NORTHWESTERN UNIVERSITY

The Relative Contribution of Syntactic and Semantic Prominence to the Salience of Discourse Entities

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Ralph L. Rose

EVANSTON, ILLINOIS

June 2005

ABSTRACT

The Relative Contribution of Syntactic and Semantic Prominence to the Salience of Discourse Entities

Ralph L. Rose

Entities realized in syntactically prominent positions are preferred antecedents for pronominal reference. This has been demonstrated in numerous psycholinguistic experiments (e.g., Hudson-D'Zmura and Tanenhaus, 1997; Mathews and Chodorow, 1988) and corpus investigations (Arnold, 1998b). However, for many verbs in English—especially eventive verbs—syntactic role and semantic role are often conflated. That is, syntactic subjects are often semantic agents and carry more PROTO-agent entailments (e.g., sentience, volition; see Dowty, 1991), while syntactic objects are often semantic patients and carry more PROTO-patient entailments (e.g., undergo change-of-state, causally-affected). Thus, it could be said that entities realized in prominent roles on a hierarchy of semantic roles (e.g., agent > patient as in Jackendoff, 1972) are preferred antecedents for pronominal reference. As such, the central question of this dissertation can be worded as follows: Is it the case that what has heretofore been observed as a result of syntactic prominence is in fact a result of semantic prominence?

I take as a starting point a generalized model of discourse processing in which the relative salience of discourse referents in the current context is seen as influencing subsequent reference to those referents as well as the form of referring expression (cf., Grosz et al., 1995; Kintsch and van Dijk, 1978) and then incorporate a detailed notion of discourse salience in which two prominence factors contribute to overall salience: syntactic and semantic prominence. Using this model as a foundation, I compare the relative effects of syntactic and semantic prominence in two investigative paradigms: psycholinguistic experimentation and corpus analysis. Results from both investigations give evidence that both syntactic and semantic prominence contribute to the salience of discourse referents. These results are analyzed with respect to two approaches to determining semantic prominence—a frame semantic (Fillmore, 1968, 1976) approach using the FrameNet system (Baker et al., 1998) and a PROTO-role approach using the PROTO-role entailments of Dowty (1991). A comparison of these two approaches and implications of the main findings for psycholinguistic models and computational implementations of discourse processing are discussed.

Acknowledgements

The title page of a doctoral dissertation contains the name of just one person in spite of the fact that the research effort which a dissertation represents involves the labor of many people. In my case, I have no small number of people whose efforts on my behalf I would like to acknowledge.

I would like to acknowledge my committee members for whom I have much respect and admiration. As I have interacted with them over the years, I can't help being reminded of Mark Twain's book, *A Connecticut Yankee in King Arthur's Court*. I'm sure each of my committee members must have felt at some point like the Yankee when he said of the King, "His head was an hour-glass; it could stow an idea, but it had to do it a grain at a time, not the whole idea at once." I thank my committee for their patience with me in this respect.

I am deeply indebted to my doctoral committee chairperson, Stefan Kaufmann. His extensive knowledge of discourse representation and mathematical methods has challenged me over these years to make my arguments more sound. I appreciate the long discussions we had in his office, at each other's homes, over the telephone, and even in Tokyo. For the numerous times I have kept him waiting for various drafts, I am sincerely apologetic and for the thoroughness with which he pored through and commented on those drafts, I am extremely grateful.

I also could not have completed the present work without the support of my other committee members. Mike Dickey has advised me on this research from its earliest days in the summer after my first year at Northwestern. I have always appreciated his talent for asking questions that forced me to be clarify my thoughts. He has been a faithful adviser, friend, and hotel roommate at conferences. I am also thankful to both Gregory Ward and Chris Kennedy for their service on my committee and for their flexibility in the scheduling of my defense. I also appreciate their support in other roles during my years as a graduate student. As department chair and director of graduate studies, both have made distinct contributions toward helping me finish my degree.

There are many others I wish to thank including Ann Bradlow who provided the laboratory space and equipment for running the psycholinguistic experiments; Katy Carlson and Catherine Anderson who allowed me to share experimental resources with them; and Erin McMahon-Leddon who provided the initial spark of insight from which this research grew. In addition there are numerous others who have given me helpful comments or useful resoursces at various stages of my research: Cleo Condoravdi, Vana Doufexi, Lyn Frazier, Peter Gordon, Hironobu Hosoi, Elsi Kaiser, Bill Ladusaw, Jeff Lidz, Barbara Partee, Colin Phillips, Jiha Sung as well as a number of experimental participants who gave some very interesting insights during post-experimental feedback sessions. I am further grateful to many other faculty and students I have had the fortune of studying with at Northwestern University including Janet Pierrehumbert, Julie Moore, Ann Bunger, Elisa Sneed, Evar Strid, Deborah Gordon, Nick Kasper, Jody Sostrin, Tessa Bent, and many others too numerous to name.

I also wish to acknowledge the Northwestern Linguistics department which has been a vibrant atmosphere in which to study. The rapport between faculty and students made for a rich academic environment. It will serve as a model for me to emulate as I begin my career in academia.

On a personal note, I wish also to thank several people for their support and encouragement over these past six years. I thank my parents, my brothers, and their families for cheering me up from time to time and always expressing their faith that I would finish. I also thank my in-laws for their emotional support during the most challenging final years of the dissertation.

Finally, I wish to thank two people who are very near and dear to me, whose love and patience with me over these years has enabled me to continue, and to whom I owe much. I thank my wife, Seika, for supporting me emotionally during these years, and also for letting me take these years to return to student life. I also thank my son, Ken, for being quiet when I needed to concentrate, for playing with me when I needed a break, for tagging along to many classes, meetings, and colloquia and sitting through them quietly. I have been extremely fortunate to have the cooperation of both my wife and son during these six years. Without them, I could not have finished; and without them, it would not have been meaningful.

Contents

List of Tables

List of Figures

Introduction

When looking closely at English texts, it is surprising just how much potential ambiguity there is, even in texts that have been extensively proofread. And yet, when we read a text, these potential ambiguities trip us up far less often than they actually occur. For instance, consider the short news article that follows.

> Houston (Reuters) A Texas woman was sentenced to 10 years in jail for running over the Manager of a McDonald's with her car because she wanted Mayonnaise on her cheeseburger.
>
> Waynetta Nolan, 37, showed no emotion Thursday as the sentence was read in court following a trial in which the McDonald's manager, Sherry Jenkins, said she gave Nolan the mayonnaise she requested, but she flew into a rage anyway.
>
> "I gave her everything she asked for—mayonnaise, no mustard, onions, everything I could possibly do for this lady. Mayo, mayo, mayo, and it's still not good enough," Jenkins told reporters outside the courtroom. ("Odds and Ends", 2003)

In this short text of about 100 words, there are two main characters who are referred to by their full names just once each. Besides that there are some 16 other references to these same characters using just last names, indefinite or definite noun phrases, or

pronouns. While names are very explicit and make it very easy for the reader to identify the intended entity, at the other extreme, pronouns are not at all explicit and may allow alternative interpretations. However, when reading a text as above, we usually feel that we understand who or what is being referred to even when referred to using a pronoun. One of the reasons for this of course lies in world knowledge. Certain things we know about the world and how events typically transpire lead us to prefer certain interpretations over others. But another reason lies in the structure of the text itself. The way the text is organized guides our interpretations. The text above begins with the noun phrase, *a Texas woman*, highlighting this character for the remainder of the sentence. This makes it easier to identify the car used in the event as belonging to the Texas woman. This interpretation is, admittedly, reachable by world knowledge alone, but here, text structure seems to underscore that interpretation.

In order to look at this notion of highlighting more closely, let's consider some cases where the texts are not as well-structured. For instance, consider the following state of affairs in (1) (using standard predicate logic notation) containing reference to three individuals (Matt, John, and book) and describing a set of events involving these individuals.

(1) $\exists x \exists y \exists z [Matt(x) \wedge John(y) \wedge book(z) \wedge give(x, z, y) \wedge$
$tell(x, y, read(y, z)) \wedge ask - about(x, y, z)]$

These facts could be expressed in many different ways. Three possibilities are listed in (2)-(4) in which the indices are intended to identify coreference—here, noun phrases which are intended to identify the same entity in the world.

(2) a. Matt$_i$ gave John$_j$ a book$_k$.

b. Then he$_i$ told him$_j$ to read it$_k$.

c. Later he$_i$ asked him$_j$ about it$_k$.

(3) a. Matt$_i$ gave John$_j$ a book$_k$.

b. Then he$_j$ was told by him$_i$ to read it$_k$.

c. Later he$_i$ asked John$_j$ about it$_k$.

(4) a. Matt$_i$ gave John$_j$ a book$_k$.

b. Then Matt$_i$ told him$_j$ to read it$_k$.

c. Later Matt$_i$ asked him$_j$ about it$_k$.

While these texts are nearly synonymous and are consistent with the state of affairs described in (1), the reader will likely notice that (2) is considerably easier to read than (3) or (4). In fact, although the coreference indices are indicated in (3) and (4), on a first reading of each of these, one is perhaps tempted to assign different indices; that is, to interpret the pronouns differently. Intuitively, there are two reasons for these observations. One is that while (2) appears to highlight a single individual, namely, Matt, (3) appears to shift our attention (from Matt to John and back to Matt). These attentional shifts can be seen as requiring more cognitive effort on the part of the reader (cf., Di Eugenio, 1997; Turan, 1995). Another reason is that although both (2) and (4) seem to tell the story more consistently from Matt's perspective, the former uses pronouns to indicate this continuity but the latter does not. Apparently, not using a pronoun when it is expected has certain processing consequences (cf., Gordon et al., 1993). In short, given the respective initial utterances and what the structure reveals about which entities are

highlighted, more effort is required to process subsequent utterances—particularly the referring expressions—in the texts in (3) and (4).

Highlighting seems to play an important role in how utterances are processed. But just how does highlighting work? Which factors determine which entities are highlighted? This question forms the basis for this dissertation. In particular, I am interested in outlining a specific model of highlighting, and investigating two factors in detail which contribute to it. Throughout this dissertation, I will refer to highlighting as *discourse salience* and give a general model of discourse salience in Chapter 1 and two specific implementations of it in Chapter 2.

As a preview to this investigation, consider the following text.

(5) a. Luke punched Max.

b. Then, he ran home.

In (5), there is some ambiguity in the interpretation of the pronoun: *he* could possibly be interpreted as referring to either Luke or Max. However, my intuition is that the more natural reading is that Luke is the one that ran home. We may account for this by saying first that Luke is the most salient entity in the preceding sentence, (5a), and second, that a pronoun should be used when refering to a salient entity. But exactly how is this salience determined here? And which factors contribute to it? One possibility is to look at the structure of the utterance: Luke appears as the subject of the verb *punch* and also is the first entity mentioned. This sort of information, as derived from the surface syntactic structure of the utterance, could be what determines the salience of entities. I will call this *syntactic prominence* (borrowing the term from McKoon et al., 1993). Hence, in (5a), Luke is the most syntactically prominent entity.

However, things are not quite so simple because there is a potential confound. While Luke appears as the syntactic subject of (5a), he takes a role of a different kind with respect to the semantics of the event. The verb in (5a) describes a punching event in which Luke is asserted to be the perpetrator of the event and Max is the victim. In commonly-used semantic terms (cf., Fillmore, 1968; Jackendoff, 1972; Speas, 1990; Baker, 1997, inter alia), Luke is the agent and Max is the patient. So another explanation for the interpretive preference in (5b) is that Luke is more salient because of his role as the semantic agent of the event described in (5a). I will call this *semantic prominence*.

But if syntactic and semantic prominence are confounded, how can we determine which of the two factors is responsible for discourse salience—or could in fact both be responsible? If both are responsible, then how do they work together to determine the overall salience of entities in a discourse? These, then, are the central questions that I will try to answer in this dissertation.

Chapter 1 builds the foundation for this study by defining a generalized model of discourse structure in which the relative salience of entities referred to in a discourse is seen as influencing pronoun resolution—the procedure by which readers determine the semantic interpretation of a pronoun.

Chapter 2 begins with a detailed but flexible model of how the discourse salience of entities is determined. The model allows for any number of factors to contribute to salience, but in this dissertation I will focus on only two: syntactic and semantic prominence. The chapter then continues with some detailed approaches for how syntactic and semantic prominence might each be determined. I consider two methods of determining syntactic prominence—a grammatical role-based method and a hierarchical tree-search

algorithmic method—as well as two approaches of determining semantic prominence—a frame semantic (Fillmore, 1968, 1976) approach using the FrameNet semantic system (Baker et al., 1998) and an approach based on the PROTO-role entailments (Dowty, 1991). The chapter concludes with a detailed illustration of how these factors might be instantiated in the discourse salience model.

Chapter 3 describes two series of psycholinguistic experiments designed to compare the relative influence of syntactic and semantic prominence on the pronoun resolution strategies of human readers. The experiments take advantage of alternating constructions (*spray*/*load* and *tough*/non-*tough*-constructions) in order to tease apart the confound described above.

Chapter 4 describes a pilot corpus analysis also designed to compare the relative influence of syntactic and semantic prominence. The 5,000+ word corpus consists of texts of narrative fiction and has been marked up to include syntactic and semantic role information information. In particular, the corpus has parallel mark-up for both the FrameNet and PROTO-role entailment systems.

Finally, Chapter 5 concludes the dissertation by tying together the results of the psycholinguistic experiments and corpus analysis. Implications for psycholinguistic models as well as computational implementations of discourse salience and pronoun resolution are discussed. Finally suggestions are given for further work.

CHAPTER 1

Background

1.1. Discourse Representation

The central domain of this dissertation is local discourse structure. I take as a discourse a string of utterances intentionally arranged to communicate more than the mere accumulation of propositions asserted by those utterances (Grosz and Sidner, 1986; Kamp and Reyle, 1993; Kehler, 1995, 2002; Kintsch and van Dijk, 1978; Halliday and Hasan, 1976; van Eijck and Kamp, 1997, inter alia).[1] There are several crucial properties of local discourse structure which must be discussed here as background and about which I must make some simplifying assumptions. In this section, I will describe these properties in some detail. In particular, I will discuss a general model of the relationship between the structure of discourse and how it interacts with the way that speakers highlight certain things in discourse as well as the way hearers interpret such things as pronouns in discourse. Finally, building on this model, I discuss the role that discourse salience plays in these processes.

I should like to note up front that I will largely be portraying these concepts from the hearer's perspective: that is, the role that the hearer has in processing discourse. However, in acknowledgement of recent work which has begun to examine the speaker's

[1]This approach is in contrast (but not unrelated to) the approach taken by those who view discourse analysis from a broader perspective (e.g., in sociolinguistic ethnography), approaching a text with respect to such factors as *setting*, *participants*, *ends*, *act sequence*, *key*, *instrumentalities*, *norms*, and *genre* (i.e., the "SPEAKING" mnemonic of Hymes, 1974).

and hearer's roles independently (e.g., bidirectional OT in Blutner, 1998, 2000; van Rooy, 2003), I will from time to time make reference to how the hearer's and speaker's roles differ in discourse production and discourse perception, respectively. One of my ultimate goals, though, in this dissertation is to find some core conceptualization of discourse salience which both speaker and hearer rely on while performing their respective tasks.

1.1.1. Discourse Referents

I assume that discourse processing involves perceiving two things in the message: who or what is being talked about, and assertions about those entities. As a hearer perceives these things, a mental representation of the discourse is constructed to reflect these perceptions (cf., Kintsch and van Dijk, 1978; Lambrecht, 1994). A fundamental part of this representation is some sort of cognitive representation of the entities which are evoked. Following Kamp (1981), Karttunen (1976), and Heim (1982, 1983), I will call these representations *discourse referents*. I also assume that all noun phrases in a discourse initiate the introduction discourse referents into this representation. However, exactly how—or rather, where—they are introduced into the representation may vary, leading to certain consequences such as their later accessibility (to be discussed in greater detail below). I further assume that these discourse referents persist throughout the duration of an incoming discourse. Thus, at the completion of a discourse, the cognitive representation of that discourse contains a cumulative set of all referents introduced in that discourse.

1.1.2. Discourse Structure: Inter-utterance Relations

As discussed above, I take a discourse to be a sequence of utterances which are intentionally arranged to communicate more than the mere sum of the propositions of its utterances. The arrangement may indicate, for example, a temporal relation between two events, or a causal relation between two propositions. For instance, compare the sequence of utterances in (6) to the sentences in (7) written by my son for a vocabulary homework assignment.

(6) a. John and Andrew went to the amusement park.

b. John rode the big roller coaster seven times.

c. Andrew wanted to do so, too, but he was a little scared.

(7) a. An elephant's body is big.

b. The mouse moved forward.

c. John frequently brushes his teeth.

d. Bob can't comprehend what the teacher is trying to say.

As one reads through (6), it is easy to discern how each sentence follows on from the sentence preceding it: (6b) can easily be seen as a narrative continuation of (6a), that is, as a subsequent event, while (6c) can be readily seen as expressing some degree of resemblance to the content of (6b). However, in (7) it is not at all clear how to connect successive sentences. One might, with great effort, be able to infer some connection between say, (7a) and (7b). After reading (7a), the reader's representation of the context thus far would contain an image of a particularly large elephant. Upon reading (7b), the reader might then infer that there exists a rather aggressive species of mouse. Then it might be plausible that one of these mice, noticing the large elephant, is tempted to

challenge it by moving toward it. In this context, we might be able to see (7a) and (7b) as related by a causal connection.

The discussion of the two texts above brushed on a very important feature of discourse: each utterance of the discourse is processed in the context of the preceding utterances (Kehler, 2002; Kintsch and van Dijk, 1978). If there is an explicit link between an utterance and its preceding context, or if it can be accommodated with minimal inferential effort, then it will be more easily accepted as a continuation of the discourse and processed as such (Kehler, 2002). In this dissertation, I am centrally interested in these explicit links among adjacent utterances and, in particular, links which involve coreferring noun phrases as in the discourse shown in (8).

(8) a. John$_i$ hit Matt.

b. He$_i$ was angry.

In (8a), there are two discourse referents, let's say X and Y, which refer to John and Matt respectively. Then in (8b), there is just one discourse referent, let's say, Z, which refers to John. Thus between these two adjacent utterances there is one coreferential link involving a pair of discourse referents—X in (8a) and Z in (8b)—both of which refer to John. For the remainder of this dissertation, when I talk about coreference, I mean it to be of this type: when two discourse referents refer to the same entity in the real world. In particular, in this dissertation, I will be centrally concerned with inter-utterance coreference.

1.1.3. Discourse Representation Theory (DRT)

In order to illustrate the two properties of discourse discussed so far, discourse referents and structure, here I review one formalism of discourse representation widely used in computational linguistics. In Discourse Representation Theory (DRT: Kamp, 1981; Kamp and Reyle, 1993), the current state of the discourse is represented in data structures called Discourse Representation Structures (DRS) consisting of two parts: a set of discourse referents and a set of semantic conditions on those discourse referents. Each utterance then constitutes a transformation to be applied to the current DRS in order to update these two sets. Construction rules process each utterance, introducing new items to the set of discourse referents, as necessary, and adding propositions to the set of conditions. At any time, there is only one top-level DRS—called the principal DRS—which is taken to be the truth-conditional representation of the discourse so far: The representation evaluates as true if there are entities in the world being discussed which correspond to the discourse referents in the principle DRS, and the set of conditions below it hold true of those entities.

As an illustration, consider the discourse shown in (9).

(9) a. A man bought a new book.

b. He liked it.

To simplify this example, we can regard the state of the discourse at the beginning of this DRS as empty: that is, there is no shared knowledge between the speaker and hearer (except, of course, for presumed knowledge of the language, world, etc). Hence, the initial DRS is empty as illustrated in Figure 1.1 as DRS_0. The result of the application of the first sentence of the discourse, *A man bought a new book*, to DRS_0 results in DRS_1 in

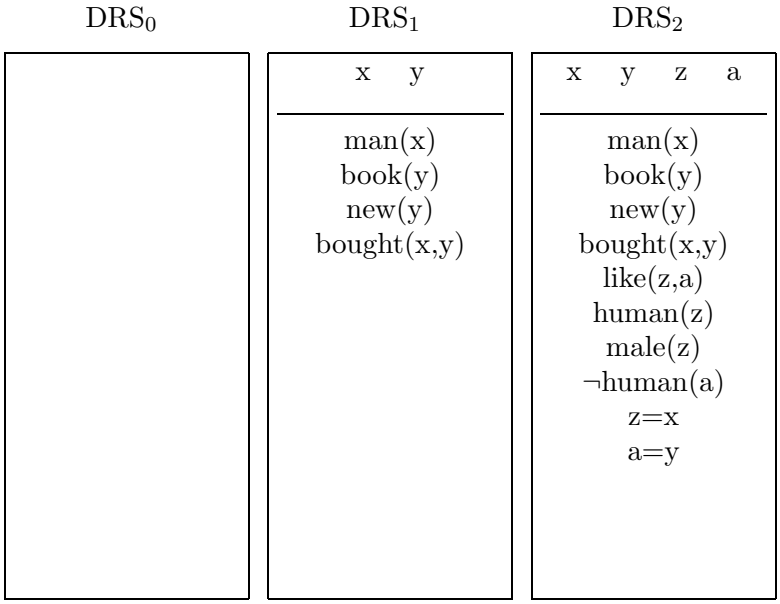


Figure 1.1. A sample discourse in Discourse Representation Theory. The context DRS is progressively updated as each new utterance is processed

which two discourse referents, X and Y have been introduced and four semantic conditions on X and Y have been added. Similarly, after applying *He liked it*, DRS_2 results. Two more discourse referents are added, Z and A, as well as the respective conditions. The truth value of DRS_2 can then be determined by checking to see if there are real-world referents (in the relevant domain) for the discourse referents X, Y, Z, and A who also satisfy all of the listed semantic conditions.

One of the strengths of DRT (and in fact, also one of the motivations for its development) is the way that it handles some of the difficulties of intra- and inter-sentential coreference. For example, consider the anomalous discourse in (10).

(10) a. John doesn't own a donkey.

b. #It is grey.

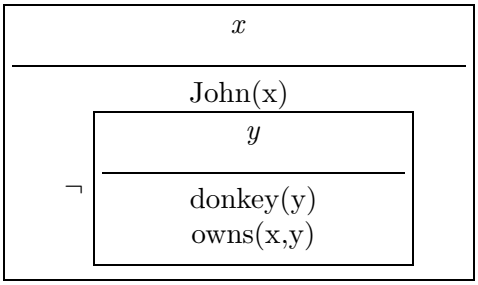


Figure 1.2. Discourse Representation Theory accounts for inter-sentential coreference restrictions with complex representations: Referents in the embedded DRS are not accessible for subsequent pronominal reference.

While both utterances in (10) are syntactically well-formed, the (10b) is semantically anomalous in this context. This is accounted for in the DRT formalism in the way that DRT handles negation and quantification. In these cases, *complex* sub-DRSs are constructed. Thus, when the DRS is updated after (10a) (see Figure 1.2), a single discourse referent X is added to the principal DRS along with two semantic conditions: one which constrains the reference of X to John, and another condition which is itself a DRS. This sub-DRS is considered complex in that it is within the scope of an operator (here, negation) but otherwise has a standard DRS structure with a discourse referent set and accompanying semantic conditions. In this case, there is one new discourse referent, Y, and two conditions.

The principal DRS shown in Figure 1.2 is then the current context when the next sentence is encountered. In order to resolve the pronoun *it* in (10b), the system must find a suitable antecedent in an accessible location. In DRT, accessibility is defined in leftward and upward terms: discourse referents in leftward DRSs (e.g., as in the antecedent of a consequent of a logical entailment relation) or upward (i.e., containing) DRSs are accessible. Thus, in the present example, the only accessible referent for *it* in

(10b) is that in the principal DRS: x. For the given context, (10a), the continuation (10b) is therefore anomalous.

The ease with which DRT handles intersentential coreference is a great advantage of the formalism. However, some parts of the coreference process have not yet been spelled out. For example, referring back to DRS_2 in Figure 1.1, note the equivalence relations, $z = x$ and $a = y$. How were these relations determined? In other words, how did the system manage to establish a coreferential relationship between *a man* and *he* (via their respective discourse referents) and also between *a book* and *it*? Or perhaps a more challenging case would be the discourse in (11).

(11) a. Luke$_i$ hit Max.

b. Then, he$_i$ ran home.

When sentence (11b) is encountered, the context (i.e., principal DRS) would contain two discourse referents corresponding to Luke and Max. In the original description of DRT, there is no procedure for deciding which of these two discourse referents the pronoun *he* should be resolved to. Furthermore, the present example is relatively simple: In a longer discourse in which many referents have been introduced to the representation, the difficulty of determining coreference will be compounded. To be fair, Kamp (1981) acknowledges this fact and leaves it to future work. Thus, one of the goals of this dissertation is to provide a characterization of pronominal reference resolution which could be used in an adaptation of DRT.[2] I will explain this resolution process in more detail in Chapter 2.

[2]There are other proposed adaptations of DRT to handle pronominal reference resolution already discussed in the literature. See Gordon and Hendrick (1997a,b) for one such proposal.

1.1.4. The Store and Anaphoric Expressions

In this dissertation, I will call the set of discourse referents the *store* and treat it as more than a set of items, but rather as a highly structured representation. I will describe this structure in detail in Chapter 2, but here I would like to make some comments about the relationship between the store and anaphoric expressions.

Anaphoric expressions, broadly speaking, are linguistic expressions which are referentially dependent on other linguistic expressions within a text (cf., Fiengo and May, 1994; Halliday and Hasan, 1976; Hirst, 1981). As illustrated in (12) by coindexing, this includes null pronouns, overt pronouns, reflexives, verb phrase ellipsis, *one...the other* anaphors, definite anaphors, and many others (see Hirst, 1981, for an overview of many of these types).[3]

(12) a. John$_i$ climbed the mountain and then $\emptyset_i$ took a picture.

b. John watched a ballerina$_i$. She$_i$ was very graceful.

c. John$_i$ shaved himself$_i$.

d. John [defended himself]$_i$ better than the lawer $\emptyset_i$.

e. There were [two books]$_{i+j}$ left. John bought one$_i$ and Mark bought the other$_i$.

f. John connected an oscilloscope$_i$ to the prototype circuit board, but after ten minutes of testing, the instrument$_i$ broke.

[3]There is a certain group of linguistic phenomena which has been called "antecedentless anaphors" (Cornish, 1996) in which there is no linguistic antecedent in the context to give a pronoun its full interpretation (see also Clark and Sengul, 1979; Yule, 1979). One account of these argues that the antecedents of such anaphors are taken as given and therefore already part of the mental representation of the context. Cornish (1996) further argues that the antecedent must be salient in the context—consistent with the model I am presenting here.

One advantage of the store here is that it provides a useful cognitive mechanism influencing the interpretation of a referring expression. Consider the indefinite noun phrase in (12b), *a ballerina*. In DRT terms, when this sentence is integrated into the current context, a new discourse referent in the store would be created for *a ballerina*. Then, when the next sentence is incorporated, the discourse referent introduced by the pronoun *she* should be linked up under an equivalence relation to that of the ballerina already in the store. In short then, in this model, an anaphoric expression is a cognitive indication that there is a suitable and accessible discourse referent in the context (cf., Ariel, 1988; Givón, 1983b; Gundel et al., 1993, inter alia), and thus that the search for an antecedent should be constrained to the store. Another way to think of this has been described by Geurts (1999): an anaphoric expression *presupposes* that its referent is already in the context[4] and thus interpretation of the anaphor must occur with respect to the preceding context, including the current state of the store.

In DRT, the store is an unordered set of discourse referents. However, research on intersentential anaphoric reference suggests that if there is such a thing as a store, then it must have some sort of structure or ordering imposed on it. Many approaches to anaphora assume that there is some sort of hierarchy of anaphoric expressions (e.g., the "givenness hierarchy" of Gundel et al. (1993), the "accessibility hierarchy" of Ariel (1988)) which represents the ease with which the antecedent of an anaphoric expression can be located in the context. Null pronominals as in (12a) are regarded as highest on

[4]This is somewhat of a simplification since there are some patterns of anaphoric reference which do not seem to require an explicit antecedent: for example, inferrables (cf., Prince, 1981, 1992) as in *I walked past the school and the door was open*. However, even these typically require some earlier element for their interpretation. Note that in the above example, *the door* is referentially dependent on *the school*. As such, the search for such help is still confined to the store. Assuming that inferrables require their referent to be entailed by information already in the store, then they would still be compatible with a presuppositional theory of anaphoric reference.

these hierarchies meaning that when a construction with a null pronoun is used, the intended referent should be easily interpretable. Overt pronominals, which will be the central focus of this study, are typically next on these hierarchies. I take it that this ease of interpretability is directly related to the structure of the store. If the store is taken as a linear list of items, then the accessible items should be located at or near the head of the list, while less accessible items should be located further down the list. I will propose a more complex structure in Chapter 2 for the store, but for illustrative purposes in the present discussion, I will use a list structure.

While not logically implicated by this accessibility notion, one may hypothesize that if an entity is in a highly accessible position in the store, then hearers will expect that it *must* be referred to with a reduced referring expression such as a pronoun. For instance, consider the discourse in (13) which sounds particularly odd without some pronouns.

(13) a. John went to the supermarket.

b. #John/He bought some fish.

c. #John/He went home and ate dinner.

This hypothesis might be motivated by an understanding of pragmatic cooperation (Grice, 1975): Because the hearer expects a reduced referring expression to refer to a referent in a highly accessible position in the store, it would be regarded as uncooperative for a speaker to use an unreduced referring expression. Gordon et al. (1993) and Almor (1999) show that when a repeated name or definite description is used to refer to a salient entity in a discourse, certain negative consequences result (e.g., increased reading times). This hypothesis, if true, provides a useful way to ascertain the structure of the

store at a given point in the discourse: Discourse referents which are highly accessible should preferentially be referred to pronominally.

If the form of referring expression and position in the store are related in this way and if this is, in fact, motivated by pragmatic cooperation, then one would expect to find evidence of this in both production and perception. The studies by Gordon et al. (1993) and Almor (1999) cited above show this in a perceptual paradigm. Another study by Arnold (1998a) shows this in a production paradigm, observing that participants preferred to use pronouns to refer entities which had previously been focused. Given this evidence, this is one of the tools that will be used to diagnose which referents are perceived to be more accessible in the store. in the psycholinguistic and corpus linguistic investigations in Chapter 3 and Chapter 4, respectively.

1.1.5. Summary

In this section I have outlined some basic properties of discourse which I will be assuming throughout this dissertation. Namely, these properties are that a discourse is an intentionally arranged sequence of utterances in which each successive contribution to the utterance is interpreted with respect to the preceding context. Referring expressions in a discourse refer to discourse referents—mental representations of entities. A set of discourse referents is maintained in the store comprising all entites evoked in the discourse thus far and the interpretation of anaphoric expressions is made with respect to the items in the store and the structure imposed on them (e.g., accessibility). I have given an overview of DRT as a formal model of how discourse is processed and the sorts of constraints it imposes on referring expressions in a discourse, noting, however, that

the formalism does not spell out how pronoun resolution is to proceed and particularly how the processor would decide among competing candidates.

I intend to deal with this problem by detailing a model of pronoun resolution in which the items in the store are arranged with respect to their salience in the context. I present a formal model of salience below, but before that I wish to discuss the notion of interconnectedness in discourse in order to establish some of the background to discourse salience and to give better theoretical footing to some of the experimental predictions I will make.

1.2. Discourse Coherence

In order to define what I mean by discourse coherence, first I'll take a pre-theoretic point of view. Imagine I go into a local pub and strike up a conversation with a man who has already had several drinks too many. After a while I might begin to regard his speech as incoherent in his current drunken state. But what do we usually mean here by "incoherent"? We typically do not mean that his speech is ungrammatical—that he is making subject-verb agreement errors or the like. Rather, we usually mean something like he is not making sense or that we can't understand what he's talking about. In terms of the theory of discourse structure presented above we might say that, given the preceding context, there is no way to make sense of his current utterance—no way to incorporate it into the discourse representation. It is this conceptualization of coherence that I will adopt in this dissertation.

This characterization of discourse coherence is hardly new. Kehler (1995, 2002), following similar concepts presented in Halliday and Hasan (1976), Hobbs (1979), and Mann

and Thompson (1987), outlines a detailed model of discourse coherence in which coherence is principally determined from the *coherence relations* existing between sequential utterances in a text and that these relations come from a set of three fundamental relations (from Hume, 1748): *resemblance*, *contiguity*, and *cause and effect* as illustrated in (14).

(14) a. RESEMBLANCE Dick Gephardt organized rallies for Gore, and
Tom Daschle distributed pamphlets for him.

b. CONTIGUITY George picked up the speech. He began to read.

c. CAUSE-EFFECT George is a politician and therefore he's dishonest.

In some cases (e.g., (14c)), the particular coherence relation intended between two adjacent clauses is explicitly marked with a connective (e.g., *because*, *therefore*, *next*). However, in other cases (e.g., (14a-b)), this relation is implicit in the meaning of the two clauses. In the case of the items in (14), hearers are easily able to recover the respective relations. Kehler refers to the process of recovering these explicitly or implicitly marked relations as *coherence establishment*.

I will discuss these relations and Kehler's theory in more detail in Section 1.3.3. Here, I wish to note simply that the model I present shares the fundamental notion that discourse coherence is determined in terms of how the current utterance relates to the preceding utterances, its context. This is also consistent with the DRT formalism described above. My subsequent aim is to apply this notion of coherence to reference within discourse: The degree of ease with which referring expressions in the current utterance can be interpreted with respect to the preceding context influences (along with other factors) discourse coherence. In this dissertation I will be centrally concerned with

referential coherence in this sense, and particularly how it relates to the interpretation of pronominal referring expressions.

1.2.1. Centering Theory

One framework which is designed to model discourse coherence and on which much psycholinguistic and computational linguistic research is based is Centering Theory (CT: Grosz and Sidner, 1986; Grosz et al., 1995). The prominence of CT in psycholinguistic and computational linguistic research over the last two decades makes it a good reference point for the research I describe here. In particular, I would like to show how CT captures many of the important issues discussed so far, but at the same time lacks sophistication in its treatment of (what I call) discourse salience. I present CT here as partial motivation for the psycholinguistic experiments in Chapter 3 and the corpus analysis in Chapter 4.

CT captures the notion of referential coherence by looking at local discourse structure, particularly interutterance connections, called *centers*. All of the entities which are referred to in an utterance constitute the set of forward-looking centers, or Cf. This list is ranked according to syntactic role as shown in (15).

(15) subject > object(s) > others

The highest ranked member of the set of forward-looking centers is called the preferred center, Cp. Thus, in (16), there are three entities (or in my model, discourse referents) in Cf: JOHN, MATT, and A USED CAR. Given the hierarchy in (15), JOHN is ranked highest because it appears as the syntactic subject and is therefore the Cp of the utterance.

Table 1.1. Transitions in Centering Theory

	$Cb_{Utt_n} = Cb_{Utt_{n-1}}$ or $Cb_{Utt_{n-1}} = \emptyset$	$Cb_{Utt_n} \neq Cb_{Utt_{n-1}}$
$Cb_{Utt_n} = Cp_{Utt_n}$	CONTINUE	SMOOTH-SHIFT
$Cb_{Utt_n} \neq Cp_{Utt_n}$	RETAIN	ROUGH-SHIFT

(16) John sold Matt a used car.

$Cf = \{\textsc{John}, \textsc{Matt}, \textsc{a used car}\};\ Cp = \textsc{John}$

Every utterance also contains at most one backward-looking center, Cb, which is defined as the highest-ranking member of the Cf of the preceding utterance that is realized in the current utterance. Thus, adding a continuation utterance to (16) yields (17) in which the continuation has three centers: MATT, A USED CAR, and CALIFORNIA. MATT is the most highly ranked member of the Cf of (17a) which is realized in (17b) and therefore is the Cb of (17b).

(17) a. John sold Matt$_i$ a used car.

$Cf = \{\textsc{John}, \textsc{Matt}, \textsc{used car}\}; Cp = \textsc{John}$

b. He$_i$ drove it to California.

$Cf = \{\textsc{Matt}, \textsc{used car}, \textsc{California}\}; Cp = \textsc{Matt}; Cb = \textsc{Matt}$

With these definitions, then for any given utterance, there are two questions which can be asked of any sequence of utterances Utt_{n-1} and Utt_n as follows.

(18) a. Is the Cb of Utt_n also the Cp of Utt_n?

b. Is the Cb of Utt_n coreferent with the Cb of Utt_{n-1}?

Combining these two questions into their four logical permutations defines an exhaustive set of interutterance relations called *transitions* as shown in Table 1.1.

CT also defines two basic rules as follows.

(19) Rule 1: If any center in an utterance is realized as a pronoun, then the Cb of that utterance must be realized as a pronoun.

(20) Rule 2: CONTINUE > RETAIN > SMOOTH-SHIFT > ROUGH-SHIFT

While both of these rules are simply stipulated, they are intended to capture some important intuitions. Rule 1 is intended to capture the intuition that when there is any coherence ink across utterances, then there should be no ambiguity about which entity is participating in that link. Furthermore, the questions in (18) reveal some important insights for Rule 2: For each question, discourses for which the questions can be answered affirmatively are easier to read. Thus, the discourse in (21a) is easier than the one in (21b) because John is both Cb and Cp in the latter conjunct. Similarly, the discourse sequence (22a-b-c) is easier than (22a-b'-c) because the Cb (SUSAN) remains the same throughout.

(21) a. John$_i$ hit Matt and then he$_i$ was told by Bill to stop.

b. John$_i$ hit Matt and then Bill told him$_i$ to stop.

(22) a. Susan$_i$ met Nancy at the coffee shop.

b. She$_i$ bought a tall Latte.

b' She$_j$ had been waiting there for over an hour.

c. Then she$_i$ left almost immediately.

The questions are not equally weighted: Intuitively, it is more important that a discourse maintain the center of attention (i.e., Cb of Utt$_{n-1}$ = Cb of Utt$_n$) than that the Cb and Cp of an utterance coincide (i.e., Cb of Utt$_n$ = Cp of Utt$_n$). In short, the difference between the discourses in (22i) s more important than the differences between

the discourse in (21.) Given these preferences, the ranking shown as Rule 2 in (20) results.[5]

Now we can explain the relative incoherence of the discourses in (2)-(4) from the Introduction shown below as (23)-(25) respectively with centering annotation.

(23) a. Matt_i gave John_j a book_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = $\emptyset$

b. Then he_i told him_j to read it_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = MATT, CONTINUE

c. Later he_i asked him_j how he_j liked it_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = MATT, CONTINUE

(24) a. Matt_i gave John_j a book_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = $\emptyset$

b. Then he_j was told by him_i to read it_k.

Cf = {JOHN, MATT, BOOK}, Cp = JOHN, Cb = MATT, SMOOTH-SHIFT

c. Later he_i asked John_j how he_j liked it_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = JOHN, ROUGH-SHIFT

(25) a. Matt_i gave John_j a book_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = $\emptyset$

b. Then Matt_i told him_j to read it_k.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = MATT, CONTINUE

[5]See Kibble (2001); Kibble and Power (2004) for a proposed reformulation of Rule 2 in terms of independently motivated constraints and Beaver (2003) for a reformulation of Centering Theory in optimality theory.

c. Later Matt$_i$ asked him$_j$ how he$_j$ liked it$_k$.

