to grasp, lift and turn objects of differing sizes to suit a variety of tasks.

c. The robot *then*$_i$ swivels to place parts in position at the automated lathe or milling machine.

If this were a syntactic dependency then the DLT structural integration for this dependency would have to span roughly 7 intervening discourse units (or 'EU') [1], which would be high for an English syntactic dependency.[2] These long-distance temporal relations were low in incidence[3], but not especially infelicitous, suggesting that such long-distance temporal relations are not unexpected by readers.

Experiments 4 and 5 were designed to test two hypotheses about temporal relations that follow from: (1) these corpus observations; (2) EMRE; and, (3) the hypotheses of Webber et al. (2003):

1. Experiments 4 and 5 were jointly designed to test the hypothesis advanced by Webber et al. (2003) that coherence predicates marked by sentence-level conjunctions are a radically different kind of dependency than those marked by discourse-level coherence cue phrases[4]. As such, experiments 4 and 5 are very similar in design — what differs between them is whether the temporal cue phrase used is a coordination or an adverbial.

---

[1]Cost is parametrized as the number of new discourse referents between the two dependent heads (*'robot'*, *'dominates'*, *'arm'*, *' pivoting base'*, *'fitted'*, *'grasp/lift/turn'*, *'variety of tasks'*).

[2]Though perhaps less so in head-final languages such as Japanese (Nakatani and Gibson, 2010).

[3](3–29 out of 226 total temporal relations in GraphBank depending on how 'long-distance dependencies' is operationalized.)

[4]The terms employed for these in their work were *'structural connectives'* and *'discourse adverbials'*.

2. Experiment 5, which involved discourse-level adverbial cue phrases, was designed to test the joint hypotheses that:

   (a) Participants would be sensitive to minimal manipulations of verb tense in their reports of what clause was the target of the temporal cue phrase, and;

   (b) As the items were designed such that there was no disambiguation and hence no surprisal, participants would be relatively insensitive to the tense condition in terms of their acceptability ratings of the texts.

5.2   Experiments 4 and 5

Experiment 4 tested for the magnitude of two main effects predicted by Webber et al. (2003) on the attachment preferences for sentence-level temporal cue phrases. Experiment 5 tested for one main effect predicted by EMRE on readers' reference expectations for discourse-level temporal cue phrases. All predicted effects were observed in the data.

5.2.1   Design

Experiments 4 and 5 used a shared two-condition design, in which the independent factor was a manipulation of tense. The items consisted of short texts containing three clauses. Clauses 2 and 3 were separated by either a sentence-level temporal cue (a coordination; Experiment 4) or a discourse-level temporal cue (an adverbial; Experiment 5).

(26)   EXPERIMENT 4 SAMPLE ITEM — Cue = *before*

    a. *Consistent tense:* Kenneth is laying out his beach blanket. He wants to take a quick swim before he relaxes in the sun.

    b. *Inconsistent tense:* Kenneth is laying out his beach blanket. He took a quick swim before he relaxes in the sun.

(27) EXPERIMENT 5 SAMPLE ITEM — Cue = *next*

    a. *Consistent tense:* Mary is calling her mother. She needs to finish her homework. Next, she will book a plane ticket home.

    b. *Inconsistent tense:* Mary is calling her mother. She had been finishing her homework. Next, she will book a plane ticket home.

There were two dependent measures in these experiments. First, the participants were asked which of the clauses the cue phrase referred to, as in Example 28:

(28) Kenneth is laying out his beach blanket. He took a quick swim *before* he relaxes in the sun.

    *Which does the emphasized word **"before"** refer to?*

    a. 'He took a quick swim'

    b. 'Kenneth is laying out his beach blanket'

Participants were then asked to rate the acceptability of the text on a scale from 1 to 5.

Three Latin-square sets of item conditions were generated such that each participant saw exactly one condition for each of the 8 items in Experiment 4 and the 24 items in Experiment 5. These experiments did not include filler items. Each of these sets was shuffled into two randomly ordered lists, for a total of 6 presented lists of 32 stimuli. Each stimulus was paired with a 'Yes/No' comprehension question, to check for participant compliance, but due to an error in the materials these were not used to exclude participants from analysis.

5.2.2 Participants

Each of the six lists was presented to 10 human participants in the form of an online survey, giving 60 survey responses. Participants were selected and paid via Amazon's Mechanical Turk crowd-sourcing system. Participants were screened to have internet addresses within the United States, and to be self-reported native English speakers. Responses were excluded from one participant who responded to 3 lists, and from 7

who gave full responses to less than 5% of stimuli, leaving 50 usable survey responses. Because of a confound in the comprehension questions, no participants were excluded on the basis of comprehension questions.

5.2.3 Predictions

Experiment 4 was intended as a cross-paradigm control — since the materials used a sentence-level coordinating cue phrase ('*before*'), EMRE and the behavioral model had no way to make predictions. However, predictions for this type of experiment are stated clearly in Webber et al. (2003). The predictions following from this work were that participants would almost uniformly report that the sentence-level cue phrase referred to the immediately preceding clause, and that the largest effect would be one of acceptability, with the condition in which the tense was inconsistent being rated much worse.

The predictions of EMRE and the behavioral model for Experiment 5 were that:

- Due to top-down expectation, the local antecedent would be favored as the reported target in the *consistent tense* condition.

- In the *inconsistent tense* condition, the inconsistency in tense in the local clause would lead to lower bottom-up match for that clause, and for the non-local clause to be reported to be the target more often than in the *consistent tense* condition.

- Since there was no disambiguation of the cue phrase, there should be no surprisal. As such, if ratings differences between conditions were observed, they should be quite small relative to those in Experiment 5.

5.2.4 Analysis

Data were analyzed using the R statistical language (R Development Core Team, 2010), including the lme4 package (Bates et al., 2008) for linear mixed effect regression (LMER) modeling, and the languageR package for calculating statistical

significance (p values) using Highest Posterior Density (HPD) parameter estimation and Markov Chain Monte Carlo (MCMC) sampling from the resulting mixed effects model (Baayen, 2008).

5.2.5 Results

As shown in Table 5.1, the results of both Experiment 4 and 5 came out as predicted. In Experiment 4, which used the sentence-level cue phrase *'before'*, participants' reports that the local clause was the target of the cue phrase were essentially at ceiling, with a mean locality rating of 0.95 (95%), and a null main effect of verb consistency (magnitude 0.025 (2.5%), $p > 0.1$) (Figure 5-1). The predicted main effect on acceptability was observed, with magnitude -1.8 ($p < 1\mathrm{E}^{-4}$) (Figure 5-2) — participants strongly disliked the inconsistent items in this experiment, but that was not enough to drive them to choose the other referent when the cue phrase entailed a syntactic dependency.

Experiment	Cues	Condition	Locality	StdErr	Accept.	StdErr
Exp. 4	*'before'*	consistent	0.95867	0.04910	4.6520	0.1292
		inconsistent	0.93445	0.04879	2.8453	0.1284
Exp. 5	*'Next'*	consistent	0.71327	0.03578	3.8319	0.1001
	'Later'	inconsistent	0.36762	0.03579	3.7243	0.1001

Table 5.1: Experiments 4 and 5: Reported local target by Experiment and Condition

The effects in Experiment 5 (which used the discourse-level cues *'next'* and *'later'* were, as predicted, exactly the reverse, with a large main effect of consistency on locality (magnitude 0.35 (35%); ($p < 1\mathrm{E}^{-4}$); Figure 5-1). The main effect of consistency on acceptability was, as predicted, quite small (magnitude of 0.1), though significant ($p = 0.035$) (Figure 5-2).

Binning the responses in Experiment 4 according to the dependent measure of reported target would have been relatively meaningless, since there were so few responses that reported the non-local clause as the target. However, this analysis was performed on the responses from Experiment 5. As can be seen in Table 5.2 (Figure 5-3), the acceptability ratings are quite similar between conditions. Unlike Experi-
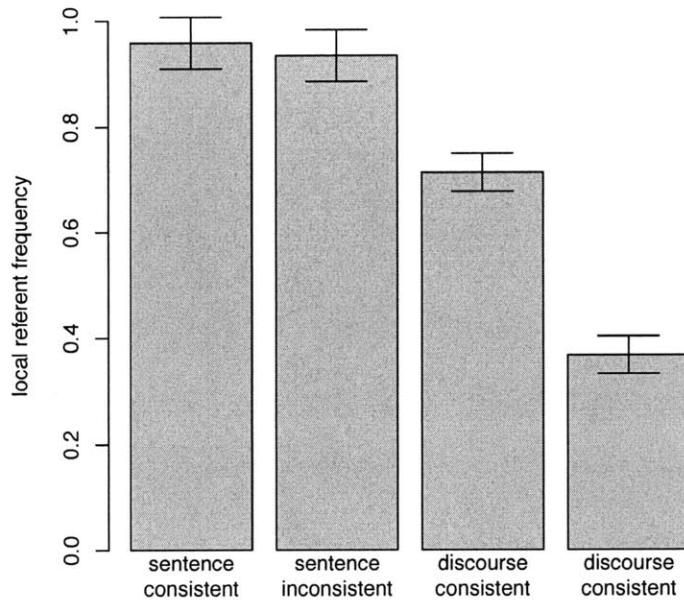
Figure 5-1: Experiments 4 and 5: Local vs. Non-local Referent
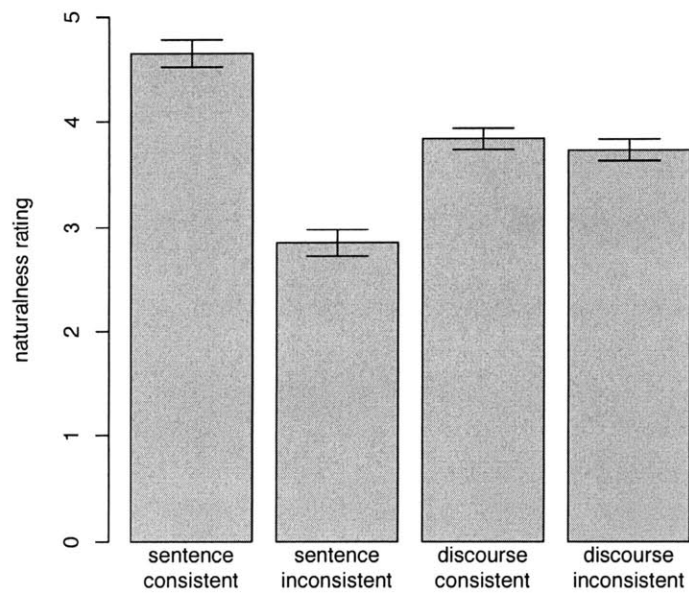
Figure 5-2: Experiments 4 and 5: Mean Acceptability / Experiment and Condition
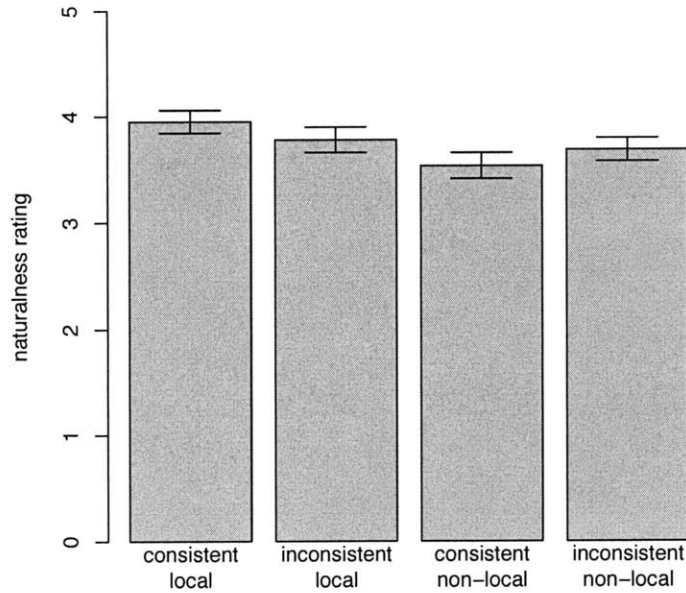
Figure 5-3: Experiment 5: Mean Acceptability / Condition × Reported Target

ment 4, where large differences in acceptability did not alter reported target locality, here the opposite is observed — large differences in reported target locality with no correspondingly large differences in reported acceptability.

Condition	Response	Accept.	StdErr
consistent	local	3.9509	0.1074
consistent	non-local	3.7804	0.1183
inconsistent	local	3.5382	0.1224
inconsistent	non-local	3.6906	0.1093

Table 5.2: Experiment 5: binned dependent measure *'locality'*

A linear mixed effects model (Table 5.3) does show that there are small but significant main effects for both consistency and locality, as well as for their interaction. These effects can be more clearly visualized by observing Figure 5-4. This figure shows that the extremely flat acceptability difference between conditions in the unbinned responses masks some moderately larger differences that can be observed when responses are binned. In particular, the difference between participants who select the

Fixed Effects	Estimate	Std. Error	p-value
(consistent + local)	3.95090	0.09965	0.0000
inconsistent	-0.17046	0.08059	0.0346
non-local	-0.41268	0.09283	0.0000
inconsistent : non-local	0.32287	0.12191	0.0082

Table 5.3: Experiment 5: Linear Fixed Effects (LMER)

local and non-local clauses in the *'consistent'* condition is roughly $^2/_3$ the size of the main effect of locality in Experiments 1 and 3.[5]

A concise way to describe this pattern is that in responses where the participant reports a local target, they are happier when there is a consistent verb than when they are not, while in responses where they select a non-local target they are happier when there is an inconsistent verb than when there is not. Although this may sound like a backwards explanation, it is entirely consistent with the theory introduced by Vul (2010) that the mind approximates inference by sampling. This will be addressed further in the discussion.

5.3 Discussion

This section opens with a discussion of how the results of Experiment 5 provide further evidence for EMRE as a model of reference (Section 5.3.1). This is followed by brief discussions of how the contrast between Experiments 4 and 5 is a clear demonstration of the distinction between the *'structural connectives'* and *'discourse adverbials'* proposed by Webber et al. (2003) (Section 5.3.2). The discussion closes by considering how the results of Experiment 5 are made clearer by considering them in terms of the theory of inference as sampling advanced by Vul (2010) (Section 5.3.3).

5.3.1 Tense in bottom-up match

One of the dependent measures in these experiments was the participants' report of which of two clauses a temporal cue phrase (*'Next'* and *'Later'*) referred to. The

---

[5]The mean acceptability ratings in the *'inconsistent'* condition are practically on top of each other, but this is a coincidence.
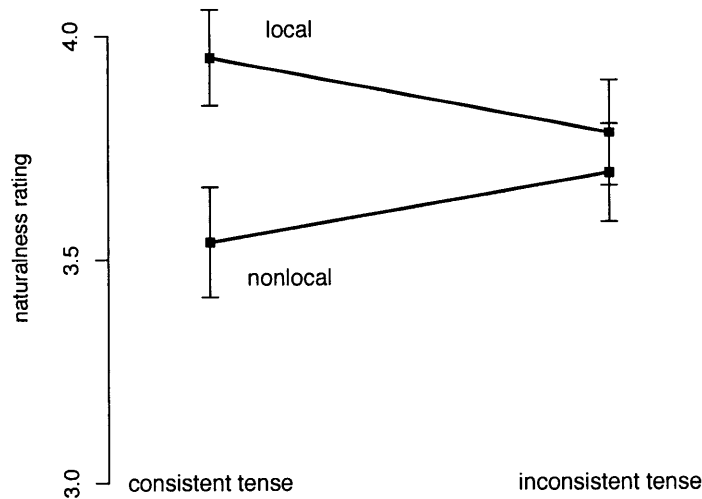
Figure 5-4: Experiment 5: Mean Acceptability / Condition × Reported Target *(alt.)*

effects predicted and observed for this measure were:

1. In the baseline condition, in which the tense of both verbs was consistent with the tense of the attached argument of the cue phrase (*e.g..*, *'he ate lunch'* in the sentence *'Next, he ate lunch'*), there was a preference for the local clause[6], but not an absolute preference (as compared to the reference reports for Experiment 4).

2. In the *'inconsistent tense'* conditions, the local clause was less favored than in the *'consistent tense'* condition.

The first of these is evidence for the obvious claim that readers have non-uniform expectations about temporal reference. The second is evidence that they are able to use local features like verb tense to update these expectations such that they have a higher expectation of the non-local clause, in a way that is similar enough to how

---

[6]This is what would be predicted any plausible theory of temporal reference.

they form expectations about pronoun reference that the same model can be used for both. Though observing the predicted effects was evidence in favor of the model, Section 5.3.3 will argue that the unpredicted effects offer deeper suggestions about computational models of reference.

5.3.2 Empirical support for discourse adverbials

This pair of experiments provides evidence in favor of the claim made by Webber et al. (2003) that while clausal connectives and adverbials can often be used to signal the same coherence predicates, they operate in fundamentally different ways. While it is true that this claim depends on between-item comparisons, the reason this approach was taken was that no way could be found to bring the designs any closer together. The semantics of the cues were tightly matched — the denotative semantics of *'before'* are very close to those of *'next'* and *'later'*[7] — but they make different pairs of tenses incompatible (see the materials in Appendix A for examples), so a simple substitution of one for the other was not possible.

The implications of this for discourse corpora are addressed in the general discussion (Chapter 7).

5.3.3 Inference as sampling

Some significant trends and effects in the data in Experiment 5 were not predicted by either the model or this researcher. One effect that was not predicted was the intercept — specifically, the extent to which, in both conditions, the distribution of reported target localities was flatter than expected (*i.e.*, nearer to $1/2$ than to 0 or 1). Another unexpected effect was on acceptability: the only prior prediction made about the effect of tense inconsistency on acceptability in Experiment 5 was that it would be much smaller than the corresponding effect in Experiment 4.

---

[7]This researcher has yet to encounter anyone who did not have to visibly expend significant cognitive resources before being able to accept this claim. In some cases, even providing a series of matched examples (*e.g.*, *'I brushed my teeth before I went to sleep.'* vs. *'I brushed my teeth. Next, I went to sleep.'*) were not fully convincing. The strength of this effect is such that it might constitute stronger evidence than the experiments that these two types of cue phrase are very different in kind.

127

The effect observed in the unbinned data was indeed extremely small, but binning according to the dependent measure of reported locality revealed that this was hiding another perplexing pair of effects. Within the responses that reported the *local* clause, those that were made in the *'inconsistent'* condition, — which favored the *non-local* clause — were ranked significantly lower than those that were in the condition that favored the local response that they gave. The same was true of the responses that reported the non-local clause. Fully rational, well-calibrated readers, if presented with conditions that were sufficiently distinguishable to drive: (1) large differences in locality preferences; and, (2) smaller but predictable differences in acceptability ratings, should re-calibrate their responses such that their responses would not be predictive of their acceptability ratings.[8]

Instead, the responses gave the appearance that the participants had been choosing a response essentially at random[9], and then later being unhappy with their choice. Under a theory recently advanced by Vul (2010) that "the mind approximates inference by sampling", this is exactly what they should be expected to do. Vul's theory is like that of Levy (2008) in postulating that the mind represents knowledge in the form of implicit probability distributions, and updates these distributions only when required to by outside data, since this requires significant cognitive resources (*i.e.*, surprisal). Under this theory, the way that a reader could arrive at the *best* response to the question of, *e.g.*, what a word referred to would be to draw many samples from the distribution. The more samples drawn, the more certainty the reader could have about having the best response.

Often though, Vul argues, getting the best answer is not of huge importance.[10] In these cases, people are likely to draw a single sample from their beliefs about the world, and use that as a sufficient approximation. This is consistent with the way that participants seemed to be choosing in a semi-random way. However, after being asked about the referent, the participants went on to answer a question about how acceptable the text was. At this point, if they had selected a target referent that was

---

[8](This assertion seems to need a lemma.)

[9]though with systematic biases corresponding neatly with the predicted effects

[10]Responding to an online survey would be a good example of this.

128

less consistent with the text, a surprisal response would be expected.

This is obviously a post-hoc explanation, but it is offered in a spirit of inquiry. Also, it is not a difficult idea to test. The first step would be to see if this pattern of responses replicates, or if the p-values are misleading and it is in fact simply a random effect. A simple way to test whether sampling from the data is indeed having an effect on reference would be to use the same materials, but reverse the order of the questions for the two dependent measures. If the observed effect were the result of the reader having sampled from an expectation, then this would be predicted to make the effect disappear.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6

Expectation of Causal Predicates

Coherence cue phrases are optional semantic markers, used by speakers and writers to emphasize or clarify coherence predicates in text. These experiments test predictions about when causal cue phrases (*e.g.*, *'therefore'*) can clarify causal predicates and when they cannot, as estimated by text acceptability ratings from human experimental participants. The hypothesis behind these predictions was located through corpus analysis of the GraphBank annotated coherence corpus ((Wolf et al., 2003); Appendix ??); their particular structure is derived from EMRE and the behavioral model. Modeling both NP anaphora and coherence predicates under the same mathematical formalism is one of the unique contributions of this work.

6.1 Background

6.1.1 Observations from corpora

As discussed in Appendix ??, analysis of the GraphBank annotated coherence corpus (Wolf et al., 2003)) was key to the development of EMRE. These experiments follow up on an observation about GraphBank's long-distance causal relations[1]: excluding

---

[1] *'Relations'* are a annotational primitive in GraphBank, used to indicate several kinds of sentence- and discourse-level dependencies. These include coherence predicates (*e.g.*, causal and temporal predicates), attributions of claims to a source, and 'elaboration' relationships, in which the interpretation of a clause hinges on the fact that the referent of some expression in the clause is the same as that of one in another clause. This work uses *'relation'* exclusively to refer to these annotations

131

degenerate cases of non-locality such as intervening attributive clauses (*e.g.*, *'the court wrote'* in Example 29), no long-distance causal relation was marked by a causal cue phrase. By comparison, approximately three in five causal relations in GraphBank is marked by a cue phrase; roughly the same is true in the PDTB (Prasad et al., 2008). In temporal and other types of relations, long-distance dependencies were commonly marked by cue phrases. (See Appendix ?? for more detail.)

(29)   a.   "If true,"

b.   the court wrote,

c.   "this contention would justify dismissal of these actions on prudential grounds."

Since the causal relations under investigation marked long-distance discourse-level predicates, they clearly could not have been marked by sentence-level cue phrases that specify causal predicates between adjacent clauses (*e.g.*, *'since'*, *'despite'*; see Experiment 4 in Chapter 5). The only cue phrases that could in theory have marked these relations are discourse-level adverbial cue phrases that can refer to material in separate sentences. However, these cue phrases were never observed to be used in this manner.

This peculiarity of causal relations raised the question of whether causal cue phrases might sometimes be unable to clarify causal predicates. The paired approaches of: (1) regarding cue phrases as a form of referring expression (as suggested by Webber et al. (2003)); and, (2) using EMRE as a model of referring expressions, suggested experiments similar to those used to test EMRE's predictions about NP anaphora.

6.1.2   Manipulation of top-down causal expectation

These experiments are based on pairs of texts that express causal predicates without using a cue phrase. The pairs differ by systematic manipulations of the locality (Experiment 6; 6.2) or the syntactic prominence (Experiment 7; 6.3) of the causal

in GraphBank.

target. The experimental question is what effect adding causal cue phrases will have on the acceptability of the texts. This is not a question that has been addressed in previous research on discourse coherence.

If the causal cue phrases are treated as referring expressions (as suggested by Webber et al. (2003) and the results of Experiment 5), EMRE models the posterior reference expectation of the clausal antecedent as the result of performing a rational belief update on a *top-down expectation* over clauses (Section 2.1.1) with the bottom-up evidence the local material (the cue phrase and the clause to which it is attached) offers for or against each clause, as determined by the *bottom-up match* between the cue phrase and the clause (Section 2.1.2).

If locality and syntactic prominence have an effect on top-down expectation over causes that is similar to the effects on anaphoric expectation of NPs suggested by the results of Experiments 1 and 3 (Section 3.5.1), then these factors should affect the causal target's reference expectation.

6.1.3   Bottom-up match and predicability

If adding a causal cue phrase has an effect on the interpretation of the causal predicate[2], then a difference in the target clause's prior top-down reference expectation between the two *'causal cue'* conditions should lead the causal cue to have different effects on acceptability in the two conditions, relative to their corresponding uncued conditions. Specifically:

- To the extent that the causal cue phrase is interpreted as referring to the target clause, it should increase expectation of the target interpretation. When the causal predicate is disambiguated by the continuation of the final sentence past the cue phrase, this will lead to reduced surprisal of the continuation (SoC), and increase the acceptability ratings.

- To the extent that the causal cue phrase is interpreted as referring to the distractor clause, it should decrease expectation of the target interpretation. This

---

[2]This seems like a safe initial working hypothesis.

should increase the SoC, and decrease the acceptability ratings.

However, the bottom-up match between the causal cue phrase with its attached clause and the two possible referent clauses was a wild card in this experimental design. Depending on the extent to which the extent to which the features of the local clause attached to the cue phrase were able to drive the bottom-up match, possible outcomes included:

1. If the match between the local clause and the target had a very strong effect, then the causal cue could have always been interpreted to refer to the target clause. In this case, adding the causal cue could have led to increased target expectation and a main effect of increased acceptability for the presence of the causal cue across both top-down conditions.

2. If the features of the local clause were not used at all in bottom-up match (only the casual cue phrase itself), then, since causal cue phrases are themselves uninformative about their arguments (see Section ??), the cue phrase would have provided little evidence for one clause over the other, and had a negligible impact, as either a main effect or an interaction.

3. If the bottom-up match information from the local clause was available, but only in a limited manner, then it could be interpreted to refer to the target clause only when the target clause exceeded some threshold for top-down expectation. This would lead to an interaction, in which in some conditions the causal cue increased expectation of the predicate, and decreased the SoC, while in others it would be interpreted to refer to the distractor clause, decreased expectation of the predicate, and increased SoC. This would produce an interaction.

This third possibility represents a 'sweet spot', in which the effect of bottom-up match is neither so large that the cue phrase always increases acceptability, nor so small that it has very little effect at all, but instead is in some middle ground. Nonetheless, the intuitions of the researchers and the evidence from GraphBank

suggested that experimental materials could be constructed to focus on this middle ground. This point will be further discussed in the context of the experimental results in Section 6.4.

6.2 Experiment 6

6.2.1 Design

Experiment 6 co-varied: (1) the locality of the clause that was the cause in a causal coherence predicate; with, (2) the presence or absence of a causal cue phrase (*'consequently'* or *'as a result'*). All items involved a person or two people engaged in unexceptional activities or events (*e.g.*, visiting a restaurant, playing sports, seeing an auto mechanic).

6.2.2 Materials

All 16 experimental items had the same format. The first sentence had a single clause, introducing the person or people and establishing the situation. The second sentence had two clauses, with the *'cause'* argument to the coherence predicate in the subordinate clause. The third sentence described the corresponding *'effect'* argument.

The expectation of the cause was varied by manipulating the locality of the phrase in which it was described (the *'target clause'*). In *'local'* conditions, the target clause appeared immediately before the clause containing the effect, while in *'nonlocal'* conditions, whereas in *'nonlocal'* conditions, there was an intervening clause. The other independent factor was the presence or absence of a causal cue phrase at the beginning of the third sentence.

(30) EXPERIMENT 6: SAMPLE ITEM

    a. *(local + no cue):* The accounting department that Serena works in is always short on cash. She takes very good care of her equipment, though her current workstation is old and obsolete. Her computer couldn't run the new accounting software she needed.

135

b. *(non-local + no cue):* The accounting department that Serena works in is always short on cash. Though her current workstation is old and obsolete, she takes very good care of her equipment. Her computer couldn't run the new accounting software she needed.

c. *(local + causal cue):* The accounting department that Serena works in is always short on cash. She takes very good care of her equipment, though her current workstation is old and obsolete. Consequently, her computer couldn't run the new accounting software she needed.

d. *(non-local + causal cue):* The accounting department that Serena works in is always short on cash. Though her current workstation is old and obsolete, she takes very good care of her equipment. Consequently, her computer couldn't run the new accounting software she needed.

Four Latin-square sets of item conditions were generated such that each participant saw exactly one condition for each experimental item. Each of these sets, along with 32 filler items, was shuffled into two randomly ordered lists, for a total of eight presented lists of 48 stimuli. Each stimulus was paired with two 'Yes/No' comprehension questions, to check for participant compliance.

6.2.3   Participants

Each of the eight lists was presented to 10 human participants in the form of an online survey, giving 80 survey responses. Participants were selected and paid via Amazon's Mechanical Turk crowd-sourcing system. Participants were screened to have internet addresses within the United States, and to be self-reported native English speakers. When a participant left more than 10% of the items blank, answered more than 25% of the comprehension questions incorrectly, or responded to more than one list in the experiment (invalidating the results due to viewing multiple versions of the same item), all data from that participant were excluded from analysis. 10 participants were excluded in this manner, leaving 70 usable survey responses.

6.2.4 Predictions

EMRE predicts a possible interaction between target clause locality and cue phrase presence. As explained in Section 6.1, if the interaction appeared, it was predicted to be a cross-over interaction in which the cue phrase increases the acceptability ratings of the text rated as more acceptable in the 'no cue' conditions, and decreases the acceptability ratings of the text rated as less acceptable.

6.2.5 Analysis

Data were analyzed using the R statistical language (R Development Core Team, 2010), including the lme4 package (Bates et al., 2008) for linear mixed effect regression (LMER) modeling, and the languageR package for calculating statistical significance (p) values using Highest Posterior Density (HPD) parameter estimation and Markov Chain Monte Carlo (MCMC) sampling from the resulting mixed effects model (Baayen, 2008).

6.2.6 Results

The interaction predicted by EMRE was observed in the data ($p = 0.03$), as can be seen in the mean acceptability ratings for each condition (Figure 6-1; Table 6.1).

Condition	Estimate
local + no cue	4.0284
local + causal cue	4.1334
non-local + no cue	3.6045
non-local + causal cue	3.4582
(Residual Std. Error)	0.142

Table 6.1: Experiment 6: Mean Acceptability / Condition

In the LMER estimates of the fixed effects (Table 6.2; Figure 6-2), the largest effect is, as usual, the main effect of locality (0.4 point difference in acceptability; $p = 1\text{E}^-4$). The interaction is approximately $^3/_5$ the magnitude of the locality main effect (0.25 points; ($p = 0.03$).
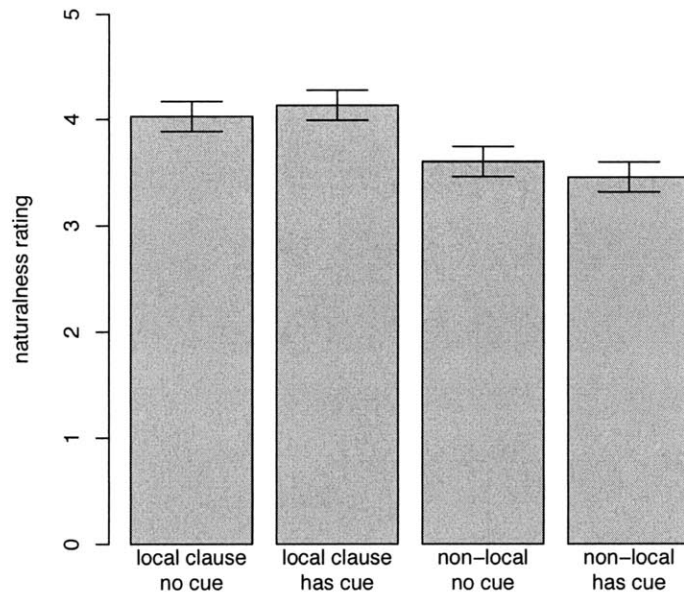
Figure 6-1: Experiment 6: Mean Acceptability / Condition
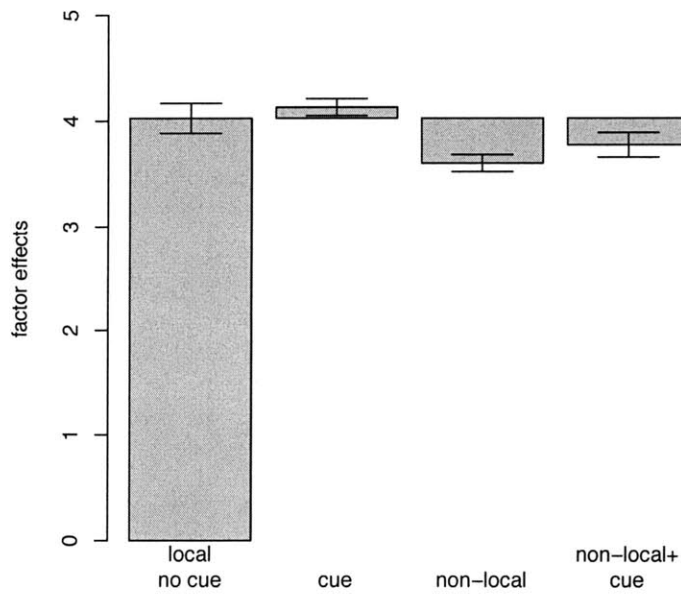
Figure 6-2: Experiment 6: Effect Sizes

Fixed Effects	Estimate	Std. Error	p-value
(local + no cue)	4.0284	0.1421	0.0001
causal cue	0.1050	0.0814	0.1930
non-local target	-0.4239	0.0809	0.0001
non-local : causal cue	-0.2513	0.1148	0.0280

Table 6.2: Experiment 6: Linear Fixed Effects (LMER)

causal–cue/locality interaction
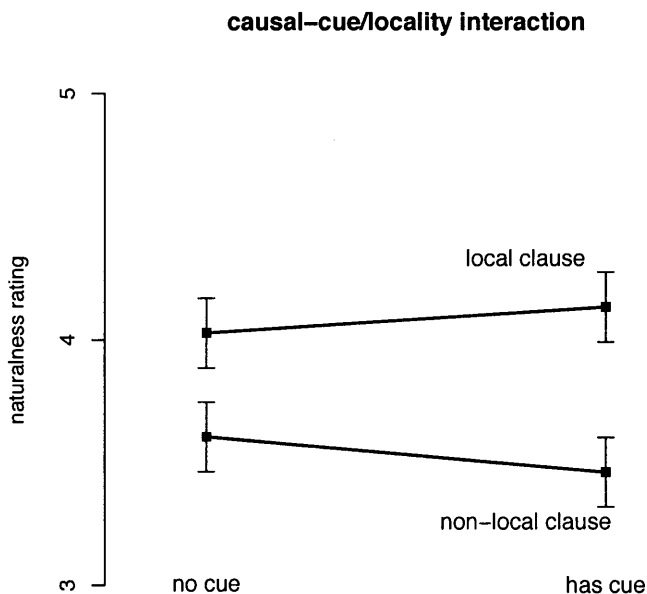
Figure 6-3: Experiment 6: Locality / Causal Cue Interaction

This interaction is also visible in a linear effects plot of the acceptability means (Figure 6-3). This plot shows that the interaction is the predicted cross-over, in which the causal cue phrase increases acceptability in the 'local' conditions and decreases it in the 'non-local' conditions.

6.3 Experiment 7

As with Experiment 6, this experiment co-varied the presence or absence of a causal cue phrase with a manipulation of a factor predicted to be a cue to the top-down expectation of the causal target. In this case, cause expectation is manipulated

by varying the antecedent's syntactic prominence. The motivation and the model predictions are identical to those in the prior experiment.

6.3.1 Design

Experiment 7 co-varied: (1) the syntactic prominence of the clause that was the cause in a causal coherence predicate; with, (2) the presence or absence of a causal cue phrase (*'consequently'* or *'as a result'*). all items involved a person or two people engaged in unexceptional activities or events (*e.g.*, visiting a restaurant, playing sports, seeing an auto mechanic).

6.3.2 Materials

(31) EXPERIMENT 7: SAMPLE ITEM

    a. *(prominent + noncued):* Mark's hobby is repairing antique clocks. Replacement parts can be impossible to find, though he has an extensive network of sources. He sometimes ends up making hard-to-locate parts himself in his machine shop.

    b. *(subordinate + noncued):* Mark's hobby is repairing antique clocks. While replacement parts can be impossible to find, he has an extensive network of sources. He sometimes ends up making hard-to-locate parts himself in his machine shop.

    c. *(prominent + cued):* Mark's hobby is repairing antique clocks. Replacement parts can be impossible to find, though he has an extensive network of sources. Consequently, he sometimes ends up making hard-to-locate parts himself in his machine shop.

    d. *(subordinate + cued):* Mark's hobby is repairing antique clocks. While replacement parts can be impossible to find, he has an extensive network of sources. Consequently, he sometimes ends up making hard-to-locate parts himself in his machine shop.

Four Latin-square sets of item conditions were generated such that each participant saw exactly one condition for each experimental item. Each of these sets, along with 32 filler items, was shuffled into two randomly ordered lists, for a total of eight presented lists of 48 stimuli. Each stimulus was paired with two 'Yes/No' comprehension questions, to check for participant compliance.

6.3.3 Participants

Each of the eight lists was presented to 10 human participants in the form of an online survey, giving 80 survey responses. Participants were selected and paid via Amazon's Mechanical Turk crowd-sourcing system. Participants were screened to have internet addresses within the United States, and to be self-reported native English speakers. When a participant left more than 10% of the items blank, answered more than 25% of the comprehension questions incorrectly, or responded to more than one list in the experiment (invalidating the results due to viewing multiple versions of the same item), all data from that participant were excluded from analysis. 9 participants were excluded in this manner, leaving 71 usable survey responses.

6.3.4 Predictions

EMRE predicts a possible interaction between target clause prominence and cue phrase presence. As explained in Section 6.1, if the interaction appeared, it was predicted to be a cross-over Interaction in which the cue phrase increases the acceptability ratings of the text rated as more acceptable in the *'no cue'* conditions, and decreases the acceptability ratings of the text rated as less acceptable.

6.3.5 Analysis

Data were analyzed using the R statistical language (R Development Core Team, 2010), including the lme4 package (Bates et al., 2008) for linear mixed effect regression (LMER) modeling, and the languageR package for calculating statistical significance (p) values using Highest Posterior Density (HPD) parameter estimation
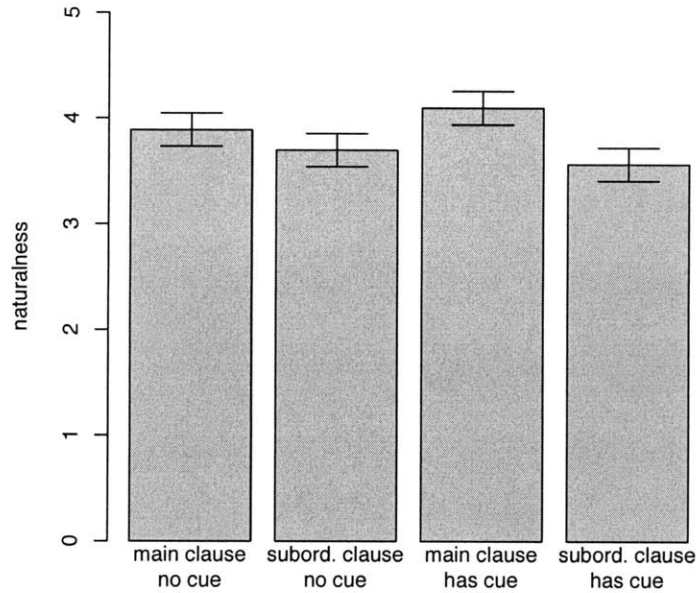
Figure 6-4: Experiment 7: Mean Acceptability / Condition

and Markov Chain Monte Carlo (MCMC) sampling from the resulting mixed effects model (Baayen, 2008).

6.3.6 Results

The interaction predicted by EMRE was observed in the data ($p = 0.03$), as can be seen in the mean acceptability ratings for each condition (Figure 6-4; Table 6.3).

Condition	Estimate
uncued + main	3.8911
uncued + subord.	3.6974
causal cue + main	4.0957
causal cue + subord.	3.5656
(Residual Std. Error)	0.157

Table 6.3: Experiment 7: Mean Acceptability / Condition

In the LMER estimates of the fixed effects (Table 6.4; Figure 6-5), the largest effect is the predicted interaction (0.3 point difference in acceptability; $p = 0.003$).

142
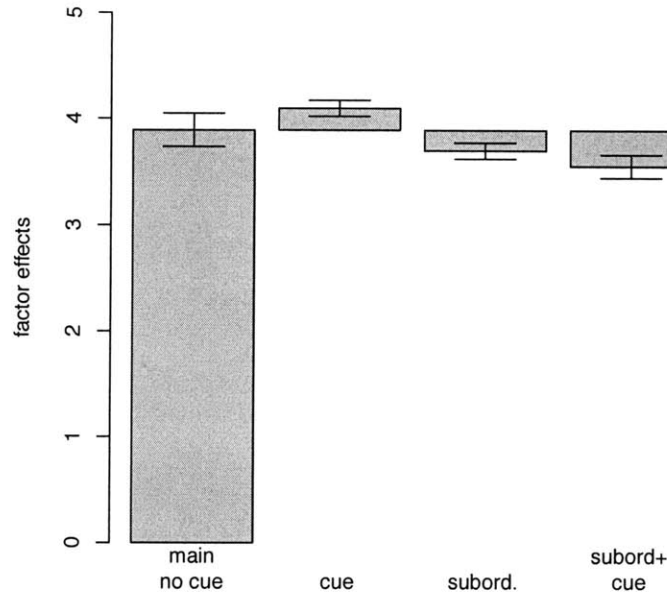
Figure 6-5: Experiment 7: Effect Sizes

There were also main effects for both non-prominence and the presence of the causal cue, both approximately $2/3$ the magnitude of the interaction effect (in both cases, magnitude 0.2 points, $(p \simeq 0.01)$.

Fixed Effects	Estimate	Std. Error	p-value
(uncued + main)	3.8911	0.15662	0.0001
causal cue	0.2046	0.07668	0.0088
subord.	-0.1936	0.07565	0.0134
cue : subord	-0.3365	0.10834	0.0034

Table 6.4: Experiment 7: Linear Fixed Effects (LMER)

This interaction is also visible in a linear effects plot of the acceptability means (Figure 6-6). This plot shows that the interaction is the predicted cross-over, in which the causal cue phrase increases acceptability in the *'prominent'* conditions and decreases it in the *'subordinate'* conditions.

143
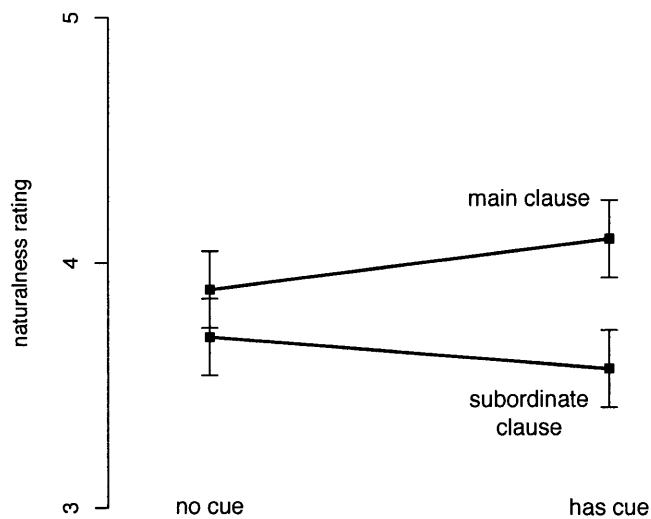
Figure 6-6: Experiment 7: Prominence / Causal Cue Interaction

6.4   Discussion

The data from Experiments 6 and 7 suggest the following propositions, which will be explained in more detail below:

1. The causal predicates in the items are key to readers' evaluations of the items even when there is no cue phrase.

2. The locality and prominence of the target clause both affect how readers interpret a cue phrase when it is present.

3. The cue phrase makes some aspect of the item's interpretation less surprising in the *'local + subordinate'* and *'non-local + main'* conditions, but more surprising in the *'non-local + subordinate'* conditions.