Cf = {MATT, JOHN, BOOK}, Cp = MATT, Cb = MATT, CONTINUE

The discourse in (23) represents an optimal discourse under the CT paradigm: the transitions are always of the preferred CONTINUE type, and the Cbs are always realized as pronouns. The discourse in (24), on the other hand, is extremely marked because the transitions are of the less preferred types according to Rule 2—SMOOTH-SHIFT and ROUGH-SHIFT. Finally, the discourse in (25) is marked because it violates Rule 1: The Cbs of utterances (b) and (c) (i.e., MATT) are not realized as pronouns although other centers are (i.e., JOHN and BOOK). As such, (24) and (25) suffer from degraded coherence.

One crucial theoretical difference between CT and the model of discourse I present in this chapter and the next is that CT is highly localized, its view limited to an adjacent pair of utterances. As a result of this, the Cf list is not cumulative—representing only the centers in the current utterance. Thus, while centers do correspond to discourse referents in my model, the Cf list is not the same as the store. However, in practice, my model resembles many aspects of CT because the factors which determine discourse salience (to be discussed below) result in effects which are most easily observable locally. Therefore, I will occasionally relate the experimental results in Chapter 3 and Chapter 4 to CT.

CT has motivated no little amount of research and the main principles of the theory are now supported by a sizable body of empirical evidence.[6] However, some questions can be raised about the breadth of its applicability in its basic form. One issue to discuss here is the ranking schema for the set of forward-looking centers. In the CT paradigm, Cf

[6]See Grosz et al. (1995) for a collection of papers on many aspects of Centering Theory as well as Poesio et al. (2004) for a recent investigation of several aspects of the theory.

ranking in English occurs on the basis of syntactic role with discourse referents realized as syntactic subjects ranked highest. Given an optimally coherent discourse with respect to CT—that is, a discourse which exhibits only the more highly preferred transition types—then sentence-initial pronouns should preferentially be coreferent with the subject of the preceding utterance. Indeed there is much evidence to support this prediction (e.g., Gordon and Chan, 1995; Gordon et al., 1999; Hudson-D'Zmura and Tanenhaus, 1997; Lappin and Leass, 1994; Mathews and Chodorow, 1988; Poesio et al., 2004). In short, then, syntactic subjects can be seen as playing a special role in a discourse: discourse referents realized as subjects stand out above other referents.

A valid question to ask now is whether grammatical role alone is sufficient to determine which entity stands out in this manner. A number of CT researchers have already observed that other factors are necessary in other languages (e.g., Di Eugenio, 1997; Kameyama, 1985, 1986; Strube and Hahn, 1996; Turan, 1995) while Cote (1997) argues for an approach based on the lexical conceptual structures of Jackendoff (1990). In the next section I will discuss other factors which may be relevant in determining which entities in an utterance are highlighted as well as try to offer a broad, formal concept of what "highlighting" really means—a concept I will refer to as *discourse salience*.

1.3. Salience of Discourse Referents

Although I have already used the term salience numerous times so far, I would like to revert briefly to the pre-theoretic term I used in the introduction: "highlighting". When communicating some information to a hearer, we often highlight different things we wish the hearer to pay attention to, and then talk about those things. Some of this highlighting is done quite explicitly as in *You know John? Well, he...* But much

of this highlighting is done in less obvious ways and yet hearers seem to have little trouble attaching propositions to the intended entities (cf., information packaging as in Halliday and Hasan, 1976; Vallduví, 1990, inter alia). When such highlighting is done as the hearer expects and each respective contribution can be easily accommodated into the mental representation of the discourse, then that contribution will be taken as a coherent continuation of the discourse. But conversely, when the highlighting is inconclusive (i.e., it is unclear what is being highlighted) or subsequent contributions fail to fit with previous highlighting then those contributions will be taken as less coherent. My goal in this section is to give an overview of how this highlighting process is done, and conclude with further specification of the discourse model presented earlier.

Before delving into this highlighting process, I would like to settle some terminology. What I am calling here highlighting has been studied widely under a variety of names and theories. The characterization I've given above is more in pragmatic terms and the same or similar notions have been discussed in literature on "focusing" in discourse (Ward, 1985), "givenness" (Gundel et al., 1993), and "topic" (Givón, 1983b). A slightly different characterization may be made in more psycholinguistic terms with respect to cognitive structure and processing efficiency. In this literature, it has been called "prominence" (Gordon and Hendrick, 1997a,b) and "salience" (Arnold, 1998a). Finally, the computational literature has used such terms as "focusing" (Grosz, 1981; Sidner, 1981), "centering" (Grosz and Sidner, 1986; Grosz et al., 1995; Walker et al., 1997), and "salience" (Lappin and Leass, 1994). Many readers will have noticed that many of the references given above actually straddle more than one of the areas of pragmatics, psycholinguistics, and computational linguistics. This is because the various research efforts

on the notion of (again, what I here call) highlighting in these three fields feeds on each other, even if greater concensus on terminology has not emerged. In this dissertation, I will not be able to propose any sort of concensus. On the contrary, I will adopt and use two of these terms for my own purposes. However, I will remain true to the spirit of the basic notion that, at least in some sense, these terms are used to describe something which is somehow highlighted within the discourse.

In the remainder of this dissertation I will use just two terms as I discuss the factors involved in highlighting: *salience* and *prominence*. The use of any of the other terms above will be either in reference to other work which has made use of those terms, or in non-theoretic senses. I will use *salience* to refer to the overall degree to which something is highlighted within a particular discourse. I assume that several factors contribute to the overall salience of an item in discourse. I will use *prominence* to refer to the degree to which an item is more salient with respect to one of these factors. For instance, below I will discuss syntactic prominence as one component which contributes to overall discourse salience. There is, therefore some independence between prominence and salience: An item may be prominent with respect to one factor, but still not be particularly salient because other prominence factors contribute to the promotion of other items. Thus, each prominence factor imposes an ordering on the discourse referents in the store and salience is a net ordering which is computed from all of the prominence factors.

In the model I am describing here, referents are evoked in a manner such that a variety of prominence factors determine the degree of their salience in the immediate discourse. Subsequent reference to salient referents is then done in a way that differs from reference to less-salient (or new) referents. Integrating this notion of salience with

the model I've described so far, salience can be seen as a ranking on discourse referents in the store. The next important question to address then is precisely how salience is determined. In CT, as described above, salience is determined solely by grammatical role. However, there are many theories about how to determine discourse salience and a number of prominence factors other than grammatical role have been shown to influence the salience of discourse referents. In the following section, I will describe a number of these factors. First, though, I must describe some factors which do *not* contribute to discourse salience.

1.3.1. Constraints

Pronominal reference is restricted by certain morphosyntactic constraints as in (26)-(27).

(26) a. The doctor$_i$ put himself$_{i/*j}$ at the front of the list.

b. The teacher$_i$ put him$_{*i/j}$ at the front of the list.

c. He$_i$ thought that the man$_{*i/j}$ was at the front of the list.

(27) a. Mrs. Smith$_i$ put [*their$_i$/her$_i$] books on the desk.

b. The woman$_i$ put [*his$_i$/her$_i$] books on the desk.

c. The student borrowed a book$_i$ from the library. [*He$_i$/It$_i$] was very dusty.

The sentences in (26a-c) exemplify configurational constraints on coreference; in terms of the government and binding formalism introduced in Chomsky (1981) there are violations of conditions A, B, and C, respectively. (27a-c), on the other hand, illustrate morphological feature mismatches: NUMBER, GENDER, and HUMANNESS, respectively. In

short, binding conditions and feature-matching conditions place inviolable constraints on coreference.[7]

In the model of discourse processing I am describing in this chapter and the next, I will regard these factors as acting as a sort of morphosyntactic filter, a common characterization in many discussions of pronoun resolution algorithms (cf., Brennan et al., 1987; Hobbs, 1978; Kennedy and Boguraev, 1996; Lappin and Leass, 1994; Mitkov, 1994, 1996; Nasukawa, 1994) and psycholinguistic models (cf., Arnold et al., 2000; Boland et al., 1998; MacDonald and MacWhinney, 1990; Nicol and Swinney, 1989). Thus, these factors influence pronoun resolution by filtering referents which are subject to these processes—that is, filtering the store with respect to particular pronouns. From a processing point of view, then, it would be efficient for such a filter to apply early in pronoun resolution so that the later stages of the process will have fewer referents to work with (although see Badecker and Straub, 1994, for evidence that binding principles act in parallel with other factors in a constraint-satisfaction model rather than as an early filter).

1.3.2. Preferences

In constrast to the constraints discussed above there are a number of factors which serve as much looser constraints on pronominalization and pronoun resolution processes.

[7]Here it should be noted that *inviolable* is being used in a relative sense: these constraints can be violated but under much more limited conditions than the preferences discussed below. For instance, consider (i).

(i) John$_i$, my drag-queen friend, told me she$_i$ had to leave the party.

Even if it is perfectly clear to both the speaker and hearer that John is male, the use of the feminine pronoun may be felicitous. As such, it may be more accurate to regard such exceptions as "unconventional" rather than "ungrammatical". However, I do not think that either characterization affects the reality of the constraints I describe in this section and how they affect processing. For instance, it seems clear to me that resolving *she* to JOHN in (i) requires considerably more cognitive effort—something which might be explained if JOHN had initially been filtered out of the candidate list and had to be recovered later. (cf., Bock and Miller, 1991)

These factors represent certain preferences (cf., Kameyama, 1995; Mitkov, 1999) which contribute to greater or lesser salience of discourse referents and include such factors as recency, syntactic prominence, semantic prominence, syntactic parallelism, and semantic parallelism. Here I discuss each one of these in some detail.

1.3.2.1. Recency. In pronoun resolution algorithms, the surface distance between a candidate antecedent and a target pronoun has been shown to be one of the strongest factors (Lappin and Leass, 1994). This distance may be measured in different ways (i.e., clause distance, word distance), but nonetheless represents the recency of the last realization in the discourse of a particular discourse referent. While the distance between a pronoun and its antecedent certainly tends to be minimal, it is not impossible to have cases of quite long-distant coreference. Mitkov (1995) notes a case in which a pronoun and its antecedent are separated by 17 sentences. Such cases are obviously quite rare. A typical computational approach to quantifying the effect of recency is to use some sort of asymptotic decay algorithm based on the clausal distance between a pronoun and candidate antecedents: For instance, with respect to a given pronoun, the salience index of a candidate antecedent might be halved for each intervening clause (cf., Kennedy and Boguraev, 1996).

1.3.2.2. Syntactic Prominence. Syntactic position or role has been shown to influence both pronominalization and pronoun resolution processes. Syntactic prominence has been measured in two different ways: In some models, (e.g., Grosz et al., 1995) syntactic prominence has been determined by surface syntactic position using a prominence hierarchy as shown in (28). Much experimental evidence is consistent with this model in showing an overall preference for pronominal resolution to a discourse referent realized

as the grammatical subject of the preceding utterance (Gordon and Chan, 1995; Gordon et al., 1999; Hudson-D'Zmura and Tanenhaus, 1997; Lappin and Leass, 1994; Mathews and Chodorow, 1988).

(28) subject > object > oblique

It is important to note here that this hierarchy has been around for some time and is the foundation of many models of discourse structure and pronoun resolution (e.g., Centering Theory of Grosz and Sidner, 1986; Grosz et al., 1995), but the lower part of the hierarchy as not been explicitly evaluated.[8]

A somewhat different method of determining syntactic prominence was employed in the pronoun resolution algorithm of Hobbs (1978). In this model, the search for a compatible antecedent proceeds by searching the syntactic tree in a left-to-right, depth-first manner. For the most part, this would give results similar to that of the surface syntactic approach described above except that it would give preference to preposed elements. This, however, may not be an undesirable result: Gernsbacher and Hargreaves (1988) show in a series of probe recognition tasks that referents evoked in preposed noun phrases are more accessible than when realized in their canonical position.

Yet another way of determining syntactic prominence comes from looking at the linear surface structure of an utterance as determining the relative prominence with respect to the order in which entities are overtly evoked. This approach is taken in Gernsbacher and Hargreaves (1988) and is referred to by them as *order-of-mention*. They present evidence and argue for a model of discourse in which order-of-mention is

[8]The pronoun resolution algorithm of Lappin and Leass (1994) uses weightings consistent with (28) and therefore gives implicit support for the whole hierarchy.

the sole factor determining the salience of discourse referents, irrespective of syntactic role (e.g., subject, object) or semantic role (e.g., agent, patient).

It is important to note here that these various techniques make essentially the same predictions for utterances structured with canonical SVO order, simplex noun phrases, and little or no subordination of clauses. However, some differences arise as utterances become more complex or noncanonical. For example, consider the sentence in (29).

(29) [[John's] mother] met [[Bill's] mother] at the park.

(30) a. JOHN'S MOTHER > BILL'S MOTHER > { JOHN, BILL, PARK }

b. JOHN'S MOTHER > JOHN > BILL'S MOTHER > BILL > PARK

c. JOHN > JOHN'S MOTHER > BILL > BILL'S MOTHER > PARK

Five different entities are evoked in (29), but each of the three theories above makes different predictions about the relative syntactic prominence of their respective discourse referents. Under the syntactic prominence hierarchy shown in (28), the subject, JOHN'S MOTHER should be most prominent followed by the object, BILL'S MOTHER. The remaining entities, JOHN, BILL, and THE PARK, are ranked lowest, but remain unranked with respect to each other as shown in (30a). Under Hobbs' algorithm, however, a left-to-right, depth-first search will find JOHN'S MOTHER first, but turn up JOHN next, and so on to produce the ranking shown in (30b). Finally, order-of-mention proposes a different ordering as shown in (30c). It is interesting to note that all three of these methods give the same respective ordering for the three syntactic roles, subject, object, and oblique. The differences arise with the introduction of entities in embedded positions.

In Chapter 2, I will discuss syntactic prominence further and suggest two different ways of determining it. These two methods will then be evaluated in the corpus analysis in Chapter 4.

1.3.2.3. Semantic Prominence. Compared to syntactic prominence, there has been relatively little formal study of how semantic information may influence discourse salience of entities for pronoun resolution. Prat-Sala and Branigan (1999) observe that animate entities are preferred antecedents for pronominal reference over inanimate entities. Another series of studies (Stevenson et al., 1994; Stevenson, 1999; Stevenson et al., 2000) looks at the implicit causality of an event as determining the relative salience of the participants in that event.

In this dissertation, I will take a different approach to how semantic information may contribute to the discourse salience of entities. This will be described in detail in Chapter 2, but I give a brief preview here. My approach is motivated by the confound between syntactic and semantic prominence which was touched upon in the Introduction: subjects are often agents and objects are often patients and so on. My approach thus looks at semantic prominence as being determined by the semantic role which is assigned to an argument by the verb which heads the clause in which that argument is evoked. To my knowledge, this particular approach has not been taken formally in studies of discourse salience and pronoun resolution.

1.3.2.4. Syntactic Parallelism. There is some evidence in studies of pronominal resolution of a preference for syntactic parallelism between a pronoun and the last realization in the discourse of the entity it is resolved to (Chambers and Smyth, 1998; Crawley et al., 1990; Smyth, 1994; Smyth and Chambers, 1996; Stevenson et al., 1995). This

effect appears to be limited to cases where the verb frames in which the pronoun and its antecedent are realized share the same basic syntactic structure. Furthermore, Kehler (2002) argues that connectives (e.g., *because*, *so*, *and then*) also function to support or cancel the effect of syntactic parallelism. Consider the sentences in (31) (from Chambers and Smyth, 1998, their (1) and (2), respectively).

(31) a. Josh criticized Paul and then he insulted Marie. (*he*=Josh)

b. Josh criticized Paul and then Marie insulted him. (*him*=Paul)

In (31a), the pronoun in the second clause, which is in subject position, is preferentially interpreted as coreferent with the referent realized as the subject of the first clause, JOSH. However, in (31b), the pronoun, which is in object position, is preferentially interpreted as coreferent with the referent realized as the object of the first clause, PAUL.

1.3.2.5. Semantic Parallelism. Semantic parallelism—when a pronoun and its antecedent are realized in parallel semantic roles—has also been hypothesized to be a significant factor in anaphoric phenomena as illustrated in (32)-(33) (from Mitkov, 1999). Under semantic parallelism, it is preferred that the pronoun in (32b) is interpreted as coreferent with SODY since they both serve as semantic goals in their respective sentences. Similarly, it is preferred that pronoun in (33b) is interpreted as coreferent with VINCENT since they both serve as semantic agents.

(32) a. Vincent gave the diskette to Sody$_i$.

b. Kim also give him$_i$ a letter.

(33) a. Vincent gave the diskette to Sody$_i$.

b. He$_i$ also give Kim a letter.

Mitkov (1999) claims that semantic parallelism is a more important factor than syntactic parallelism but offers no argumentation to back this up. While there is some research on the relevance of semantic parallelism in such phenomena as verb-phrase ellipsis (cf., Hobbs and Kehler, 1997), I have not found any study linking semantic parallelism and pronouns.

1.3.3. Coherence Relations

The soft constraints described in the sections above form a somewhat curious set in that in many respects they would seem to compete against each other. For instance, consider the abstract discourse shown in (34) in which the syntactic and symantic features of each linguistic expression are indicated by subscripting.

(34) a. $x_{\{subject,agent\}}...y_{\{object,patient\}}$

b. $z_{\{subject,agent\}}...pronoun_{\{object,patient\}}$

When trying to resolve the pronoun in (34b), syntactic prominence would give preference to X while recency and syntactic and semantic parallelism would give preference to Y. Kehler (2002) further suggests that the complexity of trying to unite these different factors into one system stands in stark contrast to the economy pronouns are often taken to represent. He proposes a model of coherence relations among adjacent utterances that accounts for much of the data in a straightforward manner. As discussed above, he uses three basic coherence relations in his model—CAUSE-EFFECT, RESEMBLANCE, and CONTIGUITY—as shown in (14), repeated here as (35).

(35) a. RESEMBLANCE Dick Gephardt organized rallies for Gore, and

Tom Daschle distributed pamphlets for him.

b. CONTIGUITY George picked up the speech. He began to read.

c. CAUSE-EFFECT George is a politician and therefore he's dishonest.

To illustrate the model, consider (36) ((367) in Kehler, 2002).

(36) Margaret Thatcher admires Hillary Clinton, and George W. Bush absolutely worships her.

(37) a. $admires(Thatcher, Clinton)$

b. $worships(Bush, her)$

The pronoun *her* in (36) is potentially ambiguous: it could conceivably refer to either Thatcher or Clinton, but intuition seems to lean toward interpreting it as referring to Clinton (in spite of the fact that our world knowledge of these figures suggests otherwise). Looking closely at the similarity between the propositions given in the two conjuncts (as shown in (37)), it is easy to see that the clauses are related to one another by a RESEMBLANCE relation. Therefore, as the hearer constructs a mental model of this utterance, it is preferable to interpret the pronoun as referring to Clinton because that sets up a coherent parallel between the two clauses. In this manner, Kehler argues, pronoun interpretation proceeds without specific reference to heuristics of syntactic prominence and so on.

As a model of pronoun interpretation, Kehler's concept of coherence relations is qualitatively different from that of the other factors described above. While the other factors can be seen as promoting some candidate discourse referents for the purpose of pronoun resolution, assignment with respect to coherence relations falls out naturally from the process of coherence establishment among utterances. I admire the simplicity of the model, but I am less confident about some of motivations for it. The argument that

a complex algorithm of prominence factors and preferences somehow belies the relative economy pronouns are taken to represent is an important point and one that needs to be answered by those who propose such algorithms. However, the fact remains that even in Kehler's system, syntactic and semantic information must be fully determined before coherence establishment, and thence, pronoun interpretation, can take place. Hence, Kehler's system still presupposes much of the same computational burden.

For the psycholinguistic experiments and corpus analysis described in Chapter 3 and Chapter 4, respectively, I will be using an algorithmic model which will be laid out in Chapter 2. However, the evidence that the type of coherence relation between two clauses affects pronoun interpretation is not ignored: The stimuli used in the psycholinguistic experiments controls for these coherence relations by using only CONTIGUITY relations.

1.3.4. Resolution Algorithms: Putting the Factors Together

In order to determine the overall salience of candidate antecedents, and thus, the relative ranking of these candidates, the influence of the several prominence factors described above must be combined under some algorithm. In Lappin and Leass (1994), for each candidate, a numerical index is calculated from each factor and then the sum of these indices constitutes the overall salience index for that candidate. On the other hand, Arnold et al. (2000) propose a constraint-based system in which candidates are evaluated with respect to a series of constraints based on factors such as those above and the optimal candidate is then proposed as the antecedent (see Bouma, 2003, for an implementation of this in Dutch). In Chapter 2, I will outline an abstract method which could be implemented in a variety of different computational models. In this method, the various

factors are seen as imposing independent orderings on the discourse referents. Overall discourse salience can then be determined with respect to how the orderings align (or do not align).

Another question to be explored in a model of discourse salience is whether salience can be seen as a finite, closed system in which the total amount of salience allotted to discourse referents remains constant, or whether total salience may vary anywhere from zero to infinity. When the salience of one discourse referent increases, is an equal amount of salience taken away from other entities in order to maintain equilibrium? Or rather, is it possible to just keep increasing the salience of an entity indefinitely, while the salience of competitors remains unchanged? For instance, consider the salience of the discourse referent associated with John at the end of the vignettes in (38)-(39).

(38) a. John participated in his first boxing match at the gym.

b. He was excited about the match and really wanted to win.

c. He hit his opponent many times in the face and arms.

d. He gave him a bloody nose, too.

e. Then, he knocked him out in the third round.

f. He couldn't believe it.

(39) a. John knocked his opponent out in the third round.

b. He couldn't believe it.

Structurally, the last two sentences of the two vignettes are the same. However, for me, there is a difference in the interpretations available for the last sentence. In (38f), I find it only plausible that the pronoun *he* refer to John. However, in (39b), although I still prefer John, I find it much easier to entertain the interpretation that the pronoun

refers to John's opponent. This might be explained by suggesting that as the discourse in (38) progresses, each successive realization of John as an agent-subject boosts the overall salience of the referent JOHN such that by the end of the discourse, it far outranks any other referent. Hypothetically, the discourse could continue in the same manner and the salience of JOHN would continue to increase indefinitely.

While this question is interesting and has some implications for discourse processing, the model of discourse salience and semantic prominence I outline in Chapter 2 does not decide between these two possibilities and remains adaptable to either position.

1.3.5. Summary

In this section, I have introduced a number of different factors which have been proposed in the literature as contributing to discourse salience for pronoun resolution. Here, a brief word on methodology in investigating these factors is warranted. It is important to note that it is difficult to observe any single one of these factors in isolation. That is, it is nearly impossible to vary one of the factors without also varying another factor. For instance, consider the sentences discussed in the Introduction, reproduced here as (40a).

(40) a. John hit Matt. He ...

b. Matt was hit by John. He ...

For the pronoun in the continuation sentence of (40a), there are two potential antecedents: JOHN and MATT. Experimental evidence shows that there is preference for the pronoun to refer to the referent introduced in subject position, JOHN. However, this is not enough evidence to conclude that syntactic prominence is the primary reason for this. JOHN is realized as a syntactic subject, but also as a semantic agent. So, it could

be the case that some notion of semantic prominence is the primary driving force. I will discuss this concept in much greater detail in Chapter 2, but for the present discussion it is important to see that syntactic and semantic prominence are (often) conflated. One way around this is to contrast sentences like those in (40a) with their passive counterparts as in (40b). If it could be shown that there is greater preference for the pronoun to refer to JOHN in (40a) but to MATT in (40b), that would be a stronger argument in favor of syntactic prominence. However, if the preference is for the pronoun to refer to JOHN in both cases, then semantic prominence would be a better explanation. It should be pointed out, though, that we still have not succeeded in viewing either syntactic prominence or semantic prominence in isolation. Rather, the best that can be done here is to observe *both* of them and compare their relative influence. This is the methodological approach which will be employed in the psycholinguistic experiments in Chapter 3 and the corpus analysis in Chapter 4.

1.4. Conclusion

In this chapter, I have described a basic model of discourse structure and representation in which the current state of the discourse is represented in terms of two data structures: a store comprising the discourse referents evoked in the discourse so far, and a set of assertions about those referents. I have given a somewhat detailed model of discourse salience in which the salience of a discourse referent for subsequent pronoun resolution is determined in some sort of cumulative fashion over several factors. This model, as described, is general enough to suit most psycholinguistic models of pronoun resolution or computational implementations thereof. I also hypothesize that the model,

as described so far, is fully compatible with both the speaker's task in language production as well as the hearer's task in language perception.

In the next chapter, I lay out a detailed model of discourse salience based on two factors: syntactic and semantic prominence. I also show two different methods for determining each of these prominence factors and give explicit illustrations of the discourse salience model.

CHAPTER 2

Discourse Salience and Syntactic and Semantic Prominence

2.1. Introduction

A variety of factors have been shown to contribute to the highlighting of entities in a discourse—what I am calling discourse salience. One of the most well-investigated factors is that of syntactic prominence. Thus, in a discourse like (41) (repeated from (5) in the Introduction), there is an overall preference to resolve the pronoun in (41b) to a discourse referent realized as the subject or first-mentioned entity of the preceding utterance; here, LUKE.

(41) a. Luke punched Max.

b. Then, he ran home.

This preference has been observed in a number of psycholinguistic investigations (Gordon and Chan, 1995; Gordon et al., 1999; Hudson-D'Zmura and Tanenhaus, 1997; Mathews and Chodorow, 1988) and is often encoded in one way or another in computational implementations of pronoun resolution algorithms (Kennedy and Boguraev, 1996; Lappin and Leass, 1994) or representations of discourse coherence (e.g., Centering Theory: Grosz and Sidner, 1986; Grosz et al., 1995).

However, as observed in the Introduction, syntactic information is often confounded with semantic information. That is, syntactic subjects are often semantic agents, while

syntactic objects are often semantic patients. Thus, it could be the case that the relative salience of syntactic subjects for pronoun resolution might be more accurately explained as resulting from the semantic prominence of agents. The next two chapters will investigate this possibility from a psycholinguistic perspective and a corpus linguistic perspective, respectively. In the present chapter, I will lay the groundwork for the investigation by laying out a detailed but flexible model of how the discourse salience of entities in an utterance may be determined. This is followed by a detailed description of both syntactic and semantic prominence and how each of these two may be measured within a given utterance. An implicit part of my model is an undertstaning of recency as a prominence factor. Therefore, I will briefly discuss it before the last two sections of the chapter. In those sections, I will illustrate the salience model with an implementation using syntactic and semantic prominence as factors and apply it to some sample discourses.

2.2. A Model of Discourse Salience

In this section I outline a generalized model of discourse processing with a pronoun resolution mechanism which relies on a detailed concept of discourse salience. This model will be general enough to permit implementation in most any mainstream theory of discourse processes. The remainder of the argumentation of this dissertation will be cast in terms of this model.

The mental representation of a discourse consists of two primary parts: a store of discourse referents and a list of semantic conditions on those referents (cf., the DRSs of DRT described in Section 1.1.3). I have little to say about the latter list of semantic conditions except that I assume the task of constructing these conditions is delegated to

the processor. For the present research, I am more concerned with the structure of the store and how it is used in discourse processing.

The set of discourse referents in the store represents a cumulative set of all entities evoked in the discourse thus far. These referents are partially ordered. For shorthand, this ordering is referred to as (discourse) salience and every referent in the store has some position (or at least some position relative to others) in this ordering; that is, every referent has some degree of salience and this ordering plays a role in how anaphoric reference is processed. I assume that for every non-anaphoric referring expression in an utterance, a new discourse referent is added to the store and its rankings with respect to a number of prominence factors are determined (this process is described in detail in the next section) and recorded with it in the store. However, for anaphoric expressions, slightly different operations will take place. Most importantly, the store will be searched for an existing compatible coreferent. Just how this search takes place and what is considered compatible depends on the kind of anaphor. For instance, a definite anaphor (simplifying greatly) will require search for a unique referent with a compatible (but more specific) semantic description. However, this search should skip the most salient referent in the current context—this referent is reserved for pronominal anaphoric reference. This is because, as discussed in Chapter 1, the default referent of a pronoun is the most salient (compatible) discourse referent in the store.

One question to ask here is how this model handles multiple pronominal reference as in (42) or certain parallelism effects as in (43).

(42) a. John$_i$ gave a book$_j$ to Matt$_k$.

b. He$_i$ told him$_k$ to read it$_j$.

(43) a. John$_i$ gave Matt$_j$ a book.

b. Mark gave him$_j$ a magazine.

My judgment of (42) is that the indicated indexing in the continuation sentence constitutes the most natural reading of the discourse. If so, then this example is explained in a relatively straightforward manner. When the pronoun at the beginning of the continuation sentence is encountered, the store contains three referents—JOHN, BOOK, MATT—which are ranked (I assume) in that order.[1] After the morphosyntactic filter is applied, only two compatible referents for the pronoun will remain: JOHN and MATT. Since JOHN is the most salient compatible referent, the pronoun will be resolved to JOHN. When the processor reaches the next pronoun, *him*, I assume that the morphosyntactic filter will return only one entity: MATT. JOHN will be filtered out of the list because resolution to JOHN would create a Condition B violation with the subject pronoun in the same clause. Hence, the pronoun will be resolved to MATT. Finally, the last pronoun, *it*, will be correctly resolved to the only compatible referent in the discourse: BOOK. Hence, the model here assumes a relatively simple processor which proceeds in a left-to-right manner, resolving each pronoun to the most salient compatible referent in the store.

The second discourse above suggests that this rather simple processor may be too simple. According to this processor, the pronoun in (43) should be resolved to the more salient referent, JOHN. However, my intuition is that the more natural referent is MATT (as indexed). This exhibits what is often called "syntactic (or arguably, semantic) parallelism" (see Section 1.3.2.4, above). In order to account for these sorts of cases, the

[1]While I think it is relatively uncontroversial that JOHN is the highest ranked referent, there may be some question as to the relative ranking of BOOK and MATT. However, for the present purposes, this will not be relevant because MATT and BOOK will never have to compete with each other for pronominal reference resolution in the examples discussed.

processor will have to be somewhat smarter, or at least have access to more information: for instance, the processor would need to know the syntactic role in which a particular referent was last realized as well as the syntactic role of the current pronoun. The syntactic role of the current pronoun within its utterance should be immediately available. In addition, in my discourse salience model, syntactic information is partially recoverable because syntactic prominence ordering are encoded in the store. Hence, provided the processor can detect an instance of parallelism, then it should be possible to identify the appropriate referent.

2.2.1. Graphical Illustration of the Store

In this subsection I present a notational convention I will use throughout this dissertation to illustrate the store. This convention uses graph theory. A graph is a mathematical structure consisting of two parts: a set of vertices which are often represented as points in the graph space, and a set of arcs (also called edges) which connect pairs of vertices. In a directed graph, the arcs are further defined to extend from one vertex to another vertex (typically represented with an arrow). Thus, the directed graph in Figure 2.1 consists of a set of four vertices, {A, B, C, D} and an ordering relation consisting of four orderings, $\{a > b, b > c, b > d, c > d\}$ (where the right caret ">" corresponds directed from the left operand to the right operand).[2]

In the discourse model, I assume that at the end of each utterance, the processor will enter the new discourse referents into the store and organize them along with existing discourse referents with respect to some orderings. These orderings are determined with

[2]Graph theory also allows for looping arcs which begin and end on the same vertex. However, the model of discourse salience I am outlining in this chapter assumes that the orderings imposed by prominence factors are irreflexive. Thus, the graphical illustrations will never contain such loops.

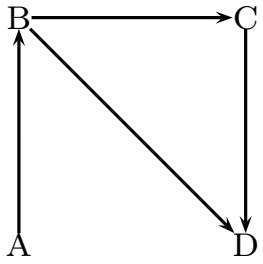


Figure 2.1. A sample directed graph with four vertices and four directed arcs.

respect to one or more prominence factors as described in Chapter 1. For expository purposes throughout this dissertation, I will illustrate this in directed graphs in which the vertices are the discourse referents, and directed arcs indicate orderings. If discourse referents X and Y are ordered with respect to some prominence factor as $X > Y$, then in a graphical illustration there will be a directed arc extending from X to Y.

The ordering relations imposed by the prominence factors are taken to be transitive in my model. Thus, for referents X, Y, and Z, if $X > Y$ and $Y > Z$, then $X > Z$. However, for the sake of simplicity, orderings which are derived by transitivity will not be illustrated in the graphs.

One final comment is in order here. Throughout the remainder of this dissertation I will be using the graphs as a convenient way of illustrating the contents of the store—the discourse referents and all of the prominence relations among them—in a digestible manner. However, I will not present any evidence that crucially depends on graph theory for explanation. In fact, all of the results and conclusions I make here could be explained using basic arithmetic operations. Furthermore, I see the discourse salience model presented in this chapter as fully adaptable to a variety of other computational

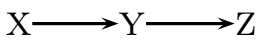


Figure 2.2. A sample illustration of the store of discourse referents. One prominence factor determines a ranking of the referents, illustrated by the directed arcs. The referent at the tail of an arc is more prominent than the referent at the head.

techniques whether it be graph theory, optimality theory, or other non-linear approaches. Nevertheless, graphical theory provides the best visual illustration of the model I have yet found. I therefore employ it here for its expository value.

2.2.2. Filtering and Salience Operations

In this section, I present and illustrate a number of abstract examples of discourse salience models as a means of introducing some other important features of the generalized discourse salience model—specifically, some operations over the store.

First let's consider a very simple model of discourse salience in which only one factor is relevant: X PROMINENCE. Let's assume that at the end of a particular utterance which introduces X, Y, and Z, X determines a prominence ranking where X > Y > Z. This relationship can be illustrated as shown in Figure 2.2. For this example, solid lines represent the X arcs.

I should note here that the store does not directly encode the discourse salience hierarchy of the current utterance, although it is directly determinable from it. Rather, I assume that there are two operations which may be applied to a store. First is a filtering operation, Filter(), which takes a copy of the store and returns it after removing discourse referents which are incompatible with the current pronoun (because of morphosyntactic

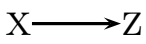


Figure 2.3. A modified copy of the store illustrated in Figure 2.2 after morphosyntactic filtering has removed one referent.

reasons) but preserving the orderings imposed by all prominence factors via transitivity. This filtered copy is then an input to the second operation, Salience(), which returns a set containing exactly one referent corresponding to the maximal referent, or if no maximal referent can be determined, then Salience() returns the empty set, $\emptyset$. The maximal referent is defined as the referent X for which there is no referent Y such that according to some prominence factor, Y > X. It is important to note that both of these operations work on a copy of the store, not the store itself. Hence, the store is left fully intact for other possible operations.

Continuing with the example shown in Figure 2.2, let's assume the next utterance contains a pronoun which needs to be resolved, and that the pronoun is morphologically compatible (in number, gender, animacy) with only discourse referents X and Z. The filtering operation should then take the store in Figure 2.2, filter out Y and return the store (copy) shown in Figure 2.3 in which the filtering operation has preserved the prominence relationship between X and Z with respect to X.

The second operation, Salience(), takes the filtered store and returns the maximal discourse referent, if one can be determined. In this case it is X. Thus, X is returned as the most salient (compatible) referent with respect to the current pronoun. The processor thus assigns X as the interpretation of the pronoun. This I take as the default procedure for pronoun resolution. Of course, if this initial assignment turns out to be

Figure 2.4. An illustration of the store after an utterance in which the hierarchical prominence relations among the referents are aligned. Solid lines represent the ordering due to one factor and dashed lines represent the orderings due to the second factor.

incompatible with, say, world knowledge, then some additional cognitive effort must be expended to recover the hierarchy and resolve to the next most salient referent.

Now let's consider a somewhat more elaborate model of discourse salience in which two prominence factors are relevant: X and Y PROMINENCE. Again, let's assume the context contains three discourse referents, X, Y, and Z. Further, let's assume that the prominence hierarchies are aligned such that both hierarchies impose a prominence ranking where X > Y > Z. The graphical illustration of the store would then be as shown in Figure 2.4 in which the solid arcs represent the X relations and the dashed arcs represent the Y relations.

Now, if a pronoun is encountered in the subsequent utterance which is not compatible with Y, then the filtering operation should return a modified copy of the store as illustrated in Figure 2.5 because transitive closure entails $x >_{\mathrm{X}} z$ and $x >_{\mathrm{Y}} z$. Then, when the salience operation is applied to the filtered store, it should be able to determine the maximal discourse referent: X. And, as before, the interpretation of the pronoun will be assigned to this referent.

Lastly, let's consider one more example using the same model of discourse salience with two prominence factors, but imagine that the utterance is one in which the two prominence factors impose different rankings on the discourse referents: in short, things

Figure 2.5. An illustration of the store shown in Figure 2.4 after filtering has removed one referent. Note that because the ordering relations are transitive, a prominence relation among X and Z is determined and represented after filtering.

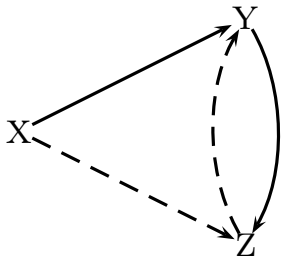


Figure 2.6. An illustration of the store after an utterance in which the prominence relations are not aligned, determining a different ordering of the referents.

are not so neatly aligned as before. According to X, the referents are ranked X > Y > Z, but according to Y, the ranking is X > Z > Y. This is illustrated in Figure 2.6.

Now, if the pronoun in the next utterance is incompatible with X, then the filtering operation should return a modified copy of the store as illustrated in Figure 2.7. However, this time, the salience operation will be unable to determine a maximal referent in the store and will therefore return the empty set, $\emptyset$. In short the system will be unable to propose a default referent for the pronoun, and other strategies will have to be used to determine an interpretation.

This model predicts two different conditions should exhibit some sort of cognitive processing load: (1) when a pronoun is used to refer to a referent which is not the most

Figure 2.7. An illustration of the store shown in Figure 2.6 after filtering has removed one referent.

salient compatible referent, and (2) when the salience operation cannot propose a single referent for a pronoun (i.e., returns $\emptyset$). I will make the case in Chapter 3 that the results of the psycholinguistic experiments therein exhibit both of these cases and verify these predictions.

As it is, this mechanism for determining discourse salience is relatively discrete and the abstract examples I have given here are well-behaved for illustrative purposes. However, natural language is considerably more complex and the model may need further refinement to fully capture the kinds of prominence relationships and salience effects seen in naturally occurring data. One method of achieving a more representative system might be to assign weights to the prominence orderings so that "net" precedence can be more precisely determined (i.e., in an additive/subtractive manner). This possible refinement is discussed in further detail in Chapter 5 as a possible extension of the corpus analysis.