4. The observed pattern of results is not consistent with either an entirely top-down or an entirely bottom-up model of causal reference.

Taken together, these propositions suggest a single explanation for both the locality and interaction effects: the latent parameter driving the acceptability differences between conditions is the expectation of *the causal predicate itself*. The effect of a cue phrase's reference expectation on acceptability is causally mediated by the predicate's expectation.

The first proposition — that the causal predicates are key to the items' interpretation even in the non-cue conditions — can be inferred from the main effect of locality that appears in these conditions. A naïve observer might question what exactly is "non-local" in the *'non-local + no cue* condition — after all, the reader does not have a certain expectation of a causal predicate in the text (most of the filler items did not involve causal predicates), and unlike with NP anaphora and the cued conditions, there is no concrete expression in the text that clearly demands an antecedent. Other than the implicit presence of the causal predicate, there is no obvious reason for participants to have rated the *'non-local'* condition as less acceptable. That they did so suggests that even absent a cue phrase, the causal predicate is a salient feature of the texts.

The second proposition — that target locality affects how readers interpret the causal cue phrases — follows from the differing effects that the cue has in the two conditions. That the cue phrase has any effect on the ratings shows that it affects the interpretation of the items. Since its effects are not only different between conditions but actually in opposite directions (the predicted cross-over), shows that it is itself interpreted differently depending on target locality.

The third proposition — that these opposing effects reflect an underlying difference in the surprisal generated by some aspect of the text's interpretation — follows on the expectation theory of Levy (2008) (Section 1.4.2). The behavioral model (Section 2.3) suggests that the pattern of acceptability ratings in Experiments 6 and 7 could be explained in terms of surprisal – some feature of the text must more surprising in the *'local+cue'* condition than in the *'non-local+cue'* condition.

One possible interpretation of this cross-over is the reader's discovery that his or her top-down expectation of the cue phrase's referent was incorrect. This, plus an

145

added main effect for locality, could suffice to account for the fixed effects observed in data.

An explanation that is both simpler and more complete however, is that the ratings differences reflect the surprisal of the causal predicate itself. Even though surprisal-based models have typically been unable to explain away many effects that are well-predicted by the presence or absence of non-local dependencies (Patil et al., 2008; Demberg and Keller, 2009; Levy and Keller, 2010), and this explanation does not include a DLT-like non-locality effect (Gibson, 1998, 2000), surprisal may still be a better explanation than dependency length for the main effect of locality on the acceptability difference between the two 'non-cue' conditions. The causal predicates in the uncued conditions of Experiment 6 do not hinge on words or phrases that can only be interpreted in the context of a non-local dependency. As argued above, the acceptability difference between them is due to the causal predicate, which involves no lexical- or syntactic-level dependencies. As such, explaining the observed main effect of locality in terms of 'dependency locality' seems like a misuse of the DLT.

Surprisal theory suggests that a reader encountering a new clause has a top-down expectation about how likely it is that there will be a causal predicate in which the arguments are the topic of the clause attached to the referring expression and any previously introduced topic[3]. If the reader encounters a causal cue phrase, two distinct stages of inference will occur. In the first stage, the reader uses a reference model (like EMRE) to form an expectation over antecedents, estimating how likely the cue phrase is to refer to previous clauses[4]. In the second stage, this expectation is used to update the reader's separate expectation of possible causal predicates[5]

---

[3]This expectation is not a proper distribution (*i.e.*, does not cover the entire set of possible continuations), because it is not certain that the interpretation of the clause will hinge on a causal predicate. Most of the filler items in Experiment 6, for instance, do not involve causal predicates. Also, more significantly, neither do most sentences in natural text (Wolf and Gibson, 2005).

[4]In pure computational (Marr level 1) terms, the domain of the cue phrase expectation could be interpreted as a proper distribution over all previous phrases in the text (or even all possible phrases). More realistically, this would involve some limited set of candidate clauses.

[5]In the simplest case, these two computations could be performed serially, but there are both theoretical and empirical reasons to postulate that surprisal theory offers a better explanation of language comprehension if uncertainty is modeled as a multi-level inference, maintaining uncertainty about many levels of language, including inputs that have already been processed (Levy et al., 2009a).

Under this explanation, the observed fixed effects would not be independent results of (1) dependency locality, and (2) the surprisal generated by the reader learning what the cue phrase (*e.g.*, *'consequently'*) actually referred to. Instead, all differences between conditions would stem from the reader's expectation of the causal predicate, and the surprisal of the continuation of the final sentence past the cue phrase, which disambiguates the predicate.

Rather than an independent effect due to the cost of processing a long-distance dependency (Gibson (2000); Section 3.5.1), the main effect of locality in the un-cued condition here would be due to the lower expectation and greater surprisal of a non-local cause. The interaction would be the result of the reference expectation of the cue phrase, which would affects surprisal by either increasing or decreasing the expectation of the target causal predicate.

However, as discussed in Section 6.1.3, and mentioned in the final point above, under this interpretation of the results, the fixed effects observed in the data are consistent with neither an entirely top-down or an entirely bottom-up model of how readers interpret the causal cue.

This claim hinges on a cross-items comparison between the two experiments — namely that the causal cue decreases SoC both in *'local + non-prominent'* and in *'non-local + prominent'* conditions. These conditions are mutual inverses: when the target clause is in one of these conditions, the distractor must be in the other.

If the causal cue was entirely uninformative about its antecedent, then its reference expectation would simply be the normalized top-down expectation over the two clauses[6]. It would then have a general tendency across subjects and items to be interpreted as referring either to the *'local + non-prominent'* clause, or the *'non-local + prominent'* clause. However, the experimental results suggest that the causal cue tends to increase the predicate expectation of whichever of these two clauses is in fact the target. This can only be the case if the relative plausibility of the two clauses as a potential cause of the local clause is having an effect on the interpretation.

---

[6]Given the experimental results, it seems likely that the top-down reference expectation for the framing sentence that introduces each item is very near zero.

However, since the presence of the cue decreases acceptability in the *'non-local + non-prominent'* condition in both experiments, it is presumably generating a reference expectation that favors the distractor clause, thus increasing SoC. If so, it cannot be the case that readers are simply considering every available clause equally and assigning reference expectation based on bottom-up match. The data are similarly inconsistent with an account in which readers are fully considering both of the possible causes and assigning reference based on a true MAP (maximum a posteriori) likelihood. Rather, they seem to be making partial, heuristic use of bottom-up evidence to form reference expectations by updating the prior top-down expectations using this bottom-up heuristic assessment.

Therefore, under the surprisal theory of Levy (2008), these data are most consistent with a model of causal cue reference that is sensitive to both top-down prior expectations, and bottom-up feature-based match.

Chapter 7

General discussion

This chapter considers the overall design of EMRE (Section 7.1), discusses the relationship between computational and resource-driven accounts of reference (Section 7.2) and offers a few closing remarks.

7.1 Model Design

This work attempts to fill a gap in the probabilistic cognitive modeling literature. The Bayesian revolution in cognitive modeling that began two decades ago (Pearl, 1988) has swept the world of computational cognitive science and remade it in a new form. Many contemporary researchers' work with computational models is focused not on immediately practical knowledge engineering applications, but instead on using probabilistic modeling to discover the abstract structure of thought.[1] However, this wave has not yet broken in the field of discourse coherence.

The field of discourse is divided between highly application-oriented work that is very much in the non-cognitive NLP tradition (*e.g.*, Marcu (2000b)), corpus-based linguistics research (*e.g.*, Prasad et al. (2008)), and a vaguely related field of the humanities. However, probabilistic modeling of coherence has not yet become an active area of research.

One reason for this is that, unlike areas of language research like acoustics, acquisi-

---

[1]An even semi-complete list of references would fill half of this page, so it will be omitted.

tion, prosody and syntax, there is no broadly accepted way of representing coherence other than RST. Although, based on personal correspondence, most researchers who apply RST are aware that it is not a viable model of the underlying structure of discourse. Similarly, the observation that GraphBank, one of the best-known coherence models other than RST, fails to distinguish between at least two very different types of discourse[2], it would seem to have little future as a cognitive model of discourse. And the PDTB, which takes a strictly empirical approach that goes to great lengths to avoid making unnecessary theoretical commitments, is not designed to be a cognitive model.

There have been no existing models for representing discourse that are designed primarily with an eye towards representing a cognitive faculty rather than a structure in written text. EMRE is an attempt to start from the beginning, representing the smallest units of languages that could reasonably be called 'discourse'. Also, EMRE is intended to do so using 'off the shelf' components — i.e., without positing any particularly novel computational features.

Of all the entries in this work's bibliography, the structure of EMRE probably owes the most to David Marr. All computational models of cognition must navigate between the twin perils of allowing too much complexity and forcing too much simplicity. Adding more complexity can allow a model to compute more, or represent more, or promise a better fit to future data. But in order to be doing cognitive science rather than applied AI, any feature of the model must represent a computation that is earnestly proposed to be analogous to some cognitive process.

A problem particular to discourse is that it is at a very 'high level' in the language abstraction stack. The answer to how some of the computations in discourse occur is necessarily 'human intelligence'[3]. One objection that could be made to the use of a model to make the predictions that are tested in the experiments is that using a

---

[2]I believe there to be at least five.

[3]For instance, in the materials to Experiments 6 and 7, in which, if they were presented in the form of a multiple-choice test, a child would be able to tell the plausible causes from the implausible causes, whereas any existing computational model that could score above chance at this task would be performing a calculation (e.g., LSA) that obviously has nothing to do with how a human would do the task.

probabilistic model to explain why the target conditions sound worse than others is over-thinking the problem — it is possible to see that the target condition is worse by simply using 'common sense'.

One presumes that researchers in low-level vision probably occasionally hear this type of objection as well, but unlike in low-level vision, any realistic cognitive model of discourse will necessarily include arrows pointing to black boxes that are labeled *'common sense'*[4]. The 'simplest' possible model, under this way of thinking, would be to forego all the arrows, and pare the model down to a single black box.[5] EMRE, like other computational models of cognition, is an attempt to factor out cognitive processes that should in theory be well-separated enough that they interact in a constrained way, and form a theory of this interaction.

The cognitive theory to which the top-down / bottom-up distinction is intended to correspond is that people *may prepare at leisure, but must decide in haste.* This is inspired by the cognitive theories proposed by both Levy (2008); Levy et al. (2009a) and Vul (2010). These theories feature a model of expectation/belief in which: (1) a constantly maintained and updated state of calibrated uncertainty is the optimal cognitive representation; and, (2) the ability to update beliefs quickly (surprisal) is a precious and limited resource.

The top-down expectation parameter in EMRE is meant to correspond to the reader's capacity to 'prepare at leisure' — maintaining a state of optimal expectation by using all available information to be maximally prepared for what is most likely to occur next. The bottom-up match parameter of the model is meant to correspond to the reader's need to 'decide in haste'. A reader with unlimited time and cognitive resources, upon encountering a referring expression, could simultaneously apply the full inferential power of human cognition to assessing all of the possible antecedents, consider all of them for a while, and calculate the best possible posterior reference expectation.

However, since this operation must be performed rapidly and with limited cogni-

---

[4]n.b., 'boxes and arrows' here is used in jest

[5]The problem with this, of course, is that it does not force you into specific predictions — a useful model must constrain expectations.

tive resources, the update of the prior top-down expectation is probably based on a cursory inspection using surface features. This is what the feature model is meant to correspond to. In some special cases, such as pronouns and verb tenses, it is clear what some of those features are and what their effects are likely to be. In the case of causal relationships, the nature of the comparison that would be performed is much less obvious.

7.2   Computation and Memory

This work has argued that the formation of rational top-down expectations for different types of referring expression can be thought of as the formation of an *optimal code* for reference (Shannon, 1951; Huffman, 1952; Reed and Solomon, 1960; MacKay, 2002). Such a code would have the paired features of making referring expressions more concise ((Huffman, 1952), and closer to uniform in information density (Levy and Jaeger, 2007). This way of considering reference offers different explanations than those usually given for a handful of commonly-known phenomena in NP anaphora. Most obviously, the use of reduced expressions to refer to the entities most likely to be mentioned is a classic characteristic of an optimal code, and would be predicted regardless of the memory properties of the receiver. Also, the Repeated Name Penalty of Gordon and Hendrick (1998) would be predicted by a second-order expectation on the part of receivers that senders would make optimal use of such an optimal code. Failure to do so would constitute evidence that the speaker and the listener had very different expectations about the discourse (*cf.* Arnold et al. (2004)).

A common attitude in the field of psycholinguistics is that statistical approaches to such relationships between frequency and referential form are the result of constraints imposed by the memory system. Take for instance this quote from Almor and Nair (2007):

> "...as is often the case with statistically based approaches, it is not clear to what extent the statistical regularities underlie the processing of referential expressions, or simply reflect patterns generated by other con-

straints." (Almor and Nair, 2007)

This work offers a third possibility: that, to the extent that an aspect of language resembles an efficient code, this is neither the result of how statistical regularities are used in calculation, nor an accidental emergent characteristic of other cognitive constraints, but rather evidence that uniform information density is a significant aspect of language design. If expectation-based theories similar to this prove to be a useful approach to non-syntactic dependencies, it is likely to be the case that some aspects of the way that memory functions in language are the result of a pressure towards optimality.

Remarks As mentioned before, I attempted to construct this model in terms of 'off the shelf' components — computational elements that will be familiar to many researchers, to make it easy to understand and change. I hope this work will encourage other researchers with an interest in the 'higher' layers of the language abstraction stack to apply probabilistic models in their own work.

I also hope this work will be useful to coherence researchers. I would encourage more of them to join the probabilistic cognitive revolution, and develop new models that can be not only practically applicable but also consistent with what is known about the cognitive linguistic structure of coherence.

My hopes for this work in the field of NP anaphora and pronoun reference are much more modest — most contemporary findings in the field of NP anaphora are too complex to be usefully represented by a model as simple as EMRE. However, given the onwards march of the probabilistic revolution I expect that in the the near future of psycholinguistics, computational models will be associated less with application-oriented engineering and more with serious attempts to understand the mind.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

Experimental Materials

A.1 Experiment 1

A.1.1 Experimental items

1. *(while / female antecedent)*

    (a) *(local + agree)* Jenny and Karen are in the same tennis club. Jenny often skips practice to sleep in, while Karen arrives early every Saturday morning. She's always eager for a game.

    (b) *(nonlocal + agree)* Jenny and Karen are in the same tennis club. While Karen arrives early every Saturday morning, Jenny often skips practice to sleep in. She's always eager for a game.

    (c) *(local + disagree)* Josh and Karen are in the same tennis club. Josh often skips practice to sleep in, while Karen arrives early every Saturday morning. She's always eager for a game.

    (d) *(nonlocal + disagree)* Josh and Karen are in the same tennis club. While Karen arrives early every Saturday morning, Josh often skips practice to sleep in. She's always eager for a game.

2. *(because / male antecedent)*

    (a) *(local + agree)* Frank is Matt's assistant at an accounting firm. Frank took detailed notes during the last meeting, because Matt was out of town on business. He often travels to meet with clients.

    (b) *(nonlocal + agree)* Frank is Matt's assistant at an accounting firm. Because Matt was out of town on business, Frank took detailed notes during the last meeting. He often travels to meet with clients.

    (c) *(local + disagree)* Liz is Matt's assistant at an accounting firm. Liz took detailed notes during the last meeting, because Matt was out of town on business. He often travels to meet with clients.

    (d) *(nonlocal + disagree)* Liz is Matt's assistant at an accounting firm. Because Matt was out of town on business, Liz took detailed notes during the last meeting.

He often travels to meet with clients.

3. *(because / female antecedent)*

   (a) *(local + agree)* Susan and Beth were making a cross-country trip together. Susan had to do all the driving, because Beth couldn't drive a stick shift. She paid the tolls and handled navigation instead.

   (b) *(nonlocal + agree)* Susan and Beth were making a cross-country trip together. Because Beth couldn't drive a stick shift, Susan had to do all the driving. She paid the tolls and handled navigation instead.

   (c) *(local + disagree)* John and Beth were making a cross-country trip together. John had to do all the driving, because Beth couldn't drive a stick shift. She paid the tolls and handled navigation instead.

   (d) *(nonlocal + disagree)* John and Beth were making a cross-country trip together. Because Beth couldn't drive a stick shift, John had to do all the driving. She paid the tolls and handled navigation instead.

4. *(while / male antecedent)*

   (a) *(local + agree)* Joe and Tom are roommates. Joe is very tidy, while Tom is a slob. He's always leaving messes around the apartment.

   (b) *(nonlocal + agree)* Joe and Tom are roommates. While Tom is a slob, Joe is very tidy. He's always leaving messes around the apartment.

   (c) *(local + disagree)* Hannah and Tom are roommates. Hannah is very tidy, while Tom is a slob. He's always leaving messes around the apartment.

   (d) *(nonlocal + disagree)* Hannah and Tom are roommates. While Tom is a slob, Hannah is very tidy. He's always leaving messes around the apartment.

5. *(since / female antecedent)*

   (a) *(local + agree)* Linda is Barbara's next door neighbor. Linda is shoveling the snow on both sidewalks, since Barbara is out of town on business. She won't be back for a week.

   (b) *(nonlocal + agree)* Linda is Barbara's next door neighbor. Since Barbara is out of town on business, Linda is shoveling the snow on both sidewalks. She won't be back for a week.

   (c) *(local + disagree)* Jack is Barbara's next door neighbor. Jack is shoveling the snow on both sidewalks, since Barbara is out of town on business. She won't be back for a week.

   (d) *(nonlocal + disagree)* Jack is Barbara's next door neighbor. Since Barbara is out of town on business, Jack is shoveling the snow on both sidewalks. She won't be back for a week.

6. *(since / male antecedent)*

   (a) *(local + agree)* Erik takes skiing lessons from Rick. Erik has been training alone, since Rick has a sprained wrist. He's expected to make a full recovery soon.

   (b) *(nonlocal + agree)* Erik takes skiing lessons from Rick. Since Rick has a sprained wrist, Erik has been training alone. He's expected to make a full recovery soon.

   (c) *(local + disagree)* Jill takes skiing lessons from Rick. Jill has been training alone, since Rick has a sprained wrist. He's expected to make a full recovery soon.

(d) *(nonlocal + disagree)* Jill takes skiing lessons from Rick. Since Rick has a sprained wrist, Jill has been training alone. He's expected to make a full recovery soon.

7. *(because / female antecedent)*

  (a) *(local + agree)* Betty and her older sister Lisa go to the same college. Betty saves a lot of money on school supplies, because Lisa is a year ahead and passes her textbooks on. She's always been generous with her family.

  (b) *(nonlocal + agree)* Betty and her older sister Lisa go to the same college. Because Lisa is a year ahead and passes her textbooks on, Betty saves a lot of money on school supplies. She's always been generous with her family.

  (c) *(local + disagree)* Mark and his older sister Lisa go to the same college. Mark saves a lot of money on school supplies, because Lisa is a year ahead and passes her textbooks on. She's always been generous with her family.

  (d) *(nonlocal + disagree)* Mark and his older sister Lisa go to the same college. Because Lisa is a year ahead and passes her textbooks on, Mark saves a lot of money on school supplies. She's always been generous with her family.

8. *(before / male antecedent)*

  (a) *(local + agree)* Donald and Jacob are volunteering at a park today. Donald needs to clean up the sporting equipment and litter, before Jacob can cut the grass on the game fields. He has a power mower that makes fast work of it.

  (b) *(nonlocal + agree)* Donald and Jacob are volunteering at a park today. Before Jacob can cut the grass on the game fields, Donald needs to clean up the sporting equipment and litter. He has a power mower that makes fast work of it.

  (c) *(local + disagree)* Maria and Jacob are volunteering at a park today. Maria needs to clean up the sporting equipment and litter, before Jacob can cut the grass on the game fields. He has a power mower that makes fast work of it.

  (d) *(nonlocal + disagree)* Maria and Jacob are volunteering at a park today. Before Jacob can cut the grass on the game fields, Maria needs to clean up the sporting equipment and litter. He has a power mower that makes fast work of it.

9. *(because / female antecedent)*

  (a) *(local + agree)* Kim is babysitting Julia on Saturday afternoon. Kim put a soccer ball and net in her car's trunk, because Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

  (b) *(nonlocal + agree)* Kim is babysitting Julia on Saturday afternoon. Because Julia always wants to play outdoor sports, Kim put a soccer ball and net in her car's trunk. She doesn't get enough exercise during her week in kindergarten.

  (c) *(local + disagree)* Kevin is babysitting Julia on Saturday afternoon. Kevin put a soccer ball and net in his car's trunk, because Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

  (d) *(nonlocal + disagree)* Kevin is babysitting Julia on Saturday afternoon. Because Julia always wants to play outdoor sports, Kevin put a soccer ball and net in his car's trunk. She doesn't get enough exercise during her week in kindergarten.

10. *(while / male antecedent)*

(a) *(local + agree)* Adam and Steve met for lunch. Adam brought food from home, while Steve had to buy something. He often forgets to bring his lunch.

(b) *(nonlocal + agree)* Adam and Steve met for lunch. While Steve had to buy something, Adam brought food from home. He often forgets to bring his lunch.

(c) *(local + disagree)* Carol and Steve met for lunch. Carol brought food from home, while Steve had to buy something. He often forgets to bring his lunch.

(d) *(nonlocal + disagree)* Carol and Steve met for lunch. While Steve had to buy something, Carol brought food from home. He often forgets to bring his lunch.

11. *(while / female antecedent)*

(a) *(local + agree)* Elizabeth and Nancy were tidying up their diving club's reading room. Elizabeth cleaned off the tables and mopped the floor, while Nancy organized the books by topic. She had once been to library school.

(b) *(nonlocal + agree)* Elizabeth and Nancy were tidying up their diving club's reading room. While Nancy organized the books by topic, Elizabeth cleaned off the tables and mopped the floor. She had once been to library school.

(c) *(local + disagree)* Roland and Nancy were tidying up their diving club's reading room. Roland cleaned off the tables and mopped the floor, while Nancy organized the books by topic. She had once been to library school.

(d) *(nonlocal + disagree)* Roland and Nancy were tidying up their diving club's reading room. While Nancy organized the books by topic, Roland cleaned off the tables and mopped the floor. She had once been to library school.

12. *(because / male antecedent)*

(a) *(local + agree)* Edward and Daniel were assigned to be partners for a school project. Edward has to do a lot of extra work, because Daniel mostly messes around during study sessions. He's always been rather lazy.

(b) *(nonlocal + agree)* Edward and Daniel were assigned to be partners for a school project. Because Daniel mostly messes around during study sessions, Edward has to do a lot of extra work. He's always been rather lazy.

(c) *(local + disagree)* Liza and Daniel were assigned to be partners for a school project. Liza has to do a lot of extra work, because Daniel mostly messes around during study sessions. He's always been rather lazy.

(d) *(nonlocal + disagree)* Liza and Daniel were assigned to be partners for a school project. Because Daniel mostly messes around during study sessions, Liza has to do a lot of extra work. He's always been rather lazy.

13. *(while / female antecedent)*

(a) *(local + agree)* Ashley and Hannah have been friends since they were children. Ashley has lived on the same street her whole life, while Hannah went away to college and never moved back. She comes back to visit every summer.

(b) *(nonlocal + agree)* Ashley and Hannah have been friends since they were children. While Hannah went away to college and never moved back, Ashley has lived on the same street her whole life. She comes back to visit every summer.

(c) *(local + disagree)* Brandon and Hannah have been friends since they were children. Brandon has lived on the same street his whole life, while Hannah went away to college and never moved back. She comes back to visit every summer.

(d) *(nonlocal + disagree)* Brandon and Hannah have been friends since they were children. While Hannah went away to college and never moved back, Brandon has lived on the same street his whole life. She comes back to visit every summer.

14. *(because / male antecedent)*

(a) *(local + agree)* Zack was going to William's birthday party. Zack couldn't find a good gift to bring, because William is difficult to shop for. He prefers to have very few material possessions.

(b) *(nonlocal + agree)* Zack was going to William's birthday party. Because William is difficult to shop for, Zack couldn't find a good gift to bring. He prefers to have very few material possessions.

(c) *(local + disagree)* Jessica was going to William's birthday party. Jessica couldn't find a good gift to bring, because William is difficult to shop for. He prefers to have very few material possessions.

(d) *(nonlocal + disagree)* Jessica was going to William's birthday party. Because William is difficult to shop for, Jessica couldn't find a good gift to bring. He prefers to have very few material possessions.

15. *(because / female antecedent)*

(a) *(local + agree)* Megan manages a softball club, and Lauren is their starting pitcher. Megan is considering finding someone else, because Lauren is having elbow problems. She's been playing so much that she developed a repetitive stress injury.

(b) *(nonlocal + agree)* Megan manages a softball club, and Lauren is their starting pitcher. Because Lauren is having elbow problems, Megan is considering finding someone else. She's been playing so much that she developed a repetitive stress injury.

(c) *(local + disagree)* Anthony manages a softball club, and Lauren is their starting pitcher. Anthony is considering finding someone else, because Lauren is having elbow problems. She's been playing so much that she developed a repetitive stress injury.

(d) *(nonlocal + disagree)* Anthony manages a softball club, and Lauren is their starting pitcher. Because Lauren is having elbow problems, Anthony is considering finding someone else. She's been playing so much that she developed a repetitive stress injury.

16. *(while / male antecedent)*

(a) *(local + agree)* Justin and James share a computer. Justin only really knows how to use email and the web, while James uses it heavily for work. He has a job as a software engineer.

(b) *(nonlocal + agree)* Justin and James share a computer. While James uses it heavily for work, Justin only really knows how to use email and the web. He has a job as a software engineer.

(c) *(local + disagree)* Emma and James share a computer. Emma only really knows how to use email and the web, while James uses it heavily for work. He has a job as a software engineer.

(d) *(nonlocal + disagree)* Emma and James share a computer. While James uses it heavily for work, Emma only really knows how to use email and the web. He

has a job as a software engineer.

A.1.2 Filler items

1. Tristan and Hannah went shopping for shoes. Tristan bought black sneakers, while Hannah opted for red pumps. They were both happy with what they found.
2. Isaac and Jenna are part of the same church. Isaac loaned Jenna his car because hers was broken. He asked her to return it by Monday.
3. Dorothy and Trent are subcontracting on the same house. After Trent finishes the wiring, Dorothy will be able to finish the plaster. He's taking a very long time to get it done.
4. Alan and Monica go rock climbing together. Since Monica is an expert climber, Alan usually takes her advice. With her advice, he is improving rapidly.
5. Keith and Bradley are brothers. Because Bradley went to the same high school, Keith always gets compared to him. They aren't actually very similar, though.
6. Cassie and Veronica both like to go trout fishing. Though Cassie sometimes brings her boyfriend, Veronica prefers the company of just one companion. She's always been a quiet person.
7. Alisha and Garrett share a driveway. Garrett often takes up too much space, while Alisha never shovels after it snows. They're often annoyed with each other.
8. Marissa and Kelly are old friends, and study martial arts at the same dojo. While Marissa just started, Kelly has been a black belt for many years. She gives her old friend a lot of extra help.
9. Henry and Melinda work at the same restaurant. Melinda is a waiter and earns decent money, while Henry washes pots in the kitchen. He's been promised a wait staff position next time one of the waiters leaves.
10. Damon works as Hunter's golf caddy on weekends. Hunter is a much weaker player, but Damon doesn't give him a lot of advice. He's planning to become a golf instructor soon so he can get paid better.
11. Ian and Evan work at the same company. Evan is always at work early, while Ian stays late. At least one of them is usually there when something urgent comes up.
12. Maggie and Janice are starting a catering company together. Once Maggie finishes filing the tax paperwork, Janice will place ads in the local paper. They're hoping to be in business by the end of the month
13. Courtney brought her boyfriend Tyler to her family reunion. After she left to visit with her cousins, he snuck off to watch football. They got in a fight when she was unable to find him.
14. Chad and Peter were in the same tour group together. Chad is a mechanical engineer, while Peter is an architect. They spent a lot of time talking about the buildings they saw on the trip.
15. Edward and his wife Monica are both avid gardeners. Edward likes to grow vegetables, while Monica enjoys ornamental gardening. They haggle with each other about space every spring.
16. Meredith and Roger are neighbors, and their kids are in the same day care. Meredith walks the kids there in the morning, then Roger walks them home for lunch. After that, he goes to work the late shift at his job.

17. Bridget and Jay run a soup kitchen together. Bridget manages the organization and money, while Jay handles the kitchen. They feed about thirty people a day.

18. Terrance and Levi play in the same football club. Terrance is part of the offensive line, while Levi plays defence. They trade off watching their sons on the sidelines.

19. Autumn and Jerome are trainers at a gym. Autumn is a swimming instructor, while Jerome is a weight training coach. They often go running together after work.

20. Ruth and Cara share art studio space. Cara's paintings cover most of the walls, while Ruth's large sculptures occupy the middle of the room. They seldom have to negotiate about display space.

21. Amy and Marvin volunteer at a before-school art program. Because Amy lives on his way, Marvin often gives her a ride there. She enjoys the exercise of walking home though.

22. Cliff and his friend Stanley are amateur astronomers. Because Cliff enjoys managing the equipment, Stanley mostly collaborates on planning and analysis. He sometimes never even leaves the warmth of his study while they're working.

23. Andy and Gloria eat at the same diner every morning. Because Andy isn't very friendly, Gloria rarely says hello to him. He's always been something of a loner.

24. Sharon and Felicia used to bike to work together. Because Sharon moved away, Felicia has been taking the bus more often. She's worried about getting out of shape.

25. Joanna and Ross are in the community theatre together. Because Ross hasn't learned all of his lines, Joanna hasn't been able to rehearse some of her scenes. She's very annoyed with him.

26. Clint and Hector drive the same bus on different shifts. Since Clint doesn't fully prep the bus afterwards, Hector has to clean up after him. He's complained to their supervisor, but nothing changed.

27. Arthur and Carolyn went to the same rural high school. After Carolyn moved to the city he lived in, she and Arthur got back in touch. They eventually ended up getting married.

28. Robin and Wendy taught at the same school for decades. After Robin retired, Wendy stopped enjoying her job. She retired the very next year.

29. Deborah and Walter are on the school board together. Walter cares only about athletics, while Deborah doesn't care about athletics at all. They often fight over the allocation of money.

30. Nicolas works for Glenn at an antique shop. Because Glenn is often on the road going to auctions, Nicolas is left alone to run the shop. He's better at cataloging the inventory than selling it, though.

31. Eugene and Paula went on a date. Because Eugene almost never stopped talking, Paula had trouble participating in the conversation. She's not planning to go out with him again.

32. Esther is on retainer as Lacey's legal counsel. After Lacey was slandered in the local paper, Esther filed a libel suit on her behalf. They think that it will be settled out of court.

A.2 Experiment 2 materials

A.2.1 Experimental items

1. *(While/while / female target)*

   (a) *(prominent + agree)* Kim and Julia are in the same fencing club. Julia arrives early every Sunday morning to warm up before practice while Kim is still asleep in bed. She's always eager for a match.

   (b) *(subordinate + agree)* Kim and Julia are in the same fencing club. While Julia arrives early every Sunday morning to warm up before practice, Kim is still asleep in bed. She's always eager for a match.

   (c) *(prominent + disagree)* Kevin and Julia are in the same fencing club. Julia arrives early every Sunday morning to warm up before practice while Kevin is still asleep in bed. She's always eager for a match.

   (d) *(subordinate + disagree)* Kevin and Julia are in the same fencing club. While Julia arrives early every Sunday morning to warm up before practice, Kevin is still asleep in bed. She's always eager for a match.

2. *(Because/so / male target)*

   (a) *(prominent + agree)* Adam is Steve's secretary at an architecture firm. Steve was visiting a building site in another city, so Adam took notes at meetings. He often travels to check on sites.

   (b) *(subordinate + agree)* Adam is Steve's secretary at an architecture firm. Because Steve was visiting a building site in another city, Adam took notes at meetings. He often travels to check on sites.

   (c) *(prominent + disagree)* Carol is Steve's secretary at an architecture firm. Steve was visiting a building site in another city, so Carol took notes at meetings. He often travels to check on sites.

   (d) *(subordinate + disagree)* Carol is Steve's secretary at an architecture firm. Because Steve was visiting a building site in another city, Carol took notes at meetings. He often travels to check on sites.

3. *(After/before / female target)*

   (a) *(prominent + agree)* Elizabeth and Nancy are volunteering at a park today. Nancy cleaned up sports equipment and litter before Elizabeth mowed the fields. She filled ten garbage bags with trash from the ground.

   (b) *(subordinate + agree)* Elizabeth and Nancy are volunteering at a park today. After Nancy cleaned up sports equipment and litter, Elizabeth mowed the fields. She filled ten garbage bags with trash from the ground.

   (c) *(prominent + disagree)* Roland and Nancy are volunteering at a park today. Nancy cleaned up sports equipment and litter before Roland mowed the fields. She filled ten garbage bags with trash from the ground.

   (d) *(subordinate + disagree)* Roland and Nancy are volunteering at a park today. After Nancy cleaned up sports equipment and litter, Roland mowed the fields. She filled ten garbage bags with trash from the ground.

4. *(Because/so / male target)*

(a) *(prominent + agree)* Edward and Daniel are bicycling through the mountains together. Daniel is a master bike mechanic with a complete roadside toolkit so Edward is short on equipment. He has to stay nearby or risk stranding Edward in the mountains.

(b) *(subordinate + agree)* Edward and Daniel are bicycling through the mountains together. Because Daniel is a master bike mechanic with a complete roadside toolkit, Edward is short on equipment. He has to stay nearby or risk stranding Edward in the mountains.

(c) *(prominent + disagree)* Liza and Daniel are bicycling through the mountains together. Daniel is a master bike mechanic with a complete roadside toolkit so Liza is short on equipment. He has to stay nearby or risk stranding Liza in the mountains.

(d) *(subordinate + disagree)* Liza and Daniel are bicycling through the mountains together. Because Daniel is a master bike mechanic with a complete roadside toolkit, Liza is short on equipment. He has to stay nearby or risk stranding Liza in the mountains.

5. *(Because/so / female target)*

(a) *(prominent + agree)* Ashley is babysitting Hannah on Saturday afternoon. Hannah always wants to play active outdoor sports, so Ashley packed a soccer ball. She doesn't get enough exercise during her week in kindergarten.

(b) *(subordinate + agree)* Ashley is babysitting Hannah on Saturday afternoon. Because Hannah always wants to play active outdoor sports, Ashley packed a soccer ball. She doesn't get enough exercise during her week in kindergarten.

(c) *(prominent + disagree)* Brandon is babysitting Hannah on Saturday afternoon. Hannah always wants to play active outdoor sports, so Brandon packed a soccer ball. She doesn't get enough exercise during her week in kindergarten.

(d) *(subordinate + disagree)* Brandon is babysitting Hannah on Saturday afternoon. Because Hannah always wants to play active outdoor sports, Brandon packed a soccer ball. She doesn't get enough exercise during her week in kindergarten.

6. *(While/though / male target)*

(a) *(prominent + agree)* Zack and William met in the park for lunch. William had to hit a restaurant for carry-out, though Zack had a packed lunch. He never brings his lunch to work.

(b) *(subordinate + agree)* Zack and William met in the park for lunch. While William had to hit a restaurant for carry-out, Zack had a packed lunch. He never brings his lunch to work.

(c) *(prominent + disagree)* Jessica and William met in the park for lunch. William had to hit a restaurant for carry-out, though Jessica had a packed lunch. He never brings his lunch to work.

(d) *(subordinate + disagree)* Jessica and William met in the park for lunch. While William had to hit a restaurant for carry-out, Jessica had a packed lunch. He never brings his lunch to work.

7. *(While/while / female target)*

(a) *(prominent + agree)* Megan and Lauren were doing an inventory at their sailing club. Lauren inspected the ropes and sails for damage while Megan cataloged

the books. She found one sail missing, and three more badly torn.

(b) *(subordinate + agree)* Megan and Lauren were doing an inventory at their sailing club. While Lauren inspected the ropes and sails for damage, Megan cataloged the books. She found one sail missing, and three more badly torn.

(c) *(prominent + disagree)* Anthony and Lauren were doing an inventory at their sailing club. Lauren inspected the ropes and sails for damage while Anthony cataloged the books. She found one sail missing, and three more badly torn.

(d) *(subordinate + disagree)* Anthony and Lauren were doing an inventory at their sailing club. While Lauren inspected the ropes and sails for damage, Anthony cataloged the books. She found one sail missing, and three more badly torn.

8. *(Because/so / male target)*

(a) *(prominent + agree)* Justin and James are on the same team at an automotive manufacturing plant. James makes a lot of mistakes while assembling the car bodies, so Justin has to fix the problems. He's always been careless on the job.

(b) *(subordinate + agree)* Justin and James are on the same team at an automotive manufacturing plant. Because James makes a lot of mistakes while assembling the car bodies, Justin has to fix the problems. He's always been careless on the job.

(c) *(prominent + disagree)* Emma and James are on the same team at an automotive manufacturing plant. James makes a lot of mistakes while assembling the car bodies, so Emma has to fix the problems. He's always been careless on the job.

(d) *(subordinate + disagree)* Emma and James are on the same team at an automotive manufacturing plant. Because James makes a lot of mistakes while assembling the car bodies, Emma has to fix the problems. He's always been careless on the job.

9. *(Because/so / female target)*

(a) *(prominent + agree)* Jenny is Karen's next door neighbor. Karen is overseas doing factory safety inspections, so Jenny is mowing both lawns. She won't be back for a month.

(b) *(subordinate + agree)* Jenny is Karen's next door neighbor. Because Karen is overseas doing factory safety inspections, Jenny is mowing both lawns. She won't be back for a month.

(c) *(prominent + disagree)* Josh is Karen's next door neighbor. Karen is overseas doing factory safety inspections, so Josh is mowing both lawns. She won't be back for a month.

(d) *(subordinate + disagree)* Josh is Karen's next door neighbor. Because Karen is overseas doing factory safety inspections, Josh is mowing both lawns. She won't be back for a month.

10. *(Because/so / male target)*

(a) *(prominent + agree)* Matt tutors Frank in algebra three times a week. Matt has been stuck at home with strep throat this week, so Frank did poorly on a quiz. He's expected to get his doctor's permission to go out soon.

(b) *(subordinate + agree)* Matt tutors Frank in algebra three times a week. Because Matt has been stuck at home with strep throat this week, Frank did poorly on a quiz. He's expected to get his doctor's permission to go out soon.

(c) *(prominent + disagree)* Matt tutors Liz in algebra three times a week. Matt has been stuck at home with strep throat this week, so Liz did poorly on a quiz. He's expected to get his doctor's permission to go out soon.

(d) *(subordinate + disagree)* Matt tutors Liz in algebra three times a week. Because Matt has been stuck at home with strep throat this week, Liz did poorly on a quiz. He's expected to get his doctor's permission to go out soon.

11. *(Because/so / female target)*

(a) *(prominent + agree)* Susan and her older cousin Beth both study chemical engineering. Beth hands down the textbooks from the classes she takes so Susan doesn't have to spend much on books. She's always generous towards her family.

(b) *(subordinate + agree)* Susan and her older cousin Beth both study chemical engineering. Because Beth hands down the textbooks from the classes she takes, Susan doesn't have to spend much on books. She's always generous towards her family.

(c) *(prominent + disagree)* John and his older cousin Beth both study chemical engineering. Beth hands down the textbooks from the classes she takes so John doesn't have to spend much on books. She's always generous towards her family.

(d) *(subordinate + disagree)* John and his older cousin Beth both study chemical engineering. Because Beth hands down the textbooks from the classes she takes, John doesn't have to spend much on books. She's always generous towards her family.

12. *(While/while / male target)*

(a) *(prominent + agree)* Joe and Tom are roommates. Tom is a slob who leaves trash around their house while Joe does the cleaning up. He never takes care of his messes.

(b) *(subordinate + agree)* Joe and Tom are roommates. While Tom is a slob who leaves trash around their house, Joe does the cleaning up. He never takes care of his messes.

(c) *(prominent + disagree)* Hannah and Tom are roommates. Tom is a slob who leaves trash around their house while Hannah does the cleaning up. He never takes care of his messes.

(d) *(subordinate + disagree)* Hannah and Tom are roommates. While Tom is a slob who leaves trash around their house, Hannah does the cleaning up. He never takes care of his messes.

13. *(While/while / female target)*

(a) *(prominent + agree)* Linda and Barbara were good friends in high school. Barbara has lived on the same street her whole life while Linda has drifted around the globe. She looks forward to hearing stories about far off places when Linda visits her.

(b) *(subordinate + agree)* Linda and Barbara were good friends in high school. While Barbara has lived on the same street her whole life, Linda has drifted around the globe. She looks forward to hearing stories about far off places when Linda visits her.

(c) *(prominent + disagree)* Jack and Barbara were good friends in high school. Barbara has lived on the same street her whole life while Jack has drifted around

the globe. She looks forward to hearing stories about far off places when Jack visits her.

(d) *(subordinate + disagree)* Jack and Barbara were good friends in high school. While Barbara has lived on the same street her whole life, Jack has drifted around the globe. She looks forward to hearing stories about far off places when Jack visits her.

14. *(Since/so / male target)*

(a) *(prominent + agree)* Erik was a guest at Rick's retirement party. Rick is extremely difficult to shop for, so Erik simply brought a card. He's very particular about what he enjoys.

(b) *(subordinate + agree)* Erik was a guest at Rick's retirement party. Since Rick is extremely difficult to shop for, Erik simply brought a card. He's very particular about what he enjoys.

(c) *(prominent + disagree)* Jill was a guest at Rick's retirement party. Rick is extremely difficult to shop for, so Jill simply brought a card. He's very particular about what he enjoys.

(d) *(subordinate + disagree)* Jill was a guest at Rick's retirement party. Since Rick is extremely difficult to shop for, Jill simply brought a card. He's very particular about what he enjoys.

15. *(Because/so / female target)*

(a) *(prominent + agree)* Betty manages a soccer club, and Lisa is their goalie. Lisa bruised her hip in a car crash, so Betty has been fielding stand-ins. She's taking physical therapy seriously so she can get back to her goal.

(b) *(subordinate + agree)* Betty manages a soccer club, and Lisa is their goalie. Because Lisa bruised her hip in a car crash, Betty has been fielding stand-ins. She's taking physical therapy seriously so she can get back to her goal.