2.3. Syntactic Prominence

As discussed in Section 1.3.2.2, several ways of determining the syntactic prominence of discourse referents have been used in different studies. Typical methods include ranking based on a syntactic role hierarchy (e.g., as in Centering Theory; Grosz and Sidner, 1986; Grosz et al., 1995), a left-to-right, breadth-first tree search algorithm (cf., Hobbs, 1978; Tetreault, 2001), and order-of-mention (Gernsbacher and Hargreaves, 1988). In this section, I will present two different methods derived from these three basic approaches. These methods will be used throughout the remainder of the dissertation. In short, the first method makes use of syntactic role information while the second combines a tree-search algorithm with order-of-mention. Both of these methods begin with the clause as a central frame of reference. I regard a clause as a syntactic unit containing at most one verb (or copula+adjectival predicate) as its head, and zero or more noun phrases which serve as direct logical arguments of the predicate. A clause may contain other noun phrases which serve to modify the predicate in some way, but are not arguments of the predicate. These include noun phrases within time or manner adverbials as well as locative expressions. These relationships are illustrated in (44), where clause boundaries are indicated by {curly braces}, verbs are in italics, and noun phrases are indicated by [brackets].

(44) a. {[John] *hit* [Matt]}.

b. {[John] *hit* [Matt] with [a stick]}.

c. {[John] *hit* [Matt] by [the river]}.

d. {[The man {[I] *saw* [∅] at [the park] [the other day]}] *hit* [Matt]}.

e. {[The book {[I] *read* [∅]}] *is bigger* than [the book {[I] didn't *read* [∅]}]}.

f. {[[John] *believes* {[Matt] *hit* [Robert]}}.

(44a) illustrates a simple case of one clause headed by the verb *hit* and having two logical arguments. (44b) shows a variation of this with an additional argument, the instrument used in the hitting event. (44c) shows a case where one of the noun phrases in the clause is not an argument, but merely a locative expression indicating where the hitting event occurred. (44d-f) show more complex sentences with embedded clauses. In (44d-e), these clauses are embedded inside noun phrases, but still are full-fledged clauses in their own right. (44f) represents a case in which the embedded clause actually serves as one of the direct arguments of *believe*. However, in the experimentation that follows, I will be concerned only with the discourse salience of discourse referents introduced by noun phrases and not those introduced by clauses (i.e., event reference). Thus, for the purposes of simplification, I will treat clauses like the *believe*-clause in (44f) as containing one argument.

I should also remind the reader here that although I will be talking below about the relative prominence of, say, subjects and objects for example, what is really at stake here is determining the syntactic prominence of discourse referents evoked by noun phrases occuring in those syntactic positions. So, throughout this discussion, I will mean an assertion like "subjects are more prominent than objects" to be shorthand for "discourse referents evoked by noun phrases in subject position are more prominent than discourse referents evoked by noun phrases in object position."

2.3.1. Role-based Method

The first method I present here uses the syntactic hierarchy shown in (45).

(45) subject > object > oblique

In English, every clause contains zero or more noun phrases which take on these roles with respect to the verb which heads the clause. Thus in the role-based method, a pre-verbal argument noun phrase which is not immediately preceded by a preposition is regarded as having the subject role. A post-verbal argument noun phrase which is not immediately preceded by a preposition is regarded as having an object role, while any other argument noun phrase which is immediately preceded by a preposition is regarded as having an oblique role. As noted above, any other noun-phrases are taken as non-arguments and are therefore considered as having the none status.

In the case of an utterance with just one main clause and no embedded clauses, the hierarchy in (45) provides a partial ordering of all the arguments (a full ordering may be obtained by assuming that ties resulting from equally-ranked arguments such as those in double-object constructions or clauses with multiple obliques are resolved by a rule of left-to-right precedence). This is the approach taken in many implementations which use this hierarchy including Centering Theory. However, I have yet to find an explicit theory of what happens when there are embedded clauses. I will assume the following rules as a means of giving a partial ordering of all the argument noun phrases in an utterance.

(46) all subjects > all objects > all obliques

One practical decision that must be made here is what elements will be counted in this hierarchy. For instance, a sentence like "{[John] told [Matt] {[∅] to read [a book]}}", Matt has one overt realization as an object in the matrix clause, but also a covert

realization as the subject of the embedded clause. I think it is uncontroversial that the overt realization should count towards the ultimate syntactic prominence of MATT. So, the question becomes whether to count the covert realization. For the sentence above, this question is crucial because if we do count covert elements, then the embedded subject realization of MATT will promote it to be on a par with the matrix subject realization of JOHN. If it is not counted, then it will merely be on a par with the embedded object, BOOK. My intuition is that JOHN is more prominent than MATT.[3] Hence, I will only count overt syntactic realizations in the computation of the syntactic prominence of discourse referents.

Another motivation for this decision comes from my own earlier work (Rose, 2002). In that experiment I looked at the relative salience of entities introduced in NP-raising constructions (e.g., *certain*, *likely*). A comparison of the salience of noun phrases introduced in NP-raising constructions ("Nancy is certain ∅ to beat Susan") to comparable non-raising constructions ("Nancy will certainly beat Susan") showed that there was no significant advantage provided to the matrix subject by its additional covert realization in the embedded clause.

In order to illustrate how this method works, the sample sentences given in (44) are repeated below as (47) with the salience ranking of their respective discourse referents shown.

(47) a. {[John] *hit* [Matt]}.

{JOHN} > {MATT}

[3]In my judgment, it is further the case that MATT is more prominent than BOOK, but that may be due to the fact that animate entities are generally more salient than inanimate entities, cf., Prat-Sala and Branigan (1999).

b. {[John] *hit* [Matt] with [a stick]}.

{JOHN} > {MATT} > {A STICK}

c. {[John] *hit* [Matt] by [the river]}.

{JOHN} > {MATT}

d. {[The man {[I] *saw* [∅] at [the park] [the other day]}] *hit* [Matt]}.

{THE MAN, I} > {THE PARK, MATT}

e. {[The book {[I] *read* [∅]}] *is bigger* than [the book {[I] didn't *read* [∅]}]}.

{THE BOOK_1, I, THE BOOK_2}

f. {[[John] *believes* {[Matt] *hit* [Robert]}}.

{JOHN, MATT} > {ROBERT}

One might question why I've chosen to rank *all* subjects higher than *all* objects and so on. This has the result, of course that an embedded subject may be more syntactically prominent than a matrix object. I have one primary motivation for this. Methodologically, the obvious alternative—ranking elements in higher clauses before elements in lower clauses—is, in practice, essentially the same as the Hierarchical Method to be discussed below. Thus, the current procedure provides a significant contrast for testability.

2.3.2. Hierarchical Method

The second method I present here combines a tree-search algorithm with left-to-right ordering and as a result gives a full ordering of the discourse referents evoked in an utterance. For the same motivations as discussed under Method 1, I will only count the overt syntactic realizations. Also, while this procedure counts only the direct arguments

of predicates, the particular syntactic role in which they are realized is not relevant. Hence, the procedure is as follows.

(48) a. For two NP arguments x and y, not necessarily arguments of the same verb, if y is in a more deeply embedded clause than x, then $x > y$.

b. For two NP arguments x and y, not necessarily arguments of the same verb, if x and y are at the same clause depth and x precedes y, then $x > y$.

The sample sentences used above are shown here again with the prominence rankings of their respective discourse referents shown according to this method.

(49) a. {[John] *hit* [Matt]}.

JOHN > MATT

b. {[John] *hit* [Matt] with [a stick]}.

JOHN > MATT > A STICK

c. {[John] *hit* [Matt] by [the river]}.

JOHN > MATT

d. {[The man {[I] *saw* [∅] at [the park] [the other day]}] *hit* [Matt]}.

THE MAN > MATT > I > THE PARK

e. {[The book {[I] *read* [∅]}] *is bigger* than [the book {[I] didn't *read* [∅]}]}.

THE BOOK_1 > THE BOOK_2 > I

f. {[[John] *believes* {[Matt] *hit* [Robert]}}.

JOHN > MATT > ROBERT

2.3.3. Further Comments on the Methods

I take these two methods to represent two near-extremes along a continuum of methods. The hierarchical method is at one end of the continuum where the procedure results in a full-ordering of all the discourse referents evoked in an utterance while the role-based method is near the other end of the continuum, giving only a partial ordering. I do not regard either of these approaches, however, as being fixed in stone. Either (or both) may require some tweaking to fit the data more accurately. For instance, it could be that the partial-order given by the role-based method is still more granular than necessary and that perhaps all that is really needed is to distinguish between subjects and non-subjects, or even simply between matrix subjects and others. The last would be more reminiscent of the approach often taken in many information packaging theories which fundamentally distinguish two elements in each utterance: THEME/RHEME (Firbas, 1964, 1966), TOPIC/COMMENT (Chomsky, 1965; Gundel, 1974; Chafe, 1976), TOPIC/FOCUS (Sgall, 1967; Büring, 1995), among others (see Kruijff-Korbayová and Steedman, 2003, for an overview of this area of the information structure literature).

2.4. Semantic Prominence

The approach I will be taking in this dissertation to determining the semantic prominence of entities is not one that I've seen formally implemented anywhere in the literature. So, as background, I'd like to give a very brief description of the approach followed by some discussion of the motivations for it. Then, the remaining subsections will discuss two particular implementations of the approach in detail.

2.4.1. Background

I view the semantic prominence of discourse referents in an utterance as being determined by the semantic roles (e.g., agent-like, patient-like, etc.) which a predicate assigns to its noun phrase arguments. I have both theoretical and practical motivations for this approach. There has been some work reported in the literature which shows that animate entities are more salient than inanimate entities. For instance, Prat-Sala and Branigan (1999) observed that experiment participants were more likely to produce syntactic structures in which entities made salient in preceding discourse preceded other entities. But, in particular, they found that this tendency was stronger when the salient entities were animate. Thus, given that the inherent animacy of entities may contribute to their overall discourse salience, it is reasonable to hypothesize that when a verbal predicate requires one or more of its arguments to be animate, then those arguments might become more salient in the discourse. For instance, consider the sentences in (50) (in which the hypothesis is represented by a question mark).

(50) a. [The child/The rock] rolled down the hill.

THE CHILD > THE HILL

b. [The child/#The rock] walked down the hill.

THE CHILD >>? THE HILL

The sentences show that *walk* requires an animate subject while *roll* does not. So, the hypothesis is that the relative salience of CHILD and HILL referents might be different between (50a) and (50b).

A similar type of motivation comes from a series of studies by Stevenson et al. (1994, 2000) and Stevenson (1999). In those experiments, the central observation is that certain

semantic roles were preferred by participants to form coherence links across utterances: patients were preferred to agents, experiencers were preferred to stimuluses, and so on. Stevenson et al. conclude that the implicit causality in an event determines the relative salience of the participants in that event. This is compatible with what I am suggesting so far: That is, the relative salience of discourse referents may be influenced by the semantic information which that predicate imposes on those referents.

One final motivation to mention here is a practical one. In applications which require some implementation of discourse salience, systems for determining the inherent animacy of noun phrases are not as well-developed as systems for which the semantic roles of verbs have been fleshed out. So the kind of approach I am recommending here may be more computationally feasible at present. Furthermore, it could capture the causality phenomena described above, yet allow other less-studied semantic information to contribute to discourse salience as necessary. A good example of this might be movement: participants which have undergone physical movement in an event (e.g., themes) might have a different salience than stationary entities.

In summary, I view the approach I take here to semantic prominence as a part of a larger attempt to capture some of the phenomena which have been observed previously (i.e., animacy, causality) in a system which is computationally feasible, and which at the same time opens doors to the inclusion of other semantic information (e.g., movement) as possible influences on the discourse salience of entities.

2.4.2. Verb Semantics

Below, I will present two different possible approaches to determining the semantic prominence of discourse referents. Both of these approaches are rooted, however, in the same basic tradition of the semantics of verbs. Hence, before describing the approaches, it will be useful to give some background discussion.

One of the theoretical challenges at the syntax-semantics interface is the development of a coherent and comprehensive explanation of the linking between predicate argument structure and surface syntax. Central evidence on which many prominent linking theories are founded includes such well-attested alternations as the active-passive alternation, locative alternation, double-object constructions, psych-verb alternations, and so on (cf., Baker, 1997; Dowty, 1991; Fillmore, 1968, 1976; Grimshaw, 1990; Gruber, 1965; Jackendoff, 1972, 1987, 1990; Levin, 1993; Rappaport and Levin, 1988; Speas, 1990; Tenny, 1994). In order to explain these alternations, most of these theories assume a set of discrete semantic (also called thematic) roles which identify the role a particular argument plays within the event denoted by a given predicate.

Consider the verb *bake* which takes two arguments. It may be represented as in (51). The notation (from Levin and Rappaport, 1986; Marantz, 1984) indicates that the agent is realized outside the verb phrase (i.e., subject position) and the arguments inside the verb phrase include the patient which is realized in object position (as indicated by underlining). Given this mapping from the semantic frame to the conceptual structure, (52) receives the expected interpretation in which John performs the baking action and the cake is the recipient of this action.

(51) BAKE: AGENT < <u>PATIENT</u> >

(52) John baked a cake.

In linking theories such as these, there are two key problems that must be solved. The first problem is what roles to include in the basic inventory of roles. There is no consensus on this problem. While most inventories tend to include some fairly obvious roles like agent, patient, theme, goal, instrument, and so on, there remains much variation in the actual makeup of the sets in terms of both which roles comprise the set as well as how many roles are in the set. Dowty (1991) proposes a system in which there are just two roles (but participants may bear these roles to different degrees) while Ostler (1979) proposes a set of 48 roles.

There is plenty of variation across theories in the degree to which these roles are taken as psychologically real or even primitive. In the Case Grammar/Frame Semantics of Fillmore (1968, 1976), the various thematic roles which arguments bear are viewed as fundamental elements of the semantic *frame* in which they participate and are believed to be derived from a small set of primitive semantic concepts. On the other hand, Dowty (1991) views the roles—or as he calls them, PROTO-roles—as labels for flexible configurations of semantic entailments assigned to a particular argument within the predicate's domain. Other theories fall somewhere between these two extremes.

The second fundamental problem for a linking theory is how to rank these roles along a thematic hierarchy. This is crucial for showing how the semantic information maps onto the syntactic representation. Here, too, there is no consensus on the exact hierarchy, but the hierarchies do seem to share many relations. For instance, the thematic hierarchy shown in (53) is largely reflected in the hierachies proposed in Fillmore (1968),

Jackendoff (1972), Larson (1988), and Speas (1990) among others (see Speas, 1990, for a useful overview of thematic hierarchies).

(53) agent > patient > others

In this dissertation, I will assume that the semantic prominence of discourse referents is determined by the particular semantic roles assigned to each referent. The actual ranking procedures differ markedly between the two approaches, so I will reserve discussion of that for their respective sections, below. However, some discussion is necessary here about which noun phrases will count in these procedures. In the procedures for determining syntactic prominence discussed above, I took the option that only overt noun phrase realizations count toward determining the prominence of their respective discourse referents. However, for semantic prominence, I will assume that *all* noun phrases which evoke a discourse referent (i.e., excluding expletives, etc.), overt or null, may count toward determining the prominence of their respective discourse referents.

This differing approach is motivate by two different reasons. First, in a sentence such as *John wants to go to Spain*, JOHN can be seen as taking on two semantic roles, one as somebody who desires a particular state of affairs, and one as a (hypothetical) traveler. It is my intuition that both of these roles are salient in the discourse, and not merely the role which also coincides with the overt expression of JOHN.

The second reason for counting all the semantic roles taken by a discourse referent is methodological in nature. Consider a sentence in which there are multiple overt and covert references to a single discourse referent. Existing evidence suggests that a hierarchy of syntactic roles exists, so subsequent analysis of syntactic prominence may be readily simplified by counting only the realization which is highest on the hierarchy.

However, for semantic prominence, although there are some existing semantic hierarchies, they do not provide comprehensive coverage of the roles/entailments employed below. Thus, a comparable simplification cannot be made. Counting all the semantic roles invoked will therefore raise the likelihood of observing which roles are truly more prominent than others. Although, admittedly, this procedure is likely to introduce more noise into the analyses.

In summary, while both Frame Semantics and Dowtian PROTO-role theories make use of semantic roles to explain the link between predicate argument structure and surface syntax, the contrast in their approaches is interesting and provides two useful contexts in which to test the effect of semantic prominence. The following two sections give more background on the two different approaches and how they will be implemented in this dissertation.

I would like to note up front that the research I am reporting in this dissertation is not intended to evaluate the theoretical adequacy of these two approaches as models of argument-linking. However, it may turn out that, with respect to determining semantic prominence, one or the other approach has the benefit of greater computational simplicity or explanatory power. After reporting on the results of the psycholinguistic experiments in Chapter 3 and the corpus analysis in Chapter 4, I will discuss this question more in Chapter 5.

2.5. Recency

I assume that the store contains a set of all the referents evoked in the discourse thus far. In each of the examples I have used up to now, this has been a fairly small set because the discourses have consisted of only two sentences, in which the first sentence

determines the context. However, in reality, the context is sure to be much larger and the store is sure to contain many more referents. recency was described by Lappin and Leass (1994) as one of the most important factors in their pronoun resolution procedure. In the present model, I assume therefore that recency is also a factor which imposes an ordering on the referents with respect to the recency of the utterance in which they were evoked. Thus, if referent X was introduced in utterance i, and referent Y was introduced in utterance j, and j is more recent than i, then Y is more prominent than X with respect to recency (i.e., Y $>_{recency}$ X).

Given this arrangement, one might wonder whether there is any point in worrying about older referents since they will always be less prominent (with respect to recency) than the current referents. One place where I see this making a difference is when the filtering operation returns a modified copy of the store in which none of the current referents remain, and recency (perhaps along with other factors not incorporated account here) must be relied upon to determine which referent is most salient.

However, in the remainder of this dissertation, I will look only at local coreference, effectively ignoring all referents "older" than one utterance. Therefore, for the sake of simplicity, I will not include these relations in the discussions or graphical illustrations. However, future study should include it as part of a full-featured model of discourse salience.

2.6. Implementation 1 Using Frame Semantics

Case grammar (Fillmore, 1968) is founded on the principle that both grammatical relations and meaning are governed by what Fillmore calls "deep case" roles which are taken as "a set of universal, presumably innate concepts which identify certain types of

judgments human beings are capable of making about the events that are going on around them" (p. 24). He proposes a minimum set of cases including AGENTIVE, INSTRUMENTAL, DATIVE, FACTITIVE, LOCATIVE, and OBJECTIVE, and suggests the possible existence of others. These roles apply to the participants in an event in a systematic way. For example, consider the verb *cook* for which the semantic entry would determinet that it has an AGENTIVE and an OBJECTIVE participant. This single entry could then be referenced to account for the fact that the sentences in (54a-c) can be used to describe the same event while syntactic transformational procedures would account for the surface variations.

(54) a. Mother is cooking the potatoes.

b. The potatoes are cooking.

c. Mother is cooking.

Frame Semantics (Fillmore, 1976, 1977) builds on case grammar by seeking greater generalizations in the semantic frames in which predicates operate. Instead of delineating for each predicate in the language the set of deep case roles which it uses, predicates are seen as members of conceptual frames which determine the participants in events within that frame. For instance, in a commercial transaction frame, there are minimally three participants—a seller, a buyer, and the goods transferred between them. A number of lexical items (in this case, both nouns and verbs) which invoke this frame include *buy*, *sell*, *purchase*, *rent*, *lease*, *vend*, *price*, *sale*, and *vendor*. These semantic frames are seen as the fundamental unit around which meaning is organized. Theories of this type have been particularly popular in artificial intelligence (e.g., schema theory (Minsky, 1975) and scripts in conceptual dependency theory (Schank and Rieger, 1974; Schank

and Abelson, 1977)) and especially in machine translation where the frames are seen as a convenient domain in which to establish the correspondence between texts in two different languages.

2.6.1. FrameNet

One practical implementation of frame semantics is embodied in the FrameNet project (Baker et al., 1998). In this ongoing project, the designers are building up a set of semantic frames with their respective case roles (called frame elements) by drawing generalizations from the syntactic contexts in which they appear in a corpus of texts. FrameNet II, released in 2001, includes 376 frames comprising approximately 500 unique frame elements. FrameNet takes advantage of a computational feature known as inheritance in its hierarchy of frames. The properties and features of abstract base frames are accessible to derived frames. For instance, as illustrated in Figure 2.8, the abstract frame TRANSITIVE_ACTION involving two frame elements, agent and patient, is inherited by INTENTIONALLY_AFFECT which adds the frame elements means and instrument. This frame is subsequently inherited by a variety of other more specific frames including ATTACK, CAUSE_MOTION, and HIRING. Frame inheritance is indicated in the graph by directed arrows (ancestor to descendant) between frames, and indices on frame elements show which elements have been passed on or inherited.

The design provides a nice means for representing the hierarchical relationships among a wide variety of both general and specific events. However, two comments are warranted. It appears that while the FrameNet system is, in theory, designed to take advantage of inheritance, this is not fully implemented as there are a number of frames

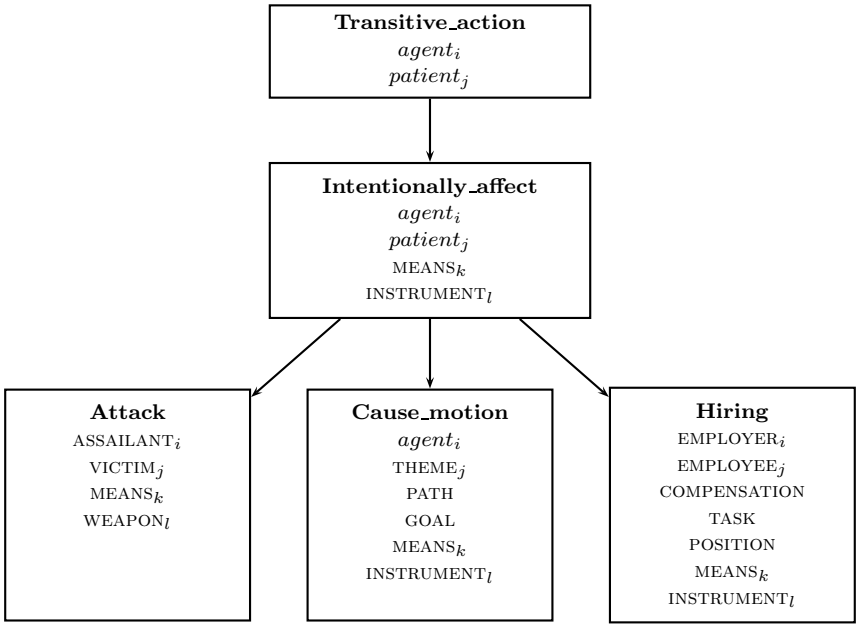


Figure 2.8. An illustration of inheritance in FrameNet II. Frames which are lower on the hierarchy inherit the frame elements of higher, more abstract frames. Element inheritance is illustrated here by subscripting.

which have not been fit into the overall hierarchy yet. Thus, there are a relatively large number of base frames and the hierarchy is not perhaps as deep as one would ultimately hope.

Furthermore, the large number of frame elements seems to betray the hope for a minimally small number of roles implicit in Fillmore (1968). That said, the FrameNet system is one of the more actively supported semantic networking projects, presumably with the hope that it will one day serve practical applications in natural language processing and artifical intelligence. Hence, I hope the present work will contribute to the

body of knowledge surrounding the FrameNet system by evaluating its usefulness in determining semantic prominence.

2.6.2. Illustration 1

In this section, I will give a detailed illustration of how syntactic prominence and semantic prominence under the FrameNet system can be implemented in the model of discourse salience described in Section 2.2. For this illustration, I will use the following context sentences and show how the store is constructed for each.

(55) The teacher tossed some prizes to the kids.

(56) The teacher tossed the kids some prizes.

Both (55) and (56) contain three discourse referents which I will refer to by name (rather than variables) for simplicity: TEACHER, PRIZES, and KIDS. For this simple example, the relative syntactic prominence of the discourse referents is relatively straightforward. The rankings under the role-based and hierarchy-based methods are as shown in (57)-(58). The primary difference between the two examples shows up in (58) where the role-based method results in only a partial ordering of the referents.

(57) The teacher tossed some prizes to the kids.

syntactic prominence $_{role}$: TEACHER > PRIZES > KIDS

syntactic prominence $_{hierarchy}$: TEACHER > PRIZES > KIDS

(58) The teacher tossed the kids some prizes.

syntactic prominence $_{role}$: TEACHER > {KIDS,PRIZES}

syntactic prominence $_{hierarchy}$: TEACHER > KIDS > PRIZES

In order to determine the semantic prominence of the discourse referents, first it is necessary to consult the FrameNet system. The verb *toss* is in the CAUSE_MOTION frame (derived from the INTENTIONALLY_AFFECT frame) in which the agent, theme, and goal frame elements describe the roles of TEACHER, PRIZES, and KIDS, respectively. For illustrative purposes here, I will assume that these three roles are ranked with respect to each other as shown in (59) (in line with the proposed thematic hierarchies of Larson, 1988; Speas, 1990). Given this ranking, then the relative semantic prominence of the discourse referents is the same for both (55) and (56). The complete set of prominence relationships is thus shown in (60) and (61). The store may then be illustrated as shown in Figure 2.9 and Figure 2.10.

(59) agent > theme > goal

(60) The teacher tossed some prizes to the kids.

syntactic prominence $_{role}$: TEACHER > PRIZES > KIDS

syntactic prominence $_{hierarchy}$: TEACHER>PRIZES>KIDS

semantic prominence: TEACHER > PRIZES > KIDS

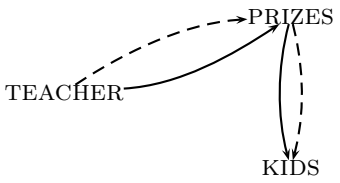
(61) The teacher tossed the kids some prizes.

syntactic prominence $_{role}$: TEACHER > {KIDS,PRIZES}

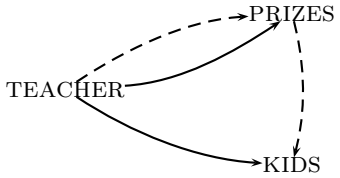
syntactic prominence $_{hierarchy}$: TEACHER>KIDS>PRIZES

semantic prominence: TEACHER > PRIZES > KIDS

Assuming the thematic hierarchy in (59) is accurate, then it is easy to see from the illustrations what different predictions can be made regarding the salience of entities for subsequent pronoun resolution. For the sentence in (60), the stores under both the role-based and hierarchy-based method for syntactic prominence are the same, and

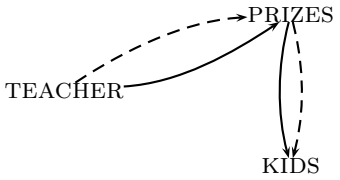


The teacher tossed some prizes to the kids.

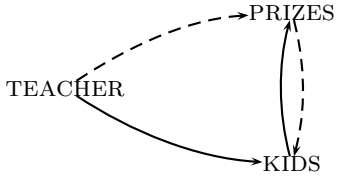


The teacher tossed the kids some prizes.

Figure 2.9. A illustration of the store using the syntactic prominence role method and the semantic prominence FrameNet approach. The syntactic prominence relations are shown with solid arcs and the semantic prominence relations are shown with dashed arcs.



The teacher tossed some prizes to the kids.



The teacher tossed the kids some prizes.

Figure 2.10. An illustration of the store using the syntactic prominence hierarchy method and the semantic prominence FrameNet approach. The syntactic prominence relations are shown with solid arcs and the semantic prominence relations are shown with dashed arcs.

the prominence hierarchies are nicely aligned. Thus, in an experimental sense, this might be viewed as the control condition. If the next utterance begins with a pronoun, "They...", then one would predict that the preferred interpretation of that pronoun would be PRIZES. Of course, this case by itself would reveal nothing about the relative influence of syntactic and semantic prominence. The double-object construction in (61), on the other hand, provides the experimental test case where predictions diverge. If

the hypothesis that syntactic prominence is actually explained by semantic prominence is true, then subsequent pronominal reference by *they* should still show an interpretive preference to PRIZES. However, if this hypothesis is false and syntactic prominence is the only relevant factor, then subsequent pronominal reference by *they* should show different interpretive preferences depending on whether the role-based method or the hierarchy-based method to syntactic prominence is more accurate: If the hierarchy-based method is better, then the preference should be toward KIDS; if the role-based method is better, then there should be no greater preference for either KIDS or PRIZES.[4]

One final possibility here is that *both* syntactic and semantic prominence contribute to discourse salience. Under this hypothesis, the factors might then compete against each other in the promotion of the various discourse referents. In the "control" case, things do not change much because the prominence hierarchies are aligned and promote the same candidates. In the "divergent" case, if the role-based method is used, then the prediction would be a preference for PRIZES, while if the hierarchy-based method is used, then the prediction would be no preference for either KIDS or PRIZES.[5]

It is important to note here that for these test sentences, the predictions do not fall into clear distinguishable categories. Thus, for example, if experimental results showed no preference for KIDS or PRIZES in the divergent case (i.e., (61) as illustrated in the right-hand figures of Figure 2.9 and Figure 2.10), we could conclude one of two things: either syntactic prominence is the only relevant factor and the role-based method is

[4]In this particular example, it is possible that the inherent animacy of KIDS might give it an inherent edge over PRIZES. So a better example of a potentially ambiguous case should control for animacy.

[5]There are further possibilities here if we consider that the factors may be weighted. If, say, syntactic prominence counts only half as much as semantic prominence then the picture will be very different. Or if the individual orderings have varying weights then many other possibilities come into play. In the present research, I do not consider these more complex scenarios, but will return to the topic briefly in later chapters.

the better method for determining it; or both syntactic and semantic prominence are relevant factors and the hierarchy-based method is better. But this hardly constitutes much progress and would require several further tests to truly answer the central question of what is the relative contribution of syntactic and semantic prominence. Therefore, while these double-object constructions seem, at first glance, to be promising here, they do not provide a very efficient test of the hypothesis. Instead, in Chapter 3, I will propose using *spray/load* constructions (*spray some paint on a wall/spray a wall with some paint*) which I argue provide a minimal contrast with clear, distinct experimental predictions with respect to the relative salience of the discourse referents.

2.7. Implementation 2 Using PROTO-roles

Dowty (1991)'s proposal for argument selection takes a very different approach than that of Frame Semantics. He defines only two "fuzzy" PROTO-roles: PROTO-agent and PROTO-patient and discusses a set of semantic entailments associated with each as shown in (62)-(63).

(62) PROTO-agent entailments

- sentience
- volition
- causing event or change-of-state
- undergoing movement

(63) PROTO-patient entailments

- undergoing change of state
- incremental theme

Table 2.1. PROTO-role Entailments for *Bake*

	baker	bakee
sentience	√	
volition	√	
cause event/c-o-s	√	
movement		
undergo c-o-s		√
incremental theme		√
causally-affected		√
stationary		

- causally-affected
- stationary

Dowty does not claim this set of entailments to be either exhaustive or accurate: some of the entailments may be superfluous while others not included may be necessary. For instance, as will be further discussed below, it may very well be the case that the PROTO-agent entailments of sentience and volition are not fully independent entailments (i.e., it's difficult to find a predicate that entails volition but not also sentience of one of its arguments). However, the model is a particularly interesting way of establishing a link between lexical conceptual structure and surface syntax.

In order to establish this link, Dowty proposes an argument selection principle which determines that in any given predicate, the argument having the most PROTO-agent entailments will be the surface subject, while the (remaining) argument with the most PROTO-patient entailments will be the surface object. Any remaining arguments will be realized in oblique position(s). For instance, consider the verb *bake*. For the sake of the present discussion I'll call the two semantic arguments the baker and the bakee. The PROTO-role entailments for each of these arguments are illustrated in Table 2.1.

The baker has more PROTO-agent entailments so it becomes the grammatical subject while the bakee with more PROTO-patient entailments becomes the object, as in (64).

(64) Marcus baked a cake.

Dowty does not discuss any sort of hierarchy for the two PROTO-roles, perhaps because it is implicit from the argument selection principle that the argument with the most PROTO-agent entailments will be inserted into a higher syntactic position than any other argument. However, what I would like to propose is the possibility that it is these individual entailments which contribute in varying degrees to the semantic prominence of the arguments. That is, it could be the case that such PROTO-agent entailments as sentience and volition contribute much to making an argument more prominent, while such PROTO-patient entailments as stationary contribute little. To illustrate, let's suppose that discourse salience is solely determined by the PROTO-agent entailment of sentience (which Dowty describes as "sentience with respect to the event or state denoted by the verb", p. 573). Hence, for the verb *admire* as in *John admired the scenery*, the referent realized as the grammatical subject will be the most salient entity in the discourse and preferred for subsequent pronominalization. However, for the verb *amaze* as in *The scenery amazed John*, the referent realized as the grammatical object will be more salient. Of course, this is an oversimplification: it is probable that several entailments are important to determining semantic prominence and further that their individual contribution varies from others.

In order to clarify how the PROTO-role approach differs from the FrameNet approach for the question of semantic prominence, a little more discussion is warranted here. If it were to turn out that *all* of the PROTO-agent entailments contribute more to semantic

prominence than *all* of the PROTO-patient entailments, then at first glance, one might be tempted to say that this is the same as agent > patient and would be the same as the FrameNet approach. However, it is important to remember that the entailments may apply to *any* argument. Thus, it is theoretically conceivable that the argument with the most PROTO-agent entailments becomes the subject but that none of these entailments contribute much to its semantic prominence. Then, concurrently, a different argument having a PROTO-agent entailment crucial to high semantic prominence becomes the object. If it could be shown that y is more semantically prominent than x in *x amazed y*, then this would be good evidence for such a view.

An important theoretical question to address here is what the entailment hierarchy might look like. For the discrete-role approaches described above, the motivation for any particular hierarchy has been the direct linkage between argument roles and structural positions. However, in the discussion above, I have put forth the idea that under the PROTO-role system, the entailment ranking which determines syntactic prominence may look entirely different: perhaps only some of the entailments contribute significantly to the salience of discourse referents which bear those entailments, while perhaps some PROTO-patient entailments contribute more than some PROTO-agent entailments. I have not found any empirical evidence in the literature to motivate a particular entailment ranking and therefore leave it as an open question to which I will try to provide a partial answer through the psycholinguistic experiments in Chapter 3 and corpus analysis in Chapter 4. However, let me try to motivate one possibility in advance. I suggest that a ranking might be determinable with respect to perceptual salience. Intuitively, when

perceiving the world around us, sentient entities are more salient than non-sentient entities and entities undergoing movement are more salient than stationary entities. Thus, one might propose a rudimentary hierarchy—or rather, set of hierarchical relations—as in (65); of course, to be refined later as a result of further study and experimentation.

(65) sentience > non-sentience

movement > non-movement

For the present research, the PROTO-role theory provides a nice contrast to the deep case roles of Frame Semantics, and I use it in this corpus analysis for its theoretical value in giving breadth to an empirical investigation of semantic prominence.

2.7.1. Linguistic Tests for PROTO-role Entailments

One task that any Dowtian type of linking theory must eventually complete is the process of actually determining for each predicate which entailments are satisfied by which arguments. While (Dowty, 1991) gives an overview of the entailments, he does not explicitly define any semantic tests for entailment satisfaction. In addition, I have not yet found any reference which does so for the full set of entailments (as noted below, however, there are proposed tests for some of the individual entailments). My goal here, then is to outline a minimal battery of semantic tests for the eight different entailments Dowty uses to distinguish the PROTO-roles. However, some of these tests are drawn directly from or based on tests used in discussions of some well-studied semantic entailments including undergo c-o-s and incremental theme. A comprehensive discussion of some of the issues surrounding these entailments (e.g., causality, change-of-state, motion) is beyond the scope of this dissertation. Using Dowty's brief descriptions of these entailments as a

starting point, I intend to provide here an overview of each entailment and some specific linguistic tests which may be used to evaluate whether an argument of a predicate bears that entailment.

One thing I would like to note in advance is that I have taken here a rather strict view of the semantic entailments of a predicate: I am only concerned with entailments imposed on arguments by the predicate alone. Modification with adverbial phrases or changes in tense can also add further entailments (as well as cancel others, like incremental theme-hood). Examples of such modification are discussed in more detail below. For the moment, I will ignore these complications and focus only on the core semantic contribution of the predicate.

2.7.1.1. Sentience. According to Dowty, the sentience of a participant in a state or event implies the awareness of that entity's participation in that particular event or state: The objects of such verbs as *elect*, *appoint*, and *nominate* are necessarily human and by inference may also be sentient. But their awareness of having been elected, appointed, or nominated is not entailed by the respective predicates. Thus it is perfectly acceptable to say something like *John was elected chairman, but he didn't know it.* This brings up one good test for sentience: If a participant in an event is entailed to be sentient, then denying that participant's awareness of that event should be semantically anomalous. This can be tested by adding a conjoined clause which denies the awareness of the participant to be tested. Or, if the participant is in subject position, a simpler method would be to modify the event with the adverb *unwittingly*.

(66) a. John broke the lamp but John didn't know it.

b. John unwittingly broke the lamp.

c. John hit Mark but John didn't know it.

d. John unwittingly hit Mark.

e. John hit Mark but Mark didn't know it.

f. John slept but John didn't know it.

g. John unwittingly slept.

h. #John recognized Mary but John didn't know it.

i. John recognized Mary but Mary didn't know it.

j. #John unwittingly recognized Mary.

The verbs *break*, *hit*, and *sleep* do not entail any of their participants to be aware of their participation in the event. Thus (66a-g) are all perfectly fine. However, *recognize* entails sentient of its subject but not its object as shown by (66h-f).

An alternative—but possibly less accurate—test of sentience is merely to insert non-sentient entities to see if anomalies are generated. Perhaps the most well-known non-sentient entity which is useful here is a rock as shown in (67).

(67) a. A rock broke the lamp.

b. A rock hit Mark.

c. #A rock slept.

c. #A rock recognized Mary.

This test results in the same conclusion for *break*, *hit*, and *recognize*, but suggests that *sleep* requires a sentient subject. However, this is precisely where this test may be inccurate. Indeed, both *sleep* and *recognize* require that their subjects be sentient (or at least be capable of sentience), but crucially, only *recognize* requires that its subject

be aware of its participation in the event. Thus, while the "rock" test may be a useful quick test of sentience, the "unwittingly" test above is more precise.

2.7.1.2. Volition. A volitional act is one that is performed willfully by an actor. Hence, it should be anomalous to negate the willful action of a participant. This can be done by conjoining a clause involving the negation of such constructions as *try to*, *choose to*, or *decide to*. Alternatively, the volition of the subject may be tested by inserting the adverb *accidentally* before the verb. These two tests are illustrated in (68).

(68) a. John dreamed but he didn't try/choose/decide to dream.

b. John accidentally dreamed.

c. John fell but he didn't try/choose/decide to fall.

d. John accidentally fell.

e. John hit Matt but he didn't try/choose/decided to hit Matt.

f. John hit Matt but Matt didn't try/choose/decide to be hit by Matt.

g. John accidentally hit Matt.

h. #John listened to Matt but he didn't try/choose/decide to listen to Matt.

i. John listened to Matt but Matt didn't try/choose/decide to be listened to by John.

j. #John accidentally listened to Matt.

The verbs *dream*, *fall*, and *hit* do not require a volitional participant as the sentences in (68a-g) show. However, *listen* requires volitional participation of its subject as shown in (68h,j) but not of its object as shown in (68i). [6]

[6]The sentences in (68h,j) do have a felicitous reading where John was listening to something and was not aware while he was listening that it was Matt. However, in this case, *accidentally* seems to be modifying not the listening event as a whole, but rather a specific part of the event: the target of the listening event, Matt. Under this interpretation, though, John is still listening volitionally.