(c) *(prominent + disagree)* Mark manages a soccer club, and Lisa is their goalie. Lisa bruised her hip in a car crash, so Mark has been fielding stand-ins. She's taking physical therapy seriously so she can get back to her goal.

(d) *(subordinate + disagree)* Mark manages a soccer club, and Lisa is their goalie. Because Lisa bruised her hip in a car crash, Mark has been fielding stand-ins. She's taking physical therapy seriously so she can get back to her goal.

16. *(While/while / male target)*

(a) *(prominent + agree)* Donald and Jacob share a computer. Jacob uses it heavily for his job as a software engineer while Donald just browses the web. He has dozens of software development tools installed.

(b) *(subordinate + agree)* Donald and Jacob share a computer. While Jacob uses it heavily for his job as a software engineer, Donald just browses the web. He has dozens of software development tools installed.

(c) *(prominent + disagree)* Maria and Jacob share a computer. Jacob uses it heavily for his job as a software engineer while Maria just browses the web. He has dozens of software development tools installed.

(d) *(subordinate + disagree)* Maria and Jacob share a computer. While Jacob uses it heavily for his job as a software engineer, Maria just browses the web. He has dozens of software development tools installed.

A.2.2 Filler items

(same as Experiment 1)

A.3 Experiment 3 materials

A.3.1 Experimental items

1. *(while / female target)*

    (a) *(local + main + disagree)* Josh and Karen are in the same tennis club. While Josh often skips practice to sleep in, Karen arrives early every Saturday morning. She's always eager for a game.

    (b) *(nonlocal + main + disagree)* Josh and Karen are in the same tennis club. Karen arrives early every Saturday morning, while Josh often skips practice to sleep in. She's always eager for a game.

    (c) *(local + subord + disagree)* Josh and Karen are in the same tennis club. Josh often skips practice to sleep in, while Karen arrives early every Saturday morning. She's always eager for a game.

    (d) *(nonlocal + subord + disagree)* Josh and Karen are in the same tennis club. While Karen arrives early every Saturday morning, Josh often skips practice to sleep in. She's always eager for a game.

    (e) *(local + main + agree)* Jenny and Karen are in the same tennis club. While Jenny often skips practice to sleep in, Karen arrives early every Saturday morning. She's always eager for a game.

    (f) *(nonlocal + main + agree)* Jenny and Karen are in the same tennis club. Karen arrives early every Saturday morning, while Jenny often skips practice to sleep in. She's always eager for a game.

    (g) *(local + subord + agree)* Jenny and Karen are in the same tennis club. Jenny often skips practice to sleep in, while Karen arrives early every Saturday morning. She's always eager for a game.

    (h) *(nonlocal + subord + agree)* Jenny and Karen are in the same tennis club. While Karen arrives early every Saturday morning, Jenny often skips practice to sleep in. She's always eager for a game.

2. *(because / male target)*

    (a) *(local + main + disagree)* Liz is Matt's assistant at an accounting firm. Liz took detailed notes during the last meeting. Matt was out of town on business. He often travels to meet with clients.

    (b) *(nonlocal + main + disagree)* Liz is Matt's assistant at an accounting firm. Matt was out of town on business, so Liz took detailed notes during the last meeting. He often travels to meet with clients.

    (c) *(local + subord + disagree)* Liz is Matt's assistant at an accounting firm. Liz took detailed notes during the last meeting because Matt was out of town on business. He often travels to meet with clients.

    (d) *(nonlocal + subord + disagree)* Liz is Matt's assistant at an accounting firm. Because Matt was out of town on business, Liz took detailed notes during the last meeting. He often travels to meet with clients.

(e) *(local + main + agree)* Frank is Matt's assistant at an accounting firm. Frank took detailed notes during the last meeting. Matt was out of town on business. He often travels to meet with clients.

(f) *(nonlocal + main + agree)* Frank is Matt's assistant at an accounting firm. Matt was out of town on business, so Frank took detailed notes during the last meeting. He often travels to meet with clients.

(g) *(local + subord + agree)* Frank is Matt's assistant at an accounting firm. Frank took detailed notes during the last meeting because Matt was out of town on business. He often travels to meet with clients.

(h) *(nonlocal + subord + agree)* Frank is Matt's assistant at an accounting firm. Because Matt was out of town on business, Frank took detailed notes during the last meeting. He often travels to meet with clients.

3. *(because / female target)*

(a) *(local + main + disagree)* John is taking Beth along on a cross-country trip. John has to do all the driving. Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

(b) *(nonlocal + main + disagree)* John is taking Beth along on a cross-country trip. Beth can't work the clutch without stalling, so John has to do all the driving. She has never driven a car with a stick shift.

(c) *(local + subord + disagree)* John is taking Beth along on a cross-country trip. John has to do all the driving because Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

(d) *(nonlocal + subord + disagree)* John is taking Beth along on a cross-country trip. Because Beth can't work the clutch without stalling, John has to do all the driving. She has never driven a car with a stick shift.

(e) *(local + main + agree)* Susan is taking Beth along on a cross-country trip. Susan has to do all the driving. Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

(f) *(nonlocal + main + agree)* Susan is taking Beth along on a cross-country trip. Beth can't work the clutch without stalling, so Susan has to do all the driving. She has never driven a car with a stick shift.

(g) *(local + subord + agree)* Susan is taking Beth along on a cross-country trip. Susan has to do all the driving because Beth can't work the clutch without stalling. She has never driven a car with a stick shift.

(h) *(nonlocal + subord + agree)* Susan is taking Beth along on a cross-country trip. Because Beth can't work the clutch without stalling, Susan has to do all the driving. She has never driven a car with a stick shift.

4. *(while / male target)*

(a) *(local + main + disagree)* Hannah and Tom are roommates. While Hannah is a very tidy person who cleans constantly, Tom is a slob. He's always leaving messes around the apartment.

(b) *(nonlocal + main + disagree)* Hannah and Tom are roommates. Tom is a slob, while Hannah is a very tidy person who cleans constantly. He's always leaving messes around the apartment.

(c) *(local + subord + disagree)* Hannah and Tom are roommates. Hannah is a very

tidy person who cleans constantly, while Tom is a slob. He's always leaving messes around the apartment.

(d) *(nonlocal + subord + disagree)* Hannah and Tom are roommates. While Tom is a slob, Hannah is a very tidy person who cleans constantly. He's always leaving messes around the apartment.

(e) *(local + main + agree)* Joe and Tom are roommates. While Joe is a very tidy person who cleans constantly, Tom is a slob. He's always leaving messes around the apartment.

(f) *(nonlocal + main + agree)* Joe and Tom are roommates. Tom is a slob, while Joe is a very tidy person who cleans constantly. He's always leaving messes around the apartment.

(g) *(local + subord + agree)* Joe and Tom are roommates. Joe is a very tidy person who cleans constantly, while Tom is a slob. He's always leaving messes around the apartment.

(h) *(nonlocal + subord + agree)* Joe and Tom are roommates. While Tom is a slob, Joe is a very tidy person who cleans constantly. He's always leaving messes around the apartment.

5. *(since / female target)*

(a) *(local + main + disagree)* Jack has been shoveling Barbara's sidewalk. Jack has had to deal with a tremendous amount of snow. Barbara has been away on business for weeks. She's getting back to town tomorrow.

(b) *(nonlocal + main + disagree)* Jack has been shoveling Barbara's sidewalk. Barbara has been away on business for weeks, so Jack has had to deal with a tremendous amount of snow. She's getting back to town tomorrow.

(c) *(local + subord + disagree)* Jack has been shoveling Barbara's sidewalk. Jack has had to deal with a tremendous amount of snow since Barbara has been away on business for weeks. She's getting back to town tomorrow.

(d) *(nonlocal + subord + disagree)* Jack has been shoveling Barbara's sidewalk. Since Barbara has been away on business for weeks, Jack has had to deal with a tremendous amount of snow. She's getting back to town tomorrow.

(e) *(local + main + agree)* Linda has been shoveling Barbara's sidewalk. Linda has had to deal with a tremendous amount of snow. Barbara has been away on business for weeks. She's getting back to town tomorrow.

(f) *(nonlocal + main + agree)* Linda has been shoveling Barbara's sidewalk. Barbara has been away on business for weeks, so Linda has had to deal with a tremendous amount of snow. She's getting back to town tomorrow.

(g) *(local + subord + agree)* Linda has been shoveling Barbara's sidewalk. Linda has had to deal with a tremendous amount of snow since Barbara has been away on business for weeks. She's getting back to town tomorrow.

(h) *(nonlocal + subord + agree)* Linda has been shoveling Barbara's sidewalk. Since Barbara has been away on business for weeks, Linda has had to deal with a tremendous amount of snow. She's getting back to town tomorrow.

6. *(since / male target)*

(a) *(local + main + disagree)* Jill takes skiing lessons from Rick. Jill has been training alone. Rick has a sprained ankle. He's expected to be off his crutches

soon.

    (b) *(nonlocal + main + disagree)* Jill takes skiing lessons from Rick. Rick has a sprained ankle, so Jill has been training alone. He's expected to be off his crutches soon.

    (c) *(local + subord + disagree)* Jill takes skiing lessons from Rick. Jill has been training alone since Rick has a sprained ankle. He's expected to be off his crutches soon.

    (d) *(nonlocal + subord + disagree)* Jill takes skiing lessons from Rick. Since Rick has a sprained ankle, Jill has been training alone. He's expected to be off his crutches soon.

    (e) *(local + main + agree)* Erik takes skiing lessons from Rick. Erik has been training alone. Rick has a sprained ankle. He's expected to be off his crutches soon.

    (f) *(nonlocal + main + agree)* Erik takes skiing lessons from Rick. Rick has a sprained ankle, so Erik has been training alone. He's expected to be off his crutches soon.

    (g) *(local + subord + agree)* Erik takes skiing lessons from Rick. Erik has been training alone since Rick has a sprained ankle. He's expected to be off his crutches soon.

    (h) *(nonlocal + subord + agree)* Erik takes skiing lessons from Rick. Since Rick has a sprained ankle, Erik has been training alone. He's expected to be off his crutches soon.

7. *(because / female target)*

    (a) *(local + main + disagree)* Mark and his older sister Lisa go to the same college. Mark saves a lot of money on school supplies. Lisa always has used textbooks. She is happy to be able to help his sister out.

    (b) *(nonlocal + main + disagree)* Mark and his older sister Lisa go to the same college. Lisa always has used textbooks, so Mark saves a lot of money on school supplies. She is happy to be able to help his sister out.

    (c) *(local + subord + disagree)* Mark and his older sister Lisa go to the same college. Mark saves a lot of money on school supplies because Lisa always has used textbooks. She is happy to be able to help his sister out.

    (d) *(nonlocal + subord + disagree)* Mark and his older sister Lisa go to the same college. Because Lisa always has used textbooks, Mark saves a lot of money on school supplies. She is happy to be able to help his sister out.

    (e) *(local + main + agree)* Betty and her older sister Lisa go to the same college. Betty saves a lot of money on school supplies. Lisa always has used textbooks. She is happy to be able to help her sister out.

    (f) *(nonlocal + main + agree)* Betty and her older sister Lisa go to the same college. Lisa always has used textbooks, so Betty saves a lot of money on school supplies. She is happy to be able to help her sister out.

    (g) *(local + subord + agree)* Betty and her older sister Lisa go to the same college. Betty saves a lot of money on school supplies because Lisa always has used textbooks. She is happy to be able to help her sister out.

    (h) *(nonlocal + subord + agree)* Betty and her older sister Lisa go to the same college. Because Lisa always has used textbooks, Betty saves a lot of money on

school supplies. She is happy to be able to help her sister out.

8. *(because / male target)*

   (a) *(local + main + disagree)* Maria and Jacob are volunteering at a park today. Maria picked up sports equipment and litter from the game fields. Jacob needed a clear path for his lawnmower. He has a fancy new riding mower that made quick work of the job.

   (b) *(nonlocal + main + disagree)* Maria and Jacob are volunteering at a park today. Jacob needed a clear path for his lawnmower, so Maria picked up sports equipment and litter from the game fields. He has a fancy new riding mower that made quick work of the job.

   (c) *(local + subord + disagree)* Maria and Jacob are volunteering at a park today. Maria picked up sports equipment and litter from the game fields because Jacob needed a clear path for his lawnmower. He has a fancy new riding mower that made quick work of the job.

   (d) *(nonlocal + subord + disagree)* Maria and Jacob are volunteering at a park today. Because Jacob needed a clear path for his lawnmower, Maria picked up sports equipment and litter from the game fields. He has a fancy new riding mower that made quick work of the job.

   (e) *(local + main + agree)* Donald and Jacob are volunteering at a park today. Donald picked up sports equipment and litter from the game fields. Jacob needed a clear path for his lawnmower. He has a fancy new riding mower that made quick work of the job.

   (f) *(nonlocal + main + agree)* Donald and Jacob are volunteering at a park today. Jacob needed a clear path for his lawnmower, so Donald picked up sports equipment and litter from the game fields. He has a fancy new riding mower that made quick work of the job.

   (g) *(local + subord + agree)* Donald and Jacob are volunteering at a park today. Donald picked up sports equipment and litter from the game fields because Jacob needed a clear path for his lawnmower. He has a fancy new riding mower that made quick work of the job.

   (h) *(nonlocal + subord + agree)* Donald and Jacob are volunteering at a park today. Because Jacob needed a clear path for his lawnmower, Donald picked up sports equipment and litter from the game fields. He has a fancy new riding mower that made quick work of the job.

9. *(because / female target)*

   (a) *(local + main + disagree)* Kevin babysat Julia on Saturday afternoon. Kevin brought along a soccer ball and net. Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

   (b) *(nonlocal + main + disagree)* Kevin babysat Julia on Saturday afternoon. Julia always wants to play outdoor sports, so Kevin brought along a soccer ball and net. She doesn't get enough exercise during her week in kindergarten.

   (c) *(local + subord + disagree)* Kevin babysat Julia on Saturday afternoon. Kevin brought along a soccer ball and net because Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

   (d) *(nonlocal + subord + disagree)* Kevin babysat Julia on Saturday afternoon. Because Julia always wants to play outdoor sports, Kevin brought along a soccer

171

ball and net. She doesn't get enough exercise during her week in kindergarten.

(e) *(local + main + agree)* Kim babysat Julia on Saturday afternoon. Kim brought along a soccer ball and net. Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

(f) *(nonlocal + main + agree)* Kim babysat Julia on Saturday afternoon. Julia always wants to play outdoor sports, so Kim brought along a soccer ball and net. She doesn't get enough exercise during her week in kindergarten.

(g) *(local + subord + agree)* Kim babysat Julia on Saturday afternoon. Kim brought along a soccer ball and net because Julia always wants to play outdoor sports. She doesn't get enough exercise during her week in kindergarten.

(h) *(nonlocal + subord + agree)* Kim babysat Julia on Saturday afternoon. Because Julia always wants to play outdoor sports, Kim brought along a soccer ball and net. She doesn't get enough exercise during her week in kindergarten.

10. *(while / male target)*

(a) *(local + main + disagree)* Carol and Steve left their offices and met for lunch. While Carol had brought leftovers from home, Steve had to buy lunch. He never brings his lunch to work.

(b) *(nonlocal + main + disagree)* Carol and Steve left their offices and met for lunch. Steve had to buy lunch, while Carol had brought leftovers from home. He never brings his lunch to work.

(c) *(local + subord + disagree)* Carol and Steve left their offices and met for lunch. Carol had brought leftovers from home, while Steve had to buy lunch. He never brings his lunch to work.

(d) *(nonlocal + subord + disagree)* Carol and Steve left their offices and met for lunch. While Steve had to buy lunch, Carol had brought leftovers from home. He never brings his lunch to work.

(e) *(local + main + agree)* Adam and Steve left their offices and met for lunch. While Adam had brought leftovers from home, Steve had to buy lunch. He never brings his lunch to work.

(f) *(nonlocal + main + agree)* Adam and Steve left their offices and met for lunch. Steve had to buy lunch, while Adam had brought leftovers from home. He never brings his lunch to work.

(g) *(local + subord + agree)* Adam and Steve left their offices and met for lunch. Adam had brought leftovers from home, while Steve had to buy lunch. He never brings his lunch to work.

(h) *(nonlocal + subord + agree)* Adam and Steve left their offices and met for lunch. While Steve had to buy lunch, Adam had brought leftovers from home. He never brings his lunch to work.

11. *(while / female target)*

(a) *(local + main + disagree)* Roland and Nancy tidied up their diving club's reading room. While Roland cleaned off the tables and mopped the floor, Nancy sorted and shelved the books by topic. She enjoyed the job of organizing and categorizing the books.

(b) *(nonlocal + main + disagree)* Roland and Nancy tidied up their diving club's reading room. Nancy sorted and shelved the books by topic, while Roland

cleaned off the tables and mopped the floor. She enjoyed the job of organizing and categorizing the books.

(c) *(local + subord + disagree)* Roland and Nancy tidied up their diving club's reading room. Roland cleaned off the tables and mopped the floor, while Nancy sorted and shelved the books by topic. She enjoyed the job of organizing and categorizing the books.

(d) *(nonlocal + subord + disagree)* Roland and Nancy tidied up their diving club's reading room. While Nancy sorted and shelved the books by topic, Roland cleaned off the tables and mopped the floor. She enjoyed the job of organizing and categorizing the books.

(e) *(local + main + agree)* Elizabeth and Nancy tidied up their diving club's reading room. While Elizabeth cleaned off the tables and mopped the floor, Nancy sorted and shelved the books by topic. She enjoyed the job of organizing and categorizing the books.

(f) *(nonlocal + main + agree)* Elizabeth and Nancy tidied up their diving club's reading room. Nancy sorted and shelved the books by topic, while Elizabeth cleaned off the tables and mopped the floor. She enjoyed the job of organizing and categorizing the books.

(g) *(local + subord + agree)* Elizabeth and Nancy tidied up their diving club's reading room. Elizabeth cleaned off the tables and mopped the floor, while Nancy sorted and shelved the books by topic. She enjoyed the job of organizing and categorizing the books.

(h) *(nonlocal + subord + agree)* Elizabeth and Nancy tidied up their diving club's reading room. While Nancy sorted and shelved the books by topic, Elizabeth cleaned off the tables and mopped the floor. She enjoyed the job of organizing and categorizing the books.

12. *(because / male target)*

(a) *(local + main + disagree)* Liza and Daniel were assigned to be partners for a school project. Liza is doing almost all the work. Daniel messes around during lab sessions. He just talks on his cell phone instead.

(b) *(nonlocal + main + disagree)* Liza and Daniel were assigned to be partners for a school project. Daniel messes around during lab sessions, so Liza is doing almost all the work. He just talks on his cell phone instead.

(c) *(local + subord + disagree)* Liza and Daniel were assigned to be partners for a school project. Liza is doing almost all the work because Daniel messes around during lab sessions. He just talks on his cell phone instead.

(d) *(nonlocal + subord + disagree)* Liza and Daniel were assigned to be partners for a school project. Because Daniel messes around during lab sessions, Liza is doing almost all the work. He just talks on his cell phone instead.

(e) *(local + main + agree)* Edward and Daniel were assigned to be partners for a school project. Edward is doing almost all the work. Daniel messes around during lab sessions. He just talks on his cell phone instead.

(f) *(nonlocal + main + agree)* Edward and Daniel were assigned to be partners for a school project. Daniel messes around during lab sessions, so Edward is doing almost all the work. He just talks on his cell phone instead.

(g) *(local + subord + agree)* Edward and Daniel were assigned to be partners for

173

a school project. Edward is doing almost all the work because Daniel messes around during lab sessions. He just talks on his cell phone instead.

(h) *(nonlocal + subord + agree)* Edward and Daniel were assigned to be partners for a school project. Because Daniel messes around during lab sessions, Edward is doing almost all the work. He just talks on his cell phone instead.

13. *(while / female target)*

(a) *(local + main + disagree)* Brandon and Hannah have been friends since they were children. While Brandon has lived on the same street his whole life, Hannah went away to college and never moved back. She comes back to visit every summer.

(b) *(nonlocal + main + disagree)* Brandon and Hannah have been friends since they were children. Hannah went away to college and never moved back, while Brandon has lived on the same street his whole life. She comes back to visit every summer.

(c) *(local + subord + disagree)* Brandon and Hannah have been friends since they were children. Brandon has lived on the same street his whole life, while Hannah went away to college and never moved back. She comes back to visit every summer.

(d) *(nonlocal + subord + disagree)* Brandon and Hannah have been friends since they were children. While Hannah went away to college and never moved back, Brandon has lived on the same street his whole life. She comes back to visit every summer.

(e) *(local + main + agree)* Ashley and Hannah have been friends since they were children. While Ashley has lived on the same street her whole life, Hannah went away to college and never moved back. She comes back to visit every summer.

(f) *(nonlocal + main + agree)* Ashley and Hannah have been friends since they were children. Hannah went away to college and never moved back, while Ashley has lived on the same street her whole life. She comes back to visit every summer.

(g) *(local + subord + agree)* Ashley and Hannah have been friends since they were children. Ashley has lived on the same street her whole life, while Hannah went away to college and never moved back. She comes back to visit every summer.

(h) *(nonlocal + subord + agree)* Ashley and Hannah have been friends since they were children. While Hannah went away to college and never moved back, Ashley has lived on the same street her whole life. She comes back to visit every summer.

14. *(because / male target)*

(a) *(local + main + disagree)* Jessica was going to William's birthday party. Jessica couldn't find a good gift to bring. William is difficult to shop for. He prefers to have very few material possessions.

(b) *(nonlocal + main + disagree)* Jessica was going to William's birthday party. William is difficult to shop for, so Jessica couldn't find a good gift to bring. He prefers to have very few material possessions.

(c) *(local + subord + disagree)* Jessica was going to William's birthday party. Jessica couldn't find a good gift to bring because William is difficult to shop for. He prefers to have very few material possessions.

(d) *(nonlocal + subord + disagree)* Jessica was going to William's birthday party. Because William is difficult to shop for, Jessica couldn't find a good gift to bring. He prefers to have very few material possessions.

(e) *(local + main + agree)* Zack was going to William's birthday party. Zack couldn't find a good gift to bring. William is difficult to shop for. He prefers to have very few material possessions.

(f) *(nonlocal + main + agree)* Zack was going to William's birthday party. William is difficult to shop for, so Zack couldn't find a good gift to bring. He prefers to have very few material possessions.

(g) *(local + subord + agree)* Zack was going to William's birthday party. Zack couldn't find a good gift to bring because William is difficult to shop for. He prefers to have very few material possessions.

(h) *(nonlocal + subord + agree)* Zack was going to William's birthday party. Because William is difficult to shop for, Zack couldn't find a good gift to bring. He prefers to have very few material possessions.

15. *(because / female target)*

(a) *(local + main + disagree)* Anthony manages a softball club, and Lauren is their pitcher. Anthony is looking for a second pitcher. Lauren is having elbow problems. She's been pitching every game, and has developed a repetitive stress injury.

(b) *(nonlocal + main + disagree)* Anthony manages a softball club, and Lauren is their pitcher. Lauren is having elbow problems, so Anthony is looking for a second pitcher. She's been pitching every game, and has developed a repetitive stress injury.

(c) *(local + subord + disagree)* Anthony manages a softball club, and Lauren is their pitcher. Anthony is looking for a second pitcher because Lauren is having elbow problems. She's been pitching every game, and has developed a repetitive stress injury.

(d) *(nonlocal + subord + disagree)* Anthony manages a softball club, and Lauren is their pitcher. Because Lauren is having elbow problems, Anthony is looking for a second pitcher. She's been pitching every game, and has developed a repetitive stress injury.

(e) *(local + main + agree)* Megan manages a softball club, and Lauren is their pitcher. Megan is looking for a second pitcher. Lauren is having elbow problems. She's been pitching every game, and has developed a repetitive stress injury.

(f) *(nonlocal + main + agree)* Megan manages a softball club, and Lauren is their pitcher. Lauren is having elbow problems, so Megan is looking for a second pitcher. She's been pitching every game, and has developed a repetitive stress injury.

(g) *(local + subord + agree)* Megan manages a softball club, and Lauren is their pitcher. Megan is looking for a second pitcher because Lauren is having elbow problems. She's been pitching every game, and has developed a repetitive stress injury.

(h) *(nonlocal + subord + agree)* Megan manages a softball club, and Lauren is their pitcher. Because Lauren is having elbow problems, Megan is looking for a second pitcher. She's been pitching every game, and has developed a repetitive stress

injury.

16. *(while / male target)*

    (a) *(local + main + disagree)* Emma and James share a computer. While Emma only really knows how to use email and the web, James uses it heavily for work. He has a job as a software engineer.

    (b) *(nonlocal + main + disagree)* Emma and James share a computer. James uses it heavily for work, while Emma only really knows how to use email and the web. He has a job as a software engineer.

    (c) *(local + subord + disagree)* Emma and James share a computer. Emma only really knows how to use email and the web, while James uses it heavily for work. He has a job as a software engineer.

    (d) *(nonlocal + subord + disagree)* Emma and James share a computer. While James uses it heavily for work, Emma only really knows how to use email and the web. He has a job as a software engineer.

    (e) *(local + main + agree)* Justin and James share a computer. While Justin only really knows how to use email and the web, James uses it heavily for work. He has a job as a software engineer.

    (f) *(nonlocal + main + agree)* Justin and James share a computer. James uses it heavily for work, while Justin only really knows how to use email and the web. He has a job as a software engineer.

    (g) *(local + subord + agree)* Justin and James share a computer. Justin only really knows how to use email and the web, while James uses it heavily for work. He has a job as a software engineer.

    (h) *(nonlocal + subord + agree)* Justin and James share a computer. While James uses it heavily for work, Justin only really knows how to use email and the web. He has a job as a software engineer.

A.3.2 Filler items

(same as Experiment 1)

A.4 Experiment 4 materials

1. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

    (a) *(consistent)* Mary is calling her mother. She needs to finish her homework. Next, she will book a plane ticket home.

    (b) *(inconsistent)* Mary is calling her mother. She had been finishing her homework. Next, she will book a plane ticket home.

2. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

    (a) *(consistent)* John is returning a library book. He plans to visit a new restaurant. Next, he will walk to the coffee shop.

    (b) *(inconsistent)* John is returning a library book. He had been visiting a new restaurant. Next, he will walk to the coffee shop.

3. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

(a) *(consistent)* Patricia is claiming her luggage. She intends to eat some fast food. Next, she will drive to her hotel.

(b) *(inconsistent)* Patricia is claiming her luggage. She had been eating some fast food. Next, she will drive to her hotel.

4. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

   (a) *(consistent)* Robert is filling his car's gas tank. He wants to have it washed and waxed. Next, he will go to pick up his date.

   (b) *(inconsistent)* Robert is filling his car's gas tank. He had been having it washed and waxed. Next, he will go to pick up his date.

5. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

   (a) *(consistent)* Linda is clearing trash and debris from the sidewalk. She is meaning to sweep her front porch. Later, she wants to cut flowers for the table.

   (b) *(inconsistent)* Linda is clearing trash and debris from the sidewalk. She has swept her front porch. Later, she wants to cut flowers for the table.

6. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

   (a) *(consistent)* Michael is writing a letter to his brother. He is planning to read a few chapters of a novel. Later, he intends to go to bed early.

   (b) *(inconsistent)* Michael is writing a letter to his brother. He has read a few chapters of a novel. Later, he intends to go to bed early.

7. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

   (a) *(consistent)* Barbara is waiting for a bus. She is planning to try out in a Broadway audition. Later, she has to hurry to get to work on time.

   (b) *(inconsistent)* Barbara is waiting for a bus. She has tried out in a Broadway audition. Later, she has to hurry to get to work on time.

8. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

   (a) *(consistent)* William is beating the dust out of his rugs. He is intending to sort through old magazines to throw out. Later, he needs to change the batteries in the smoke detectors.

   (b) *(inconsistent)* William is beating the dust out of his rugs. He has sorted through old magazines to throw out. Later, he needs to change the batteries in the smoke detectors.

9. *(later : past + [past / past-pret] + past)*

   (a) *(consistent)* Elizabeth rode the bus home. She finished that day's newpaper. Later, she went for a quick run.

   (b) *(inconsistent)* Elizabeth rode the bus home. She had finished that day's newpaper. Later, she went for a quick run.

10. *(later : past + [past / past-pret] + past)*

    (a) *(consistent)* David stacked his moving boxes in the corner. He located some basic kitchen equipment. Later, he hung a couple pictures on the wall.

    (b) *(inconsistent)* David stacked his moving boxes in the corner. He had located some basic kitchen equipment. Later, he hung a couple pictures on the wall.

11. *(later : past + [past / past-pret] + past)*

177

(a) *(consistent)* Jennifer made copies of some legal documents. She took the originals to be notarized. Later, she filed the papers with the county clerk.

(b) *(inconsistent)* Jennifer made copies of some legal documents. She had taken the originals to be notarized. Later, she filed the papers with the county clerk.

12. *(later : past + [past / past-pret] + past)*

(a) *(consistent)* Richard solved the crossword puzzle in the paper. He bought coffee and a bagel. Later, he walked to his subway stop.

(b) *(inconsistent)* Richard solved the crossword puzzle in the paper. He had bought coffee and a bagel. Later, he walked to his subway stop.

13. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Maria was removing a door's hinges and doorknob. She covered its panes of glass with tape. Next, she applied a fresh coat of paint to the door.

(b) *(inconsistent)* Maria was removing a door's hinges and doorknob. She had covered its panes of glass with tape. Next, she applied a fresh coat of paint to the door.

14. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Charles was lifting weights at the gym. He went for a three-mile run. Next, he performed some cool-down exercises.

(b) *(inconsistent)* Charles was lifting weights at the gym. He had gone for a three-mile run. Next, he performed some cool-down exercises.

15. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Susan was packing food and water. She plotted out the full length of the trail on her maps. Next, she jogged to the trail head.

(b) *(inconsistent)* Susan was packing food and water. She had plotted out the full length of the trail on her maps. Next, she jogged to the trail head.

16. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Joseph was setting up the pieces on the chessboard. He shook his opponent's hand. Next, he punched the clock to start the game.

(b) *(inconsistent)* Joseph was setting up the pieces on the chessboard. He had shaken his opponent's hand. Next, he punched the clock to start the game.

17. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

(a) *(consistent)* Margaret is pumping water out of her basement. She needs to carry papers and other fragile goods upstairs. Next, she intends to call a plumber to fix the leak.

(b) *(inconsistent)* Margaret is pumping water out of her basement. She has been carrying papers and other fragile goods upstairs. Next, she intends to call a plumber to fix the leak.

18. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

(a) *(consistent)* Thomas is purchasing lumber to replace his house's front steps. He needs to acquire the necessary hardware and fasteners. Next, he intends to drive his lumber to his workshop.

(b) *(inconsistent)* Thomas is purchasing lumber to replace his house's front steps. He has been acquiring the necessary hardware and fasteners. Next, he intends to drive his lumber to his workshop.

19. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

    (a) *(consistent)* Dorothy is putting on her shoes and jacket. She needs to check the contents of her briefcase. Next, she plans to leave to meet with a client.

    (b) *(inconsistent)* Dorothy is putting on her shoes and jacket. She has been checking the contents of her briefcase. Next, she plans to leave to meet with a client.

20. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

    (a) *(consistent)* Christopher is greeting the opposing team's captain. He wants to give his team a pep talk. Next, he intends to signal the referee that the team is ready.

    (b) *(inconsistent)* Christopher is greeting the opposing team's captain. He has been giving his team a pep talk. Next, he intends to signal the referee that the team is ready.

21. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

    (a) *(consistent)* Lisa has been skimming floating leaves out of the pool. She needs to adjust the level of the chemicals in the water. Later, she will swim some practice laps.

    (b) *(inconsistent)* Lisa has been skimming floating leaves out of the pool. She has adjusted the level of the chemicals in the water. Later, she will swim some practice laps.

22. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

    (a) *(consistent)* Daniel has been cleaning all his hand tools. He wants to discard some unwanted scraps of lumber. Later, he will sweep up the debris that's collected in the corners.

    (b) *(inconsistent)* Daniel has been cleaning all his hand tools. He has discarded some unwanted scraps of lumber. Later, he will sweep up the debris that's collected in the corners.

23. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

    (a) *(consistent)* Nancy has been cutting her picture's frame to the right lengths. She plans to trim the mat and backing to fit it. Later, she will fasten the frame and glass around the layers.

    (b) *(inconsistent)* Nancy has been cutting her picture's frame to the right lengths. She has trimmed the mat and backing to fit it. Later, she will fasten the frame and glass around the layers.

24. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

    (a) *(consistent)* Paul has been replying to a few emails. He intends to fill out his travel expense reports. Later, he will check in with his business partners.

    (b) *(inconsistent)* Paul has been replying to a few emails. He has filled out his travel expense reports. Later, he will check in with his business partners.

A.5 Experiment 5 materials

1. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

    (a) *(consistent)* Mary is calling her mother. She needs to finish her homework. Next, she will book a plane ticket home.

    (b) *(inconsistent)* Mary is calling her mother. She had been finishing her homework. Next, she will book a plane ticket home.

2. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

    (a) *(consistent)* John is returning a library book. He plans to visit a new restaurant. Next, he will walk to the coffee shop.

    (b) *(inconsistent)* John is returning a library book. He had been visiting a new restaurant. Next, he will walk to the coffee shop.

3. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

    (a) *(consistent)* Patricia is claiming her luggage. She intends to eat some fast food. Next, she will drive to her hotel.

    (b) *(inconsistent)* Patricia is claiming her luggage. She had been eating some fast food. Next, she will drive to her hotel.

4. *(next : pres-prog + [modal-pres / past-pret-prog] + fut)*

    (a) *(consistent)* Robert is filling his car's gas tank. He wants to have it washed and waxed. Next, he will go to pick up his date.

    (b) *(inconsistent)* Robert is filling his car's gas tank. He had been having it washed and waxed. Next, he will go to pick up his date.

5. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

    (a) *(consistent)* Linda is clearing trash and debris from the sidewalk. She is meaning to sweep her front porch. Later, she wants to cut flowers for the table.

    (b) *(inconsistent)* Linda is clearing trash and debris from the sidewalk. She has swept her front porch. Later, she wants to cut flowers for the table.

6. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

    (a) *(consistent)* Michael is writing a letter to his brother. He is planning to read a few chapters of a novel. Later, he intends to go to bed early.

    (b) *(inconsistent)* Michael is writing a letter to his brother. He has read a few chapters of a novel. Later, he intends to go to bed early.

7. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

    (a) *(consistent)* Barbara is waiting for a bus. She is planning to try out in a Broadway audition. Later, she has to hurry to get to work on time.

    (b) *(inconsistent)* Barbara is waiting for a bus. She has tried out in a Broadway audition. Later, she has to hurry to get to work on time.

8. *(later : pres-prog + [pres-modal-prog / pres-pret] + modal-pres)*

    (a) *(consistent)* William is beating the dust out of his rugs. He is intending to sort through old magazines to throw out. Later, he needs to change the batteries in the smoke detectors.

(b) *(inconsistent)* William is beating the dust out of his rugs. He has sorted through old magazines to throw out. Later, he needs to change the batteries in the smoke detectors.

9. *(later : past + [past / past-pret] + past)*

(a) *(consistent)* Elizabeth rode the bus home. She finished that day's newpaper. Later, she went for a quick run.

(b) *(inconsistent)* Elizabeth rode the bus home. She had finished that day's newpaper. Later, she went for a quick run.

10. *(later : past + [past / past-pret] + past)*

(a) *(consistent)* David stacked his moving boxes in the corner. He located some basic kitchen equipment. Later, he hung a couple pictures on the wall.

(b) *(inconsistent)* David stacked his moving boxes in the corner. He had located some basic kitchen equipment. Later, he hung a couple pictures on the wall.

11. *(later : past + [past / past-pret] + past)*

(a) *(consistent)* Jennifer made copies of some legal documents. She took the originals to be notarized. Later, she filed the papers with the county clerk.

(b) *(inconsistent)* Jennifer made copies of some legal documents. She had taken the originals to be notarized. Later, she filed the papers with the county clerk.

12. *(later : past + [past / past-pret] + past)*

(a) *(consistent)* Richard solved the crossword puzzle in the paper. He bought coffee and a bagel. Later, he walked to his subway stop.

(b) *(inconsistent)* Richard solved the crossword puzzle in the paper. He had bought coffee and a bagel. Later, he walked to his subway stop.

13. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Maria was removing a door's hinges and doorknob. She covered its panes of glass with tape. Next, she applied a fresh coat of paint to the door.

(b) *(inconsistent)* Maria was removing a door's hinges and doorknob. She had covered its panes of glass with tape. Next, she applied a fresh coat of paint to the door.

14. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Charles was lifting weights at the gym. He went for a three-mile run. Next, he performed some cool-down exercises.

(b) *(inconsistent)* Charles was lifting weights at the gym. He had gone for a three-mile run. Next, he performed some cool-down exercises.

15. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Susan was packing food and water. She plotted out the full length of the trail on her maps. Next, she jogged to the trail head.

(b) *(inconsistent)* Susan was packing food and water. She had plotted out the full length of the trail on her maps. Next, she jogged to the trail head.

16. *(next : past-prog + [past / past-pret] + past)*

(a) *(consistent)* Joseph was setting up the pieces on the chessboard. He shook his opponent's hand. Next, he punched the clock to start the game.

(b) *(inconsistent)* Joseph was setting up the pieces on the chessboard. He had shaken his opponent's hand. Next, he punched the clock to start the game.

17. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

   (a) *(consistent)* Margaret is pumping water out of her basement. She needs to carry papers and other fragile goods upstairs. Next, she intends to call a plumber to fix the leak.

   (b) *(inconsistent)* Margaret is pumping water out of her basement. She has been carrying papers and other fragile goods upstairs. Next, she intends to call a plumber to fix the leak.

18. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

   (a) *(consistent)* Thomas is purchasing lumber to replace his house's front steps. He needs to acquire the necessary hardware and fasteners. Next, he intends to drive his lumber to his workshop.

   (b) *(inconsistent)* Thomas is purchasing lumber to replace his house's front steps. He has been acquiring the necessary hardware and fasteners. Next, he intends to drive his lumber to his workshop.

19. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

   (a) *(consistent)* Dorothy is putting on her shoes and jacket. She needs to check the contents of her briefcase. Next, she plans to leave to meet with a client.

   (b) *(inconsistent)* Dorothy is putting on her shoes and jacket. She has been checking the contents of her briefcase. Next, she plans to leave to meet with a client.

20. *(next : pres-prog + [modal-pres / pres-pret-prog] + modal-pres)*

   (a) *(consistent)* Christopher is greeting the opposing team's captain. He wants to give his team a pep talk. Next, he intends to signal the referee that the team is ready.

   (b) *(inconsistent)* Christopher is greeting the opposing team's captain. He has been giving his team a pep talk. Next, he intends to signal the referee that the team is ready.

21. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

   (a) *(consistent)* Lisa has been skimming floating leaves out of the pool. She needs to adjust the level of the chemicals in the water. Later, she will swim some practice laps.

   (b) *(inconsistent)* Lisa has been skimming floating leaves out of the pool. She has adjusted the level of the chemicals in the water. Later, she will swim some practice laps.

22. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

   (a) *(consistent)* Daniel has been cleaning all his hand tools. He wants to discard some unwanted scraps of lumber. Later, he will sweep up the debris that's collected in the corners.

   (b) *(inconsistent)* Daniel has been cleaning all his hand tools. He has discarded some unwanted scraps of lumber. Later, he will sweep up the debris that's collected in the corners.

23. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

(a) *(consistent)* Nancy has been cutting her picture's frame to the right lengths. She plans to trim the mat and backing to fit it. Later, she will fasten the frame and glass around the layers.

(b) *(inconsistent)* Nancy has been cutting her picture's frame to the right lengths. She has trimmed the mat and backing to fit it. Later, she will fasten the frame and glass around the layers.

24. *(later : pres-pret-prog + [modal-pres / pres-pret] + fut)*

(a) *(consistent)* Paul has been replying to a few emails. He intends to fill out his travel expense reports. Later, he will check in with his business partners.

(b) *(inconsistent)* Paul has been replying to a few emails. He has filled out his travel expense reports. Later, he will check in with his business partners.

A.6 Experiment 6 materials

A.6.1 Experimental items

1. *(consequently / though)*

(a) *(local + noncue)* After her soccer game, Johanna went straight to a restaurant. She'd had a solid breakfast, though she hadn't eaten for hours since then. She was ravenously hungry now.

(b) *(nonlocal + noncue)* After her soccer game, Johanna went straight to a restaurant. Though she hadn't eaten for hours since then, she'd had a solid breakfast. She was ravenously hungry now.

(c) *(local + hascue)* After her soccer game, Johanna went straight to a restaurant. She'd had a solid breakfast, though she hadn't eaten for hours since then. Consequently, she was ravenously hungry now.

(d) *(nonlocal + hascue)* After her soccer game, Johanna went straight to a restaurant. Though she hadn't eaten for hours since then, she'd had a solid breakfast. Consequently, she was ravenously hungry now.

2. *(as a result / though)*

(a) *(local + noncue)* This afternoon while Tom was in class there was an unexpected blizzard. He usually drives on snowy days like this, though he biked to school today. His only way home is to take the bus.

(b) *(nonlocal + noncue)* This afternoon while Tom was in class there was an unexpected blizzard. Though he biked to school today, he usually drives on snowy days like this. His only way home is to take the bus.

(c) *(local + hascue)* This afternoon while Tom was in class there was an unexpected blizzard. He usually drives on snowy days like this, though he biked to school today. As a result, his only way home is to take the bus.

(d) *(nonlocal + hascue)* This afternoon while Tom was in class there was an unexpected blizzard. Though he biked to school today, he usually drives on snowy days like this. As a result, his only way home is to take the bus.

3. *(consequently / while)*

(a) *(local + noncue)* In January Laura went to a conference in Australia, where it was summertime. Her colleagues were in shorts, while she had only packed heavy clothing. She was uncomfortably warm for most of the week.

(b) *(nonlocal + noncue)* In January Laura went to a conference in Australia, where it was summertime. While she had only packed heavy clothing, her colleagues were in shorts. She was uncomfortably warm for most of the week.

(c) *(local + hascue)* In January Laura went to a conference in Australia, where it was summertime. Her colleagues were in shorts, while she had only packed heavy clothing. Consequently, she was uncomfortably warm for most of the week.

(d) *(nonlocal + hascue)* In January Laura went to a conference in Australia, where it was summertime. While she had only packed heavy clothing, her colleagues were in shorts. Consequently, she was uncomfortably warm for most of the week.

4. *(consequently / though)*

(a) *(local + noncue)* Before a dinner party, Beth forgot to answer an email asking about food preferences. She strongly dislikes corn, though she tries hard to be a polite guest. She ate her corn chowder without saying anything.

(b) *(nonlocal + noncue)* Before a dinner party, Beth forgot to answer an email asking about food preferences. Though she tries hard to be a polite guest, she strongly dislikes corn. She ate her corn chowder without saying anything.

(c) *(local + hascue)* Before a dinner party, Beth forgot to answer an email asking about food preferences. She strongly dislikes corn, though she tries hard to be a polite guest. Consequently, she ate her corn chowder without saying anything.