One further note should be made here about the relationship between sentience and volition. Dowty (1991) hypothesizes that the entailments he proposes are semantically independent. To illustrate this, for each entailment, he presents some sentences in which he claims the predicates require only one entailment. For volition, he presents the sentence shown in (69) (his (29a), p. 572).

(69) John is being polite to Bill/is ignoring Mary.

I agree with Dowty's judgment that *polite* and *ignore* require a volitional subject, but I do not share his judgment that this is the only entailment on the subject. I would argue that for both of these verbs, sentience is also entailed. It is hard for me to see how a person who tries/chooses/decides to be polite to or to ignore somebody cannot also be aware of that. Consider the sentience tests for these verbs in (70).

(70) a. #John was polite to Bill but John didn't know he was polite to Bill.

b. #John ignored Mary but John didn't know he ignored Mary.

There is an interpretation of the sentences in (70) where Bill and Mary each, from their own perspectives, interpret John's actions as being polite or cold-shoulder-like, respectively. However, in that case, it is hard then to see John as a volitional participant in the first place: He is not trying/choosing/deciding to be polite or to ignore in such a case. The other participants are only interpreting his actions as such.

Every verb I can find which does entail volition of one of its participants (e.g., *listen*, *watch*, *decide*) also entails sentience of that same participant. In other words, it appears that in English, volition entails sentience. Thus, it seems that the PROTO-agent entailments are not independent of each other. One consequence of this is that it may not be necessary to distinguish between sentience and volition. It might turn out

to be theoretically simpler and practically more efficient to collapse the PROTO-agent entailments into some smaller set.

2.7.1.3. Causing Event or Change-of-state. The PROTO-agent entailment of cause event/c-o-s is actually a rather complex entailment which is not independent of some other PROTO-role entailments. In order to determine whether a particular participant has caused an event or a change-of-state in another participant it is first necessary to determine whether there is any participant that has undergone a change-of-state. This is actually one of the PROTO-patient entailments, but because it is related to the causation entailment, I will deal with it here. Furthermore, because both of these entailments then influence the causally-affected PROTO-patient entailment, I will also deal with that in this section.

Undergo Change-of-state. One fairly straightforward way to test whether a participant has undergone a change-of-state is to add a conjoined clause which denies that that participant has changed. One specific technique that does this is to use the phrase *but nothing happened to...* (Shibatani, 1976b). This is illustrated in (71).

(71) a. #John broke the window with a hammer but nothing happened to the window.

b. Mark saw the eclipse through a telescope but nothing happened to the eclipse.

The results of the test in (71) indicate that *break* entails that its object (here, the window) undergoes a change-of-state, but *see* does not. For some verbs, the "nothing happened to" test just isn't appropriate because the participants may be somewhat more abstract or because they are events. In these cases, it may be more useful to deny or

assert the existence or occurrence of the event. For instance, consider the sentences in (72).

(72) a. #Luke made a mistake but there wasn't any mistake.

b. #Philip stopped biting his nails but there was a nail-biting event afterward.

c. #The virus caused the computer malfunction but there was no computer malfunction.

The resulting sentences here, if slightly unnatural, show that *make*, *stop*, and *cause* all entail that one of their participants undergo a change-of-state. (72b), in particular, illustrates that a participant here can even be an event: Events may undergo a change-of-state in the sense that an event may begin or end or its continuity may otherwise be altered. Thus, knowing whether there is any entity or event which undergoes a change-of-state, it is now possible to evaluate which participant, if any, is the cause of that change-of-state. I take this up in the next section.

Determining the Causer. Causation is a notion that has been widely studied, particularly as part of the investigation of causative constructions. I will not attempt to review the extensive literature here. See Shibatani (1976b) for a good, if perhaps dated, overview of generative syntactic approaches, and Parsons (1990) for a more recent, though less comprehensive overview. Here, I will adopt the following view of causation from Shibatani (1976a):

> The relation between the causing and the caused event is such that the speaker believes that the occurrence of the caused event is wholly dependent on the occurrence of the causing event.... (p. 1)

The affected participant (or caused event) must therefore be disjoint from the causer. One way to test this is by ruling out that the change of state undergone by the affected participant was entirely internal. If the affected participant was causally affected by another participant then asserting that the participant underwent the change by himself/herself/itself should result in a semantic anomaly (cf., Taranto, 2002). Consider the sentences in (73).

(73) a. #The soldier murdered the prisoner by the prisoner murdering himself.

b. #The boss fired Trudy by Trudy firing herself.

c. John fell over by making himself fall over.

A murdering event requires at least two disjoint participants, one of whom is responsible for murdering the other. Thus, in (73a), it is anomalous to assert that the prisoner murdered himself. Similarly, in (73b), Trudy cannot fire herself. Rather, her loss of employment must have been caused by another participant, namely her boss. However, it is conceivable for a falling-over event to occur entirely internal to one participant. Therefore it is felicitous in (73c) to assert that John fell over by himself.[7] Hence, the evidence points to the soldier and the boss as being the causers of the changes-of-state in (73a-b), respectively. This can be confirmed by explicitly denying their responsibility in the event. The result should be infelicitous as in (74).

[7]It might be noted that it is possible to say something like (i).

(i) Trudy was fired by Trudy firing herself.

This might be used to describe a situation in which Trudy was in a position to be fired but before the boss could take such action Trudy quit. However, this seems to bend the meaning of *fire* in a metaphorical direction. It seems unlikely that people would actually regard the event as a true firing event but rather a quitting event.

(74) a. #The soldier murdered the prisoner but the soldier had nothing to do with the murder.

b. #The boss fired Trudy but the boss had nothing to do with the firing.

A further test for causation then is to directly attribute causation to an entity by asserting some means by which the hypothetical causer brought about a change-of-state. For instance, for the *break* and *stop* sentences in (75) and (76), the test is performed by conjoining an assertion of the change-of-state (e.g., *the window broke*) and then appending a *by* or *as a result of* clause which identifies the causing event, and thereby, the causer (cf., Talmy, 1976).

(75) John broke the window with a hammer.

a. The window broke by/as a result of John's hitting it with a hammer.

b. The window broke by/as a result of the hammer's hitting it.

(76) Philip stopped biting his nails.

a. Philip's nail-biting events stopped by/as a result of Philip's efforts.

(77) Bill drained the water from the tank.

a. The tank became empty by/as a result of Bill's pulling the plug.

The data here suggest that both John and the hammer are causers of the window-breaking event, Philip is a causer of the ceasing of the nail-biting, and Bill is a causer of the tank becoming empty. However, as (75) shows, there may be more than one causer. This brings up a secondary issue. Looking closely at the event described in (75), many will probably agree that the hammer is a more direct cause of the window-breaking event than is John. John and the hammer participate in a causal chain which results in the window-breaking event. This raises a question as to how far down such a chain is it

acceptable to regard participants as causers? I am not yet sure I have a good answer to this question, but for the present, one practical solution is to use a secondary test with enabling constructions with *let* (Talmy, 1976) which implies a passive (i.e., non-causative) participation in an event. Participants which pass the "by"/"as a result of" test above, but fail the "let" test, may be rejected as causers, as follows.

(78) a. #John let the window break by hitting it with a hammer.

b. #John let the nail-biting stop by his efforts.

c. Bill let the water drain from the tank by pulling the plug.

This secondary test therefore shows that Bill's causal participation in the water-draining event is too distant to regard him as a causer.

Causally Affected. Once a predicate has been identified as entailing a change-of-state in one of its participants and also entails that another one of its participants is the cause of that change-of-state, then deciding whether any participant is causally-affected is trivial: the participant which undergoes a change-of-state must also be identified as bearing the causally-affected entailment. This, of course, brings up once again the question of the independence of the entailments. Dowty (1991) suggests that in *smoking causes cancer*, the object is entailed to be causally-affected only. I do not share this judgment. In order to determine whether or not it is true that smoking causes cancer, we must look at a number of smokers and determine whether or not a change-of-state occurred in those smokers: that is, that at one time they were cancer-free, and at a later time, they were cancer-stricken. Thus, the sentence actually does entail a change-of-state, or rather in this case, a large number of changes-of-state.

So, in short, the entailments of cause event/c-o-s, undergo c-o-s, and causally-affected are not independent entailments but are intertwined, similar to sentience and volition as discussed above. I will return to the issue of entailment independence below.

2.7.1.4. Movement. The PROTO-agent entailment of movement is described by Dowty (1991) as the movement of one participant relative to another participant. However, as stated, this definition is too loose. Consider (79).

(79) John threw the ball to Matt.

In a throwing event, one participant causes some entity to move toward some stated or unstated goal. Under the definition given above, movement of the ball in (79) is entailed because it is understood to move relative to both John and Matt. However, in a purely relativistic sense, movement is also entailed of both John and Matt relative to their co-participant in the event, the ball. So we need some means to test which participant is moving and which is stationary. This distinction can be discovered by appending an adjoined locative prepositional phrase as in (80).

(80) a. John threw the ball to Matt from the second-story window.

b. John threw the ball to Matt over the jungle gym.

Movement can be be understood in terms of two parameters: a fixed point and a path (cf., the *ground* and *path* in Talmy, 1985). Movement is thus an increase (or decrease) in the separation from the fixed point and a path along which this separation is measured. The locative phrase in (80a) defines the fixed point while the locative phrase in (80b) defines the path. Since both of these parameters apply to the ball, it is therefore entailed that the ball undergoes movement. The locative in (80a) further defines the point in space where John is. Hence, it is entailed that John is stationary—the corresponding

PROTO-patient entailment (see below). The entity evoked in the locative phrase in both sentences in (80)—MATT—is also entailed to be stationary in that it provide a fixed point toward which the ball moves. It is more difficult to test for this independently but can be determined by deduction: If the ball is moving toward a specific goal, then that goal must be stationary relative to the moving ball.

Of course it is certainly possible that an observer totally independent of the events in (80) might remark that all three participants are moving, and possibly even that Matt is running toward/away from/oblique to John as the ball travels toward him. However, the perspective that Dowty apparently wants to take is entirely relative to the event and specifically to whichever participant is perceived as being the primary mover. One of his own examples includes *the bullet overtook the arrow* in which both arguments can clearly be seen as moving from an outsider's perspective, but within the event, the faster-moving participant, the bullet, is regarded as having the movement entailment while slower-moving participant, the arrow, is regarded as having the stationary entailment.

So when, one might ask, is stationariness not entailed in a movement predicate? If the preposition introducing the oblique argument is *towards* as in (81), then Matt is not entailed either to undergo motion or to be stationary.

(81) John threw the ball towards Matt.

This does not contradict the "faster-than" notion, because it is entirely possible that Matt is moving faster than the the ball. For instance, in (82a), it is conceivable that the train is moving faster than the feather. Yet movement is entailed of the feather. The adjectival modifer *speeding* of *bullet train* entails that the train is moving and therefore

this generates an anomaly if the same NP is used in a case where throw entails it to be stationary as in (82b).

(82) a. The child threw a feather toward the speeding bullet train.

b. #The child threw a feather to the speeding bullet train.

If the movement and stationary entailments are always determined relative to other participants in the event, then this raises the question of whether movement can ever be entailed by verbs which take only one obligatory argument such as, say, *walk*.

(83) a. John walked.

b. John walked to school.

c. John walked along the river.

d. John walked away.

Dowty (1991) does not discuss this case. Intuitively, the sentences in (83b-d) all seem to entail movement because the two parameters of movement are either indicated or implied. In (83b), the school is a fixed point and the path can be assumed to be some relevant path between John's previous location and the school. In (83c), the path is the river, and the point is some relevant point which John is moving relative to. Finally, in (83d), the path is some relevant path away from some assumed fixed point. However, in (83a), there is no explicit participant which helps us to fix either of the key parameters of movement. However, resorting to intuition again, I feel that John is moving relative to some salient parameters and if it were necessary to check the truth of the statement, these parameters could be located in the area where John is supposed to have walked.

2.7.1.5. Undergo Change-of-state. [See discussion of the PROTO-agent entailment of cause event/c-o-s in Section 2.7.1.3, above.]

2.7.1.6. Causally Affected. [See discussion of the PROTO-agent entailment of cause event/c-o-s in Section 2.7.1.3, above.]

2.7.1.7. Incremental Theme. When a predicate describes a durative event, the event may be measured out with respect to one of the participants. Such a participant is called the incremental theme (Baker, 1997; Dowty, 1991). One way to determine which participant (if any) serves as the incremental theme is to modify the predicate with *partially* (or some other adverbial indicating some degree of completion of the event) and then determine which participant is most directly affected by the partiality of the event. For instance, consider the sentences in (84)-(85).

(84) a. John partially sprayed some paint on a wall.

b. # ... and there is no paint remaining.

c. ... and the wall is completely covered.

(85) a. John partially sprayed a wall with some paint.

b. ... and there is no paint remaining.

c. # ... and the wall is completely covered.

The assertion in (84b) is semantically inconsistent with the event described in (84a) which entails that part of the paint rather than part of the wall is used up. The opposite is the case in (85) in which part of the wall is used up.[8] Hence, the events in (84) and (85) are measured out by the paint and the wall, respectively. Therefore, the evidence points to the direct object argument of a *spray* predicate as the incremental theme.

However, it is important to note some caveats in the use of the "partially" test. There is an alternative interpretation of the sentences available in which *partially* can be

[8]There is some debate about these totality effects on *spray/load* verbs in the literature. See Rappaport and Levin (1988) and Tenny (1994) for opposing viewpoints. Also see Section 3.3.1.1 for some discussion.

seen as modifying the degree of the subject's participation in the event being described. The sentence in (86a) is anomalous with the meaning that his car is somehow measuring out the driving event. Rather it seems to be fine when it means that John's attention was somehow divided as in (86b).

(86) a. John partially drove his car.

b. John partially drove his car and partially chatted with his friends.

Another caveat is that while the "partially" test is fairly reliable, it is important to make sure that the arguments are of a type that can be measured out in smaller parts. Consider the sentences in (87).

(87) a. John partially loaded a cart with some hay.

→ cart partially full

b. John partially loaded some hay onto a cart.

→ there is some leftover hay

c. John partially loaded some rocks onto a cart.

→ there are leftover rocks.

d. John partially loaded a cart with some rocks.

→ cart partially full

e. John partially loaded a rock onto a cart.

→ ?

f. John partially loaded a cart with a rock.

→ ?

It's not very clear to me what (87e-f) might mean, but a possible interpretation is that the rock is only halfway onto the cart. But this could apply to both sentences, in

my judgment. In any case, when performing the "partially" test, the arguments being tested should preferably be mass or plural count nouns.

2.7.1.8. Stationary. [See discussion of the PROTO-agent entailment of movement in Section 2.7.1.4, above.]

2.7.1.9. PROTO-role dependencies. Throughout the discussion above, it was noted that the various PROTO-role entailments are not independent of each other. Below is a summary of these dependencies.

- If a predicate P entails volition of one of its arguments, x, then P entails sentience of x.
- If a predicate P entails movement of one of its arguments, x, then P entails that another of its arguments, y, (possibly implicit) is entailed to be stationary.
- If a predicate P entails that one of its arguments, x, is causally-affected, then P entails undergo c-o-s of x.
- If and only if a predicate P entails that one of its arguments, x, is causally-affected, then P entails cause event/c-o-s of another of its arguments, y.

If the eight binary entailments were totally independent, then the number of possible entailment combinations for any given participant would be $2^8 = 256$. In short, these 256 different combinations would constitute independent thematic roles. However, because of the dependencies described above, this number is considerably reduced. By my calculations, there are only 108 different possible combinations.[9] Although this may still

[9]The calculation is as follows: Between sentience and volition there are only three possible combinations (because volition entails sentience, the combination [not sentience, volition] is ruled out). Between undergo c-o-s and causally-affected there are three combinations. Between movement and stationary there are only three combinations. Then with each of cause event/c-o-s and incremental theme there are two combinations. Thus, $3 \times 3 \times 3 \times 2 \times 2 = 108$.

seem like a somewhat large number, it is an improvement over FrameNet's 500+ frame elements. For the present research, it may be that even some of these combinations can be collapsed. For instance, as suggested above, sentience and volition might be collapsed into one category, say animacy, thereby reducing the number of combinations to 72—a more manageable size for an inventory of semantic roles to be hierarchically arranged with respect to semantic prominence.

2.7.2. Illustration 2

To conclude the discussion of the Dowtian PROTO-roles here, I provide a basic illustration of how the model of discourse salience might be implemented in terms of PROTO-role entailments using the same set of sentences as for the FrameNet illustration in Section 2.6.2 above. These sentences are repeated here, along with the syntactic prominence rankings of the three discourse referents: TEACHER, PRIZES, KIDS.

(88) The teacher tossed some prizes to the kids.

syntactic prominence $_{role}$: TEACHER > PRIZES > KIDS

syntactic prominence $_{hierarchy}$: TEACHER > PRIZES > KIDS

(89) The teacher tossed the kids some prizes.

syntactic prominence $_{role}$: TEACHER > {KIDS,PRIZES}

syntactic prominence $_{hierarchy}$: TEACHER > KIDS > PRIZES

In order to determine the semantic prominence of the respective discourse referents in these sentences, we first need an analysis of this ditransitive construction under in terms of Dowtian PROTO-role entailments. Using the entailment tests outlined in Section 2.7.1, I propose that the analysis of *toss* and its three arguments is as shown in Table 2.2. I

Table 2.2. PROTO-role Entailments for *Toss*

	x	y	z
sentience	√		
volition	√		
cause event/c-o-s	√		
movement		√	
undergo c-o-s		√	
causally-affected		√	
incremental theme		√	
stationary	√		√

assume, following Baker (1997), that the two *toss*-variants represent only a single lexical entry from which only the NP-PP variant is base-generated. The NP-NP variant is then derived from the NP-PP variant.

According to the argument selection principle, then, x should be realized in subject position, y in object position, and z in oblique position in the NP-PP variant in (88). Thus, for this example, x = TEACHER, y = PRIZES, and z = KIDS. These entailment assignments will remain the same for the NP-NP variant in (89) because the latter sentence is derived from the first.

The next step toward determining the semantic prominence of the discourse referents is to consult the entailment hierarchy. For the purpose of the present illustration, I will assume the hierarchy suggested above in (65) and repeated here as (90).

(90) sentience > non-sentience

movement > non-movement

I will also assume that each of these hierarchical relations imposes an independent ordering on the discourse referents (as opposed to the single ordering resulting under the FrameNet system) so that all sentience-entailed referents are more prominent than all

non-sentience-entailed referents, and so on. Given these assumptions and the hierarchy as shown, the sample sentences are shown in (91) and (92) with their respective syntactic and semantic prominence rankings. The graphical illustration of the store would then be as shown in Figure 2.11 and Figure 2.12 where the different line types represent orderings imposed by different syntactic and semantic prominence hierarchies.

(91) The teacher tossed some prizes to the kids.

syntactic prominence $_{role}$: TEACHER > PRIZES > KIDS

syntactic prominence $_{hierarchy}$: TEACHER>PRIZES>KIDS

semantic prominence: TEACHER>$_{sentience}${PRIZES,KIDS},

PRIZES>$_{movement}${TEACHER,KIDS}

(92) The teacher tossed the kids some prizes.

syntactic prominence $_{role}$: TEACHER > {KIDS,PRIZES}

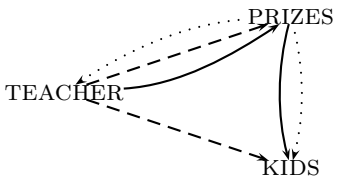
syntactic prominence $_{hierarchy}$: TEACHER>KIDS>PRIZES

semantic prominence: TEACHER>$_{sentience}${PRIZES,KIDS},

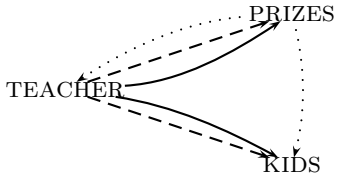
PRIZES>$_{movement}${TEACHER,KIDS}

If the next utterance then begins with a pronoun, "They...", it turns out here that the predictions are the same as under the FrameNet approach described in Section 2.6.2. For the NP-PP sentence in (91), the predicted preference should be for PRIZES. In the NP-NP sentence in (92), however, the predictions vary depending on the approach (see discussion in Section 2.6.2).

What is more interesting to focus on here is the ordering relationship(s) between TEACHER and PRIZES in both Figure 2.11 and Figure 2.12. In the illustration, there are several orderings between these two referents, but they are not all directed the same

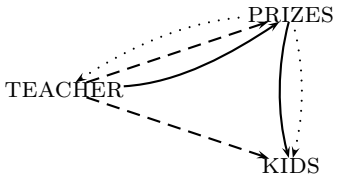


The teacher tossed some prizes to the kids.

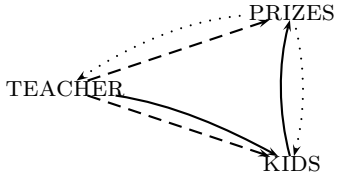


The teacher tossed the kids some prizes.

Figure 2.11. An illustration of the store using the syntactic prominence role method and the semantic prominence PROTO-role approach. The syntactic prominence relations are shown with solid arcs and the semantic prominence relations are shown with non-solid arcs (sentience with dashed arcs, movement with dotted arcs).



The teacher tossed some prizes to the kids.



The teacher tossed the kids some prizes.

Figure 2.12. An illustration of the store using the syntactic prominence hierarchy method and the semantic prominence PROTO-role approach. The syntactic prominence relations are shown with solid arcs and the semantic prominence relations are shown with non-solid arcs (sentience with dashed arcs, movement with dotted arcs).

way. For this particular example, this may not be much of a problem because TEACHER and PRIZES will never compete against each other for pronominal antecedence: one or the other will always be filtered out due to morphological feature-matching constraints (i.e., as +human and +singular, TEACHER is compatible only with *he*/*she*, while PRIZES

is compatible only with THEY). However, it is not too difficult to imagine a case where discourse referents realized in these positions might compete. For instance, consider (93).

(93) a. The mother tossed the baby girl to the fireman.

b. The mother tossed the fireman the baby girl.

For these sentences, the graphical illustration would look the same where TEACHER corresponds to MOTHER and PRIZES to BABY GIRL. If the next utterance then began with a pronoun, "She...", the system would have to have some strategy for dealing with the many relations between the two. I envision two theoretical possibilities here. One possibility is that because there are orderings pointing in opposite directions, the relation between the two is effectively indeterminate, and therefore neither is more salient than the other. The other possibility is that the more directed relations there are in one direction, the stronger is the (net) precedence in that direction. In other words, orderings extending in opposite directions might cancel each other out and remaining orderings would determine a net prominence relation. Under this possibility, then, MOTHER should be more salient than BABY GIRL. Needless to say, this is all conjecture at this point and requires experimental verification. The evidence I present the corpus analysis in Chapter 4 begins to point toward an answer to this question, suggesting that the latter possibility is more likely and that orderings are combined in some fashion to produce an overall salience relationship among referents.

2.8. Conclusion

In this chapter, I have laid out a relatively detailed, yet flexible, model of determining the salience of discourse referents for subsequent pronominal resolution. The

model incorporates detailed notions of two prominence factors: syntactic and semantic prominence. These two factors have not been clearly distinguished in the literature and are arguably confounded in many studies. In this chapter I have further shown how the relative influence of these two prominence factors on discourse salience can be evaluated. In short, this can be done by using argument reordering constructions which have the property that the syntactic roles and semantic roles/entailments are carefully controlled across the two variants. I will argue in the next chapter that so-called *spray/load* constructions have this property and use them in a series of psycholinguistic experiments.

CHAPTER 3

Psycholinguistic Experiments

3.1. Introduction

The relative discourse salience of entities for subsquent pronominal reference is known to depend on a number of different factors. One of the most well-studied of these is syntactic prominence. That is, all else being equal, there is a preference for an utterance-initial pronoun to be interpreted as coreferent with an entity which was realized in a syntactically prominent position. Hence, in (94), there is a preference to interpret the pronoun in (94b) as referring to the referent realized as the subject of the preceding utterance, LUKE.

(94) a. Luke hit Max.

b. Then, he ran home.

In this dissertation, I am concerned with investigating whether the effects attributed to syntactic prominence are in fact better explained as an effect of semantic prominence. As a step toward this goal, this chapter presents the results of several psycholinguistic experiments showing that semantic prominence has an influence on the discourse salience of entities and that this influence is systematically observable and is distinct from that of syntactic prominence. Furthermore, this chapter is also concerned with evaluating the hierarchies which are thought to determine syntactic and semantic prominence. Ultimately, I will show that while a proposed semantic prominence hierarchy

can be confirmed, the syntactic prominence hierarchy that many researchers have assumed may be in need of revision. The results will be discussed with respect to both the FrameNet and Dowtian PROTO-role approaches to determining semantic prominence and the analysis I present here will make use of the model of discourse salience presented in Chapter 2.

Before beginning some of the theoretical background discussion of these experiments, I would like to make note of one terminological convention. In Chapter 2, I discussed two basic approaches to theories of mapping from argument structure to syntax: frame semantics (Fillmore, 1968, 1976) and thematic PROTO-roles (Dowty, 1991). For convenience, in many of the discussions that follow, I will use the frame semantics approach using discrete-role labels as provided in the FrameNet system (Baker et al., 1998). It turns out that these role labels—even if not very well defined—are very handy in discussions of the lexical semantics of verb constructions. However, I do not wish this decision of convenience to be seen as prejudicial. At relevant places in the chapter, I will discuss and analyze the results from the perspectives of both approaches.

3.2. Background

In this chapter, I am concerned with showing that some notion of semantic prominence is a psychologically real and observable feature of discourse processing mechanisms. The effect of semantic prominence is compared to the well-attested effect of syntactic prominence (see Section 1.3.2.2 for detailed discussion of these effects). However, as noted in the introduction to Chapter 1 and thereafter, syntactic and semantic prominence are often conflated: syntactic subjects are often semantic agents, and so on. Therefore, in order to be able to observe differences between the influence of syntactic and semantic

prominence, it is necessary to find some paradigm in which the two are distinguishable; in other words, contexts which differ minimally in their propositional content, but in which syntactic and semantic prominence make different predictions about the salience of discourse referents. One such context are constructions which reorder clausal arguments as in (95)-(99).

(95) a. John bought a book for Matt.

b. John bought Matt a book.

(96) a. Nancy took Susan to the party.

b. Susan was taken by Nancy to the party.

(97) a. The champ beat the challenger easily in the first round.

b. The challenger was easy for the champ to beat in the first round.

(98) a. Philip sprayed some paint on a wall.

b. Philip sprayed a wall with some paint.

(99) a. The audience enjoyed the musicians.

b. The musicians pleased the audience.

In (96), for example, there are two human discourse referents: NANCY and SUSAN. With respect to syntactic prominence, there is a difference between the two alternates. According to the syntactic prominence hierarchy in (100), employed in the role-based method for determining syntactic prominence introduced in Section 2.3, NANCY is more syntactically prominent than SUSAN in (96a), but SUSAN is more syntactically prominent in (96b). However, under a FrameNet semantic analysis, the event described fits in the CARRYING frame which would assign the agent role to NANCY and the theme role to SUSAN. Hence, assuming the semantic hierarchy in (101) (consistent with the thematic

role hierarchies proposed in Larson, 1988; Speas, 1990, inter alia), semantic prominence would predict that NANCY is the most salient entity in both (96a) and (96b).

(100) subject > object > oblique

(101) agent > theme

Passive constructions reorder arguments, but are not ideal for the present study because there is a structural confound between the active and passive alternates. In the passive version, the semantic agent is relegated to an optional oblique syntactic position. If, as has been long assumed, the lower part of the syntactic prominence hierarchy is accurate, then objects are more prominent than obliques. As such, there is no direct comparison between the active and passive alternates except at the subject position. In other words, one might compare the relative salience of agent-subjects to non-agent-subjects. While this would begin to answer the question at hand, there is another construction which will do more methodological work in the present investigation.

The central question of this dissertation is the question of whether syntactic prominence is might actually be explained by semantic prominence. The first step I will take towards answering this question is to seek to validate the syntactic prominence hierarchy while looking for effects of semantic prominence. The best construction for this from the list of argument-reordering constructions above is *spray/load* constructions. These constructions allow a reordering of their internal arguments with very little change in meaning. They will form the basis for the first set of experiments described in this chapter. To preview, the results of these experiments give converging evidence of a clear effect of semantic prominence as distinct from that of syntactic prominence.

If the effects of semantic prominence can be observed in this narrow and relatively nicely controlled context, then the next question is to see if these effects can also be observed in wider, less-controlled contexts. One argument-reordering construction which has been somewhat difficult to pin down in terms of a coherent syntactic and semantic analysis is the *tough*-construction as in (97). Differences between the salience of the entities in the *tough* and non-*tough* alternates may give a clearer picture of the syntactic and semantic nature of the *tough*-construction. Therefore, this construction will form the basis for a second set of experiments parallel to those with *spray/load* constructions. Results continue to show a clear and distinguishable effect of semantic prominence. Subsequently, taken together, the results of the two sets of experiments lead to some interesting conclusions about the prominence hierarchies.

3.2.1. Experimental Methodology

The experiments presented in this chapter consist of two parallel sets of five experiments. As such, it will be useful to discuss the basic experimental methodology here as a preview, and also to give some justification for the use of these particular experimental designs in this investigation. Very broadly speaking, each set of experiments consists of two different types: *off-line* and *on-line* experiments. In psycholinguistics, these terms are used to distinguish between the degree to which a direct observation of linguistic processes take place. Of course, it is impossible to, literally, directly observe language production or processing in cognition, but some methods are more direct than others. So-called off-line experimental tasks may rely more on participants' meta-linguistic judgments and are thus more indirect. They may also lead to exaggerated (or muted) effects of different

factors on linguistic performance. For example, one very common tool for such off-line experiments is pencil-and-paper surveys. However, with such experimental tools, having unlimited time and the opportunity to re-read a sentence could conceivably exaggerate the effects of the sentence's surface form on judgments about it.

On-line tasks, on the other hand, seek to observe the basic linguistic intuitions of participants by measuring different behaviors that occur during the real-time processing of language. These include such techniques as reading-time studies, eye-tracking studies, and event-related potential (ERP) studies. It is also true that off-line experiments can be performed more easily than on-line experiments. The latter may require far more strict institutional review (in order to, say, comply with regulations protecting the rights of human participants in experiments) and also may require the use of more expensive machinery and facilities. As such, off-line tasks are a commonly used starting point for long-term investigations. However, given their less direct nature, they may also lead to different conclusions than on-line experiments. Thus, converging evidence from both off-line and on-line experiments is regarded as ideal. The present investigation therefore makes use of both off-line and on-line tasks to test the hypothesis that semantic prominence is psychologically real.

The next two sections will describe these experimental procedures in a templatic form. The specifics of the actual stimuli will be discussed further below within each set of experiments.

3.2.1.1. Off-line Experiments. Two of the experiments in each set—the first and the last—were off-line tasks in the form of a pencil-and-paper survey. Experiment 1 was designed to get a relatively reliable, but not particularly sensitive, test of the main

hypothesis using a forced-choice task in which participants are asked to choose from a limited set of pre-defined options. Stimuli were presented in a format as shown in (102).

(102) John bought Matt a pair of tickets to *Les Miserables*.

a. He paid quite a lot of money for them.

b. He was so grateful to receive them.

In each item, the context sentence introduced two entities. The two continuation options, (a) and (b), each began with a pronoun and were followed by disambiguating information such that the first option made sense when the pronoun was resolved to one of the entities introduced in the context and the other option made sense only when the pronoun was resolved to the other entity. Experimental participants were then asked to choose which option made a more natural continuation of the discourse. In terms of the discourse model presented in Chapter 1, after reading the context sentence, the discourse representation (or DRS in DRT terms) contains two discourse referents in the store compatible with the subsequent pronoun (i.e., JOHN and MATT in the example above). Then, whichever of these two referents is more salient, if referred to in the next utterance, should be referred to with a pronoun. Therefore, the task used here is a useful way to begin to find out which entities in the context sentences are being interpreted by readers as most salient.

This experiment and all of the on-line experiments discussed below observed participants' reactions to sentences which begin with a pronoun. These sentences were written to make sense under only one interpretation of the pronoun. Therefore, it was important to test that participants were, in fact, likely to make this expected interpretation. In

other words, using (102) as an example, in order to have any conclusive power, it is important to know that participants resolved the pronoun in (102a) to JOHN and in (102b) to MATT. As such, the final experiment, Experiment 5, was designed as a follow-up test after all the other experiments. Once again, a forced-choice task was used and stimuli were presented as in (103).

(103) John bought Matt a pair of tickets to *Les Miserables*. He paid quite a lot of money for them.

a. He = John

b. He = Matt

Participants were asked to choose which of the two assignments of the pronoun in the second context sentence made the most sense *within the given context*.

For both of these experiments, stimuli were organized into blocks such that every block had the same number of items in each condition. However, crucially, the order of the options was also balanced in order to remove any effect of order or presentation. In other words, it is conceivable that somebody could read the context sentence and then always feel satisfied with the first option, no matter which entity it refers to. Balancing the order of presentation corrects for this possibility.

3.2.1.2. On-line Experiments. The off-line experiments described above are designed to get a relatively coarse picture of which of the two discourse referents introduced in the context sentence is deemed to be more salient by the experimental participants. The complementary on-line experiments are designed to get a more sensitive measure of the participants' judgments. This is done by using a common task in psycholinguistic investigations: a self-paced reading task with reading-time measurements. The underlying

assumption of this task is basically that given a particular context, a clause or other element which can be more readily processed in that context should be read more quickly than those which cannot. One factor which is assumed to influence this is pronominal reference: An utterance with a pronoun which is interpreted as referring to the most salient discourse referent in the context should be processed more easily—hence, read faster—than a comparable sentence in which the pronoun is interpreted as referring to a less salient discourse referent. To illustrate, consider the following somewhat exaggerated examples.

(104) a. John arrived home promptly at 5pm as usual.

b. His dog ran up carrying the newspaper.

(105) a. Mary approached the Senate floor and began her speech.

b. His dog ran up carrying the newspaper.

While the (b) sentences in these two "discourses" are equivalent, for a number of obvious reasons, (105b) seems particularly out of place, or more accurately, incoherent (cf., Kehler, 1995, 2002). Intuitively, one can see that (104b) is easier to read and thus should be processed *faster* than (105b) in their respective contexts. Self-paced reading tasks attempt to capitalize on this by taking reading-time measurements to locate such predicted processing difficulties. Participants are seated in front of a computer screen and view a discourse one chunk (whole sentences or sometimes smaller units) at a time. Participants read through the discourse at their own pace, pressing a button to make the next chunk appear (typically replacing the previous chunk). The time between each press of the button is taken as an estimate of the reading time.

In this investigation, Experiments 2-3 used a self-paced reading task. Stimuli, embedded in five-sentence vignettes, were presented one sentence at a time on a computer screen and whole-sentence reading times were measured. To illustrate, consider the following vignette.

(106) a. John and Matt went to the grocery store.

b. The store was having a sale on breakfast cereals.

c. In the cereal aisle, Matt found John mumbling to himself.

d. He couldn't decide between Cap'n Crunch and Cookie Crisp.

d'. He saw John looking at both Cap'n Crunch and Cookie Crisp.

e. Finally, because they were on sale, they bought both.

In the investigations, the critical sentence was always the fourth sentence. In the example in (106), if readers regard JOHN as more salient in the context than MATT, then it is predicted that they should read (106d) faster than (106d') because the pronoun is interpreted as referring to JOHN. In this way, the on-line experiments are designed to be a more sensitive gradient measure of readers' percpetions of discourse salience than the categorical measures used in the off-line experiments.

Experiment 2 therefore uses the paradigm just described as a clearer test of the hypothesis. However, there is a limitation in this design. The astute reader will have noticed that the relevant comparison is being made across *different* sentences: d and d', above. These two sentences make different assertions about different sets of discourse referents. While the length of these sentences may be controlled (and they are controlled to within one ascii character), controlling for such things as structural complexity or

lexical frequency effects is much more difficult. Therefore, Experiment 3a and Experiment 3b employ a design which overcomes this limitation. In Chapter 1, I made a crucial assertion about the relationship between the store and anaphoric expressions: Reference in the current utterance to the most salient discourse referent in the context should be made with a reduced-referring expression (see Section 1.1.4 for detailed discussion). This has been observed in a number of studies (cf., Arnold, 1998b; Stevenson et al., 1994). In particular, Gordon et al. (1993) observed what they call a *repeated-name penalty.*Using reading-time experiments, they observed that overspecified references (e.g., using a name instead of a pronoun) to an already highly salient entity results in longer sentence reading times. In the present research, this repeated-name penalty is useful to overcome the problem of comparing reading times across different sentences as mentioned above. The pronouns in the fourth sentence are thus alternated with names as in (107).

(107) d. He couldn't decide between Cap'n Crunch and Cookie Crisp.

d'. John couldn't decide between Cap'n Crunch and Cookie Crisp.

d". She saw John looking at both Cap'n Crunch and Cookie Crisp.

d"'. Mary saw John looking at both Cap'n Crunch and Cookie Crisp.

The reading times of the pronoun version of each continuation can be substracted from those of the name version in order to derive a repeated-name penalty. In this way, any structural or lexical effects are held constant, and the penalties reflect just the processing effects due to reference to a salient or non-salient referent. Comparing the penalties across the relevant conditions should then constitute an accurate test of the hypothesis. The prediction is that if the intended antecedent is salient, then the

repeated-name penalty will be larger. Thus, this method is useful to indicate which referents in the context are seen as salient.

The on-line experiments were all performed using Superlab Version 2.0 by Cedrus Corporation. Stimuli, organized into blocks with balanced conditions, were presented in pseudo-random order to every participant. Furthermore, in order to ensure that participants were reading each vignette in full, each stimulus item was followed by one *true/false/indeterminate* comprehension question for which feedback ("correct" or "incorrect") was given.

3.2.1.3. On-line versus Off-line Experiments. Above, I described the difference between on-line and off-line experiments in abstract terms, but here I note how I see the two experimental approaches complementing each other in the context of the present investigation. As I discussed, because off-line experiments involve less direct observation of phenomena than on-line experiments do, crucial results may be exagerrated (or even muted). In the present investigation, the off-line experiments involve a static questionnaire which participants must read and respond to. The task requires them to consider two different representations of the discourse, and to decide which of these is more "natural". While this decision certainly must tap into the intuitions which are crucially being investigated here, their ultimate decision is based on some meta-linguistic judgment, and this is what the experiment measures (i.e., its less direct nature). Thus other things—in addition to syntactic and semantic prominence—may influence their decisions: prescriptive notions about pronoun-antecedent relationships or even style preferences.

This is a theoretical limitation to the off-line experiments in this investigation, but there is also a methodological limitation. Although the instructions explicitly ask participants to mark their first intuitions and not to change any responses, there is nothing to prevent participants from re-reading each stimulus item several times before choosing a response. It is unclear to me how a reader might (re)process a discourse for which they already have some representation. I presume that the main reason a participant might re-read is because they did not feel confident about their comprehension after the initial pass. I can then imagine two possible but opposite effects. On the one hand, they may spend extra time concentrating on the structure of the utterance in order to reach comprehension. This could have the effect of exagerrating the structural information, or more specifically here, the syntactic prominence of the referents. On the other hand, concentrating on comprehension could mean—for some participants—concentrating on the semantic coherence of the discourse and thus exaggerated attention to the semantic information (roles, entailments, etc.). This could then result in exaggerated effects of semantic prominence. It might even be the case that there is so much variation in the approaches participants use that the results are too noisy to discern any reliable effects.

In short, then, for the present investigation, it is hoped that the off-line experiments give a clear indication of whether semantic prominence is a real, observable factor in discourse processing, but that where there is variance between the off-line and on-line experiments, the on-line experiments should be given greater value and should help to clarify any tentative conclusions from the off-line results.

3.2.1.4. Other Notes on Experiments. Two other experimental methodology notes are as follows. Participants in all experiments were Northwestern University undergraduates who were native speakers of North American English. Participants received course credit in return for their participation.

All statistics were performed using Systat version 6.0 by SPSS, Inc. The probability values for all planned comparisons reported here are Bonferroni-adjusted—a statistical procedure used to protect against Type I errors (i.e., a so-called false-positive conclusion) when several post-hoc tests are performed. The standard for rejecting the null hypothesis is raised by effectively lowering the alpha-level for significance testing.

3.3. Testing the Low End of the Prominence Hierarchies

3.3.1. *Spray/Load* Constructions

3.3.1.1. Overview. *Spray/load* constructions exhibit what has been called the *locative alternation* (Levin, 1993). In (98), repeated here as (108), the goal to which the theme argument (sometimes referred to as the *location* and *patient* arguments, respectively) is moved or applied is realized in either oblique position as in (108a) or in object position as in (108b).[1]

(108) a. Philip sprayed some paint on a wall.

b. Philip sprayed a wall with some paint.

[1]All members of the *spray/load* construction class exhibit the locative alternation, and as such, for the purposes of the present investigation I will treat it as a coherent, uniform class of verbs. However, there is some variation among the members of the class as to the acceptability of such variant forms as intransitive constructions or constructions without an oblique argument (see Boas, 2003; Goldberg, 1995; Pinker, 1989, for details). I assume that these variations are not relevant here because I will make use of only the locative variants with all three arguments overtly expressed.

The two variants of a *spray/load* verb have been regarded as nearly synonymous, differing slightly with respect to certain properties of the argument in object position. The first of these properties is that the argument in object position is subject to certain totality effects. Consider the following data from Tenny (1994, her (100) and (102), respectively).

(109) a. ??Jeremiah sprayed the paint on the wall (in five minutes) and there was half a can left over.

b. Jeremiah sprayed the wall with the paint (in five minutes) and there was half a can left over.

(110) a. Jeremiah sprayed the paint on the wall (in five minutes) and the wall was only half covered.

b. ??Jeremiah sprayed the wall with the paint (in five minutes) and the wall was only half covered.

The first conjunct of (109a) and (110b) seem to entail that the object argument is totally used up or covered, respectively. Hence the second conjunct is odd as a direct contradiction of this entailment.[2]

The second property of the object of *spray/load* verbs is that they can be interpreted as measuring-out the event described by the verb. The entity realized in object position

[2]Rappaport et al. (1987); Rappaport and Levin (1988) argue that the totality effect applies only in the *with*-variant (i.e., *spray the wall with the paint*) of these verbs, attributing the apparent totality effect in the *locative*-variant (*spray the paint on the wall*) to pragmatic and linguistic effects. For instance, they claim that when the theme is definite, then the totality constraint applies whether it is in object or oblique position.

(i) Bill loaded the hay onto the truck.
(ii) Bill loaded the truck with the hay. (from their foonote 16)

My personal judgment (and apparently that also of Tenny's—see (109b) above) is that there is no totality constraint on the theme-oblique in (ii).

has been called the incremental theme—the argument that measures out the event being described (cf., Baker, 1997; Dowty, 1991). Consider (111):

(111) a. John loaded a cart with some hay in an hour.

b. John loaded some hay onto a cart in an hour.

In (111a), the incrementality of the event can be seen with respect to how full the cart is during subintervals of the hour-long event. Furthermore, at the end of the hour, the cart is understood to be loaded by some standard of loadedness while on the other hand, there may still be leftover hay (i.e., the above-mentioned totality effect). However, in (111b), the event is still telic, but the entity which serves to measure out the event is the hay rather than the cart. That is, the progression of the event can be seen with respect to how much of the hay has been loaded, not how full the cart is. In the PROTO-role system of Dowty (1991), being an incremental theme is one of the PROTO-patient entailments and is therefore one of the determiners of which argument ends up in object position. Thus, under this analysis of *spray/load* constructions it is the crucial factor which distinguishes between the two alternates.

In the next two sections, I present accounts of the *spray/load* verbs under the frame semantics and PROTO-role theories.

3.3.1.2. A Frame Semantic Analysis. In a frame semantic account of the locative alternation, the *spray/load* verbs may be seen as members of a particular semantic frame which describes events in which movable entities are applied to surfaces or put into containers. It is this frame which imposes the totality constraint and incrementality of the argument in object position. In the FrameNet system, *spray/load* verbs are found in the FILLING frame which provides three frame elements: agent, theme, and goal.

In this system, then, one analysis would be to suppose that there are two entries (cf., Boas, 2003) for a *spray/load* verb as illustrated in (112) (using notation from Levin and Rappaport, 1986; Marantz, 1984)

(112) a. LOAD$_1$: agent < <u>theme</u> goal >

b. LOAD$_2$: agent < <u>goal</u> theme >

These two entries both map the external agent role into subject position, but differ in how the internal arguments are mapped: LOAD$_1$ maps the theme to object position and the goal to an oblique position while LOAD$_2$ maps the goal to object position and the theme to an oblique position. Goldberg (1995, 2002) proposes a slightly different analysis in which there is only one entry for the verb *load*, but it maps onto two different syntactic "constructions" which in turn correspond to the two alternates. For the present investigation, I am not as concerned with distinguishing between these two analyses. What is crucial for my hypothesis is the fact that the two different analyses assign *the same* semantic roles to their respective arguments.

3.3.1.3. A PROTO-role Analysis. The Dowtian PROTO-role analysis of *spray/load* verbs crucially depends on the status of the entity in object position as the incremental theme of the event. While Dowty (1991) discusses *spray/load* verbs in this respect, he does not give a full analysis. Shown in Table 3.1 is my own analysis of the PROTO-role entailments for the two variants of *load*. My judgments here are based on the linguistic tests for PROTO-role entailments which I described in Section 2.7.1.

The PROTO-role analysis captures the differing semantic entailments on the arguments, yet neatly explains how the arguments are mapped: the argument with the most

Table 3.1. PROTO-role Entailments for *Load*

	The farmer	loaded	**some hay**	onto	**a cart**
sentience	√				
volition	√				
cause event/c-o-s	√				
movement			√		
undergo c-o-s			√		
causally-affected			√		
incremental theme			√		
stationary					√

	The farmer	loaded	**a cart**	with	**some hay**
sentience	√				
volition	√				
cause event/c-o-s	√				
movement					√
undergo c-o-s			√		
causally-affected			√		
incremental theme			√		
stationary			√		

PROTO-agent entailments becomes the subject, the (remaining) argument with the most PROTO-patient entailments becomes the object and so on.

3.3.1.4. Methodological Usefulness. In this section, I discuss why *spray/load* verbs are particularly useful for testing the central hypotheses of this dissertation. For the purposes of this discussion I will use the skeletal discourses shown in (113)-(114).

(113) a. John sprayed some paint on a wall.

b. It ...

(114) a. John sprayed a wall with some paint.

b. It ...

After the (a) sentences in both (113) and (114), the store in the discourse representation will contain three discourse referents corresponding to JOHN, PAINT, and WALL. The processor must resolve the pronoun at the beginning of the (b) sentence to one of these three referents. In order to do so, the most salient compatible discourse referent must be determined. According to the salience model presented in Chapter 2, the store is first filtered of incompatible referents. In the present example, that would leave two competing discourse referents: PAINT and WALL. So the next question is which of these two is more salient. In order to determine this, we must determine the syntactic and semantic prominence of the referents.

First, let's consider their relative syntactic prominence. Two methods for determining syntactic prominence were introduced in Section 2.3. The role-based method is based on the syntactic hierarchy shown in (115). According to this method, in (113a), PAINT should be more prominent than WALL, while in (114a), WALL should be more prominent than PAINT.

(115) subject > object > oblique

The second method, the hierarchical method, is based on a top-to-bottom, left-to-right clausal search algorithm. However, since there is only one main clause in these sentences, the ranking reduces to a simple left-to-right order. In (113a), PAINT is more prominent, while in (114a), WALL is more prominent. Thus, the two methods make the same predictions.

Next let's consider the relative semantic prominence of the discourse referents. For the frame semantic approach, I will assume, as a starting point for this research, a

hierarchy of thematic roles as shown in (116) which is consistent with the thematic hierarchies proposed in Larson (1988) and Speas (1990).

(116) agent > theme > goal

In terms of this hierarchy, the discourse referents should have the same relative semantic prominence in both (113a) and (114a): that is, PAINT should be more prominent than WALL.

For the PROTO-role approach, I have no a priori evidence as to the relative ranking of the entailments. However, one ranking scheme which would parallel that of the thematic hierarchy under the frame semantics approach is shown in (117).

(117) movement > not movement

The argument analyzed as a theme in the frame semantic analysis is the argument which bears a movement entailment in the PROTO-role system, while the argument analyzed as a goal in the frame semantic analysis is the argument which bears a stationary entailment in the PROTO-role system. Thus, the entailment ranking shown in (117) reflects the corresponding ranking of arguments in (116). Under this entailment ranking, then, PAINT should be more prominent than WALL in both discourses—the same prediction as under the frame semantic approach.[3]

Now consider the summary of the predicted rankings as shown in (118)-(119).

[3]An alternative ranking schema not investigated here but probably worth considering in the future is ranking via the number of satisfied entailments. For instance, it might be the case that the semantic prominence of a referent is directly related to the number of PROTO-agent entailments it bears. Thus, referring to Table 3.1, the prominence relations for both *the farmer loaded hay into the cart* and *the farmer loaded the cart with hay* would be FARMER > HAY > CART. One motivation for such an approach might come from considering how the semantic entailments are processed. In DRT terms (see Section 1.1.3), the entailments would presumably be added as further semantic conditions on the each discourse referent. It might be the case that processing more of these entailments, particularly PROTO-agent entailments, causes the associated referent to be more salient in the discourse. A procedure something like this is employed in Chapter 4 as part of representing semantic prominence as a single, discrete value.

(118) a. John sprayed some paint on a wall. CONTROL

syntactic prominence: PAINT > WALL

semantic prominence: PAINT > WALL

b. It ...

(119) a. John sprayed a wall with some paint. SPLIT

syntactic prominence: WALL > PAINT

semantic prominence: PAINT > WALL

b. It ...

In (118a), the syntactic and semantic prominence rankings are nicely aligned: that is, the object-theme, PAINT, is more syntactically and semantically prominent than the oblique-goal, WALL. In the experiments which follow, I will thus refer to this configuration as the CONTROL condition. However, in (119a), the syntactic and semantic prominence rankings are not aligned: the goal, WALL, is syntactically more prominent, but the theme, PAINT, remains semantically more prominent. I will refer to this configuration as the SPLIT condition.

These two different conditions yield different predictions about the relative salience of the (compatible) discourse referents for reference by the pronoun *it*. If syntactic prominence is the only factor that determines discourse salience (and semantic prominence has no effect) then the referent realized as the object should always be the preferred referent for subsequent pronominal reference; PAINT in the CONTROL condition, WALL in the SPLIT condition. On the other hand, if semantic prominence is the only relevant factor, then the theme, PAINT, should be the preferred referent in both the CONTROL and SPLIT conditions. Hence, the results of the experiments should be able to give a

clear picture of the relative effects of syntactic and semantic prominence. This is the first methodological usefulness of *spray/load* constructions in the present investigation.

The second usefulness of *spray/load* constructions is that they permit a test of the lower end of the prominence hierarchies as in (120)-(121) as well as the entailment hierarchy in (122).

(120) subject > object > oblique

(121) agent > theme > goal

(122) movement > not movement

The upper part of the syntactic prominence hierarchy (i.e., subject > object) has been demonstrated in many experiments (Gordon and Chan, 1995; Gordon et al., 1999; Hudson-D'Zmura and Tanenhaus, 1997; Lappin and Leass, 1994; Mathews and Chodorow, 1988). However, the lower end of the hierarchy has only been assumed and has not been explicitly verified, experimentally. The semantic prominence hierarchy has not been directly investigated at all, although one might make the case that the results of previous investigations comparing subjects and objects give implicit evidence of the upper end of the hierarchy. However, the present investigation (as far as I can tell) is the first to explicitly investigate the lower ends of the syntactic and semantic prominence hierarchies.

3.3.1.5. Items. Finally, before describing the experiments in detail, in this section I will give a brief description of the actual *spray/load* verbs used and why they were used.

Ten putative *spray/load* verbs listed in Levin (1993) and having a lemma frequency index of at least 100 (i.e., occurrences per 17.9 million words) in the COBUILD English corpus were selected for the following experiments. These verbs are shown with their respective frequencies in Table 3.2.

Table 3.2. *Spray/Load* Predicates Used in Experiment 1

	predicate	frequency index*
"best"	load	493
	pack	688
	rub	729
	spread	1578
	stuff	383
"good"	drape	133
	heap	100
	inject	110
	pile	371
	sprinkle	204

*number of occurrences per 17.9 million words in COBUILD English Corpus

Verbs in the "best" group had both of their alternates (e.g., *load x into y* and *load y with x*) attested in the 500-text sample of the Brown Corpus.[4] Verbs in the "good" group had one of their alternates attested and the other attested in a relative clause form (e.g., *the hay (which was) loaded into the cart* or *the cart (which was) loaded with hay*).[5]

These ten *spray/load* verbs were used in Experiments 1-4 reported below. For reference, the full set of stimulus items used in these experiments is provided in the Appendix.

3.3.2. Experiment 1: Discourse-completion Questionnaire

3.3.2.1. Design. The first experiment I describe in this chapter was designed to give a rough estimate of the relative influence of syntactic and semantic prominence through the use of an off-line pencil-and-paper task. In short, the goal of the experiment is to get an idea of which discourse referents are viewed as more salient in different contexts.

[4]This corpus is a sample of the full Brown Corpus comprising 500 extended samples.

[5]It is interesting to note that while *spray* is undeniably a member of this class of verbs and is often regarded as a paradigmatic member, it exhibited the alternation relatively infrequently in the 500-text sample of the Brown Corpus and therefore was not used in this study.

According to the discourse processing model outlined in Chapter 1, the most salient entity in the current context, if referred to in the next utterance, should be referred to pronominally. Thus, readers should prefer an utterance in which a pronoun is most easily interpreted as referring to the most salient entity over an utterance which does not.

3.3.2.2. Method.

Participants. Participants in this experiment include 32 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. Stimuli for Experiment 1 were designed as shown in (123)-(124).

(123) John sprayed some paint$_i$ on a wall$_j$. CONTROL

a. It$_i$ dribbled down and made a mess. theme

b. It$_j$ was big and needed two coats. goal

(124) John sprayed a wall$_j$ with some paint$_i$. SPLIT

a. It$_i$ dribbled down and made a mess. theme

b. It$_j$ was big and needed two coats. goal

The discourse continuation sentences ((a) and (b) in the examples) were written such that only one interpretation of the pronoun was semantically coherent with the context sentence (i.e., in (a), the intended referent is the theme; in (b), the intended referent is the goal).

Fifty such stimili were prepared using the *spray/load* verbs mentioned above and were randomized into a paper survey along with 102 filler items (from an investigation unrelated to the present one). Stimuli were presented in the survey without coindexation, as shown in (125)-(126).

(125) John sprayed some paint on a wall.

a. It dribbled down and made a mess.

b. It was big and needed two coats.

(126) John sprayed a wall with some paint.

a. It dribbled down and made a mess.

b. It was big and needed two coats.

Procedures. For each item, participants were asked to mark the discourse continuation sentence (a or b) which seemed more natural to them given the context. As such, the experiment had only one main factor with two levels, CONTROL and SPLIT, which represent the differing context sentences. Hereafter I will call this the CONTEXT factor.

Scores were tallied with respect to the proportion of responses in which the theme continuation was chosen. For this experiment, the specific predictions are that in the CONTROL condition, there should be a clear preference for the theme continuation. However, their preferences in the SPLIT condition (if any) will indicate the relative influence of syntactic and semantic prominence.

3.3.2.3. Results. The results of Experiment 1 are shown in Table 3.3. A one-sample *t*-test shows that in the CONTROL condition, when the syntactic and semantic prominence rankings are aligned, participants prefer the continuation which refers to the predicted more syntactically and semantically prominent entity, the theme, 72% of the time, significantly greater than chance. However, in the SPLIT condition, when the syntactic and semantic prominence rankings are not aligned, there is still a numerical preference for the syntactically prominent entity, but this is marginally significant only by participants.

Table 3.3. Experiment 1: Proportion of Choices for Theme Continuation

	theme	*by participants*	*by items*
CONTROL	72%	$t(31) = 8.7\ p < 0.001$	$t(49) = 7.6\ p < 0.001$
SPLIT	44%	$t(31) = 2.2\ p = 0.066$	$t(49) = 1.99\ p < 0.1$

Table 3.4. Experiment 1: Proportion of Choices for Theme Continuation - *Rub* Stimuli

	theme	*by participants*	*by items*
CONTROL	46%	$t(31) < 1.0\ n.s.$	$t(4) < 1.0\ n.s.$
SPLIT	21%	$t(31) = 6.0\ p < 0.001$	$t(4) = 7.7\ p < 0.001$

Closer inspection of the results in terms of the individual *spray/load* verbs revealed some very interesting results. While most of the verbs pattern the same as the mean, showing a clear preference in the CONTROL condition and a much more vague preference in the SPLIT condition, one verb, *rub*, shows a very different pattern as illustrated in Table 3.4.

At first glance, it seems that the prominence hierarchy in the *rub* case seems to be reversed (i.e., goal > theme) such that the CONTROL and SPLIT conditions show the opposite pattern of results. However, closer inspection reveals a more likely explanation: *rub* is not a *spray/load* verb. While *rub* certainly does exhibit an apparent alternation, its properties differ from that of the locative alternation. Note first that the totality effects do not hold in the same way. Consider (127).

(127) a. Mark rubbed the wax on the surfboard (in five minutes) and there was still half a can of wax left over.

b. Mark rubbed the surfboard with the wax (in five minutes) and the surfboard was only half covered.

In addition, while *rub* is clearly a three-place predicate, there is another sense of rub which is a two-place predicate assigning agent and patient roles. This different sense is shown in (128).

(128) Mark rubbed the surfboard (lovingly).

In short, what might have happened is that there was temporary parsing ambiguity which caused confusion about the roles of the arguments (and hence, their semantic prominence). To explain this in more detail, let's consider how the parser handles the SPLIT condition. When the parser encounters the NP *the surfboard*, it does not yet have enough information to disambiguate which sense of *rub* being used here (either (127b) or (128)). But these two senses assign completely different roles to THE SURFBOARD—one, a patient role, the other, a goal role. If the parser temporarily maintained *both* parses until disambiguating information is encountered then, hypothetically, it could be the case that even after disambiguation, the effect of THE SURFBOARD as a patient lingers.[6] That is, once the surfboard has been considered as a possible patient, it remains semantically prominent as a patient. Therefore, participants saw it as a more salient entity than would be predicted under a three-place analysis of *rub*.

Another interesting piece of data in the current experiment concerns the relative salience of animate entities. Two of the stimuli included animate entities as goal arguments (*inject some poison into a dog* and *inject some serum into a rat*). Given the evidence that animate entities are inherently more salient than inanimate entities (see Prat-Sala and Branigan, 1999), it is not surprising to find that in these two stimuli,

[6] A similar scenario is described by Christianson et al. (2001) as a failure to reanalyze.

Table 3.5. Experiment 1: Proportion of Choices for Theme Continuation - Revised Results

	theme	*by participants*	*by items*
CONTROL	77%	$t(31) = 10.6\ p < 0.001$	$t(42) = 12.2\ p < 0.001$
SPLIT	48%	$t(31) < 1.0\ n.s.$	$t(4) < 1.0\ n.s.$

participants consistently preferred the goal continuation across both CONTROL (78.1% choices for goal) and SPLIT (87.5% choices for goal) conditions.

Removing the *rub* stimuli and the two stimuli with animate entities yields the revised results shown in Table 3.5. One-sample t-tests show that the preference for the theme in the CONTROL condition is significantly greater than chance, but in the SPLIT condition, there is no significant preference.

3.3.2.4. Discussion. The revised off-line experimental results in the CONTROL condition are consistent with earlier findings suggesting a syntactic prominence hierarchy as discussed above. However, the results in the SPLIT condition are particularly surprising and cannot be explained by the influence of either syntactic prominence or semantic prominence alone. Rather, it appears that *both* syntactic prominence and semantic prominence influence the salience of discourse referents. That is, in the SPLIT condition, the syntactic prominence hierarchy promotes the goal while the semantic prominence hierarchy promotes the theme in equal degrees such that readers cannot decide which of the two is more salient for subsequent pronominal reference. Hence, choice between the two continuations is essentially at chance.

The model of discourse salience I described in Chapter 2 illustrates this situation very well as shown in Figure 3.1. Given the syntactic and semantic prominence hierarchies shown above, the store looks like the left-most graph after the context sentence. After

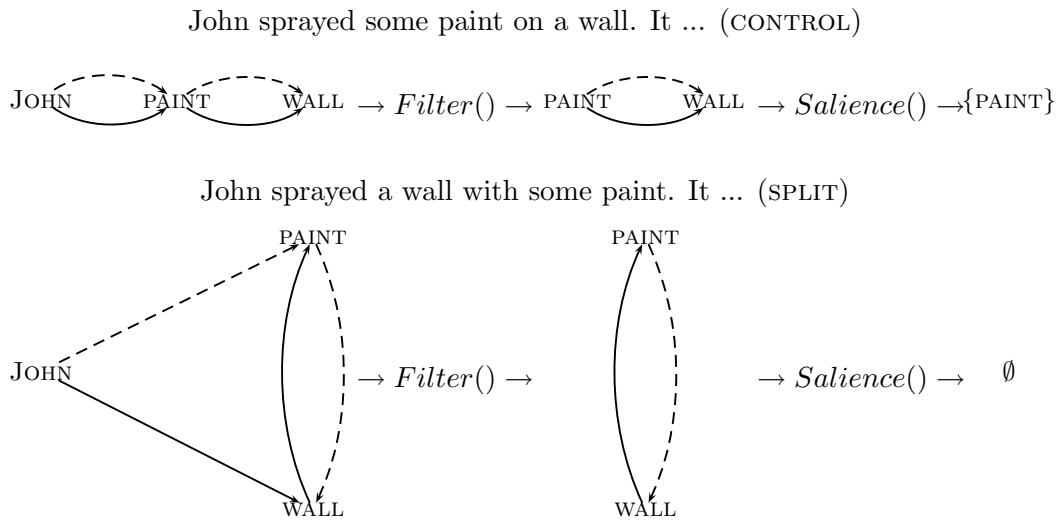


Figure 3.1. A graphical illustration of the context store based on the results of Experiment 1. The figure shows the process from the beginning structure of the store to the (default) resolution of the pronoun. syntactic prominence relations are shown with solid edges and semantic prominence relations are shown with dashed edges.

filtering, the store is reduced to the middle picture. From these filtered stores, the salience operation can determine the maximal discourse referent in only the CONTROL condition: In the SPLIT condition, the operation returns the empty set, $\emptyset$.

These results have implications for models of discourse salience. In particular, it is problematic for relatively simple approaches to discourse salience which attribute salience solely to such surface features as linear order (cf., Gernsbacher and Hargreaves, 1988) or grammatical role (as in Centering Theory). Instead, it appears that discourse salience is determined by a more complex set of features: the syntactic position as well as the semantic role in which a particular discourse referent is realized contribute independently to that referent's salience for subsequent pronominal reference. Furthermore, if my

analysis here is correct, then it is not the case that for every grammatical construction, a fully-ordered salience ranking of the entities in that utterance obtains. Rather, it seems that some constructions result in only partial orderings as predicted by the model of discourse salience outlined in Chapter 2.

Finally, the results lend initial support to the validity of the lower parts of the syntactic prominence hierarchy (i.e., object > oblique) and the semantic prominence hierarchy (i.e., theme > goal).

3.3.3. Experiment 2: Self-paced Reading Task

3.3.3.1. Design. Experiment 1 confirmed one of the predictions of the model: In cases where the most salient discourse referent (i.e., the maximal vertex in the graph) cannot be determined, certain consequences result. In the pencil-and-paper survey—an off-line task—these consequences showed up as the lack of an advantage for either the theme or goal discourse continuations in the SPLIT conditions. In an on-line task such as a reading time study, the model then predicts that there should be no reading time advantage for either of these continuations. The second experiment is designed to evaluate the relative effects of syntactic and semantic prominence by testing this prediction in such an on-line task.

3.3.3.2. Method.

Participants. Participants in this experiment included 32 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. After removing the problematic *rub* items and the items with animate entities, the remaining stimuli from the first experiment were revised and extended with some new items to make a set of 48 stimuli.[7] These were embedded in five-sentence vignettes as shown in (129). The first sentence of the vignette introduces a single human character and two inanimate entities. The human character continues as the subject through the third sentence, while the two inanimate entities are introduced in comparably prominent syntactic positions (typically object) and semantic positions (typically causally-affected patients). Therefore, while the human entity is clearly the most salient referent after the first sentence, the referents for the inanimate entities are arguably comparably ranked. The second sentence does not mention either of the two inanimate entities at all. The third sentence, here the CONTEXT sentence, varied between the CONTROL and SPLIT conditions while the fourth sentence, the target sentence for reading time measurements, varied between the intended REFERENT of the pronoun: the theme or goal. It is important to note that while the two variants of the target sentences used different syntactic constructions and words to establish coreference between the pronoun and different entities, the ascii character length of the sentences was kept constant, differing at most by one character.[8] Finally, the fifth sentence concluded the vignette. Each vignette was followed by one *true/false/indeterminate* question. The questions were deliberately made very difficult in order to encourage participants to read each vignette carefully for comprehension.

[7] In addition to the five *rub* items and two items with animate entities, there were two other items which which seemed problematic upon further review: In particular, for one or more of the continuation options in these two items, it was slightly ambiguous what the intended antecedent of the pronoun was. Hence a total of nine items were removed. New items were added to bring the total up to 48.

[8] See Ferreira and Clifton (1986) for evidence that ascii character length is the most reliable predictor of sentence reading times.

(129) a. John brought some paint into his living room and stared at one of the walls.

b. He was trying to decide how to redecorate his home.

c. John sprayed the paint$_i$ on the wall$_j$. CONTROL

c'. John sprayed the wall$_j$ with the paint$_i$. SPLIT

d. It$_i$ dribbled down and made a mess. PATIENT

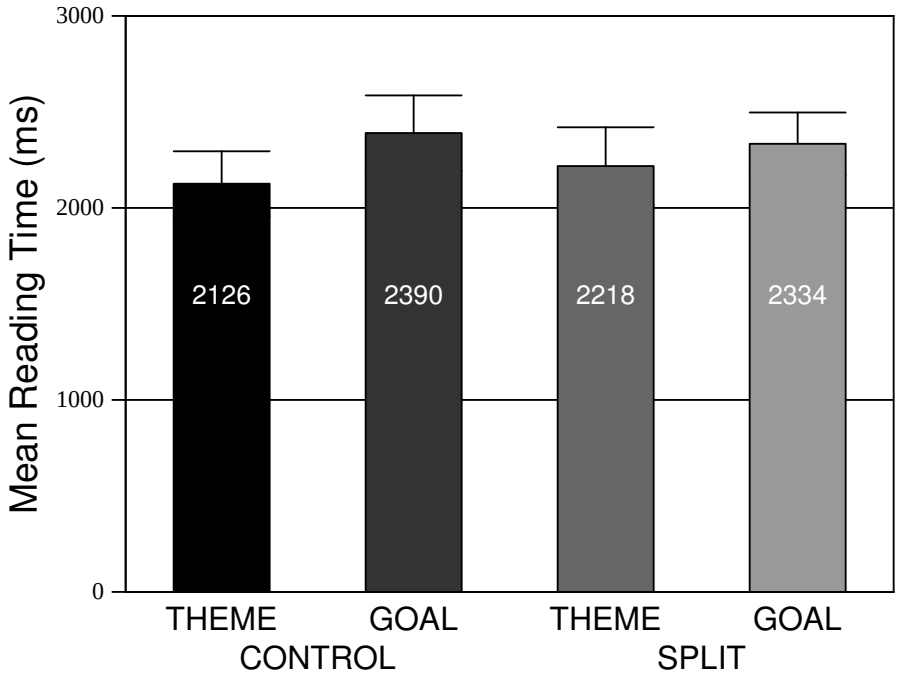
d'. It$_j$ was big and needed two coats. LOCATION

e. The job took John several days to finish.

question: John whitewashed his living room in several days.

Procedures. The vignettes were presented on a computer screen one sentence at a time using Superlab, Version 2.0 by Cedrus Corp. The experiment was self-paced with participants pressing a button to proceed to the next sentence. Reading time measurements for all sentences were recorded although only the reading times of the target sentence (d) are reported here.

3.3.3.3. Results. Results, as shown in Figure 3.2, indicate a main effect of REFERENT but no main effect of CONTEXT and no significant interaction between the two. In short, participants found it easier to read the continuation which referred to the semantic theme of the context sentence, irrespective of whether it was realized in object or oblique syntactic position. In contrast with the off-line results above which showed an effect of both syntactic and semantic prominence, there appears to be no clear effect of syntactic prominence in these on-line reading time measurements. Planned comparisons, however, reveal a slightly diferent picture: Direct comparison of the target sentence reading times with each of the two context conditions shows that in the CONTROL condition, the difference between the theme and goal continuations is significant [by participants,



	by participants		*by items*	
CONTEXT	$F(1,31) < 1.0$	n.s.	$F(1,47) < 1.0$	n.s.
REFERENT	$F(1,31) = 11.1$	$p < 0.005$	$F(1,47) = 5.6$	$p < 0.05$
CONTEXT*REFERENT	$F(1,31) = 2.0$	$p = 0.17$	$F(1,47) = 3.1$	$p = 0.09$

Figure 3.2. Mean reading times with 95% confidence intervals for the continuation sentences, (129d-d'), for participants ($n = 32$) in Experiment 2. Two main factors were tested: CONTEXT (CONTROL, SPLIT) and intended REFERENT of pronoun (theme, goal).

$t(31) = 3.6$ $p < 0.01$; by items, $t(47) = 3.0$ $p < 0.05$] while in the SPLIT condition, this difference is not significant. Hence if there is any effect of syntactic prominence in the current experiment, it is weaker than that of semantic prominence.

3.3.3.4. Discussion. The on-line results provide an interesting contrast to those of the off-line experiment. While the results may initially appear to be contradictory about the effect of syntactic prominence, closer examination of the results and the nature of the experiments reveals that they are, in fact, quite consistent. In the off-line experiment,

participants could continue to see the context sentence as well as both continuation options while making a decision about which option was more natural. Thus, participants' judgments about the naturalness of the continuations could have been based on the mental representation of the discourse *plus* the visual representation on the page. This may have exaggerated the influence of the syntactic information in their conceptualization of the discourse salience of the relevant entities. Hence, upon reflection, neither continuation in the SPLIT condition may have seemed more natural than the other. However, in the on-line experiment, participants saw only one sentence at a time on the screen, so when they read the continuation sentences, they had access only to the rankings of the discourse referents in the mental representation—the visual representation was no longer available. Therefore, the latter on-line results are likely to be a more sensitive—and therefore more accurate—measure of the relative effects of syntactic and semantic prominence on the discourse salience of discourse referents.

These results suggest that a revision to the earlier conclusions about the prominence hierarchies may be necessary. The results provide clear evidence for the lower part of the semantic prominence hierarchy, but weaker evidence for the lower part of the syntactic prominence hierarchy. One way to analyze this is to suggest that objects are not much more prominent than obliques. This is in contrast to what has previously been assumed. Perhaps it is the case that the syntactic prominence hierarchy is less ordered than has been thought and that the only real difference is that between external and internal syntactic arguments.

However, there is some ambiguity in the results of this experiment. There seems to be an overall preference for the theme continuation, suggesting that participants saw both

context conditions in the same way with (hypothetically) the same salience representation. But closer inspection shows that the theme continuation preference is significant only in the CONTROL condition. Yet, alas, this difference between the contexts does not translate into a significant interaction between CONTEXT and REFERENT. While the on-line design here is more sensitive than that of the off-line questionnaire in Experiment 1, it may still not be quite sensitive enough to detect the subtle differences between the two context conditions. Thus, further experimentation is required to look at these two conditions carefully. This is the aim of the following two experiments.

3.3.4. Experiments 3a-b: Repeated-description Penalty

3.3.4.1. Design. The emerging picture seems to be that both SYNTACTIC and semantic prominence play a role in determining the discourse salience of entities, but that semantic prominence may play a greater role at the lower end of the prominence hierachies. These conclusions depend crucially on the results of Experiment 2. However, the design of that experiment has one crucial problem which was discussed in the methodological background section above: The reading times which are being compared are reading times on different sentences. Experiment 3a and Experiment 3b are designed to address this problem by taking advantage of the *repeated-name penalty* effect of Gordon et al. (1993). In their experiments, increased reading times are observed for continuation sentences in which a repeated name instead of a pronoun was used to refer to a salient discourse referent in the context. This provides a useful test to determine which referents are seen by readers as being more salient. In order to make use of this approach in the current investigation, I would like to logically extend the repeated-name penalty

to what I will call the repeated-definite-description penalty. While the repeated-name penalty has not been explicitly extended to the repeated-definite-description penalty in any experiment that I know of, Guindon (1985) shows results that are consistent with such an extension (however, see Almor, 1999, for different evidence on repeated-definite-description penalty).

In the discourse model discussed in Chapter 1, I assume that when a name is encountered in a discourse, a referent is introduced into the store. The store is not consulted to search for an antecedent and therefore the salience of existing referents does not play a role in the immediate procesing of names. Thus, in reading-time experiments, one would expect to see no difference within a NAME condition.[9] On the other hand, my model does predict that definite descriptions will initiate a search in the store. The relative salience of referents in the store matters only as a basis for removing (or ignoring) the most salient referent (which is reserved for pronominal reference). Thereafter, remaining referents are all equally available (although this may be one place where recency could make some difference). Therefore, in the present reading-time experiments in which the agent is the most salient referent, but not a target of the pronouns, there should be no difference within the DESCRIPTION condition. So methodologically, in the current set of experiments, the repeated-definite-description penalty should generate the same results as the repeated-name penalty although theoretically, the discourse processes differ.

[9]Gordon and Hendrick (1997a,b) describe an extension of DRT which they call Discourse Prominence Representation in which the rule (called construction rules in DRT formalism) by which names are processed involves merely introducing a discourse referent into the store, but no consultation of the existing state of the store. A further construction rule is triggered to establish equivalence among the referents introduced by repeated names, but in the experiments which follow, the NAME conditions will always be repeated names. Therefore, no difference is anticipated within the NAME condition. Therefore, theoretically, it may more accurate in the present context to characterize the effect as a "pronominal advantage" rather than a "repeated-name penalty".

For the current investigation with *spray/load* verbs, this method is therefore predicted to provide a means of obtaining a more gradient measure of how readers perceive the relative salience of the theme and goal arguments. If, as mentioned in the discussion for Experiment 2, the two contexts actually have the same salience representation, then the theme and goal conditions should show the same repeated-definite-description penalty across both conditions. Any difference between the two, however, would be attributable to different salience representations.

3.3.4.2. Method.

Participants. Participants in each of experiments 3a and 3b included 32 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. Stimuli were in the format as shown in (130).

(130) a. John brought some paint into his living room and stared at one of the walls.

b. He was trying to decide how to redecorate his home.

c. John sprayed the paint$_i$ on the wall$_j$. CONTROL

c'. John sprayed the wall$_j$ with the paint$_i$. SPLIT

d. It$_i$/the paint$_i$ dribbled down and made a mess. AGENT

d'. It$_j$/the wall$_j$ was big and needed two coats. THEME

e. The job took John several days to finish.

question: John whitewashed his living room in several days.

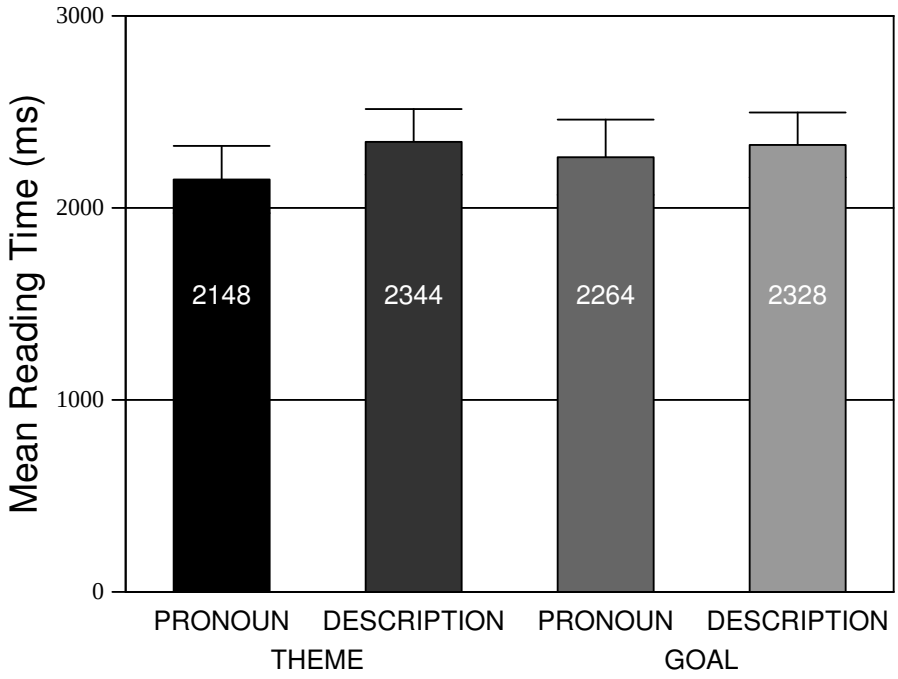
In order to increase the statistical power of the experiment, each sub-experiment focused on only one of the two main context conditions: Experiment 3a used the SPLIT condition (i.e., (130c')) while Experiment 3b used the CONTROL condition (i.e., (130c)).

Then the continuation sentences were varied between the pronoun and the definite-description condition.

Each experiment is therefore a 2 × 2 design varying between the intended REFERENT of the pronoun (theme, goal) and its referential FORM (PRONOUN, definite DESCRIPTION). The interaction between REFERENT and FORM is used to evaluate the relative repeated-definite-description penalty between the goal and theme continuation sentences.

Procedures. The procedure for these experiments was exactly the same as for Experiment 2. Participants read the vignettes one sentence at a time on a computer screeen in a self-paced reading task, and reading times (elapsed time between button-presses) were recorded. While the experiments are identical in design (differing only in the materials used), they were run independently of each other.