(d) *(nonlocal + hascue)* Before a dinner party, Beth forgot to answer an email asking about food preferences. Though she tries hard to be a polite guest, she strongly dislikes corn. Consequently, she ate her corn chowder without saying anything.

5. *(as a result / while)*

(a) *(local + noncue)* Kate borrowed a calculator for her physics exam. She understood the physics ideas, while the calculator was unfamiliar and confusing. She wasn't able to finish all the problems.

(b) *(nonlocal + noncue)* Kate borrowed a calculator for her physics exam. While the calculator was unfamiliar and confusing, she understood the physics ideas. She wasn't able to finish all the problems.

(c) *(local + hascue)* Kate borrowed a calculator for her physics exam. She understood the physics ideas, while the calculator was unfamiliar and confusing. As a result, she wasn't able to finish all the problems.

(d) *(nonlocal + hascue)* Kate borrowed a calculator for her physics exam. While the calculator was unfamiliar and confusing, she understood the physics ideas. As a result, she wasn't able to finish all the problems.

6. *(consequently / though)*

(a) *(local + noncue)* Bob visited his cousin in New York City over the holidays. He'd been driving packages all around Pennsylvania for years, though he'd never driven his delivery van into the city before. The city traffic made him very nervous.

(b) *(nonlocal + noncue)* Bob visited his cousin in New York City over the holidays. Though he'd never driven his delivery van into the city before, he'd been driving packages all around Pennsylvania for years. The city traffic made him very nervous.

(c) *(local + hascue)* Bob visited his cousin in New York City over the holidays. He'd been driving packages all around Pennsylvania for years, though he'd never driven his delivery van into the city before. Consequently, the city traffic made him very nervous.

(d) *(nonlocal + hascue)* Bob visited his cousin in New York City over the holidays. Though he'd never driven his delivery van into the city before, he'd been driving packages all around Pennsylvania for years. Consequently, the city traffic made him very nervous.

7. *(as a result / though)*

(a) *(local + noncue)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. He did get interviews with several of the executive's associates, though he wasn't able to do the prison interview. His article didn't make the front page of the paper.

(b) *(nonlocal + noncue)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. Though he wasn't able to do the prison interview, he did get interviews with several of the executive's associates. His article didn't make the front page of the paper.

(c) *(local + hascue)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. He did get interviews with several of the executive's associates, though he wasn't able to do the prison interview. As a result, his article didn't make the front page of the paper.

(d) *(nonlocal + hascue)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. Though he wasn't able to do the prison interview, he did get interviews with several of the executive's associates. As a result, his article didn't make the front page of the paper.

8. *(consequently / while)*

(a) *(local + noncue)* Rebecca wasn't sure she was going to pass her physics midterm. She fell behind in her other classes, while she studied for the test every night for a week. She did well on the midterm.

(b) *(nonlocal + noncue)* Rebecca wasn't sure she was going to pass her physics midterm. While she studied for the test every night for a week, she fell behind in her other classes. She did well on the midterm.

(c) *(local + hascue)* Rebecca wasn't sure she was going to pass her physics midterm. She fell behind in her other classes, while she studied for the test every night for a week. Consequently, she did well on the midterm.

(d) *(nonlocal + hascue)* Rebecca wasn't sure she was going to pass her physics midterm. While she studied for the test every night for a week, she fell behind in her other classes. Consequently, she did well on the midterm.

9. *(as a result / though)*

(a) *(local + noncue)* Before Daniel moved into his current apartment he'd never lived alone. He did do his laundry regularly, though he never swept or tidied up

185

his trash. His place was always a complete mess.

(b) *(nonlocal + noncue)* Before Daniel moved into his current apartment he'd never lived alone. Though he never swept or tidied up his trash, he did do his laundry regularly. His place was always a complete mess.

(c) *(local + hascue)* Before Daniel moved into his current apartment he'd never lived alone. He did do his laundry regularly, though he never swept or tidied up his trash. As a result, his place was always a complete mess.

(d) *(nonlocal + hascue)* Before Daniel moved into his current apartment he'd never lived alone. Though he never swept or tidied up his trash, he did do his laundry regularly. As a result, his place was always a complete mess.

10. *(consequently / though)*

(a) *(local + noncue)* The accounting department that Serena works in is always short on cash. She takes very good care of her equipment, though her current workstation is old and obsolete. Her computer couldn't run the new accounting software she needed.

(b) *(nonlocal + noncue)* The accounting department that Serena works in is always short on cash. Though her current workstation is old and obsolete, she takes very good care of her equipment. Her computer couldn't run the new accounting software she needed.

(c) *(local + hascue)* The accounting department that Serena works in is always short on cash. She takes very good care of her equipment, though her current workstation is old and obsolete. Consequently, her computer couldn't run the new accounting software she needed.

(d) *(nonlocal + hascue)* The accounting department that Serena works in is always short on cash. Though her current workstation is old and obsolete, she takes very good care of her equipment. Consequently, her computer couldn't run the new accounting software she needed.

11. *(consequently / though)*

(a) *(local + noncue)* Andrew misplaces the the auto maintenance reminders he gets in the mail. He always had visiting his mechanic on his to-do list, though he hadn't changed oil in his car for almost a year. His car broke down on a hot summer afternoon.

(b) *(nonlocal + noncue)* Andrew misplaces the the auto maintenance reminders he gets in the mail. Though he hadn't changed oil in his car for almost a year, he always had visiting his mechanic on his to-do list. His car broke down on a hot summer afternoon.

(c) *(local + hascue)* Andrew misplaces the the auto maintenance reminders he gets in the mail. He always had visiting his mechanic on his to-do list, though he hadn't changed oil in his car for almost a year. Consequently, his car broke down on a hot summer afternoon.

(d) *(nonlocal + hascue)* Andrew misplaces the the auto maintenance reminders he gets in the mail. Though he hadn't changed oil in his car for almost a year, he always had visiting his mechanic on his to-do list. Consequently, his car broke down on a hot summer afternoon.

12. *(as a result / while)*

(a) *(local + noncue)* Robert and Karen's family all live far away. Karen wants fly around visiting far away people and places, while Robert is unwilling to get on an airplane. They stay home and rarely see their grandchildren.

(b) *(nonlocal + noncue)* Robert and Karen's family all live far away. While Robert is unwilling to get on an airplane, Karen wants fly around visiting far away people and places. They stay home and rarely see their grandchildren.

(c) *(local + hascue)* Robert and Karen's family all live far away. Karen wants fly around visiting far away people and places, while Robert is unwilling to get on an airplane. As a result, they stay home and rarely see their grandchildren.

(d) *(nonlocal + hascue)* Robert and Karen's family all live far away. While Robert is unwilling to get on an airplane, Karen wants fly around visiting far away people and places. As a result, they stay home and rarely see their grandchildren.

13. *(consequently / though)*

(a) *(local + noncue)* It was a sunny day, so Linda went to the beach. She was careful to drink plenty of water, though she completely forgot to re-apply her sunscreen. She ended up with a nasty sunburn.

(b) *(nonlocal + noncue)* It was a sunny day, so Linda went to the beach. Though she completely forgot to re-apply her sunscreen, she was careful to drink plenty of water. She ended up with a nasty sunburn.

(c) *(local + hascue)* It was a sunny day, so Linda went to the beach. She was careful to drink plenty of water, though she completely forgot to re-apply her sunscreen. Consequently, she ended up with a nasty sunburn.

(d) *(nonlocal + hascue)* It was a sunny day, so Linda went to the beach. Though she completely forgot to re-apply her sunscreen, she was careful to drink plenty of water. Consequently, she ended up with a nasty sunburn.

14. *(consequently / though)*

(a) *(local + noncue)* Rick was on a long overseas tour. Electrical adapters were on his packing list, though he forgot to actually put them in his suitcase. He wasn't able to charge any of his electronic gadgets.

(b) *(nonlocal + noncue)* Rick was on a long overseas tour. Though he forgot to actually put them in his suitcase, electrical adapters were on his packing list. He wasn't able to charge any of his electronic gadgets.

(c) *(local + hascue)* Rick was on a long overseas tour. Electrical adapters were on his packing list, though he forgot to actually put them in his suitcase. Consequently, he wasn't able to charge any of his electronic gadgets.

(d) *(nonlocal + hascue)* Rick was on a long overseas tour. Though he forgot to actually put them in his suitcase, electrical adapters were on his packing list. Consequently, he wasn't able to charge any of his electronic gadgets.

15. *(as a result / though)*

(a) *(local + noncue)* Betty's work day at her bakery starts hours before sunrise. The hard work sometimes leaves her exhausted and sleepy, though she can leave the shop to her employees by noon. She often spends sunny afternoons hiking and playing sports.

(b) *(nonlocal + noncue)* Betty's work day at her bakery starts hours before sunrise. Though she can leave the shop to her employees by noon, the hard work

187

sometimes leaves her exhausted and sleepy. She often spends sunny afternoons hiking and playing sports.

(c) *(local + hascue)* Betty's work day at her bakery starts hours before sunrise. The hard work sometimes leaves her exhausted and sleepy, though she can leave the shop to her employees by noon. As a result, she often spends sunny afternoons hiking and playing sports.

(d) *(nonlocal + hascue)* Betty's work day at her bakery starts hours before sunrise. Though she can leave the shop to her employees by noon, the hard work sometimes leaves her exhausted and sleepy. As a result, she often spends sunny afternoons hiking and playing sports.

16. *(consequently / though)*

(a) *(local + noncue)* Mark's hobby is repairing and reconstructing antique clocks. He and other clockmakers swap supplies by mail, though some replacement parts are impossible to find. He sometimes ends up making hard-to-locate parts himself in his machine shop.

(b) *(nonlocal + noncue)* Mark's hobby is repairing and reconstructing antique clocks. Though some replacement parts are impossible to find, he and other clockmakers swap supplies by mail. He sometimes ends up making hard-to-locate parts himself in his machine shop.

(c) *(local + hascue)* Mark's hobby is repairing and reconstructing antique clocks. He and other clockmakers swap supplies by mail, though some replacement parts are impossible to find. Consequently, he sometimes ends up making hard-to-locate parts himself in his machine shop.

(d) *(nonlocal + hascue)* Mark's hobby is repairing and reconstructing antique clocks. Though some replacement parts are impossible to find, he and other clockmakers swap supplies by mail. Consequently, he sometimes ends up making hard-to-locate parts himself in his machine shop.

A.6.2   Filler items

1. Hannah was at the sporting goods store shopping for a new set of skis. Because she had just gotten a raise, she was looking at skis that were much more expensive than her last pair. As a result, she bought some very pricey ultralight carbon fiber skis.

2. Isaac loaned his car to Trent last week. After Trent's axle was broken in an accident, his mechanic had told him the repair would take a few days. Isaac has been getting annoyed at how long the mechanic is taking.

3. Dorothy is making a warm quilt for her best friend. She wants to give it to her as a wedding present, so she needs to get it done by July. Later, she hopes to make a nice bag for storing it in the summer.

4. Monica is planning for a year abroad after college. She's especially looking forward to exploring Spain, since she has dozens of Spanish relatives. Consequently, she's planning to spend at least few months there.

5. Alan wants to be an expert poker player. Before spending money on lessons, he wants to brush up his skills by improving on his own. So, he's been reading a lot of poker books and playing games online.

6. Keith and Veronica have a daughter who's about to enter kindergarten. They're

nervous about sending her, because she's very small and shy. As a result, they're considering keeping her home for another year.

7. Bradley's construction business is taking on a new contract. Because the building is larger than usual, he needs to hire some new workers. So, he's been asking his employees to refer people they'd like to be working with.

8. Some of the equipment at Alisha's lab is broken. The supplier that sold her the equipment went out of business, so the warranty she bought is worthless. Consequently, she's buying replacements from a well-established supplier.

9. Garrett replaced the faucet on his kitchen sink. He didn't read the instructions very carefully, so he assembled it with some pieces backwards. Afterwards, the faucet leaked whenever he ran the cold water.

10. Melinda is trying to figure out how she should save money. After weeks of reading books, brochures and websites, she's still unsure how to get started. Until now, she's just kept her money in a checking account.

11. Henry owns a fancy little roadster that he loves to drive. After he moved into a house in the country, he discovered that in winter he needed a second car that could deal with the snow. Eventually, he bought an affordable truck with four-wheel-drive.

12. After dinner, Ian wanted some ice cream, so he headed to the corner store. When he got there, the store was closed. An electrical fault had shut down all their refrigeration.

13. Marvin needed some exercise, so he walked to the city pool. He wanted to swim laps, but the pool was totally full. He ended up going for a run instead.

14. Joanna moderates an online discussion group. She usually blocks all the spam and inappropriate messages, but sometimes there are too many to handle. Then, she gets a friend to help her.

15. Andy entered a boomerang throwing competition. Though he'd just started throwing boomerangs recently, he wanted to see really good throwers in action. He found lots of advice and interesting conversation there.

16. Sharon manages a small taxi company. While she mostly takes care of the office, she sometimes has to help with taxi dispatch when things get really busy. Dispatch makes her much more tense and unhappy than office work.

17. Ross used to work at a global scientific consulting firm. Though he enjoyed the high salary and the challenges, he got tired of not having any free time. Now, he's the research coordinator for a small local biotech company.

18. After her vacation, Deborah needed to catch up at work. She spent a quiet Sunday afternoon at the office, while no one else was around. The next day, she was back up to speed and ready to go.

19. When it snows, Carolyn likes to walk around the city late at night. Because few drivers venture out at such times, the city seems still and serene. Also, the snow muffles other sounds and makes it very silent.

20. A neighborhood bar was showing the baseball game Arthur wanted to watch. Since he wanted an unobstructed view of the screen, he made sure to arrive well before the game started. Later, the bar was packed, with sports fans in every seat.

21. A few months before his wedding, Walter hired a dancing coach. Though he was very clumsy at first, he worked hard at it for several hours every week. Consequently, he and his new wife enjoyed dancing the night away on their wedding night.

22. Nicola offers violin lessons out of her house in the afternoon. She used to be part of a touring symphony orchestra, but she got tired of the busy travel schedule. Now, she prefers to mostly stay at home.

23. Glenn had to write a research paper for high school biology class. After spending several hours doing online research, he had gathered a list of sources he thought would be useful. He headed to the city library with some money to pay for photocopies.

24. Natalie likes to buy a mocha at the coffee shop on the way to work. This month, though, she's making herself coffee at home, because she doesn't have enough spare cash. She had to pay a large auto repair bill when her car broke down.

25. After work today, Noah went rock climbing. Though he prefers to climb outside, the rock gym is much more convenient. Next weekend, he's taking a weekend trip to do some outdoor climbing.

26. Allison is hosting a large dinner party tonight. While she was in the middle of cooking, she discovered that she had run out of butter. She asked her best friend if she could show up early and bring more butter.

27. Dylan makes and sells high-quality hardwood furniture. Though much of his product line is traditional in style, the works that sell fastest are unconventional pieces that incorporate found objects. Nonetheless, many stores don't want to carry those items.

28. Danielle is an amateur beekeeper. She keeps several hives on the roof of her urban townhouse, although beekeeping is technically against city regulations. She gives the other building residents more honey than they can use.

29. Edward works two shifts for his freelance engineering business. During school hours he visits clients and answers email, and after his wife takes their girls off to bed he does his design work. In between, he and his daughters cook and clean and play.

30. Jay and Meredith have the lead roles in a community theatre production. When they met up do do some extra rehearsal, Jay complained that he didn't understand his character's motivation. As a result, they've decided to talk about the characters rather than rehearsing the script.

31. Bridget is driving to a party at a house she's never been to before. She has trouble navigating at night, so she keeps having to pull over and look at her GPS. It's taking her a very long time to get to the party.

32. Kyle bought a more comfortable car to use on a new job that requires lots of driving. After considering selling his old car, he decided to donate it instead. He just signed it over to a charitable nonprofit that needed a vehicle.

A.7 Experiment 7 materials

A.7.1 Experimental items

1. *(consequently / though/though)*

    (a) *(prominent + noncued)* After her soccer game, Johanna went straight to a restaurant. She hadn't eaten for hours, though she'd had a solid breakfast. She was ravenously hungry now.

    (b) *(subordinate + noncued)* After her soccer game, Johanna went straight to a restaurant. Though she hadn't eaten for hours, she'd had a solid breakfast. She was ravenously hungry now.

(c) *(prominent + cued)* After her soccer game, Johanna went straight to a restaurant. She hadn't eaten for hours, though she'd had a solid breakfast. Consequently, she was ravenously hungry now.

(d) *(subordinate + cued)* After her soccer game, Johanna went straight to a restaurant. Though she hadn't eaten for hours, she'd had a solid breakfast. Consequently, she was ravenously hungry now.

2. *(as a result / though/though)*

(a) *(prominent + noncued)* This afternoon while Tom was in class there was an unexpected blizzard. He biked to school today, though he usually drives when it snows. He has to take the bus home.

(b) *(subordinate + noncued)* This afternoon while Tom was in class there was an unexpected blizzard. Though he biked to school today, he usually drives when it snows. He has to take the bus home.

(c) *(prominent + cued)* This afternoon while Tom was in class there was an unexpected blizzard. He biked to school today, though he usually drives when it snows. As a result, he has to take the bus home.

(d) *(subordinate + cued)* This afternoon while Tom was in class there was an unexpected blizzard. Though he biked to school today, he usually drives when it snows. As a result, he has to take the bus home.

3. *(consequently / while/while)*

(a) *(prominent + noncued)* In January Laura went to a conference in Australia, where it was summertime. She had only packed heavy clothing, while her colleagues were in shorts. She was uncomfortably warm for most of the week.

(b) *(subordinate + noncued)* In January Laura went to a conference in Australia, where it was summertime. While she had only packed heavy clothing, her colleagues were in shorts. She was uncomfortably warm for most of the week.

(c) *(prominent + cued)* In January Laura went to a conference in Australia, where it was summertime. She had only packed heavy clothing, while her colleagues were in shorts. Consequently, she was uncomfortably warm for most of the week.

(d) *(subordinate + cued)* In January Laura went to a conference in Australia, where it was summertime. While she had only packed heavy clothing, her colleagues were in shorts. Consequently, she was uncomfortably warm for most of the week.

4. *(consequently / though/though)*

(a) *(prominent + noncued)* Before a dinner party, Beth forgot to answer an email asking about food preferences. She tries hard to be polite, though she strongly dislikes corn. She ate her corn chowder without saying anything.

(b) *(subordinate + noncued)* Before a dinner party, Beth forgot to answer an email asking about food preferences. Though she tries hard to be polite, she strongly dislikes corn. She ate her corn chowder without saying anything.

(c) *(prominent + cued)* Before a dinner party, Beth forgot to answer an email asking about food preferences. She tries hard to be polite, though she strongly dislikes corn. Consequently, she ate her corn chowder without saying anything.

(d) *(subordinate + cued)* Before a dinner party, Beth forgot to answer an email asking about food preferences. Though she tries hard to be polite, she strongly dislikes corn. Consequently, she ate her corn chowder without saying anything.

5. *(as a result / though/though)*

   (a) *(prominent + noncued)* Kate borrowed a calculator for her physics exam. The calculator turned out to be unfamiliar, though she understood the physics material. She wasn't able to finish all the problems.

   (b) *(subordinate + noncued)* Kate borrowed a calculator for her physics exam. Though the calculator turned out to be unfamiliar, she understood the physics material. She wasn't able to finish all the problems.

   (c) *(prominent + cued)* Kate borrowed a calculator for her physics exam. The calculator turned out to be unfamiliar, though she understood the physics material. As a result, she wasn't able to finish all the problems.

   (d) *(subordinate + cued)* Kate borrowed a calculator for her physics exam. Though the calculator turned out to be unfamiliar, she understood the physics material. As a result, she wasn't able to finish all the problems.

6. *(consequently / though/though)*

   (a) *(prominent + noncued)* Bob visited his cousin in New York City over the holidays. He'd never driven his truck into the city before, though he'd been delivering mail around rural Pennsylvania for years. The city traffic made him very nervous.

   (b) *(subordinate + noncued)* Bob visited his cousin in New York City over the holidays. Though he'd never driven his truck into the city before, he'd been delivering mail around rural Pennsylvania for years. The city traffic made him very nervous.

   (c) *(prominent + cued)* Bob visited his cousin in New York City over the holidays. He'd never driven his truck into the city before, though he'd been delivering mail around rural Pennsylvania for years. Consequently, the city traffic made him very nervous.

   (d) *(subordinate + cued)* Bob visited his cousin in New York City over the holidays. Though he'd never driven his truck into the city before, he'd been delivering mail around rural Pennsylvania for years. Consequently, the city traffic made him very nervous.

7. *(as a result / though/while)*

   (a) *(prominent + noncued)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. He wasn't able to do the prison interview, though he did get interviews with several of the executive's associates. His article didn't make the front page of the paper.

   (b) *(subordinate + noncued)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. While he wasn't able to do the prison interview, he did get interviews with several of the executive's associates. His article didn't make the front page of the paper.

   (c) *(prominent + cued)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. He wasn't able to do the prison interview, though he did get interviews with several of the executive's associates. As a result, his article didn't make the front page of the paper.

   (d) *(subordinate + cued)* Zack's editor tried hard to get him access to interview an imprisoned corporate executive. While he wasn't able to do the prison interview,

he did get interviews with several of the executive's associates. As a result, his article didn't make the front page of the paper.

8. *(consequently / though/while)*

(a) *(prominent + noncued)* Rebecca wasn't sure she was going to pass her physics midterm. She studied for the test every night for a week, though she fell behind in her other classes. She did well on the midterm.

(b) *(subordinate + noncued)* Rebecca wasn't sure she was going to pass her physics midterm. While she studied for the test every night for a week, she fell behind in her other classes. She did well on the midterm.