3.3.4.3. Results. Results of Experiment 3a are shown in Figure 3.3. First, it is useful to note that the results of this experiment replicate part of the results of Experiment 2. In the SPLIT condition of Experiment 2, there was a nonsignificant numerical trend in favor of the theme over the goal continuation. Those two conditions reappear in the present experiment as the PRONOUN conditions and show the same pattern of results: a nonsignificant trend in favor of the theme. As for the overall results, there was a main effect of only FORM and no significant interaction between REFERENT and FORM. Overall, participants found the continuation sentences easier to read when a pronoun was used instead of a definite description. In short, there is an overall reading-time penalty for definite descriptions. However, planned comparisons suggests that this penalty was not constant across the REFERENT conditions. The specific penalties are shown in Table 3.6. One-sample t-tests show that the 188ms penalty in the theme condition is significant



	by participants		*by items*	
REFERENT	$F(1,31) = 1.8$	*n.s.*	$F(1,47) < 1.0$	*n.s.*
FORM	$F(1,31) = 7.0$	$p < 0.05$	$F(1,47) = 4.6$	$p < 0.05$
REFERENT*FORM	$F(1,31) = 1.0$	*n.s.*	$F(1,47) < 1.0$	*n.s.*

Figure 3.3. Mean reading times with 95% confidence intervals for the continuation sentences, (129d-d'), for participants ($n = 32$) in Experiment 3a—the SPLIT condition from Experiment 2. Two main factors were tested: intended REFERENT (theme, goal) and referential FORM (PRONOUN, definite DESCRIPTION).

although the 77ms penalty in the goal condition is not. A pairwise t-test between the two penalties was also not significant [by participants, $t(31) = 1.7$ *n.s.*; by items, $t(47) < 1.0$ *n.s.*].

Results of Experiment 3b are shown in Figure 3.4. Again, it is important to note that the PRONOUN conditions in the present experiment replicate the results of the CONTROL condition in Experiment 2, showing significantly faster reading times in the

Table 3.6. Experiment 3a: SPLIT Condition Penalties

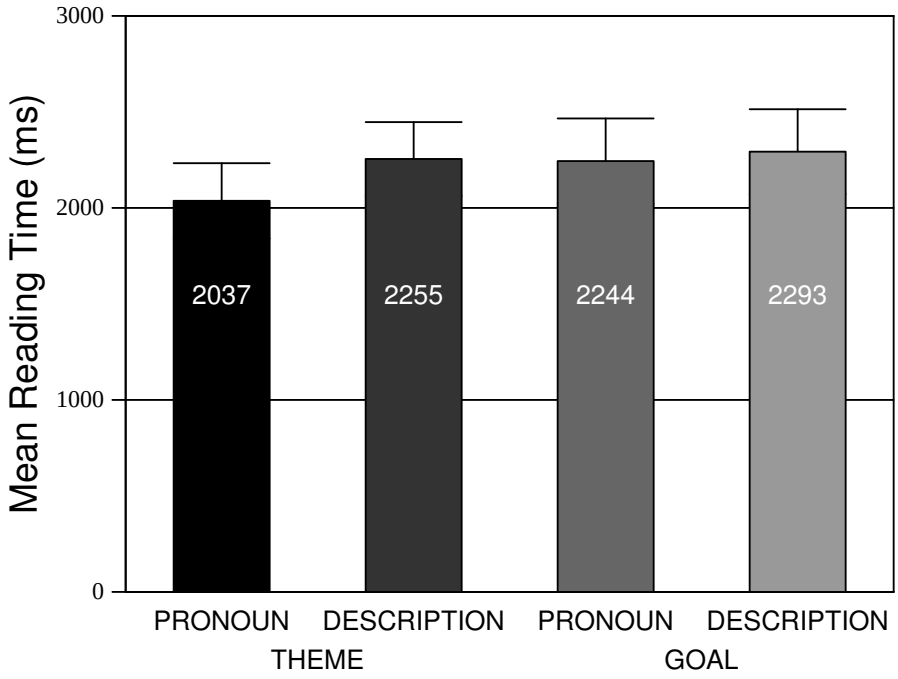
	penalty	*by participants*	*by items*
theme	188ms	$t(31) = 3.6\ p < 0.005$	$t(47) = 2.4\ p < 0.05$
goal	77ms	$t(31) < 1.0\ n.s.$	$t(47) < 1.0\ n.s.$

Table 3.7. Experiment 3b: CONTROL Condition Penalties

	penalty	*by participants*	*by items*
theme	218ms	$t(31) = 3.5\ p < 0.005$	$t(47) = 3.5\ p < 0.005$
goal	49ms	$t(31) < 1.0\ n.s.$	$t(47) < 1.0\ n.s.$

theme condition than in the goal condition [by participants, $t(31) = 3.1\ p < 0.05$; by items, $t(47) = 2.6\ p = 0.067$]. As for the overall results, there were significant main effects of both REFERENT and FORM as well as a significant interaction between the two. Overall, participants found the theme continuation easier to read than the goal continuation. Participants also found the PRONOUN continuations easier to read than the DESCRIPTION continuations. This is consistent with the results of Experiment 3a. Finally, the significant interaction between the two gives a clearer picture of the penalties across the conditions. The calculated penalties are shown in Table 3.7. One-sample t-tests show that the 218ms penalty in the theme condition is statistically significant while the 49ms penalty in the goal condition is not. The difference between the penalties is marginally significant [by participants, $t(31) = 1.7\ p = 0.1$; by items, $t(47) = 2.4\ p < 0.05$].

Finally, it is notable that there was no significant difference across the DESCRIPTION conditions in either the SPLIT or CONTROL cases. This is consistent with one of the assumptions I made about the repeated-definite-description penalty as a logical extension of the repeated-name penalty in the introduction to these experiments, above.



	by participants		*by items*	
REFERENT	$F(1,31) = 8.2$	$p < 0.01$	$F(1,47) = 2.8$	$p = 0.1$
FORM	$F(1,31) = 11.5$	$p < 0.005$	$F(1,47) = 6.4$	$p < 0.05$
REFERENT*FORM	$F(1,31) = 2.8$	$p = 0.1$	$F(1,47) = 5.6$	$p < 0.05$

Figure 3.4. Mean reading times with 95% confidence intervals for the continuation sentences, (129d-d') for participants ($n = 32$) in Experiment 3b—the CONTROL condition from Experiment 2. Two main factors were tested: intended REFERENT (theme, goal) and referential FORM (PRONOUN, definite DESCRIPTION).

3.3.4.4. Discussion. The results of Experiment 2 had suggested that there was a slight difference in the relative prominence of the theme and goal discourse referents between the CONTROL and SPLIT conditions. However, that difference was not statistically significant (i.e. a non-significant interaction between CONTEXT and REFERENT. The repeated-name penalty design—a more sensitive measure—in Experiment 3a and Experiment 3b reveals this difference to be relevant. The presence of a (near-significant) interaction

between REFERENT and FORM in the CONTROL condition (Experiment 3b) but not in the SPLIT condition (Experiment 3a) suggests the conclusion that there is a different salience representation between the two context conditions, as illustrated graphically in Figure 3.1, above.

Nevertheless, the data suggest that there is a slight advantage for the theme continuation even in the SPLIT condition. This was evidenced in both experiments 2 and 3a by slightly faster reading times for the theme continuations. If this advantage is to be taken seriously, then it might constitute some evidence that on the lower part of the syntactic prominence hierarchy, objects are not much more prominent than obliques. Thus, in the SPLIT condition, the fact that the theme argument is oblique diminishes its discourse salience, but the net result is that it is still slightly more salient than the goal. In short, this amounts to permitting a salience representation in which the prominence relations are weighted. I will discuss this theoretical possibility in greater detail in Chapter 5.

The results of Experiment 3a and Experiment 3b, taken together, also shed some light on the applicability of the repeated-name/definite-description penalty approach. The penalty is largest and most distinct when there is clearly a single, most-salient discourse referent (after filtering) in the current context. When there is some ambiguity, as in the SPLIT condition, then the penalties are diminished.

One further comment regarding the use of the repeated-definite-description penalty in these experiments is warranted. Although my assumption that there should be no difference in the reading times within the DESCRIPTION condition was borne out, it is still a null result. Clearer confirmation of the validity of the repeated-definite-description penalty in the present context may be necessary by comparing the present data to cases

when the repeated definite description refers to the most salient referent in the discourse: The model predicts that there should be a difference between these cases. However, I will have to put this line of investigation off to future work.

3.3.5. Experiment 4: Pronoun Resolution Judgment Questionnaire

3.3.5.1. Design. The last experiment reported here with *spray/load* stimuli is an off-line forced-choice task designed to test the interpretations of the pronouns at the beginning of the continuation sentences. In theory, if these pronouns actually have only one pragmatically coherent interpretation, then this task should be relatively easy for participants and items should be nearly unambiguous. Deviations should arise only from occasional careless errors. However, there is another potential source of variation. The picture which emerges from Experiments 1-3 is that when the syntactic and semantic prominence hierarchies are not aligned as in the SPLIT context condition, participants seem to exhibit some indecisiveness about which referent they perceive as being most salient in the discourse. In the present experiment, this indecisiveness hypothetically could affect their pronoun resolution decisions. While the chance that participants will carelessly make a mistake is likely small, that chance is thus apt to be greater in the SPLIT condition. It is predicted, then, that if there is any deviation from the norm, then it should be in the SPLIT condition.

3.3.5.2. Method.

Participants. Participants in this experiment include 20 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. A questionnaire was prepared using the same stimuli from the preceding experiments: 48 items from Experiments 2-3 and 9 items removed from Experiment 1 (see footnote 7). Items were presented in a format as shown in (131) where a four-sentence vignette (sentences (a)-(d) of previous experiments) was followed by two possible interpretations of the pronoun at the beginning of the fourth sentence. The vignettes varied in a 2×2 design: CONTEXT sentence (CONTROL, SPLIT) versus intended REFERENT of the pronoun in the continuation sentence (theme, goal).

(131) John brought some paint into his living room and stared at one of the walls. He was trying to decide how to redecorate his home. John sprayed the paint on the wall. It dribbled down and made a mess.

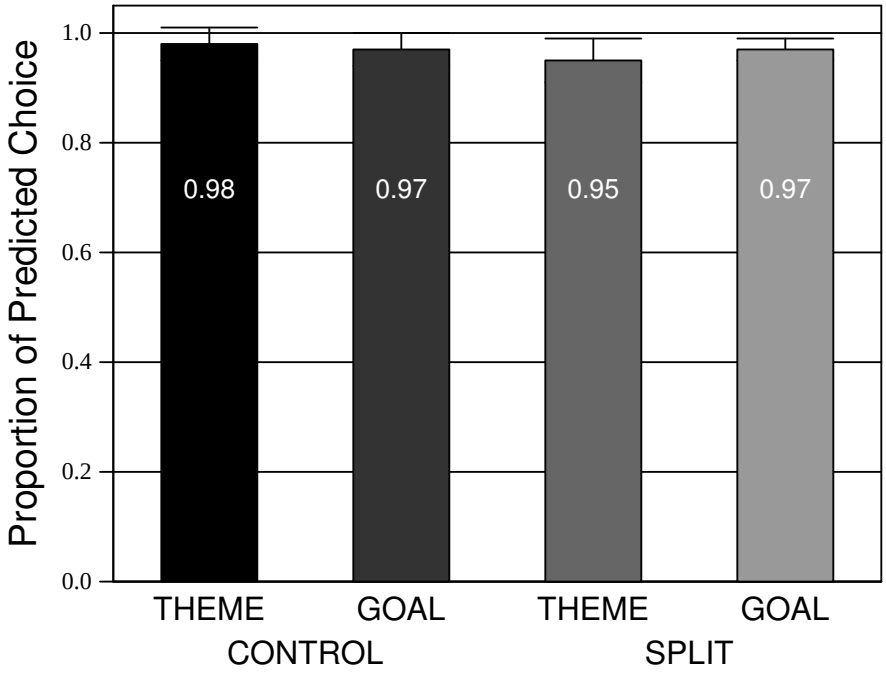
a. It = paint

b. It = wall

Procedures. Participants were asked to indicate which of the two interpretations made the most sense given the context.

3.3.5.3. Results. The results of Experiment 4 are shown in Figure 3.5. There was a marginal main effect of CONTEXT, but no effect of REFERENT and no interaction between the two. The CONTEXT effect can probably be regarded as insignificant given that the p-value approaches significance by participants only. However, even if this is regarded as a significant effect, it is exactly as predicted. As the graph illustrates, participants showed more variance in the SPLIT condition.

3.3.5.4. Discussion. Results of this experiment show that participants resolved the potentially ambiguous pronouns in the intended direction. While there was some slight deviation resulting in a marginally (at best) significant main effect of CONTEXT, this was



	by participants		*by items*	
CONTEXT	$F(1,19) = 3.8$	$p = 0.065$	$F(1,56) = 1.5$	$p = 0.2$
REFERENT	$F(1,19) < 1.0$	n.s.	$F(1,56) < 1.0$	n.s.
CONTEXT*REFERENT	$F(1,19) = 1.9$	n.s.	$F(1,56) = 1.5$	n.s.

Figure 3.5. Proportion with 95% confidence intervals of choices for the intended interpretation (theme or goal) of the pronoun in the target sentence (fourth sentence in (131)) for participants ($n = 20$) in Experiment 4. Two main factors were tested: CONTEXT (CONTROL, SPLIT) and intended REFERENT of pronoun (theme, goal).

in precisely the predicted manner: The indecisiveness in the SPLIT condition caused the chance for error to increase. In short, then, the results of Experiment 4 indicate that the results of Experiments 1-3 cannot have arisen from problems in the stimuli. These results thus underline the conclusion that syntactic and semantic prominence contribute independently to the discourse salience of discourse referents for subsequent pronominal reference.

3.3.6. Low End of the Prominence Hierarchies: Discussion

Before discussing Experiments 1-4, let me give a brief review of the major observations of the experiments and the primary conclusions they support. Overall, participants showed clear preferences in the CONTROL condition, favoring a continuation sentence in which the initial NP (either as a pronoun or definite description) was interpreted as coreferent with the theme of the preceding utterance. However, in the SPLIT condition, these preferences were less clear and (at best) only marginally favored the theme. These results are best explained by drawing the conclusion that *both* syntactic and semantic prominence influence the discourse salience of referents; and possibly also that the prominence difference between themes and goals is larger than that between objects and obliques.

One of the implications of these results is that the lower part of the syntactic prominence hierarchy (i.e., object > oblique)—which has long been assumed but never experimentally verified—may be in need of revision perhaps to something like the hierarchy shown in (132) or possibly even to that in (133). However, among these two possibilities, the evidence from the repeated-definite-description penalty experiments point toward (132), because at least some ranking differential between objects and obliques seems necessary to account for the slight variations between Experiment 3a and Experiment 3b.

(132) subject >> object > oblique

(133) subject > others

Given these revisions, the graphical illustrations of the store and the subsequent operations on it are shown in Figure 3.6 with respect to a frame semantic approach. Note that, in line with the experimental results showing a mild preference for the theme in the

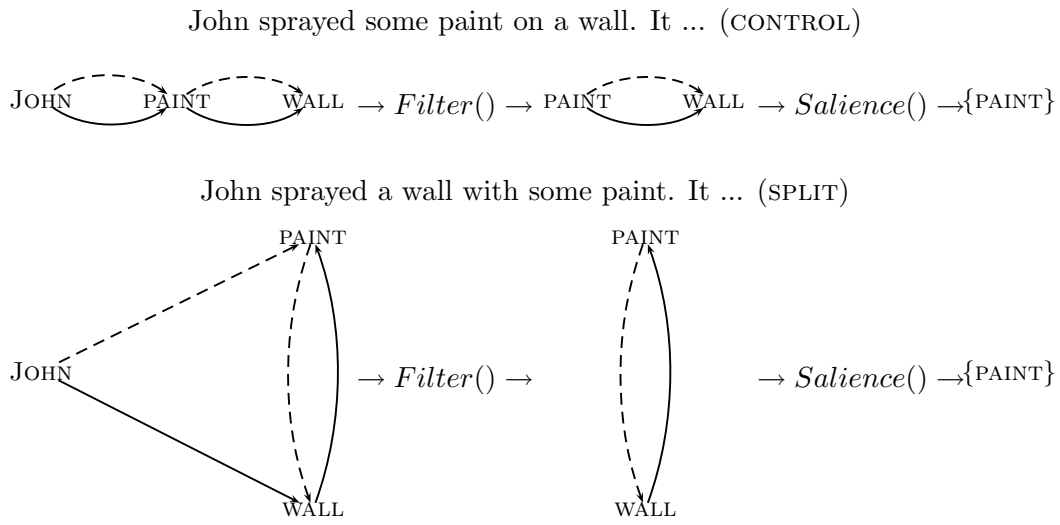


Figure 3.6. Illustration of the store for the *spray/load* stimuli under the frame semantic approach. syntactic prominence relations are represented by solid lines and semantic prominence relations are represented by dashed lines.

SPLIT condition, I have indicated that the salience operation in this condition returns the theme, PAINT. In terms of the model, it seems that there is just enough imbalance in the filtered store for the salience operation to regard PAINT as the maximal referent. A more sophisticated model of discourse salience might introduce weighted ranking relations (as opposed to the discrete rankings used here) to capture this imbalance.

Next, let me illustrate how the graphical illustrations of the store might look under a PROTO-role analysis of *spray/load* verbs. Motivated by the evidence that animate entities are more salient than inanimate entities (cf., Prat-Sala and Branigan, 1999), I assume that the PROTO-agent entailment of sentience on the subject of a *spray/load* verb imposes a semantic prominence ranking on it over the other arguments. Thus, the store is as shown in Figure 3.7. One will immediately notice that the predictions are the same

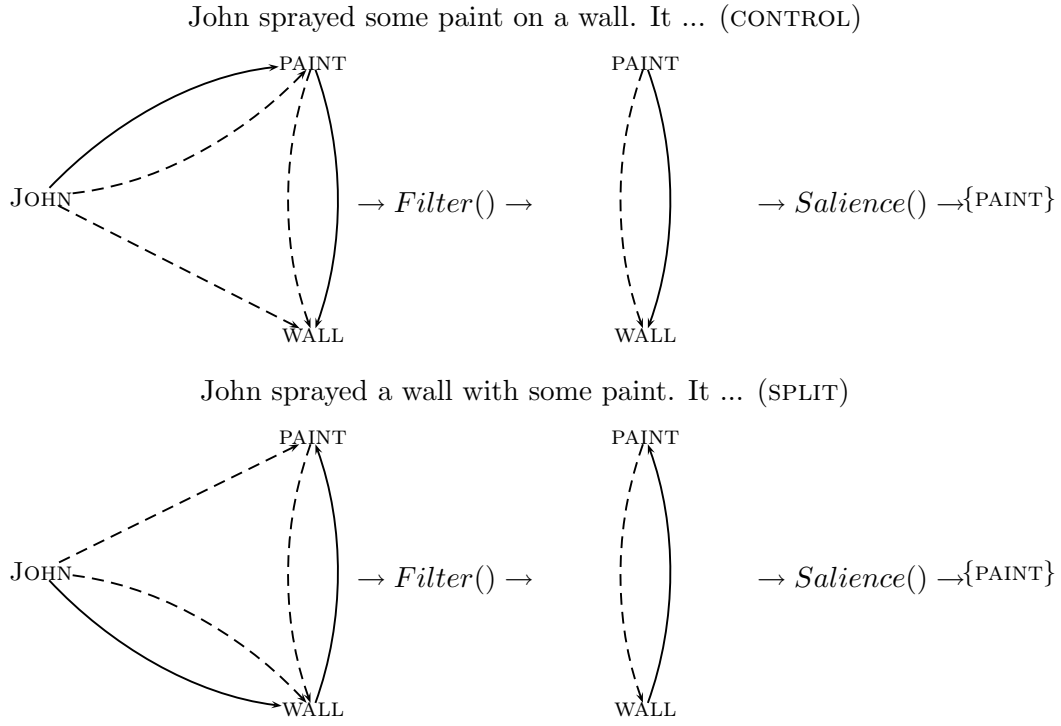


Figure 3.7. Illustration of the store for the *spray/load* stimuli under the PROTO-role approach. syntactic prominence relations are represented by solid lines and semantic prominence relations are represented by dashed lines.

here as with the frame semantic approach, above: after filtering, the same configuration of the discourse referents in the store leads to the promotion of PAINT in the CONTROL context, but some indecisiveness (assuming some sort of imbalance in the filtered store) in the SPLIT condition.

The detailed story that I've told here about what *spray/load* constructions tell us about syntactic and semantic prominence rests fundamentally on the assumption that the two variants assign the same two thematic roles (i.e., theme and goal) in the Frame

Semantics approach to their internal arguments but in different syntactic positions. However, one might question this assumption. It could be the case that the thematic roles are different between the two variants. Consider the sentences in (134.)

(134) a. John sprayed paint on the wall.

b. John sprayed the wall with paint.

Rappaport et al. (1987), for instance, make the argument that PAINT in the *with*-variant as in (134b) is actually a means. That is, the paint is used as a means of bringing about a change of state in the wall from being unpainted to being fully painted. I will not attempt to challenge this analysis here. In fact, it may even be the correct analysis. However, I would argue that it is not relevant to the basic point that I am arguing in this dissertation: that syntactic prominence alone cannot account for the data and that (some notion of) semantic prominence is necessary. If the two variants of a *spray/load* verb do not assign the same thematic roles, the real problem is not whether or not semantic prominence is relevant or not, but rather what must the relative prominence of the relevant thematic roles be in order to account for the differences between the CONTROL and SPLIT conditions.

3.4. Testing the High End of the Prominence Hierarchies

3.4.1. *Tough*/non-*Tough* Constructions

3.4.1.1. Overview. The *spray/load* stimulus items have provided a window on the contributions of both syntactic and semantic prominence to the salience of discourse referents for subsequent pronominal reference. The next step is to move beyond this relatively narrow, fairly uniform context to see if the effect of semantic prominence can

be observed in environments which are more variable. With this goal in mind, the second set of experiments centers on *tough*-constructions, another construction which re-orders the canonical order of arguments.

3.4.1.2. Analysis. A coherent and comprehensive account of the syntax and semantics of *tough*-constructions as shown in (135b) has been elusive in the literature. Debate on the syntax of these constructions has centered on a number of questions including whether the matrix subject of the *tough* adjective is base-generated in the embedded clause and moves to surface subject position via a series of movement operations (Hornstein, 2001), or whether it is simply base-generated in the matrix clause and somehow controls the gap in the embedded clause (Chomsky, 1977; Lasnik and Fiengo, 1974). I will take the position that, in the present investigation, the difference between these two analyses is not relevant because what I am most concerned with is the surface structure of these constructions. In this respect, I regard the discourse referents evoked in the *tough*/non-*tough* alternation as taking on differing syntactic roles. In the non-*tough* example in (135a), JOHN appears as a subject while MATT appears as an object. However, in the *tough* variant in (135b), both of these noun phrases are subjects. This difference has certain consequences for predictions about the syntactic prominence of these entities which I will discuss below.

(135) a. John easily hit Matt during the boxing match.

b. Matt$_i$ was easy for John to hit $\emptyset_i$ during the boxing match.

The semantic analysis of these constructions also requires some attention. It does seem clear that the referent evoked in matrix subject position (i.e., MATT in (135b),

above) does bear the thematic role or PROTO-role entailments which are normally assigned by the verb in the embedded clause to its object. However, it is less clear which, if any, semantic restrictions are assigned by the *tough* predicate itself. Some (cf., Kim, 1996) have argued that *tough* predicates are three-place predicates assigning the role of stimulus to the matrix subject, experiencer to the embedded subject, and event to the embedded infinitival clause. Most thematic hierarchy proposals take experiencers as higher than stimuluses (cf., Jackendoff, 1972). If so, then this would mirror the SPLIT condition in Experiments 1-4 in which the syntactic prominence hierarchy promotes one referent while the semantic prominence hierarchy promotes another. However, intuitively, I find that after reading (135b), the role that Matt plays in the beating event is far more salient than any role he plays in the relative ease of that event. Therefore, while the three-place predicate analysis of *tough*-constructions may be plausible, I will assume that for the purposes of determining semantic prominence, only the thematic roles assigned by the verb in the embedded clause matter.

3.4.1.3. Methodological Usefulness. Given the analysis described above, then *tough*-constructions fall into the same basic experimental paradigm as *spray/load* verbs, though with some crucial differences. In this section I'll give an overview of how the *tough*/non-*tough* constructions fit into the current experimental methodology. I will use the abbreviated discourse shown in (136)-(137) for this discussion.

(136) a. John easily hit Matt during the boxing match.

b. He ...

(137) a. Matt$_i$ was easy for John to hit $\emptyset_i$ during the boxing match.

b. He ...

After the (a) sentences, there are three discourse referents in the store: JOHN, MATT, and BOXING MATCH. However, BOXING MATCH appears in an optional adjunct position and is not a verbal argument. So, both syntactically and semantically it is predicted to be non-salient. Furthermore, in my model of pronoun resolution, it is filtered out. Therefore, I will simply ignore it in this and following discussion. In Chapter 2, I described two methods for determining the relative syntactic prominence of the discourse referents. According to the hierarchical method, in (136), JOHN is more prominent than MATT because of left-to-right ordering. However, in (137), MATT is more prominent than JOHN because *Matt* is in a higher clause. On the other hand, according to the role-based method, in (136), JOHN is more prominent than MATT because the former is evoked in subject position and the latter in object position. But in (137), both JOHN and MATT are evoked as subjects and so they are unranked with respect to each other. This is summarized in (138)-(139).

(138) a. John easily hit Matt during the boxing match.

syntactic prominence: JOHN > MATT

b. He ...

(139) a. Matt$_i$ was easy for John to hit $\emptyset_i$ during the boxing match.

syntactic prominence $_{role}$: MATT = JOHN

syntactic prominence $_{hierarchy}$: MATT > JOHN

b. He ...

As for semantic prominence, first it is necessary to briefly consider the semantics of the verb *hit*. In the FrameNet system, *hit* is in the CAUSE_IMPACT frame which

assigns two roles: an agent role which is inherited from the parent frame INTENTIONALLY_AFFECT and an impactee role which is also inherited by derivation from the patient role of the parent frame INTENTIONALLY_AFFECT (see Section 2.6.1 for discussion of frame and role/element inheritance). Assuming a semantic prominence hierarchy as in (140) (consistent with hierarchies proposed by Jackendoff, 1972; Larson, 1988; Speas, 1990), then John is more prominent than Matt in both (138) and (139).

(140) agent > patient

(141) cause event/c-o-s > [undergo c-o-s or causally-affected]

With respect to the PROTO-role system, the PROTO-agent entailment of cause event/c-o-s applies to JOHN while the PROTO-patient entailments undergo c-o-s and causally-affected apply to MATT. Thus, if we assume these entailments force a ranking on the entities which bear them according to the hierarchy in (141), then JOHN should be more prominent than MATT. Hence, both the frame semantics and PROTO-role approaches make the same predictions for semantic prominence. The full set of predictions is thus outlined in (142)-(143).

(142) a. John easily hit Matt during the boxing match. CONTROL

syntactic prominence: JOHN > MATT

semantic prominence: JOHN > MATT

b. He ...

(143) a. Matt$_i$ was easy for John to hit $\emptyset_i$ during the boxing match. SPLIT

syntactic prominence $_{role}$: MATT = JOHN

syntactic prominence $_{hierarchy}$: MATT > JOHN

semantic prominence: JOHN > MATT

b. He ...

In (142), the prominence hierarchies are aligned and therefore this is the CONTROL condition for the present series of experiments. The situation in (143), however, is slightly confusing because of the differing predictions made by the role and hierarchy methods of syntactic prominence. For the sake of consistency, I will call this the SPLIT condition. However, results will indicate whether this is warranted, and subsequently which of these two methods is a better predictor of syntactic prominence.

Now, let's consider the specific predictions for these experiments by looking at the graphical illustrations of these relations as shown in Figure 3.8. For simplicity, I have omitted the pre-filtered illustration of the store.

Given the evidence from the *spray/load* experiments that both syntactic and semantic prominence influence the discourse salience of entities, the experimental predictions are as follows. In the CONTROL condition, observations should show that participants perceive JOHN to be the most salient entity in the context. In the SPLIT condition, there are two possibilities: If the role-based method is correct, then JOHN should again be perceived as the most salient entity—the same as in the CONTROL condition. However, if the hierarchical method is correct, then neither JOHN nor MATT should be perceived as more salient than the other.

The *tough*/non-*tough* alternation offers several advantages in the current investigation for several reasons. First, as an argument re-ordering consruction, it allows a continued test of the relative contribution of syntactic and semantic prominence. In addition, it allows a test of the upper parts of the prominence hierarchies. And finally,

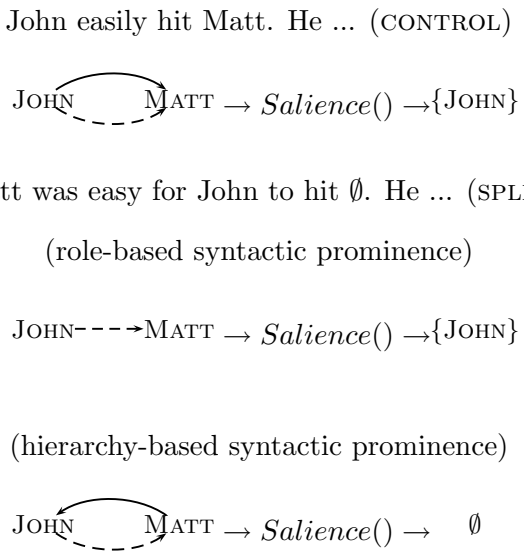


Figure 3.8. Hypothetical illustration of the store for *tough*/non-*tough* experimental stimuli using the frame semantic approach. syntactic prominence relations are shown with solid lines and semantic prominence relations are shown with dashed lines.

it allows a comparison of two different methods for determining syntactic prominence. Finally, by varying the main verb used in the embedded clause, the effects of semantic prominence can be observed in a wider variety of syntactic and semantic contexts.

3.4.1.4. Items. For the experiments which follow, 32 stimuli were prepared using six different adjectives which exhibit the *tough* alternation (*hard*, *easy*, *fun*, *a cinch*, *difficult*, and *tough*) as well as 32 different transitive non-stative verbs. The semantic analyses of these verbs with respect to a frame semantic or PROTO-role approach was not explicitly controlled, allowing for a more liberal test of the hypothesis. However, many verbs come from FrameNet frames which assign agent and patient roles (or their derivatives via frame inheritance). Some examples include *hit*, *catch*, *move*, and *capture*. Other verbs

come from frames which assign roles that, in their canonical surface order, arguably reflect various proposed thematic hierarchies (cf., Jackendoff, 1972; Larson, 1988; Speas, 1990). These include such verbs as *trick*, *help*, *tutor*, and *support*.[10]

The wider variety of verbs used in this study means that there is some variation in the thematic roles and PROTO-role entailments on arguments. However, for expository purposes, throughout the discussion of the experiments, I will simply refer to the two primary verb arguments as agent and patient (i.e., JOHN and MATT, respectively in (142) and (143)).

The six *tough*-predicates were used in Experiments 5-8 reported below. For reference, the full set of stimulus items used in these experiments is provided in the Appendix.

The following set of experiments parallel the *spray/load* experiments using the same experimental design and procedure. However, results show some interesting contrasts with those above and lead to more detailed conclusions about syntactic and semantic prominence, the prominence hierarchies, and the nature of the repeated-name penalty.

3.4.2. Experiment 5: Discourse-completion Questionnaire

3.4.2.1. Design. This experiment is designed to get a rough idea of how readers perceive the relative salience of the discourse referents introduced in *tough*/non-*tough* constructions. The design is the same as Experiment 1 with *spray/load* constructions. By looking at the participants' choices of which continuation sentences they judge to complete the discourse better, it is possible to determine which referents they perceive to be more salient in the context.

[10]One notable exception is the verb *tease* from the EXPERIENCER_OBJ frame which is a stimulus-experiencer verb—the reverse order of the thematic hierarchy of, for example, Jackendoff (1972). Only one stimulus item uses this verb and results do not differ if it is excluded.

3.4.2.2. Method.

Participants. Participants in this experiment included 32 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. Short discourses were prepared using the template shown in (144)-(145). The first sentence was the context sentence using either a *tough* or non-*tough* construction. This was followed by two optional discourse continuations, both beginning with a pronoun and written so that the pronoun's interpretation was consistent with only one of the entities introduced in the context. When this experiment was performed, only 24 items had been completed. These items were included in a questionnaire along with filler items from an unrelated investigation. Items were arranged into blocks and randomized. Different versions of the test were prepared so that each participant saw each item in only one form (*tough* or non-*tough*).

(144) John$_i$ easily hit Matt$_j$ in the boxing match. CONTROL

a. He$_i$ even managed to land a knockout punch. agent

b. He$_j$ became bruised and bloodied all over. patient

(145) Matt$_j$ was easy for John$_i$ to hit $\emptyset_j$ in the boxing match. SPLIT

a. He$_i$ even managed to land a knockout punch. agent

b. He$_j$ became bruised and bloodied all over. patient

Procedures. As in Experiment 1, participants were asked to choose the more natural continuation to the discourse.

3.4.2.3. Results. Results of Experiment 5 are shown in Table 3.8. These results parallel those of Experiment 1 showing a clear preference for the agent in the CONTROL

Table 3.8. Experiment 5: Proportion of Choices for Agent Continuation

	agent	*by participants*	*by items*
CONTROL	75%	$t(31) = 5.1\ p < 0.001$	$t(23) = 5.7\ p < 0.001$
SPLIT	49%	$t(31) < 1.0\ n.s.$	$t(23) < 1.0\ n.s.$

condition, choosing the agent continuation 75% of the time—significantly more often than chance. In the SPLIT condition, participants chose the agent continuation 49% of the time—not significantly different from chance. Furthermore, a pairwise t-test between the CONTROL and SPLIT conditions shows a significant difference [by participants, $t(31) = 3.7\ p < 0.005$; by items, $t(23) = 3.9\ p < 0.005$]. Finally, it is important to note that results in the CONTROL condition replicate the many earlier studies cited in Section 1.3.2.2 which show subject-preference in pronoun resolution.

3.4.2.4. Discussion. The results in Experiment 5 continue to support the analysis that both syntactic and semantic prominence influence the salience of discourse referents. In addition, it begins to show that these influences also occur at the upper end of the prominence hierarchies. The near-fifty-fifty split in the SPLIT condition further suggests that the hierarchical method of determining syntactic prominence may be more reliable than the role-based method. Or, alternatively, perhaps the role-based method needs a slight revision. One possibility which would account for the present results is a ranking where the matrix subject is more prominent than other subjects (and thus any other arguments). This possibility is particularly interesting in that it begins to resemble the binary nature of many information-packaging theories (e.g., *theme-rheme* in Firbas 1964, 1966, *topic-comment* in Gundel 1974, *topic-focus* in Sgall, 1967). I will discuss this in greater detail in Section 3.5 below.

There is one reason, though to be somewhat tentative about the conclusions in the present experiment. While the parallel off-line results with the *spray/load* stimuli also showed a nearly fifty-fifty split in the SPLIT condition, later on-line evidence showed a mild preference for the theme. The difference between the off-line and on-line results was explained by taking into account the methodology of the off-line experiment: Because participants could still see the context sentence printed on the page while making their judgments, the influence of syntactic prominence might have been exaggerated. If the same phenomenon occurred in this experiment, then the subsequent experiments should clarify this by showing some preference for the agent over the patient in the SPLIT condition.

3.4.3. Experiment 6: Self-paced Reading Task

3.4.3.1. Design. Experiment 6 is designed to use a more sensitive on-line measure of the participants' perceptions of the relative discourse salience of the agent and patient. If the results of the discourse completion task in Experiment 6 are reliable, then in this experiment, there should be a reading-time advantage for the agent continuation in the CONTROL condition, but no advantage for either continuation in the SPLIT condition.

3.4.3.2. Method.

Participants. Participants in this experiment included 32 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. For the self-paced reading task of Experiment 6, the context sentences and their respective continuations were embedded in five-sentence vignettes as in (146).

The first sentence introduces two named human referents in a conjoined construction which research has shown does not promote either referent to greater salience than the other (Gordon et al., 1999). The second sentence focused on some third referent introduced (or related to a referent introduced) in the first sentence. Hence, according to the discourse model, after the second sentence, the two named referents remain as non-salient members in the store. The third sentence is the context sentence and the referents were reinvoked by name in either the CONTROL or SPLIT alternates, followed by an adjunct phrase intended to cohere with sentence two. The fourth sentence began with a pronoun and varied between two conditions depending on the intended REFERENT of the pronoun (agent, patient). Finally, a fifth sentence concluded the vignette. Each vignette was followed by one *true/false/indeterminate* comprehension question for which feedback ("correct"/"not correct") was given.

(146) a. John and Matt took part in an important boxing match.

b. It was twelve rounds long.

c. John$_i$ easily hit Matt$_j$ in the final round. CONTROL

c'. Matt$_j$ was easy for John$_i$ to hit $\emptyset_j$ in the final round. SPLIT

d. He$_i$ even managed to land a knockout punch. agent

d'. He$_j$ became bruised and bloodied all over. patient

e. The judges had no trouble deciding the winner.

question: Matt lost the match by a considerable margin.

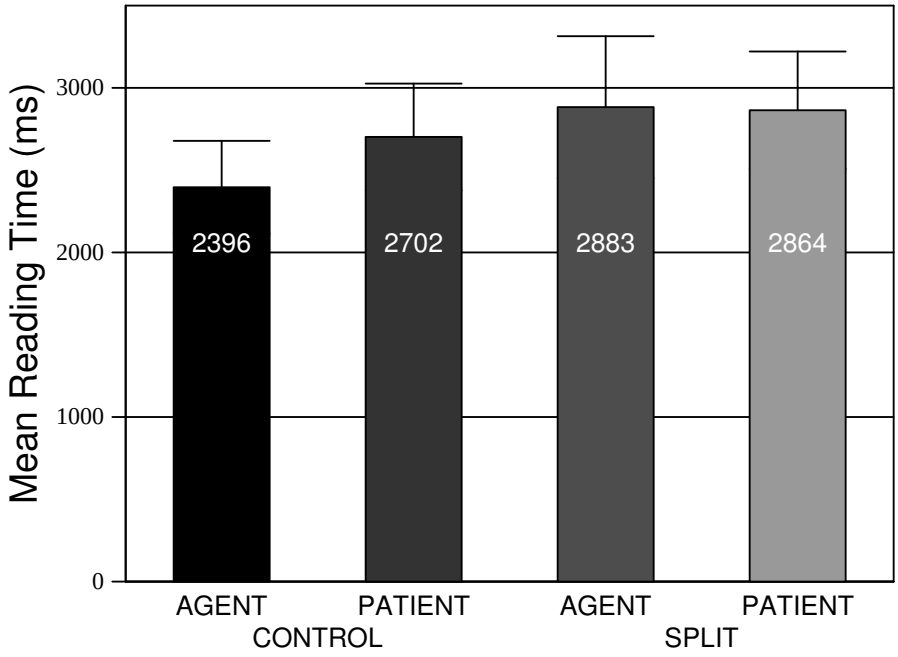
The experiment is thus a 2×2 design pitting CONTEXT (CONTROL, SPLIT) against the intended REFERENT of the pronoun (agent, patient). Stimuli were placed in blocks, and

presented to participants in pseudo-random order. Each participant saw each stimulus item in only one of its possible forms.

Procedures. Procedures were the same as for the self-paced reading tasks described in Experiments 2-3.

3.4.3.3. Results. Results of Experiment 6 are shown in Figure 3.9. There was a main effect of only CONTEXT sentence as well as a marginal interaction (by items only) between CONTEXT and REFERENT. Statistically, these results are almost the complete opposite of those of Experiment 2 in which there was a significant main effect of only REFERENT and no interaction. In the present experiment, participants found the continuation sentences more difficult to read in the SPLIT condition than in the CONTROL condition suggesting that there was even more indecisiveness than in Experiment 2. Finally, planned comparisons show that in the CONTROL condition, there is a significant difference between the mean reading times in the agent and patient conditions [by participants, $t(31) = 3.0$ $p < 0.05$; by items, $t(31) = 2.7$ $p = 0.06$]. As in Experiment 6, this replicates earlier experiments showing a subject-antecedent preference.

3.4.3.4. Discussion. The experimental results continue to be in line with the analysis that both syntactic and semantic prominence contribute to the salience of discourse referents. In the CONTROL condition, one discourse referent, the one realized as an agent-subject, was seen as being clearly more salient while in the SPLIT condition, neither the agent nor the patient seems to be more salient than the other. However, the key result which needs explanation in this experiment is the fact that participants had a greater difficulty reading the continuation sentences in the SPLIT condition as evidenced by longer reading times overall. How can this be reconciled with the results of the



	by participants		*by items*	
CONTEXT	$F(1,31) = 12.5$	$p < 0.005$	$F(1,31) = 7.8$	$p < 0.01$
REFERENT	$F(1,31) = 2.6$	$n.s.$	$F(1,31) = 1.1$	$n.s.$
CONTEXT*REFERENT	$F(1,31) = 2.2$	$p = 0.15$	$F(1,31) = 4.3$	$p < 0.05$

Figure 3.9. Mean reading times with 95% confidence intervals for the continuation sentences, (146d-d') for participants ($n = 32$) in Experiment 6. Two main factors were tested: CONTEXT (CONTROL, SPLIT) and intended REFERENT of pronoun (agent, patient).

spray/load reading-time experiment (see Experiment 2) in which such a difference was not observed? In the *spray/load* experiments, one consistency was that there was a mild preference for the theme in the SPLIT condition. This was analyzed by taking a revised view of the prominence hierarchies such that the difference between themes and goals is somewhat larger than that between objects and obliques. Hence, when the syntactic and semantic prominence hierarchies are aligned, the referent realized as a theme-oblique argument has a mild advantage over the referent realized as a goal-object argument. In

the present experiment, it seems to be the case that such a mild advantage does not result. This indicates that the prominence difference between agents and patients is the same as that between the first and second most syntactically prominent arguments under the hierarchical method (or possibly matrix subjects and non-matrix subjects, under the revised role-based method discussed above). Thus, in the SPLIT condition, the two competing discourse referents have essentially equal salience. It therefore takes readers much longer to overcome this indecision and find a satisfactory analysis. This results in longer reading times as a whole in this condition.

Another explanation for the difference between the *spray/load* and *tough* results may stem from certain differences in the referring expressions. After each administration of the experiment, there was a debriefing session with each participant. One participant noted that he found it a lot easier to read and resolve a pronominal reference to a "thing" than to a named person. Such things as paint and walls have inherent properties that are immediately accessible upon evoking an instance of them in a discourse. On the other hand, the properties which distinguish one named human from another must be determined within the discourse itself. In other words, it takes far less processing to resolve the pronoun in *it dribbled down* to PAINT over WALL than it takes to resolve the pronoun in *he became bruised* to MATT over JOHN. Thus the additional time in the SPLIT condition in this experiment may be due to the additional processing effort required to perform these operations for both discourse referents. In the CONTROL condition, on the other hand, there is a single salient referent, so the pronoun is resolved immediately to this default referent. Even when later information proves that this assignment was wrong, reanalysis and recovery is relatively simple because there is only one other compatible

entity. In the confusing situation of the SPLIT condition, both referents must be evaluated much more extensively leading to greater processing time overall.

3.4.4. Experiments 7a-b: Repeated-name Penalty

3.4.4.1. Design. The next two experiments with *tough*/non-*tough* constructions take advantage of the repeated-name penalty to attempt to get a more detailed picture of participants' perceptions of which entity is more salient. Given the trends in Experiment 5 and Experiment 6, the prediction here seems to be that in the SPLIT condition, there should be no repeated-name penalty difference between the agent and patient discourse referents. However, in the CONTROL condition, the agent referent should show an advantage.

3.4.4.2. Method.

Participants. Participants in each of Experiments 7a and 7b included 32 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. The stimuli used in these experiments were the same as those used in Experiment 6, illustrated here in (147). However, the referring expressions at the beginning of the target sentence (the fourth sentence) were varied between pronouns and names. In order to increase the statistical power of the experiments, the SPLIT and CONTROL conditions were evaluated in separate experiments, Experiment 7a and Experiment 7b, respectively. Each experiment therefore had a 2×2 design pitting the intended REFERENT of the referring expression in the target sentence (agent, patient) against the FORM of the referring expression (PRONOUN, NAME).

(147) a. John and Matt fought each other in a boxing match.

b. It was twelve rounds long.

c. John$_i$ easily hit Matt$_j$ in the final round. CONTROL

c'. Matt$_j$ was easy for John$_i$ to hit $\emptyset_j$ in the final round. SPLIT

d. He$_i$ even managed to land a knockout punch. agent-PRONOUN

d'. John$_i$ even managed to land a knockout punch. agent-NAME

d". He$_j$ became bruised and bloodied all over. patient-PRONOUN

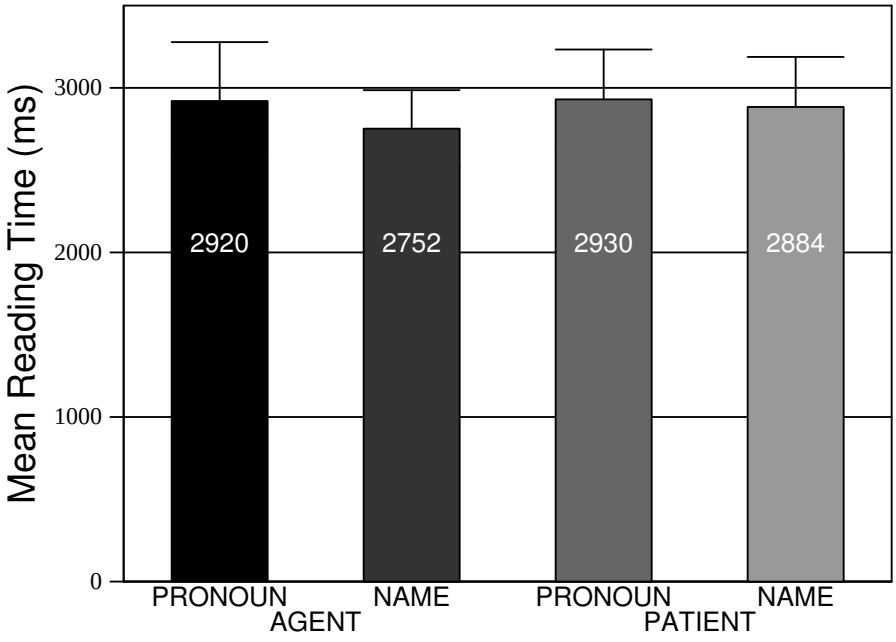
d"'. Matt$_j$ became bruised and bloodied all over. patient-NAME

e. The judges had no trouble deciding the winner.

question: Matt lost the match by a considerable margin.

Procedures. For both experiments, the procedures were the same as in Experiment 6.

3.4.4.3. Results. The results of Experiment 7a are shown in Figure 3.10 and the penalties in the two coreference conditions are shown in Table 3.9. There were no main effects and no interaction and all of the planned comparisons were insignificant. Futhermore, one-sample t-tests show that the penalties in neither the agent nor patient condition were significantly different from 0ms. While there are some apparent numerical trends, a glance at the size of the error bars (95% confidence intervals) in the graph reveals that this is a very noisy data set. One remarkable result though is that the reading time penalties are negative—continuations in the NAME condition were *faster*, not slower, than in the PRONOUN condition. Hence, these penalties might be more appropriately referred to as "advantages". However, these advantages are not significant and are, at best, numerical trends.



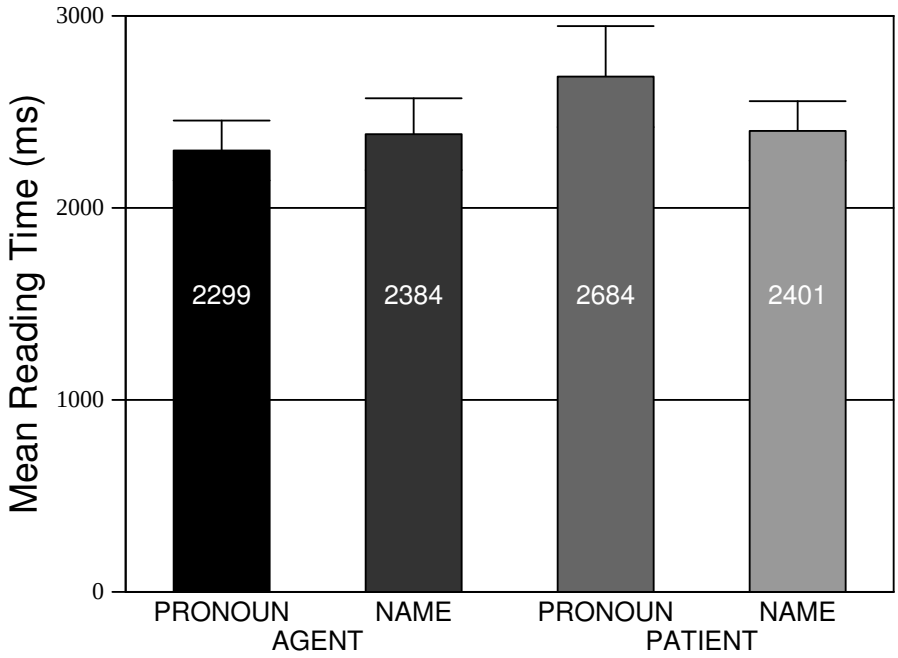
	by participants		*by items*	
REFERENT	$F(1.31) < 1.0$	*n.s.*	$F(1.31) < 1.0$	*n.s.*
FORM	$F(1, 31) < 1.0$	*n.s.*	$F(1.31) = 1.9$	*n.s.*
REFERENT*FORM	$F(1.31) < 1.0$	*n.s.*	$F(1.31) < 1.0$	*n.s.*

Figure 3.10. Mean reading times with 95% confidence intervals for the continuation sentences, (147d-d') for participants ($n = 32$) in Experiment 7a—the SPLIT condition from Experiment 6. Two main factors were tested: intended REFERENT (theme, goal) and referential FORM (PRONOUN, definite DESCRIPTION).

Table 3.9. Experiment 7a: SPLIT Condition Penalties

	penalty	*by participants*	*by items*
agent	-168ms	$t(31) < 1.0$ *n.s.*	$t(31) = 1.4$ *n.s.*
patient	-46ms	$t(31) < 1.0$ *n.s.*	$t(31) < 1.0$ *n.s.*

Results for Experiment 7b are shown in Figure 3.11 and penalties are shown in Table 3.10. In the agent condition, there is an 83ms penalty, while in the patient condition, there is a 270ms advantage, a significant difference: that is, the agent condition shows a



	by participants		*by items*	
REFERENT	$F(1,31) = 7.9$	$p < 0.01$	$F(1,31) = 3.6$	$p = 0.067$
FORM	$F(1.31) = 1.6$	*n.s.*	$F(1,31) = 1.5$	*n.s.*
REFERENT*FORM	$F(1.31) = 6.3$	$p < 0.05$	$F(1,31) = 8.5$	$p < 0.01$

Figure 3.11. Mean reading times with 95% confidence intervals for the continuation sentences, (147d-d') for participants ($n = 32$) in Experiment 7a—the CONTROL condition from Experiment 6. Two main factors were tested: intended REFERENT (theme, goal) and referential FORM (PRONOUN, definite DESCRIPTION).

Table 3.10. Experiment 7b: CONTROL Condition Penalties

	penalty	*by participants*	*by items*
agent	83ms	$t(31) < 1.0$ *n.s.*	$t(31) < 1.0$ *n.s.*
patient	-270ms	$t(31) = 2.5$ *n.s.*	$t(31) = 2.7$ $p = 0.07$

353ms *greater* penalty than the patient condition [by participants, $t(31) = 2.5, p = 0.05$; by items, $t(31) = 2.9, p < 0.01$].

3.4.4.4. Discussion. Statistically, results—falling in line with predictions—show that in the SPLIT condition, neither the agent nor the patient appear to be more salient for subsequent pronominal reference while in the CONTROL condition there is an advantage for the agent, suggesting that it is perceived as more salient in the context.

One curious result though is the apparent trend towards a repeated-name advantage for the patient in the CONTROL condition and for both referents in the SPLIT condition. If this trend is real, it makes sense within the emerging picture here. The analysis of the repeated-definite-description penalty results in Experiments 3a-b suggests that the penalty only shows up when the processor can easily return one salient (compatible) referent in the current context (or in terms of the discourse salience model, is able to determine one maximal element). This seems to be the same pattern in the results for Experiment 7b. The significant interaction between REFERENT and FORM in the CONTROL condition—when the agent is predicted to be the most salient referent—is evidence of this situation.

What is curious, though is why the patient shows a 270ms advantage. In order to understand this, it is necessary to reconsider the nature of the repeated-name penalty. In the original experiments which established the repeated-name penalty, Gordon et al. (1993) did not look at pronoun-initial utterances in which there was any realistic possibility of ambiguous reference. All of their experiments used gender-disambiguated reference except for their first experiment in which a strong effect of syntactic parallelism (and possibly topichood) prevented any potential ambiguity. Hence the results of Experiment 7b might be explained as follows. When the intended REFERENT of the pronoun is the patient, additional processing cost is necessary because, in part, reference is being

made to a non-salient discourse referent. In addition, participants would have to recall and make inferences from much more previously introduced information about the characters to deduce who the intended referent is. But in the NAME condition, the reference is disambiguated.

Finally, there is the problem of the SPLIT case when *both* referents show a numerical reading time advantage. If the advantage is to be given any credibility, then one way to explain it is in terms of the advantage given by the disambiguating information in contrast with the disadvantage of an ambiguous pronoun (when there is no salient referent) which must be resolved by recall of much more information about the respective characters.

In short, the only case which has the right configuration—pronominal/repeated-name reference to the entity which is the most salient (compatible) entity in the current context—is the only case which shows a positive penalty.

3.4.5. Experiment 8: Pronoun Resolution Judgment Questionnaire

3.4.5.1. Design. The last experiment is an off-line task designed to check that participants were likely to have resolved the pronouns to the intended referents. As in Experiment 4 with the *spray/load* stimuli, there are two possibilities here. In theory, participants should always choose the intended referent and there should be no significant effect of either CONTEXT or intended REFERENT. However, Experiment 4 showed that the somewhat mild confusion participants seemed to have experienced in the SPLIT condition resulted in an approximately 5% error rate. In the current set of experiments,

the results suggest that there is even greater confusion in the SPLIT condition with *tough*-constructions. Therefore, if there is some deviation in the SPLIT condition, it is predicted to be even larger in this experiment.

3.4.5.2. Method.

Participants. Participants in this experiment included 20 Northwestern University undergraduates who are native speakers of English. Participants received course credit in return for their participation.

Materials. Stimuli from Experiments 6-7 were taken and put into a form as shown in (148) where the third sentence of the vignette, the CONTEXT, varied between the CONTROL and SPLIT forms, and the fourth sentence varied in whether the intended REFERENT of the initial pronoun was the agent or patient. The experiment was therefore a 2×2 design. The stimuli were organized into a questionnaire in random order. Each participant saw each stimulus item in only one of its forms.

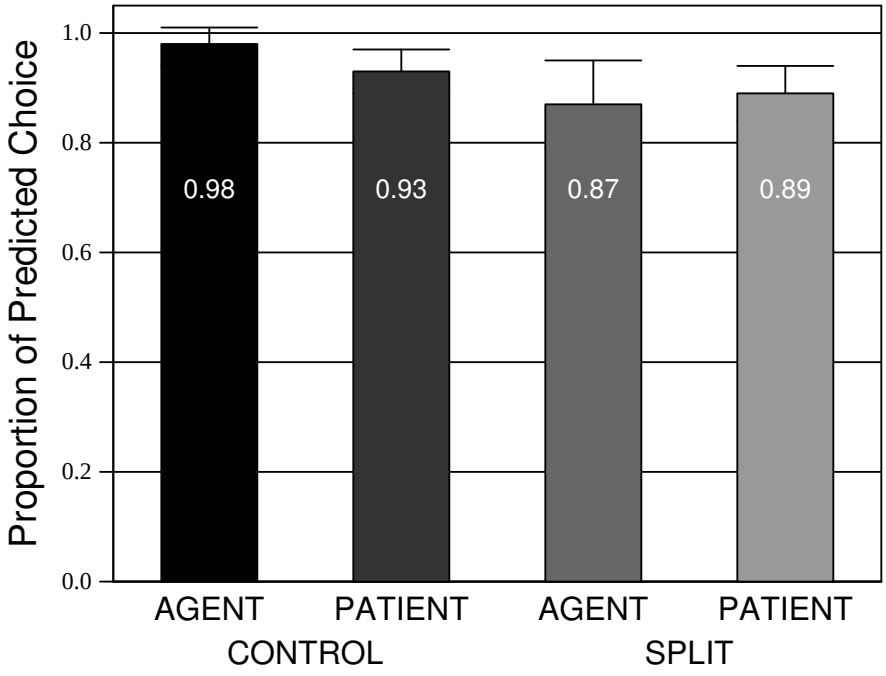
(148) John and Matt fought each other in a boxing match. It was twelve rounds long. John easily hit Matt in the final round. He even managed to land a knockout punch.

a. He = John

b. He = Matt

Procedures. Participants were instructed to choose which of the two assignments of the pronoun (option a or b) seemed the most natural within the given the context.

3.4.5.3. Results. Results for this experiment are shown in Figure 3.12. As in Experiment 4 with *spray/load* verbs, there is a significant main effect only of CONTEXT and no interaction between CONTEXT and REFERENT. However, what is remarkable here is



	by participants		*by items*	
CONTEXT	$F(1,19) = 6.3$	$p < 0.05$	$F(1,31) = 9.4$	$p < 0.005$
REFERENT	$F(1,19) < 1.0$	*n.s.*	$F(1,31) = 3.3$	*n.s.*
CONTEXT*REFERENT	$F(1,19) = 3.4$	$p = 0.08$	$F(1,31) = 2.5$	n.s.

Figure 3.12. Proportion with 95% confidence intervales of choices for the intended interpretation (agent or patient) of the pronoun in the target sentence (fourth sentence in (148) for participants ($n = 20$) in Experiment 8. Two main factors were tested: CONTEXT (CONTROL, SPLIT) and intended REFERENT of pronoun (agent, patient).

that the degree of the deviation from 100% in the SPLIT condition is much larger than in the *spray/load* case.

3.4.5.4. Discussion. The results of Experiment 8 fall right in line with the preceding experiments with *tough*-constructions showing that when the syntactic and semantic prominence hierarchies are not aligned as in the SPLIT condition, there is no clearly more salient discourse referent and confusion results. In fact, it seems that the closer

to equal the salience of the two competing references is, the greater the indecisiveness during subsequent pronominal reference resolution.

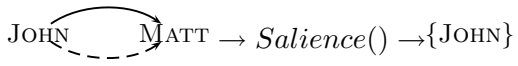
3.4.6. The High End of the Prominence Hierarchies: Discussion

First let me give a review of the main observations from Experiments 5-8 and the conclusions I draw from them. Participants in these experiments show that when the syntactic and semantic prominence hierarchies are aligned, there is a strong preference for the pronoun in a continuation sentence to refer to a discourse referent in the context which was realized as an agent-subject. This result is consistent with many previous investigations which show a preference for a pronoun to take a subject-antecedent. However, when the syntactic and semantic prominence hierarchies are not aligned, participants do not show any greater preference for either a syntactically non-prominent agent or a syntactically prominent patient. This is interpreted here as evidence of a context in which there isn't a single discourse referent more salient than others. This condition further seems to result in confusion for readers, causing them to take much longer to process and accommodate continuation sentences, or to make more errors in resolving the pronoun. In summary, the results of these experiments show that *both* syntactic and semantic prominence influence the salience of discourse referents for subsequent pronominal reference. Furthermore, the results suggest the validity of the upper parts of the prominence hierarchies as shown in (149)-(152), where (149) represents a hierarchy under a revised role-based approach.

(149) matrix subject > non-matrix subject

(150) 1 > 2

John easily hit Matt. He ... (CONTROL)



Matt was easy for John to hit ∅. He ... (SPLIT)

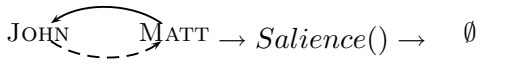


Figure 3.13. Illustration of the store for the *tough*/non-*tough* stimuli. syntactic prominence relations are represented by solid lines and semantic prominence relations are represented by dashed lines.

(151) agent > patient

(152) cause event/c-o-s > [undergo c-o-s or causally-affected]

The results suggest a store as for the CONTROL and SPLIT conditions as illustrated in Figure 3.13. Because the frame semantics and the PROTO-role approaches make the same predictions about the relative prominence of the agent and patient in the context sentence (based on the prominence hierarchies in (151)-(152)), I use the same graph here to illustrate both. Furthermore, since the syntactic prominence hierarchies shown in (149) and (150) also make the same predictions about the relative prominence of the two referents I use the same graph for both.

These experiments used the *tough*/non-*tough* alternation in order to obtain an argument re-ordering paradigm in which the relative effects of syntactic and semantic prominence can be tested. However, unlike Experiments 1-4 for which the stimuli came from one semantic context (i.e., *spray/load* verbs), the verbs used in these experiments varied much more in terms of the thematic roles assigned or the PROTO-role entailments on arguments. Thus, it is perhaps not surprising that the data in Experiments 5-8 is

rather noisier than that in Experiments 1-4. Yet, in spite of this fact, the effects of both syntactic and semantic prominence could be seen clearly. This suggests that the effects are robust and may be useful in even more uncontrolled investigative environments such as a corpus analysis. This hypothesis will be tested in Chapter 4.

3.5. General Discussion

The series of psycholinguistic experiments described in this chapter have provided converging evidence that *both* syntactic and semantic prominence contribute to the salience of discourse referents. The results have also suggested prominence hierarchies as shown in (153)-(155).

(153) matrix subject > non-matrix subject
object ≥ oblique

(154) agent > patient/theme > goal

(155) movement > not movement
cause event/c-o-s > [undergo c-o-s or causally-affected]

The syntactic prominence hierarchies shown in (153) represent the revised version of the role-based method. The evidence so far suggests the relationships shown, but perhaps further work would show that the two rankings could be reduced to just one ranking: matrix subject > other. As discussed above, this would be a very interesting result because it would then mimic the two-part nature of many information-packaging theories such as *theme-rheme* (Firbas, 1964, 1966), *topic/comment* (Chomsky, 1965; Gundel, 1974; Chafe, 1976), *topic-focus* (Sgall, 1967; Büring, 1995), among others. This would then open a new line of inquiry and possibly another way to look distinguish

some of the many factors which determine discourse salience. My work here has largely been motivated by what I have described as the conflation of syntactic prominence and semantic prominence: that in English, subjects are often agents, objects are often patients, and so on. However, one could also make the case that there is a conflation between syntactic prominence and topicality: that subjects are often discourse topics. Thus, further exploration of discourse salience might be very useful in such languages as Japanese and Korean in which subject and topic are distinguished (cf., subject-prominent and topic-prominent languages in the typology of Li and Thompson, 1976) and can be compared and contrasted to semantic prominence.

In the present research, while the hierarchical method of determing syntactic prominence has been somewhat more adequate than the role-based method at explaining the results, the hierarchical method does not predict the more gradient nature of the syntactic hierarchy which shows that the difference between subjects and objects (or others) is larger than the difference between objects and obliques. So, perhaps even the hierarchical method should be revised to provide a more gradient ranking of the discourse referents in an utterance. This would necessitate a more sophisticated discourse salience model in which weightings would be assigned to the various edges to show the strength of the ranking relationship among discourse referents. A proposal for how weightings might be employed in the discourse salience model is discussed in greater detail in Chapter 5.

One of the primary implications of the present study is that the relative semantic prominence of discourse referents contributes to the discourse salience of those referents in ways that are independent of the syntactic roles or positions in which a referent is realized. Furthermore, the results can be explained in terms of two different semantic paradigms,

with little extra theoretical baggage. The frame semantics approach explains the data by reference to a prominence hierarchy derived from existing proposed thematic hierarchies. Alternatively, the Dowtian PROTO-role system also explains the data with reference to a hierarchy of entailments which could be derived from one simple relationship: PROTO-agent entailments > PROTO-patient entailments. While the results of these experiments do not say anything about the validity of either of these two theories as linking theories, the fact that the results can be explained under both of these theories evidences the adaptability of the discourse salience model I present here. This is one of the central goals of the dissertation: to outline a flexible model of discourse salience which could be implemented in most any psycholinguistic or computational linguistic model of language processing.

Before concluding the discussion of the psycholinguistic experiments described in this chapter, it is necessary here to describe one weakness in my conclusions. The statistically significant results in these experiments have all occurred in the CONTROL condition, while the results in the SPLIT condition have been null or marginal. My conclusions rest largely on the contrast between the CONTROL and SPLIT conditions but ultimately, the conclusion that both syntactic and semantic prominence have an effect on discourse salience rests on these null results. As such, future work to clarify and confirm the present investigation should be designed such that conclusions may be based on positive experimental findings.

Finally, other work remains to be done to give the model more empirical weight. Investigating other linguistic contexts (e.g., passive-active alternations, psychological verb alternations, double-object constructions) to see how syntactic and semantic prominence

interact is a requisite next step. Furthermore, some psycholinguistic models of language processing may require a more refined notion of the store. I will consider these issues and outline some possibilities and proposals in Chapter 5.

3.6. Conclusion

In this chapter, I have presented the results of two series of on-line and off-line psycholinguistic experiments. The first series used a single rather highly-controlled linguistic environment (*spray/load* constructions) while the latter series used a less-controlled environment (*tough*/non-*tough*-constructions with various non-stative verbs in embedded clauses). Converging results from these experiments show that both syntactic and semantic prominence contribute in independent fashion to the relative salience of discourse referents for subsequent pronominal reference. The results can be explained under both a frame semantic or a Dowtian PROTO-role approach to determining semantic prominence, but suggest that for determining syntactic prominence, the hierarchical method—which ranks discourse referents by their overt realizations in clausal order, top-to-bottom, left-to-right—is more reliable. Finally, results also suggest that the repeated-name penalty experimental paradigm has some limitations in its applicability: The penalty is most observable when there is a single most salient discourse referent in the current context. Otherwise, the penalty may disappear or even become an advantage.

In the next chapter, I seek to build on the present results by describing a pilot corpus investigation designed to test the discourse salience model in the even less-controlled context of naturally occurring language.

CHAPTER 4

Corpus Investigation

4.1. Introduction

Communication is a cooperative enterprise (Clark and Haviland, 1977; Grice, 1975) in that speakers structure their utterances while keeping in mind the hearer's ability to process it, while hearers interpret an utterance based on what they know (or believe) about the speaker's state of mind (cf., Chafe, 1976; Prince, 1986; Blutner, 1998, 2000). In Chapter 3, a series of experiments using controlled language samples were performed to observe the reader's behavior while interpreting sentences. The data show that readers' pronoun resolution strategies make use of semantic information—specifically, the information assigned to the arguments of a verb—in a manner that suggests a hierarchical prominence relation among semantic roles or PROTO-role entailments. In this chapter, I investigate the writer's side of this cooperation. If, indeed, communication is cooperative, then we should expect to see writers structuring (or, packaging; see, e.g., Vallduví, 1990) their utterances in a manner that reflects the kinds of strategies they expect readers to use when interpreting those contributions. For the present research, this leads to the hypothesis that we should be able to find evidence of the same sort of semantic influence on the relationship between pronouns and their antecedents in discourse production as in discourse perception. Towards this end, I present evidence from a pilot corpus analysis which shows a connection between the semantic information provided by a verb about

its arguments and the form of referring expression used in subsequent reference to those referents those arguments introduce. As such, this information can be used by the hearer to judge what discourse referent (if any) is most salient in the current context and is thus predicted to determine the interpretation of subsequent pronominal reference.

Section 4.2 provides some background for the corpus analysis by discussing some of the existing literature on discourse salience using corpus investigative methods and establishing how the model of discourse salience outlined in Chapter 2 is predicted to be evidenced in a corpus of English texts. Section 4.3 describes the corpus and how it was analyzed. Section 4.4 gives an overview of some fundamentals of Information Theory (Weaver and Shannon, 1949) and how it will be employed as a means toward understanding the results of the corpus investigation. Preliminary results are given in Section 4.5 and provide evidence that which parallels the results of the psycholinguistic experiments in Chapter 3, suggesting that both syntactic and semantic prominence play a role in determining the salience of discourse referents.

4.2. Background

In recent years, there has been a proliferation of corpus-based linguistic investigation. Indeed, some early linguists did use corpora of linguistic data—maintained on paper—in their research, but the rapid increase and availability of computing resources has led to broader, computer-based methods for both corpus design and analysis (McEnery and Wilson, 1996; Oakes, 1998). Corpus investigations permit the testing of hypotheses over a wide variety of naturally-occurring linguistic data in which the language is far less controlled than it often needs to be in other investigative paradigms. That said,

a corpus investigation need not necessarily be the ultimate litmus test of a valid linguistic theory. However, any theory that claims to have a broad applicability across a variety of linguistic contexts is likely to make some predictions which are testable (and might only be testable) in a corpus investigation. The present investigation meets this requirement. The model of discourse salience outlined in Chapter 2 is presented as a generalized model intended to capture one aspect of the relationship between discourse structure and referring expressions. The psycholinguistic experiments described in Chapter 3 present results which are explained by the salience model in the rather narrow domain of *spray/load* constructions as well as the somewhat broader domain of *tough* constructions. However, the the strongest version of the model predicts that the kind of effects observed with *spray/load* and *tough* constructions should be observable broadly across the language (and even cross-linguistically). In this background section I will discuss these predictions in detail and also discuss how the corpus investigation may evaluate these predictions. The corpus used in the present investigation is the first part of a larger investigation which I will principally describe. At relevant points I will note some of the simplifying assumptions I make for the purpose of the present analysis.

First, I will briefly review the discourse salience model I discussed in Chapter 2 and the central hypothesis of this investigation. Where my model begins to differ (or extend) from most mainstream models is in how discourse salience is determined. After each utterance, the store and list of semantic conditions are updated. Referents are then rank-ordered in terms of salience, and this ranking then influences the processing of anaphoric noun phrases in the subsequent utterance.

In this corpus investigation, I will consider the following prediction of the model of discourse salience described in Chapter 2: If there is a coreferential link between a referring expression, r, in the current utterance and a discourse referent, X, in the context and X is the most salient (compatible) referent in the store, then r should be realized as a pronoun (cf., Gordon et al., 1993; Arnold, 1998a). In a corpus of extended texts, one would therefore expect to find a significant trend toward the presence of pronominal reference in the latter element of links between adjacent utterances.

The central question for the present investigation is then, given the nature of such coreferential links, what are the factors that determine which entity in the context is most salient? I have proposed that what has heretofore been observed as the salience of discourse referents introduced in prominent syntactic positions (e.g., as observed in Lappin and Leass, 1994; Arnold, 1998a, inter alia) might be better explained in terms of the prominence of the semantic roles or entailments imposed on these referents by the predicates for which they serve as arguments, or by some interaction between these two prominence factors. The psycholinguistic experiments in Chapter 3 show that both prominence factors influence the salience of discourse referents. This corpus investigation continues to evaluate this question by looking at the syntactic and semantic information in the discourse referents which are coreferent with subsequent pronominal expressions. The basic approach taken in this investigation is governed by three questions: One, how well does the syntactic prominence of discourse referents explain subsequent reference to those referents; two, how well does the semantic prominence of discourse referents explain subsequent reference to those referents; and three, how well do the syntactic and

semantic prominence (together) of discourse referents explain subsequent reference to those referents?

Ideally, in order to answer these questions, this study would look at each utterance and see how often elements with certain features are more salient than elements with other features. For the purposes of the present discussion, let's consider elements which are subsequently referred to pronominally as salient. Thus, this study would proceed by looking at, for example, how often subjects are more salient than objects, how often subjects are more salient than obliques, how often agents are more salient than themes, how often themes are more salient than goals, how often sentience-entailed arguments are more salient than non-sentience-entailed arguments, how often movement-entailed arguments are more salient than stationary-entailed arguments, and so on for all the different possible pairings of roles/entailments. From these comparisons, ranking hierarchies could be determined and thus, also, the effectiveness of these hierarchies in explaining subsequent pronominal reference. However, it should be clear that for this procedure to have much conclusive power, a very large corpus is required.[1]

Such a large corpus with the relevant mark-up information is not available, so some simplifying assumptions had to be made. In the present analysis, I assume that the occurrence of each syntactic role and each semantic role/entailment is *independent* of the occurrence of any other syntactic role or semantic role/entailment, respectively. That

[1]An estimate of a suitable corpus size might be made as follows. For the FrameNet roles (frame elements)—which number more than 500—there are more than 250,000 possible pairings. A corpus would have to have 1,250,000 instances of coreference in order to get the minimum 5 instances necessary for each case required to perform a χ-square test of significance. However, since some pairings are sure to be much rarer than others, many more instances would be required. Assuming 5,000,000 instances would be enough to get sufficient coverage, but that coreference among adjacent utterances occurs only half of the time and that the average utterance is ten words long, then a corpus of 100,000,000 words would be required.

is, I assume, for instance, that the syntactic prominence of a discourse referent realized as a subject in the current context is independent of whether there was a competing discourse referent in the context that was realized as an object or even as an embedded subject. In short, the corpus analysis will therefore be based on a simple comparison of the different syntactic roles and the different semantic roles and entailments. Admittedly, this is a critical assumption which is surely open to some debate. However, for the present purposes, it is a necessary assumption in order to draw any broad conclusions from the analysis. Future work will certainly be necessary to confirm or clarify the conclusions I offer in the present work.

In the discourse salience model outlined in Chapter 2, I discussed two methods for determining syntactic prominence: a role-based method determined from the grammatical function of arguments (subject, object, oblique) and a hierarchical method determined through a top-to-bottom, left-to-right clausal search algorithm. Further, I also described two methods for determining semantic prominence: a frame semantic approach and Dowtian PROTO-role approach. The psycholinguistic experiments in Chapter 3 gave some insight into the validity of these approaches for determining the salience of discourse referents. However, one of the goals of the corpus analysis for the present investigation is that to give much clearer insight into this question, showing the validity and/or usefulness of these various methods for future computational models or implementations.

Finally, one other issue to discuss is the nature of the texts used in this corpus. One of the underlying assumptions of the discourse model is that texts in which coreferential links among utterances are constructed in certain ways are more easily processed by readers than texts in which they are not (for evidence, see inter alia Arnold, 1998a;

Gordon et al., 1993; Hudson-D'Zmura and Tanenhaus, 1997, as well as Chapter 3 in this dissertation). Hence, the corpus should contain texts in which writers have taken some care to maximize the ease of readability. I suggest that works of narrative fiction are good candidates for this (cf., Brown, 1983): Authors are focused on writing planned texts clearly with little or no (unintended) ambiguity.

In comparison, news articles may be said to share these same features. However, news articles have other limitations that make them less than optimal for this investigation. Articles are generally written with shorter paragraphs so that copy editors easily know where to make last-minute cuts according to space limitations. This has two effects: Authors are less inclined to use pronominal reference because the removal of paragraphs may break these cohesive links. Furthermore, readers may be more likely to read a new paragraph as a new discourse segment, with possibly differing expectations with respect to the accessibility of previously introduced discourse referents (cf., Grosz and Sidner, 1986; Vonk et al., 1992).

A further advantage of fictional texts is that they parallel the texts used in the psycholinguistic experiments described in Chapter 3. In those experiments, participants read short five-sentence vignettes with varying degrees of intersentential coherence, varied by the use of pronouns. Fictional texts therefore mirror the same basic writing style, but can be assumed to emulate the ideals in terms of inter-utterance coreference.

4.3. Corpus Design

The goal of this corpus analysis is to investigate and compare the effects of syntactic as well as semantic prominence with respect to both FrameNet case roles and Dowtian PROTO-role entailments on the discourse salience of referents for subsequent pronominal

reference, The corpus therefore needs mark-up with respect to syntactic and semantic information (both roles and entailments) in addition to coreference relationships. I can find no publicly available corpora which have all of these features. As a result, I have opted to build my own corpus. In this section, I will describe the specific texts I have chosen to include in the corpus as well as the mark-up procedures.

4.3.1. Texts

The corpus consists of contemporary works of fiction by semi-professional authors. The texts were selected from *InterText* (http://www.intertext.com/), an online, refereed journal of fiction. The InterText corpus between 1991 and 2004 includes 55 issues with typically four or five short stories in each. This corpus was sampled by proceeding through the issues chronologically and choosing the first five stories which met the following criteria.

- third-person narrative
- not longer than 2000 words
- sparse quotations

Texts which were not narratives were excluded as well as texts written in a first or second-person perspective. The length of the texts was initially limited to 1,000 words but this was found to be too restrictive, there were only a couple of stories that met such requirements. So, the word limit was raised to 2,000. The limit was set in order to try to use entire texts, but at the same time use as many different authors as possible.

Finally, texts which had too much quoted material[2] were also excluded (cf., Arnold, 1998a). While quotations probably do not interfere with the overall coherence of a text, it is not clear what influence they have on pronominal reference. For instance, it is highly unlikely that a character in the story would be quoted as using a pronominal reference to a character the author referred to prominently in the preceding sentence of the narration.

4.3.2. Annotation Procedures

4.3.2.1. Extensible Mark-up Language (XML). The texts were annotated using eXtensible Mark-up Language (XML)—a protocol for defining the hierarchical structure of a document in terms of relations among elements of the document and optional attributes of those elements as defined in a document template definition (DTD). The XML standard is maintained by the World Wide Web Consortion (http://www.w3c.org/). In a well-formed XML document, content is delimited by tags in the format <mytag> some content </mytag>. The DTD is a grammar for determining the proper hierarchical relationships among tags. For instance, imagine we wish to define the generic structure of a document we will call <message>. The DTD for this might be as shown in (156).