(c) *(prominent + cued)* Rebecca wasn't sure she was going to pass her physics midterm. She studied for the test every night for a week, though she fell behind in her other classes. Consequently, she did well on the midterm.

(d) *(subordinate + cued)* Rebecca wasn't sure she was going to pass her physics midterm. While she studied for the test every night for a week, she fell behind in her other classes. Consequently, she did well on the midterm.

9. *(as a result / though/though)*

(a) *(prominent + noncued)* Before Daniel moved into his current apartment he'd never lived alone. He never swept or tidied up his trash, though he did do his laundry regularly. His place was always a complete mess.

(b) *(subordinate + noncued)* Before Daniel moved into his current apartment he'd never lived alone. Though he never swept or tidied up his trash, he did do his laundry regularly. His place was always a complete mess.

(c) *(prominent + cued)* Before Daniel moved into his current apartment he'd never lived alone. He never swept or tidied up his trash, though he did do his laundry regularly. As a result, his place was always a complete mess.

(d) *(subordinate + cued)* Before Daniel moved into his current apartment he'd never lived alone. Though he never swept or tidied up his trash, he did do his laundry regularly. As a result, his place was always a complete mess.

10. *(consequently / though/though)*

(a) *(prominent + noncued)* The accounting department that Serena works in is always short on cash. Her computer was really old, though she took very good care of it. It couldn't run the new accounting software she needed.

(b) *(subordinate + noncued)* The accounting department that Serena works in is always short on cash. Though her computer was really old, she took very good care of it. It couldn't run the new accounting software she needed.

(c) *(prominent + cued)* The accounting department that Serena works in is always short on cash. Her computer was really old, though she took very good care of it. Consequently, it couldn't run the new accounting software she needed.

(d) *(subordinate + cued)* The accounting department that Serena works in is always short on cash. Though her computer was really old, she took very good care of it. Consequently, it couldn't run the new accounting software she needed.

11. *(consequently / though/while)*

(a) *(prominent + noncued)* Andrew loses the maintenance reminders his mechanic sends. He hadn't gotten an oil change for almost a year, though he'd been

193

intending to take care of it soon. His car broke down on a hot summer afternoon.

(b) *(subordinate + noncued)* Andrew loses the maintenance reminders his mechanic sends. While he hadn't gotten an oil change for almost a year, he'd been intending to take care of it soon. His car broke down on a hot summer afternoon.

(c) *(prominent + cued)* Andrew loses the maintenance reminders his mechanic sends. He hadn't gotten an oil change for almost a year, though he'd been intending to take care of it soon. Consequently, his car broke down on a hot summer afternoon.

(d) *(subordinate + cued)* Andrew loses the maintenance reminders his mechanic sends. While he hadn't gotten an oil change for almost a year, he'd been intending to take care of it soon. Consequently, his car broke down on a hot summer afternoon.

12. *(as a result / while/while)*

(a) *(prominent + noncued)* Robert and Karen's family all live far away. Robert refuses to travel by airplane, while Karen wants to tour the globe. They stay home and never visit their grandchildren.

(b) *(subordinate + noncued)* Robert and Karen's family all live far away. While Robert refuses to travel by airplane, Karen wants to tour the globe. They stay home and never visit their grandchildren.

(c) *(prominent + cued)* Robert and Karen's family all live far away. Robert refuses to travel by airplane, while Karen wants to tour the globe. As a result, they stay home and never visit their grandchildren.

(d) *(subordinate + cued)* Robert and Karen's family all live far away. While Robert refuses to travel by airplane, Karen wants to tour the globe. As a result, they stay home and never visit their grandchildren.

13. *(consequently / though/while)*

(a) *(prominent + noncued)* It was a sunny day, so Linda went to the beach. She didn't wear any sunscreen, though she was careful to drink plenty of water. She ended up with a nasty burn.

(b) *(subordinate + noncued)* It was a sunny day, so Linda went to the beach. While she didn't wear any sunscreen, she was careful to drink plenty of water. She ended up with a nasty burn.

(c) *(prominent + cued)* It was a sunny day, so Linda went to the beach. She didn't wear any sunscreen, though she was careful to drink plenty of water. Consequently, she ended up with a nasty burn.

(d) *(subordinate + cued)* It was a sunny day, so Linda went to the beach. While she didn't wear any sunscreen, she was careful to drink plenty of water. Consequently, she ended up with a nasty burn.

14. *(consequently / though/though)*

(a) *(prominent + noncued)* Rick was on a long overseas tour. He forgot to bring electrical adapters, though he had put them on his list of things to pack. He wasn't able to charge any of his electronic gadgets.

(b) *(subordinate + noncued)* Rick was on a long overseas tour. Though he forgot to bring electrical adapters, he had put them on his list of things to pack. He wasn't able to charge any of his electronic gadgets.

(c) *(prominent + cued)* Rick was on a long overseas tour. He forgot to bring electrical adapters, though he had put them on his list of things to pack. Consequently, he wasn't able to charge any of his electronic gadgets.

(d) *(subordinate + cued)* Rick was on a long overseas tour. Though he forgot to bring electrical adapters, he had put them on his list of things to pack. Consequently, he wasn't able to charge any of his electronic gadgets.

15. *(as a result / though/while)*

(a) *(prominent + noncued)* Betty's work day at her bakery starts hours before sunrise. She can leave the shop to her employees by noon, though the hard work sometimes leaves her exhausted and sleepy. She often spends sunny afternoons hiking and playing sports.

(b) *(subordinate + noncued)* Betty's work day at her bakery starts hours before sunrise. While she can leave the shop to her employees by noon, the hard work sometimes leaves her exhausted and sleepy. She often spends sunny afternoons hiking and playing sports.

(c) *(prominent + cued)* Betty's work day at her bakery starts hours before sunrise. She can leave the shop to her employees by noon, though the hard work sometimes leaves her exhausted and sleepy. As a result, she often spends sunny afternoons hiking and playing sports.

(d) *(subordinate + cued)* Betty's work day at her bakery starts hours before sunrise. While she can leave the shop to her employees by noon, the hard work sometimes leaves her exhausted and sleepy. As a result, she often spends sunny afternoons hiking and playing sports.

16. *(consequently / though/while)*

(a) *(prominent + noncued)* Mark's hobby is repairing antique clocks. Replacement parts can be impossible to find, though he has an extensive network of sources. He sometimes ends up making hard-to-locate parts himself in his machine shop.

(b) *(subordinate + noncued)* Mark's hobby is repairing antique clocks. While replacement parts can be impossible to find, he has an extensive network of sources. He sometimes ends up making hard-to-locate parts himself in his machine shop.

(c) *(prominent + cued)* Mark's hobby is repairing antique clocks. Replacement parts can be impossible to find, though he has an extensive network of sources. Consequently, he sometimes ends up making hard-to-locate parts himself in his machine shop.

(d) *(subordinate + cued)* Mark's hobby is repairing antique clocks. While replacement parts can be impossible to find, he has an extensive network of sources. Consequently, he sometimes ends up making hard-to-locate parts himself in his machine shop.

A.7.2 Filler items

(same as Experiment 6)

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

János Aczél. *Lectures on Functional Equations and their Applications.* Academic Press, New York, 1966.

Edward Adelson. The checkershadow illusion. http://web.mit.edu/persci/people/adelson/checkershadow₁llusion.html, 1995.

A. Almor. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review,* 106:748–765, 1999.

Amit Almor and Veena Nair. The form of referential expressions in discourse. *Language and Linguistics Compass,* 1(2):84–89, 2007.

Amit Almor, Daniel Kempler, Maryellen C. MacDonald, Elaine S. Andersen, and Lorraine K. Tyler. Why do alzheimer patients have difficulty with pronouns? working memory, semantics, and reference in comprehension and production in alzheimer's disease. *Brain and Language,* 67:202–227, 1999.

JR Anderson and LJ Schooler. Reflections of the environment in memory. *Psychological Science,* 1991.

Mira Ariel. *Accessing Noun Phrase Antecedents.* Routledge, London and New York, 1990.

J. E. Arnold. *Reference Form and Discourse Patterns.* PhD thesis, Stanford University, 1998.

J. E. Arnold. The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes,* 31(2):137–162, 2001.

J. E. Arnold, J. G. Eisenband, S. Brown-Schmidt, and J. C. Trueswell. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition,* 76:B13–B26, 2000.

J. E. Arnold, R. Altmann, M Fagnano, and M. K. Tanenhaus. The old and thee, uh, new. *Psychological Science,* pages 578–582, 2004.

N. Asher. Reference to abstract objects, 1993.

R.H. Baayen. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press, Cambridge, UK., 2008.

D. Bates, M. Maechler, and B. Dai. *lme4: Linear mixed-effects models using S4 classes.* http://lme4.r-forge.r-project.org/, r package version 0.999375-34 edition, 2008.

David Beaver. The optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1):3–56, 2004.

Klinton Bicknell, Roger Levy, and Vera Demberg. Correcting the incorrect: Local coherence effects modeled with prior belief update. In *Proceedings of the 35th Annual Meeting of the Berkeley Linguistics Society (BLS)*, 2009.

L. Bloomfield. *An Introduction to the Study of Language.* Henry Holt and Co., New York, 1914.

Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. A centering approach to pronouns. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics*, pages 155–162, Morristown, NJ, USA, 1987. Association for Computational Linguistics. doi: http://dx.doi.org/10.3115/981175.981197.

L. Carlson, D. Marcu, and M. Okurowski. *Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory.* Kluwer Academic Press, 2002.

Lewis Carroll. *Alice's Adventures in Wonderland.* Macmillan, 1865.

W. Chafe. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li, editor, *Subject and Topic*, pages 25–56. Academic Press Inc., New York, 1976.

P.W. Cheng. From covariation to causation: A causal power theory. *Psychological Review*, 104:367–405, 1997.

Noam Chomsky. *Syntactic Structures.* Mouton, The Hague, 1957.

H. Wind Cowles and Alan Garnham. Antecedent focus and conceptual distance effects in category noun-phrase anaphora. *Language and Cognitive Processes*, 20(6):725–750, 2005.

R. T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14:1–13, 1946.

R. T. Cox. *The Algebra of Probable Inference.* Johns Hopkins University Press, Baltimore, MD, 1961.

M. Crocker and T. Brants. Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, 29(6):647–669, 2000.

V. Demberg and F. Keller. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109:193–210, 2008.

Vera Demberg and Frank Keller. A computational model of prediction in human parsing: Unifying locality and surprisal effects. In Niels Taatgen and Hedderik van Rijn, editors, *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 1888–1893, Amsterdam, 2009. Cognitive Science Society.

E. Fedorenko and E. Gibson. The interaction of syntactic and lexical information sources in language processing: the case of the noun-verb ambiguity. In *20th CUNY Conference on Human Sentence Processing*, UC San Diego, March 2007.

F. Ferreira and C. Clifton. The independence of syntactic processing. *Journal of Memory and Language*, 25:348–368, 1986.

Stephani Foraker and Brian McElree. The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, 56:357–383, 2007.

M. C. Frank, S. Goldwater, T. Griffiths, and J. Tenenbaum. Modeling human performance in statistical word segmentation. *Cognition*, in press.

S. C. Garrod and A J Sanford. Resolving sentences in a discourse context: how discourse representation affects language understanding. In M Gernsbacher, editor, *Handbook of Psycholinguistics*. Wiley, New York, 1994.

M.A. Gernsbacher. Mechanisms that improve referential access. *Cognition*, 32:99–156, 1989.

E. Gibson. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76, 1998.

E. Gibson. The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O'Neil, editors, *Image, Language, Brain*, pages 95–126. MIT Press, 2000.

E. Gibson. The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54:363–388, 2006.

E. Gibson and E. Fedorenko. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, 14(6):233–234, 2010.

Noah D. Goodman, Vikash K. Mansinghka, Daniel M. Roy, Keith Bonawitz, and Joshua B. Tenenbaum. Church: a language for generative models. *Uncertainty in Artificial Intelligence*, 2008.

P. Gordon, B. Grosz, and L. Gilliom. Language comprehension and probe-list memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26:766–775, 2000.

P. C. Gordon, B. J. Grosz, and L. A. Gilliom. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17:311–347, 1993.

Peter C. Gordon and Randall Hendrick. The representation and processing of coreference in discourse. *Cognitive Science*, 22(4):389–424, October–December 1998.

Peter C. Gordon, Randall Hendrick, and William H. Levine. Memory-load interference in syntactic processing. *Psychological Science*, 13:425–430, September 2002.

H. P. Grice. Logic and conversation. In Peter Cole and J. L. Morgan, editors, *Syntax and semantics: Speech acts*, volume 3, pages 41–58. Academic, New York, 1975.

J. Grimes. *The thread of discourse*. Mouton, The Hague, 1975.

Daniel Grodner and Edward Gibson. Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, 29(2):261–290, 2005.

B. J. Grosz, A. K. Joshi, and S. Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.

Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 69:274–307, 1993.

John Hale. A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, volume 2, pages 159–166, Pittsburgh, PA, 2001. Association for Computational Linguistics.

M. A. K. Halliday and R. Hasan. *Cohesion in English*. Longman, London, 1976.

Jorge Hankamer and Ivan Sag. Deep and surface anaphora. *Linguistic Inquiry*, 7: 391–426, 1976.

Jerry R. Hobbs. On the coherence and structure of discourse. Technical Report 85-37, Center for the Study of Language and Information (CSLI), Stanford, CA, 1985.

Norbert Hornstein. Towards a theory of tense. *Linguistic Inquiry*, 8(3):521–557, Summer 1977.

D.A. Huffman. A method for the construction of minimum-redundancy codes. In *Proceedings of the I.R.E.*, pages 1098–1102, September 1952.

Kenneth E. Iverson. *A Programming Language*, volume New York. Wiley, 1962.

T. Jaeger and R. Levy. The structural basis of probabilistic syntactic production. In *In Proceedings of Architectures and Mechanisms of Language Processing*, AMLaP, Nijmegen, Netherlands, 2006.

T. F. Jaeger. *Redundancy and Syntactic Reduction in Spontaneous Speech*. PhD thesis, Stanford University, 2006.

E. T. Jaynes. Information theory and statistical mechanics ii. *Physical Review*, 108 (2):171–190, 1957.

E. T. Jaynes. *Probability Theory: The Logic of Science*, volume 1. Cambridge University Press, 2003.

A.K. Joshi, L.S. Levy, and M. Takahashi. Tree adjunct grammars. *Journal of Computer and System Sciences*, 10, March 1975.

Andrew Kehler. *Coherence, Reference, and the Theory of Grammar*. Stanford University Press, Stanford, CA, 2002.

W. Kintsch. The role of knowledge in discourse comprehension: a construction-integration model. *Psychol Rev*, 95(2):163–182, April 1988.

Alistair Knott. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. PhD thesis, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, 1996. URL citeseer.ist.psu.edu/article/knott96datadriven. html.

Donald Knuth. Two notes on notation. *American Mathematical Monthly*, 99(5): 403–422, May 1992.

S.-Y. Kuroda. Classes of languages and linear-bounded automata. *Information and Control*, 7(2):207–223, 1964.

Alan Lee, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. Departures from tree structures in discourse: Shared arguments in the penn discourse treebank. In *Proceedings of the Constraints in Discourse III Workshop*, Potsdam, Germany,, July–August 2008.

R. Levy. Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177, 2008.

R. Levy and T.F. Jaeger. Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, and T. Hoffman, editors, *Advances in neural information processing systems*, volume 19 of *NIPS*, pages 849–856, Cambridge, MA, MIT Press 2007.

Roger Levy and Frank Keller. Expectation and locality effects in german verb-final structures. In Submission, 2010.

Roger Levy, Klinton Bicknell, Tim Slattery, and Keith Rayner. Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. In *Proceedings of the National Academy of Sciences 106(50):21086-21090.*, 2009a.

Roger Levy, Evelina Fedorenko, Mara Breen, and Ted Gibson. The processing of extraposed structures in english. In Submission, 2009b.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, March 1963.

R. D. Luce. *Individual Choice Behavior: A Theoretical Analysis.* Wiley, New York, 1959.

David J. C. MacKay. *Information Theory, Inference & Learning Algorithms.* Cambridge University Press, 2002.

DJC. MacKay and Peto. L. A hierarchical dirichlet language model. *Natural Language Engineering*, 1(3):1–19, 1995.

William C. Mann and Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

D. Marcu. Extending a formal and computational model of rhetorical structure theory with intentional structures la grosz and sidner, 2000a. URL `citeseer.csail.mit.edu/article/marcu00extending.html`.

Daniel Marcu. The rhetorical parsing, summarization, and generation of natural language texts. unpublished ph.d. dissertation, 1997.

Daniel Marcu. The theory and practice of discourse parsing and summarization, 2000b.

Daniel Marcu and Abdessamad Echihabi. An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*, 2002. URL `http://www.isi.edu/~marcu/papers/relations-acl02.pdf`.

David Marr. *Vision: A Computational Approach.* Freeman & Co., San Francisco, 1982.

Eleni Miltsakaki, C. Creswell, K. Forbes, R. Prasad, A. Joshi, and B. Webber. The discourse anaphoric properties of connectives. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2002)*, Lisbon, 2002.

K. Nakatani and E. Gibson. An on-line study of japanese nesting complexity. *Cognitive Science*, 2010.

S. Narayanan and D. Jurafsky. Bayesian models of human sentence processing. In *Proceedings of the twelfth annual meeting of the cognitive science society*, 1998.

S. Narayanan and D. Jurafsky. A bayesian model predicts human parse preference and reading time in sentence processing. *Advances in Neural Information Processing Systems*, 14:59–65, 2002.

D. M. Oppenheimer and M. C. Frank. A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, 2007. doi:10.1016/j.cognition.2007.05.010.

Umesh Patil, Shravan Vasishth, Brian Bartek, and Richard Lewis. The interplay of locality and surprisal. In *CUNY Sentence Processing Conference*, Chapel Hill, NC, 2008.

Judea Pearl. *Probabilistic reasoning n intelligent systems: networks of plausible inference*. Morgan-Kaufmann, 1988.

C. Pollard and I. Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, IL, 1994.

Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Livio Robaldo, Bonnie Webber, and Aravind Joshi (The PDTB Research Group). The penn discourse treebank 2.0 annotation manual. Technical report, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA, 2007.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. The penn discourse treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, 2008.

R Quirk and S Greenbaum. *A concise grammar of contemporary English*. Longman, 1973.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL http://www.R-project.org. ISBN 3-900051-07-0.

I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, June 1960. ISSN 0368-4245.

H. Reichenbach. *Elements of symbolic logic*. The Free Press, 1966.

B. Roark. Probabilistic top-down parsing and language modeling. *Computational Linguistics*, 27(2,):249–276, 2001.

Hannah Rohde, Roger Levy, and Andrew Kehler. Anticipating explanations in relative clause processing. In Submission, 2010.

Peter S Rosenbaum. *The grammar of English predicate complement constructions*. MIT Press, Cambridge, MA, 1967.

L.A. Schmidt, C. Kemp, and J.B. Tenenbaum. Nonsense and sensibility: Inferring unseen possibilities. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, Vancouver, BC., 2006.

M. Seidenberg, M. Tanenhaus, J. Leiman, and M. Bienkowski. Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14:489–537, 1982.

Claude E. Shannon. Prediction and entropy of printed english. *The Bell System Technical Journal,*, 30(50-64), 1951.

A. J. Stewart, M. J. Pickering, and A. J. Sanford. The role of implicit causality in language comprehension: Focus versus integration accounts. *Journal of Memory and Language*, 42:423–443, 2000.

John Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18:643–662, 1935.

Tamara Y. Swaab, C. Christine Camblin, and Peter C. Gordon. Electrophysiological evidence for reversed lexical repetition effects in language processing. *J Cogn Neurosci*, 16(5):715–726, Jun 2004.

M. Taboada and W.C. Mann. Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies*, 8(3):423–459, 2006a.

M. Taboada and W.C. Mann. Applications of rhetorical structure theory. *Discourse Studies*, 8(4):567–588, 2006b.

W. Tabor, C.J. Juliano, and M.K. Tanenhaus. Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12:211–271, 1997.

W. Tabor, B. Galantucci, and D. Richardson. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50:355–370, 2004.

M.K. Tanenhaus, J.M. Leiman, and M.S. Seidenberg. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18:427–441, 1979.

J.C. Trueswell and M.K. Tanenhaus. Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes*, 6:303–338, 1991.

J.C. Trueswell, M.K. Tanenhaus, and C. Kello. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory & Cognition*, pages 528–553, 1993.

J.C. Trueswell, M.K. Tanenhaus, and S.M. Garnsey. Semantic effects in parsing: Thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, pages 285–318, 1994.

E. Tulving, G. Mandler, and R. Baumal. Interaction of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology*, 18:62–71, 1964.

Edward Vul. *Sampling in Human Cognition*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, June 2010.

Bonnie L. Webber, Matthew Stone, Aravind K. Joshi, and Alistair Knott. Anaphora and discourse structure. *Computational Linguistics*, 29(4):545–587, 2003.

Bonnie Lynn Webber. Tense as discourse anaphor. *Comput. Linguist.*, 14(2):61–73, 1988. ISSN 0891-2017.

Terry Winograd. Understanding natural language. *Cognitive Psychology*, 3(1):1 – 191, 1972. ISSN 0010-0285. doi: DOI:10.1016/0010-0285(72)90002-3.

Florian Wolf and Edward Gibson. Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2):249–288, June 2005.

Florian Wolf, Edward Gibson, Amy Fisher, and Meredith Knight. A procedure for collecting a database of texts annotated with coherence relations. Technical report, Massachusetts Institute of Technology, Cambridge, MA, 2003.

Florian Wolf, Edward Gibson, and Timothy Desmet. Discourse coherence and pronoun resolution. *Language and Cognitive Processes*, 19(6):665–675, 2004.