(156)
```
<!DOCTYPE message [
    <!ELEMENT message    sender,recipients,date,subject?,body >
    <!ELEMENT sender     #PCDATA >
    <!ELEMENT recipients recipient+ >
    <!ELEMENT recipient  #PCDATA >
```

[2]I had no special algorithm for determining whether a story contained too much quoted material. Using a simple eyeball scan of the story, if it seemed to me that there were a lot of quotations, the text was rejected.

```
        <!ELEMENT date       #PCDATA >
        <!ELEMENT subject    #PCDATA >
        <!ELEMENT body       #PCDATA >
    ]>
```

This structure shows that a <message> document contains five elements: <sender>, <recipients>, <date>, <subject>, and <body>, in that order. However, the "?" after the <subject> element means that this element is optional. Furthermore, the <recipients> element contains one or more <recipient> elements (indicated by the "+" operator). The #PCDATA symbol means that that element simply contains "character data" (plain text). Given this definition, we can then determine that (157) is a well-formed XML document and is a valid example of the <message> type. The root node is <message> and it contains five child nodes. These nodes contain either further element nodes (e.g., <recipients> contains two <recipient> nodes) or text nodes.

(157)
```
<message>
    <sender>Ken</sender>
    <recipients>
        <recipient>Miles</recipient>
        <recipient>Jaime</recipient>
    </recipients>
    <date>April 1st</date>
    <subject>Party</subject>
    <body>
        I'm having a birthday party next week.
```

```
        Please come!
    </body>
</message>
```

XML has become a common protocol for document representation in data storage and internet applications as more and more standardized DTDs become available and tools for marking up and working with XML documents become easier and more widely available. As such, XML is also being more widely used as a mark-up protocol for linguistic corpora.

The following subsections describe the structure of the XML mark-up used in my pilot corpus investigation. The actual mark-up was performed using the CL@RK System, Version 2.0, jointly developed by Seminar fuer Sprachwissenschaft (Tuebingen, Germany) and Linguistic Modelling Laboratory (Sofia, Bulgaria).

4.3.2.2. Elements. The body of each text in the corpus is initially delimited with respect to simple typographical features: the <body> consists of a sequence of paragraphs, <p>, each terminated by a blank line in the InterText publishing format, and each <p> consists of a sequence of sentences, <s>, each terminated by a period (not including periods in abbreviations, etc.), one member of punctuation marks, <punc>.

Sentences are subsequently broken down and marked-up with a relatively flat syntactic analysis based on clauses <c>. Each clause contains (at most) one <verb> with its argument noun phrase <np> nodes as siblings. Noun phrases may further contain

other noun phrases or clauses. In addition, a noun phrase may contain one <pronoun> node.[3]

The lowest level of mark-up is <w> which marks words, as delimited by whitespace or punctuation in the text. The crucial part of the DTD for the hierachical substructure of sentences is as shown in (158).

(158)
```
<!ELEMENT s        ( w | c | punc )+ >
<!ELEMENT c        ( w | c | np | punc )*, verb?,
                   ( w | c | np | punc )* >
<!ELEMENT np       ( ( w | np | c | punc )+ | pronoun ) >
<!ELEMENT pronoun w* >
<!ELEMENT verb     w* >
<!ELEMENT w        #PCDATA >
```

Most of the syntax of these rules should be clear from the discussion above but a couple of other conventions need explanation. A set of elements surrounded by parentheses and delimited by vertical bars "|" indicates an option set: any one of the elements in the set will satisfy (that part of) the rule. Hence, <s> elements may contain a sequence of one or more elements of the option set {<w>, <c>, <punc> }. A "*" indicates that the preceding item may occur zero or more times. Thus, in a <c> element, the (optional) verb element may be preceded by a sequence of at least zero elements from the option set {<w>, <c>, <np>, <punc> }. Finally, a sequence of items delimited by "," indicates an ordered set of items. Thus, a <c> element must contain three parts, in order: a

[3]The decision to include a separate pronoun element in the document template definition was motivated purely by practical concerns. I found that having a separate pronoun element made for easier use of certain corpus analysis methods and tools.

sequence of {<w>, <c>, <np>, <punc> }, a (optional) <verb>, and another sequence of {<w>, <c>, <np>, <punc> }.

To illustrate the XML mark-up for the DTD, consider the sentence in (159) for which the mark-up is given in (160). The mark-up also illustrates the use of attributes on nodes, shown as additional mark-up inside the opening tag of each node. Here, the id attribute is shown for all <np>, <verb>, and <pronoun> elements. In the XML specification, unique identifier attributes may be defined (as is id here).

(159) John told Linda in an excited voice that he was going to Spain.

(160)
```
<s><c><np id="1">John</np>
      <verb id="2">told</verb>
      <np id="3">Linda</np>
      in
      <np id="4">an excited voice</np>
      that
      <c><np id="5">
           <pronoun id="6">
             he</pronoun></np>
         <verb id="7">was going</verb>
         to
         <np id="8">Spain</np>
      </c>
   </c>
```

```
    <punc>.</punc>
</s>
```

While all the sentences in the corpus are produced by this grammar, it should be noted that this grammar overgenerates. For example, the <np> element definition allows any combination of words, noun phrases, and clauses in any order to form a noun phrase. This is surely not correct. Unfortunately, the XML specification places some restrictions on the kinds of regular expressions which may produce elements. However, for the present investigation, the descriptive accuracy of this grammar is sufficient: The crucial feature is that the domain of each verbal predicate be defined such that its direct noun phrase arguments can be determined.

In addition to the unique `id` attribute given to <verb>, <np>, and <pronoun> elements, <np> elements were given another identifier. For the analysis described below it was important to be able to determine coreference among items in the corpus. Therefore, every <np> element was assigned a non-unique `referent` identifier. For any given identifier within a marked-up text, every <np> which has that identifier as its `referent` attribute is regarded as being coreferential. Thus, assuming that a text contains only one character named *John*, then every <np> node corresponding to an occurence of *John* in the text would have its `referent` attribute set to the same identifier string, say "JOHN", or even "09823".[4] Thus, the sentence in (161) was represented as in (162).

(161) The teacher scolded Max. The boy had been naughty.

(162) `<s><c><np id="1" referent="TEACHER">The teacher</np>`

[4]As far as the XML processors are concerned, the content of an id string doesn't matter. However, for greater user readability, I tried to use identifier strings that bore some relation to the referent of the referential expression. Thus, if a <np> node representing JOHN had the `referent` value "JOHN", then a <np> node representing JOHN'S HEAD had the `referent` value "JOHN:HEAD".

```
      <verb id="2">scolded</verb>
      <np id="3" referent="MAX">Max</np>
   </c>
   <punc>.</punc>
</s>
<s><c><np id="4" referent="MAX">The boy</np>
      <verb id="5">had been</verb>
      naughty
   </c>
   <punc>.</punc>
</s>
```

In short, then, these identifier strings correspond to the notion of discourse referents in the discourse model I outlined in Chapter 2. I should also note that because some referential expressions share the same identifiers, these strings further represent my judgments of coreference information within each entire text.

4.3.2.3. Pronouns. The <pronoun> elements in the corpus fell into one of five categories: expletive pronouns as in (164); pronouns referring to events as in (165); pronouns which refer to extra-textual entities as in (166); overt pronouns which refer to referents introduced elsewhere in the text as in (167), including both bound and free pronouns; and covert pro-forms as in (168).[5]

[5]It should be noted that this is not an exhaustive categorization of pronoun occurrences in natural text. For instance, one other type of pronoun involves reference to discontinuous constituents as in (163).

(163) John$_i$ shoved Matt$_j$ before they$_{(i+j)}$ got into a fight.

These other categories simply did not occur in this corpus, possibly because of the relatively small sample size.

(164) *It* seems that John bought a new car.

(165) John bought a new car.

That infuriated his wife.

(166) *They* say that practice makes perfect.

(167) John bought a new car.

He drove *himself* around in *it* all day long.

(168) John wanted ∅ to buy a new car.

The overt pronouns were marked as described above, while the covert pro-forms were represented as empty <`pronoun`> elements. Coreference among various elements was then determined by looking at the identifier strings encoded in the `referent` attribute of the parent <`np`> elements.

4.3.2.4. Syntactic Role Information. As a result of the tagging method, all of the noun phrase arguments of a verb were represented in the XML tree as <`np`> siblings of the respective <`verb`> node (and thus each other, as well). However, this did not mean that all <`np`> siblings of a <`verb`> represented noun phrase arguments since noun phrases in adjuncts were not structurally distinguished. In order to take advantage of syntactic information in later analysis, it was necessary to insert syntactic role information directly. The syntactic role of the noun phrase arguments of the verb was directly encoded in the mark-up as `synrole` attributes of the <`np`> node. These roles included subject, object, oblique, and none, where none is taken to mean "none of the others" (not "having no syntactic role"). The standard for determining such roles requires some discussion. Only one pre-verbal <`np`> was marked as the subject. Any post-verbal

noun-phrase not preceded by a preposition was marked as an object. Any arguments preceded by a preposition were marked as oblique.[6]

All argument noun phrases had their syntactic role marked in the XML mark-up. The status of a noun phrase as an argument of a particular verb was judged with respect to the FrameNet system. After determining which frame the verb was a member of (discussed in detail below), and thus the relevant semantic roles for that frame, if a noun phrase bore one of the roles, then it was considered an argument and its syntactic role was indicated as above. Otherwise, it was marked as none. For example, consider the noun phrase *at the White House* in (169). Here, the verb *throw* fits in the CAUSE_TO_MOVE frame. In this frame, the role goal is available and is a suitable role for WHITE HOUSE. Thus, the <np> node representing WHITE HOUSE would have its syntactic role attribute marked as oblique. However, the verb *discuss* in (170) is in the SPEAK_ON_TOPIC frame which does not have any suitable role to describe WHITE HOUSE. Thus here, the <np> node representing WHITE HOUSE would have its syntactic role attribute marked as none.

(169) The protester threw the eggs at the White House.

(170) The president discussed the bill at the White House.

The mark-up for (169) is thus illustrated in (171) (leaving out irrelevant details).

(171)
```
<s><c><np synrole="subject">The protester</np>
      <verb>threw</verb>
      <np synrole="object">the eggs</np>
```

[6]Marking all arguments preceded by a preposition as obliques is a somewhat questionable choice. The internal arguments accompanying such phrasal verbs as in *Jane is fond of tennis* could arguably be called objects rather than obliques because post-verb material is arguably not a constituent (i.e., **It is [of tennis]$_i$ that Jane is fond* e_i). To simplify the present research, I decided not to try to distinguish such cases, and instead used prepositions as a marker of the oblique status of arguments.

```
        at
        <np synrole="oblique">the White House</np>
      </c>
      <punc>.</punc>
    </s>
```

4.3.2.5. Semantic Role Information. The noun phrase arguments of a verb were marked with respect to two semantic schemas: the frame elements of FrameNet semantics and the Dowtian PROTO-role entailments. These systems and their respective roles were described in detail in Chapter 2, but some practical comments with respect to this corpus are necessary here.

Verbs. Crucial to determining the semantic role information for noun phrases was first to identify one crucial attribute of the verb heading each clause. This attribute was the `lexeme`, here taken as the dictionary form of the verb. The `lexeme` was important because as the corpus mark-up progressed, continuous reference was made to earlier occurrences of a verb's lexeme in order to ensure consistency.

FrameNet frames and elements. In the 2001 release of FrameNet II, the FrameNet system covers 6,800 lexical units in 376 semantic frames. However, it is not exhaustive. A number of predicates in the corpus could not be found in the FrameNet list of lexical units. In such cases, the following steps were taken.

- The WordNet database (Fellbaum, 1998; Miller, 1995) was consulted for synonyms in the relevant sense (as judged by my own intuitions) of the original `lexeme`. If one of these synonyms could be found in the FrameNet database, then the semantic frame in which that synonym fit was used.

- If no suitable synonym from WordNet could be found, then the FrameNet list of semantic frames was browsed to find a suitable frame for the verb (again, according to my own intuition).

For instance, the verb *fathom* was not found in the FrameNet database. However, one of its synonyms, *understand*, is in the database in the frame AWARENESS, which I judged to be a suitable fit. However, the verb *charm*, as in *The glow charmed Annabella with its steady light*, was not in the FrameNet database and neither were its relevant WordNet synonyms, *bewitch* and *entrance*. However, searching through the frames yielded a suitable if somewhat broad frame as EXPERIENCER_OBJ. For a few items, there was no suitable frame. For instance, while there is a frame that covers obligations, BEING_OBLIGATED, which includes such lexical units as *have* as in *John has to go home now*, there was no comparable frame to capture ability as in *John is able to play tennis*. Separately, one might argue that it wouldn't be necessary in this case since *be able to* expresses modality and the only really relevant predicate is *play*. However, since the FrameNet system extricates obligation, it seems—for the sake of consistency—that it should also do so for ability. In such cases as these, I chose a frame that was as close as possible to the desired frame. In the case of ability, I settled on SUITABILITY as a reasonably good compromise except that while ability expressions seem to say something about the properties of agents with respect to performing some activity, SUITABILITY evaluates an entity with respect to some purpose for which it may be used. Nonetheless, these sorts of classification difficulties occurred with only a small number of items.

One important decision that had to be made was how to handle the kinds of cases shown in (172)-(173).

(172) a. The problem was easy (for Jill).

b. John is just a regular boring guy.

(173) Mark is the one in the red jacket.

The sentences in (172) are copular expressions in which some property is predicated of the subject via an adjective in (172a) or a noun phrase in (172b). The FrameNet documentation (Johnson et al., 2002) discusses these sorts of cases and calls them *predicative* expressions. However, Johnson et al. do not define any particular frame to handle these. In some cases, where the predication is relatively straightforward as in (172a), a suitable frame could be found. Thus, for the adjective *easy* (or perhaps more accurately, for the verb phrase, *be easy*), the frame DIFFICULTY could be used where the subject bears an activity role and the noun phrase following *for* bears an experiencer role. However, more complex predicative expressions as in (172b) had no suitable frame. Therefore, for such cases as these, I devised a new frame called PREDICATION in which the subject bears a described role and the object bears a description role.

The situation in (173), in which two referential expressions are asserted to have the same real-world referent is called a *specifying* case in the FrameNet documentation. However, as above, Johnson et al. do not provide any specific frame to describe this situation. Therefore, I decided to refer to these cases as SPECIFICATION frames in which the subject bears an entity1 role and the object bears an entity2 role.

After determining the relevant FrameNet frame for a particular verb `lexeme`, frame element labels were applied to the participating noun phrase arguments. Currently, in the corpus, only noun phrases are counted as arguments and so marked. Hence, arguments of other category types (e.g., clauses as in *He wanted [PRO to visit Alaska]*) are not

currently marked although this information may be added at a later date in order to look at event reference in contrast to entity reference.

All other noun phrases (in adjuncts, nominal modifiers, genitives, etc.) were marked with a semantic role, NONE.

Dowtian PROTO-role entailments. Every noun phrase argument which was marked with a frame element under the FrameNet system was also marked with respect to the PROTO-role entailments placed on it by the verb clause in which it resides. The tests used to determine these entailments are described in Section 2.7.1. However, some practical issues with respect to mark-up require discussion. While every obligatory noun phrase argument received a frame element label as described above, many arguments did not pass any of the entailment tests and therefore had no markings for their PROTO-role entailments. For instance, in the sentence *He held his head in his hands*, it is not clear to me that any of the entailments apply to any of the arguments. There is some temptation to regard the subject as sentient and volitional, but this is probably a result of inference from the human nature of the participant. Consider *The tree held the bird's nest in its branches.* This seems perfectly fine to me and not a metaphorical use of *hold*. There are a small number of such verbs in which none of the arguments have any PROTO-role entailments, but a much larger number in which not all arguments have some entailments. At a minimum, this is unfortunate, because certainly there is more semantic information in the verb about these arguments than the Dowtian system currently yields. This limitation means that the corpus analysis with respect to PROTO-role entailments will not have quite the conclusive strength it would if all the noun phrases in the corpus contributed some information. At worst, the analysis may be skewed toward

some less representative sub-group of verbs. In the discussion section below I will discuss this issue further and the degree to which it may have influenced the results.

4.3.2.6. Recovery of Implicit Information. One feature of language is that not all information is communicated explicitly. Such phenomena as ellipsis, gapping, and control structures lead to surface structures in which noun phrase arguments or even verbs are missing. A good illustration of this is with what has been referred to as "inferrables" (Prince, 1992). In (174), it is clear that Bill ate some hot dogs, even though the surface structure of the discourse does not explicitly provide this information. Furthermore, the amount of hot dogs that Bill ate is even accessible (although there may be some disagreement on this) for subsequent linguistic reference as the pronoun in (174c) illustrates.

(174) a. John and Bill took part in a hot dog eating contest.

b. John didn't eat anywhere near as many as Bill did.

c. It was a record-breaking amount.

Just as every human has some sort of cognitive strategy for recovering such implicit information (for evidence of such strategies see Carlson and Tanenhaus, 1988), any discourse understanding system must have some procedure for recovering this information (cf., Palmer et al., 1986). While such a procedure is not an integral part of the present research, it was necessary to decide which information was truly implicit and then to manually make note of this in the mark-up for the purpose of later analysis. The following four general cases are described: verb-phrase ellipsis, control structures, gaps in relative clauses, and conjoined elements.

Verb-phrase ellipsis. Elided verb phrases as in (175) were recovered by manually adding in a null <verb> node at the ellipsis site with the `lexeme` attribute set equivalent to that of the antecedent verb phrase. This was followed by any relevant <np> arguments, each inserted together with a null <pronoun> element as its only child. Thus, the mark-up for (175) is as in (176) (leaving out some currently irrelevant details). The ellided verb phrase, *see Matt* is represented in the latter clause by the sequence of nodes: <verb id="5"> and <np id="6>.

(175) John saw Matt and Bill did ∅, too.

(176)
```
<s><c><np id="1" referent="JOHN">John</np>
      <verb id="2" lexeme="see">saw</verb>
      <np id="3" referent="MATT">Matt</np>
   </c>
   and
   <c><np id="4" referent="BILL">Bill</np>
      did
      <verb id="5" lexeme="see"></verb>
      <np id="6" referent="MATT">
        <pronoun id="7">
          </pronoun></np>
      too
   </c>
   <punc>.</punc>
</s>
```

In (176,) the `referent` attribute of the inserted <np> was set equal to that of the relevant <np> in the antecedent verb phrase. However, in some cases a noun phrase argument was not necessarily interpretable as coreferent with the corresponding noun phrase in the antecedent clause as in (177).

(177) John read a book and Matt did ∅, too.

Here it is entirely possible (and perhaps even likely) that John and Matt did not read the same book. In these cases, the `referent` attribute was assigned a different identifier.

Control structures. Structures such as those in (178) which are typically analyzed as having PRO in the subject position of the complement clause (Chomsky, 1981) were marked by inserting a <np> with a null <pronoun> child. The `referent` attribute of the <np> was set equal to that of the <np> corresponding to its controller, as shown in (179). Thus, the *PRO* element is represented by <np id="3" referent="JOHN">.

(178) John wanted *PRO* to eat lunch.

(179)
```
<s><c><np id="1" referent="JOHN">John</np>
      <verb id="2" lexeme="want">wanted</verb>
      <c><np id="3" referent="JOHN">
           <pronoun id="4">
             </pronoun></np>
         to
         <verb id="5" lexeme="eat">eat</verb>
         <np id="6" referent="LUNCH">lunch</np>
      </c>
   </c>
```

```
    <punc>.</punc>
  <s>
```

Gaps in relative clauses. A third element which requires information recovery are gaps resulting from *wh*-movement in relative clause constructions (Chomsky, 1981) as in (180). These constructions were marked by inserting a <np> element with a null <pronoun> child at the gap site. The `referent` attribute of the inserted <np> was then set equal to that of the <np> which contains the relative clause. Thus, the mark-up of (180) is illustrated in (181).

(180) John read the book that Matt bought ∅.

(181)
```
<s><c><np id="1" referent="JOHN">John</np>
      <verb id="2" lexeme="read">read</verb>
      <np id="3" referent="BOOK">the book that
        <c><np id="4" referent="MATT">Matt</np>
           <verb id="5" lexeme="buy">bought</verb>
           <np id="6" referent="BOOK">
             <pronoun id="7"></pronoun></np>
        </c>
      </np>
   </c>
   <punc>.</punc>
</s>
```

Conjoined elements. Another case where implicit information had to be inserted was in the treatment of some types of conjunctions. Consider the sentences in (182)-(184).

(182) John finished graduate school and then he took a trip to Europe.

(183) John went to graduate school and studied linguistics.

(184) John built and moved into his new home in just three months.

Cases such as (182) in which two fully complete clauses are conjoined were represented as <c> node children of a <s> node as shown in (185) (with irrelevant details omitted). Cases such as (183) however, in which only verb phrases were conjoined, were treated somewhat differently. The two verb phrases were represented as two adjacent <c> nodes. The subject of the latter clause was inserted as a <np> node with a null <pronoun> child. The referent attribute of the <np> was set equal to that of the <np> element corresponding to the overt subject of the first <c> conjunct. Finally, these two <c> nodes were then made children of one parent <c> node as in (186).

(185)
```
<s><c>John finished graduate school</c>
   and then
   <c>he took a trip to Europe</c>
   <punc>.</punc>
</s>
```

(186)
```
<s><c><c><np id="1" referent="JOHN">John</np>
         <verb id="2">went</verb>
         to
         <np id="3" referent="GRAD_SCHOOL">graduate school</np>
```

```
        </c>
        and
        <c><np id="4 referent="JOHN">
             <pronoun id="5"></pronoun></np>
           <verb id="6">studied</verb>
           <np id="7" referent="LINGUISTICS">linguistics</np>
        </c>
    </c>
    <punc>.<\punc>
</s>
```

The different treatments of conjoined clauses and verb phrases not only reflect the different linguistic analyses of these sentences, but also has consequences for the present investigation. As discussed below, the basic unit of analysis in this study is pairs of adjacent utterances, where an utterance is regarded as the immediate child of an <s> node. Hence, the conjoined clause case in (182) is taken as an utterance pair and thus the coreference between *John* and *he* is taken as an instance of intersentential coreference and is included in the analysis. However, in the conjoined verb phrase case in (183), there is only one utterance. Thus, the coreference between *John* in the first conjunct and the null <np>-<pronoun> element in the second conjunct are taken only as an instance of intrasentential coreference, thus not included in the analysis.

A more complex case of conjunction is as shown in (184) in which the initial subject *John* and the object *his new home* are interpreted as being arguments of both *build* and *move into*. This also was represented as a sequence of adjacent <c> nodes where null

<np>-<pronoun> nodes were added as necessary to fill out the interpretation as shown in (187). This has the result that the null <np>-<pronoun> has as its antecedent a node which actually follows it (in the second conjunct)—a case of cataphoric (forward-directed) reference.

(187)
```
<s><c><c><np id="1" referent="JOHN">John</np>
          <verb id="2">built</verb>
          <np id="3" referent="JOHN:HOME">
            <pronoun id="4"></pronoun></np>
        </c>
        and
        <c><np id="5" referent="JOHN">
            <pronoun id="6"></pronoun></np>
          <verb id="7">moved</verb>
          into
          <np id="8" referent="JOHN:HOME">his new home</np>
        </c>
        in
        <np id="9" referent="THREE_MONTHS">three months</np>
      </c>
      <punc>.</punc>
    </s>
```

Although there might be some theoretical questions about these representations as syntactic analyses of these constructions, I opted for these representations as a means of

ensuring that every <c> node in the mark-up had at most one <verb> child with its <np> arguments (explicit or implicit) as siblings. Practically, this allowed for a simpler DTD, as well as made many aspects of validation and subsequent data extraction and analysis easier.

4.3.3. Method of Analysis

In terms of the discourse model outlined in Chapter 1, the context—the mental representation of the discourse thus far—is determined by the cumulative representation of the preceding utterances. It is against this context that the current utterance—in particular, the referential expressions therein—is interpreted. However, in order to simplify the analysis, I will assume a model of discourse salience in which recency is a very strong factor, and therefore, investigate only highly local cases of coreference between the current utterance and the context: Thus, the main unit of analysis here will be pairs of adjacent utterances where an utterance is operationally defined as a <c> node child of a <s> node. Hereafter I will refer to the first utterance in each pair as the context. The focal point of the study is then the set of discourse referents introduced in the context and the syntactic and semantic information those referents bear, as well as instances in which a discourse referent in the context is evoked again in the second utterance and the form of that reference (i.e., pronoun or not).

For each context utterance, <c>, a set of discourse referents was extracted by determining all of the unique entities evoked in that utterance. This involved searching all of the descendant nodes of the root <c> node for <np> nodes with a unique `referent` attribute. This list thus constitutes the store for the current context. Then, for each

discourse referent, its syntactic and semantic information were determined. Two methods for determining the syntactic prominence of discourse referents were presented in Chapter 2: a role-based method, based on a hierarchy of grammatical functions and a hierarchical method based on a top-to-bottom, left-to-right clausal search of the syntactic parse tree. How these methods were actually employed in the corpus analysis requires some comment. For the role-based method, the syntactic prominence of a discourse referent was determined from the highest role on the syntactic prominence hierarchy for which that referent had an overt noun phrase realization in the utterance. Thus, if a referent was realized as a subject somewhere in the utterance—even in an embedded clause—then the referent's syntactic prominence was determined by this subject realization. For the hierarchical method, the syntactic prominence of a discourse referent was determined from the highest ranking of its realizations in a list of all overt noun phrases ordered according to the top-to-bottom, left-to-right clausal search algorithm described in Section 2.3. As an illustration of these two methods, consider (188) with its mark-up in (189).

(188) John asked Bill to pay for his dinner.

(189)
```
<s><c><np id="1" referent="JOHN">John</np>
      <verb id="2">asked</verb>
      <np id="3" referent="BILL">Bill</np>
      <c><np id="4" referent="BILL">
           <pronoun id="5"></pronoun></np>
         <verb id="6">to pay</verb>
         for
```

```
            <np id="7" referent="JOHN:DINNER">
              <np id="8" referent="DINNER">
                <pronoun id="9">his</pronoun></np>
              dinner
            </np>
          </c>
        </c>
        <punc>.<punc>
      </s>
```

The discourse referent corresponding to John is realized as a subject and a genitive. Because genitives are not counted in the syntactic prominence methods, the prominence of John (the referent) is determined from its subject realization. Bill, however, is realized as an object in the matrix clause and a subject in the embedded clause. But the latter is not an overt realization is thus not counted toward syntactic prominence. The prominence of Bill is therefore determined by its object realization. Finally, John's dinner is realized only as an oblique and therefore its syntactic prominence is determined from this role. As such, for the present analysis, the syntactic prominence values of the referents John, Bill, and John's dinner was taken as subject, object, and oblique respectively. Under the hierarchical method, the ranking of the discourse referents is {John, Bill, John's dinner}. Thus, the syntactic prominence values were taken as 1, 2, and 3, respectively.

With respect to either method, the syntactic prominence of each referent was a single, discrete value: by the role-based method, the value was one of the role labels,

while by the hierachical method, the value was a numeric integer greater than or equal to 1. However, the situation with semantic prominence is rather more complex. First, unlike syntactic prominence, there is very little a priori evidence as to what the ranking hierarchy will look like. Thus, if a discourse referent has two realizations in an utterance, one as an instrument, and another as a goal, there is no a priori reason to select one over the other as the determining that referent's semantic prominence (However, I do suggest a procedure for this in the results section, below). Therefore, the semantic prominence of the referents was not a single discrete value, but rather a set of values. Under the frame semantics approach, a referent's semantic prominence was a set of all the semantic roles assigned to its realizations (overt or null) within the utterance. Under the Dowtian PROTO-role approach, the semantic prominence was a set comprising the union of all the entailments placed on its realizations within the utterance. To illustrate, consider the representation of (188) in (190).

(190)
```
<s><c><np id="1" referent="JOHN" synrole="subject"
        semrole="speaker" sentience="yes" volition="yes">
    John</np>
  <verb id="2" frame="request">asked</verb>
  <np id="3" referent="BILL" synrole="object"
      semrole="addressee" sentience="yes">
    Bill</np>
  <c><np id="4" referent="BILL"synrole="subject
         semrole="buyer" sentience="yes" volition="yes">
       <pronoun id="5"></pronoun></np>
```

```
            <verb id="6" frame="commerce_pay">to pay</verb>
            for
            <np id="7" referent="JOHN:DINNER" synrole="oblique"
                semrole="goods">
              <np id="8" referent="JOHN" synrole="none"
                  semrole="specifier">
                <pronoun id="9">his</pronoun></np>
              dinner
            </np>
          </c>
        </c>
        <punc>.<punc>
      </s>
```

Under the frame semantics approach, the discourse referent JOHN has the semantic prominence of {speaker}, BILL has the semantic prominence of {addressee, buyer}, and JOHN'S DINNER has the semantic prominence of {goods}. Under the PROTO-role approach, these semantic prominence values are, respectively, {sentience, volition }, {sentience, volition }, and { } (the empty set). In the analyses below, all of the items in a discourse referent's semantic prominence set are counted. In other words, when calculating how often a discourse referent with a goal realization is referred to in the subsequent utterance, I will not take into account whether that discourse referent was also realized as, say, an agent or an experiencer. As discussed above, the relatively small size of the corpus and the large number of possible semantic values (e.g., agent,

patient, etc.) makes such an in-depth analysis difficult. This simplification has some consequences for the analysis. It could be the case that a particular role, x, shows a high correlation between its occurrence and subsequent pronominal reference, but that it is simply a reflex of x's co-occurrence with some highly prominent role, y.

4.4. Information Theory

In the results section which follows, after presenting some basic descriptive statistics of the corpus showing the frequencies of the various syntactic and semantic features, I will give an analysis of the data using some fundamental concepts in information theory (Weaver and Shannon, 1949). Therefore, in this section, I give an overview of these concepts.

Given a probability space, P, divided into a number of possible outcomes, o_1, o_2, ... o_n (also called *cells* in P), if we know that all the outcomes are equally probable, then our uncertainty about which outcome will occur is very high. A good example of this is a fair, six-sided die. We know that on any given throw, each of the six faces is equally likely to end up on top. However, if the cells of P are not equally divided, say, one cell takes up most of the space, then our uncertainty about the outcome is much reduced. This might be exemplified by a die which is loaded. In this case, we can be much more certain about the outcome.

Entropy is a concept in information theory used to estimate the degree of uncertainty in a given probability space. It can be calculated using the formula shown in (191).

(191) $\quad H = -\sum_{i=1}^{n} p(o_i) log_2(p(o_i))$

For example, consider the probability space, C, which represents the toss of a fair coin. There are two possible outcomes—heads or tails—and each is equally likely to occur. Thus, the probability of heads, $p(heads)$, is $1/2 = 0.5$. Similarly, the probability of tails, $p(tails)$, is 0.5. The entropy in this situation is therefore $H_C = 1$, as calculated in (192). This is the maximum possible entropy for a two-cell probability space.

(192) $$H_C = -[p(heads) * log_2(p(heads)) + p(tails) * log_2(p(tails))]$$
$$H_C = -[0.5 * log_2(0.5) + 0.5 * log_2(0.5]]$$
$$H_C = -[0.5 * (-1) + 0.5 * (-1)]$$
$$H_C = 1$$

If we had a weighted coin for which the outcome probabilities are not equal, then the entropy will be less. If we found we had a coin in which *both* sides were heads, the probability space, C1 would be very different with the probability of the outcome being heads $p(heads) = 1$. The entropy in this case would then be $H_{C1} = 0$ as shown in (193) (Note that $p(tails) = 0$ and is no longer part of the probability space and is therefore not included in the calculation of entropy). There would be no uncertainty in the system because we know what the outcome will always be: heads.

(193) $$H_C1 = -[p(heads) * log_2(p(heads))]$$
$$H_C1 = -[1 * log_2(1)]$$
$$H_C1 = -[0.5 * 0]$$
$$H_C1 = 0$$

Conditional entropy is the entropy in a probability space given that we know some other information. For instance, consider a probability space, U, representing a trial in which we open an English book to some random page and point at some random letter

and then ask whether or not that letter is "u". This probability could be estimated by the prior probability of the occurrence of "u" in the book as a whole, $p(\text{“u”})$. We could then calculate the entropy for this question from $p(\text{“u”})$ and $p(\neg\text{“u”})$. However, what if we learned that the letter before the letter in question is "q"? Then, of course, the probability that the letter is "u" will be much higher, nearly $p(\text{“u”} \mid \text{“q”}) = 1$, and our uncertainty would decrease considerably. Hence, the conditional entropy—the entropy in U given that we know the preceding letter is "q", $H_U(\text{“q”})$—would reduce to nearly 0. So, learning that the preceding letter is "q" results in a reduction in entropy. On the other hand, logically, learning that the preceding letter is not "q" would result in an increase in entropy (though in this case probably only a slight increase). The entropy reduction is calculated as the difference between the prior entropy and the conditional entropy: $H_U - H_U(\text{“q”})$. It is important to note that entropy reduction may be negative: It is possible that learning some information causes us to become *more* uncertain about the outcome.

In order to illustrate entropy reduction, consider a probability space, D, representing the roll of a fair, six-sided die. The probability for any one of the six sides is $p(o_i) = 1/6 = 0.167$. The entropy, H_D is therefore calculated as shown in (194). However, if we learned that the outcome was an even number, then there would be three possible outcomes, and the probability of each would be $p(o_i) = 1/3 = 0.33$. The conditional entropy, $H_D(even)$, would then be calculated as in (195). The reduction in entropy—which is called "information value" (IV)—of learning that the outcome was even can then be calculated as the difference between the prior entropy, H_D, and the conditional entropy, $H_D(even)$ as shown in (196).

(194) $H_D = -\sum_{i=1}^{6} p(o_i) * log_2(p(o_i))$

$H_D = -[6 * 0.167 * log_2(0.167)]$

$H_D = 2.58$

(195) $H_D(even) = -\sum_{i=1}^{3} p(o_i) * log_2(p(o_i))$

$H_D(even) = -[3 * 0.33 * log_2(0.33)]$

$H_D(even) = 1.58$

(196) $IV(even) = H_D - H_D(even) = 2.58 - 1.58 = 1.00$

One more important concept in information theory is expected information value (EIV). This is simply the weighted sum (with respect to probabilities) of the entropy reductions across all the various conditions. In the present example these conditions would be learning that the outcome is even and learning that the outcome is not even. Since the numerical values for the not even case are equivalent to those for the even case, $H_D(\neg even) = 1.58$ and $IV(\neg even) = 1.00$. Therefore, $EIV(even)$ can be calculated as shown in (197.)

(197) $EIV(even) = p(even)IV(even) + p(\neg even)IV(\neg even)$

$EIV(even) = 0.5 * 1.00 + 0.5 * 1.00$

$EIV(even) = 1.00$

In the results section which follows, I will evaluate the relative contribution of syntactic and semantic prominence to discourse salience first by looking at the EIVs of the syntactic and semantic information information of a discourse referent with respect to subsequent pronominalization of that referent. This means, for instance, looking at the EIV of learning that a particular referent was a subject, EIV(subject), or an object, EIV(object), and so on for all the various syntactic and semantic features. These are

basically measures of the value of learning whether or not a referent has the respective feature. However, a further measure, the total EIV, EIV_{tot}, will be calculated as a sum of all the EIVs for a particular prominence factor, *as a whole*. that is, for the role-based syntactic prominence method, the total EIV is calculated as EIV(subject) + EIV(object) + ... This then is a measure of the value of learning *which* particular syntactic or semantic feature a referent bears. This total calculation presupposes that its terms represent the entire probability space, but are mutually exclusive. This holds true for the syntactic prominence methods, but not for the semantic prominence approaches. I will discuss this in greater detail and present a transformation of the data which meets this requirement in the respective sections below.

If the results here were to precisely parallel the results of the psycholinguistic experiments in Chapter 3 then it is predicted that syntactic and semantic information should show comparable total EIVs. Furthermore, if the revised syntactic prominence hierarchy (i.e., subjects > non-subjects) is valid, then there should be a much larger reduction in entropy for learning whether or not a referent was realized as a subject or not.

4.5. Results and Discussion

In this section I present some results from the corpus analysis interleaved with discussion, saving some more general discussion for the Section 4.6. Some basic descriptive statistics for the corpus are shown in Table 4.1.

One important general result from the corpus analysis is a replication of earlier studies which show that repeated reference to a discourse-salient entity is usually done with a reduced referring expression such as a pronoun (Almor, 1999; Arnold, 1998a; Gordon et al., 1993, e.g.,). In the corpus, there are 498 utterances, though the more

Table 4.1. Basic Descriptive Statistics for Corpus

Words		5480
Sentences		408
Utterances		498
Overt Pronouns		618
	intersentential	224
	intrasentential	394
Covert Pronouns		331
Noun phrases		2072
	overt	1741
	covert	331
Verb	tokens	900
	types	654
Discourse Referents		1483
Inter-utterance links		291

relevant count here is the number of adjacent utterance pairs: 493. Within these pairs, there were 291 links between a discourse referent in the context (as determined by the first utterance) and a referring expression in the second. 224 of these links were pronouns (the remainder were of various types from definite anaphors to repeated names, but these types were not broken down). This proportion, 77%, is significantly greater than chance ($\chi^2 = 84.7, p < 0.001$).

Another statistic which I must report here is the prior entropy of the question of whether a discourse referent is pronominalized or not, H_P. This is based on the prior probabilities and the calculation is shown in (198). This forms a baseline against which the conditional entropies (for the various syntactic and semantic features) will be compared. Hence, the basic question to be examined here is whether learning some piece of information reduces (or increases) this baseline entropy.

(198) $$H_P = -[p(pronom) * log_2(pronom) + p(\neg pronom) * log_2(\neg pronom]$$
$$H_P = -[0.77 * log_2(0.77) + 0.33 * log_2(0.33)]$$
$$H_P = 0.778$$

The goal of comparing the effects of syntactic and semantic prominence is met in this section by extracting from the corpus an answer to two main questions. The first question reflects certain assumptions that the discourse salience model makes about coreference across adjacent utterances: If a discourse referent is the most salient referent in the context, then reference to that referent in the subsequent utterance should take the form of a pronoun. For shorthand here, I will refer to this situation as pronominalization. Thus, the question may be thought of as what is the influence of syntactic and semantic information on the pronominalization of reference to discourse referents already in the context. The second question is rooted in the discussion of information theory above and asks what is the value of knowing some syntactic or semantic information about an existing discourse referent to determining whether subsequent reference to that referent is pronominalized or not.

First, I'll look at these questions from the syntactic point of view and then from the semantic point of view, and then tie the two views together to see what they jointly say about discourse salience and pronominalization.

4.5.1. Syntactic Role Information

The tables on the next few pages all use the same format, so some comment here on their format will be helpful. Each table shows results in response to the two central questions noted above. For the tables, these questions might be reworded as follows: Given a set

Table 4.2. Corpus Results for Syntactic Prominence (Role-based Method)

x	$n(pron, x)$	$n(x)$	$p(pron \mid x)$	$EIV(x)$
subject	158	183	0.86	0.059
object	17	31	0.55	0.021
oblique	15	25	0.60	0.010
none	34	52	0.65	0.011
			$EIV_{tot} =$	0.101

of discourse referents with the syntactic or semantic feature x, how many of these are coreferent with some referring expression in the following utterance, $n(x)$; then, for how many of those occurrences is the subsequent reference pronominalized, $n(pron, x)$; and then finally, what is the prior probability that subsequent reference to a discourse referent with a particular syntactic or semantic feature x will be pronominalized, $p(pron \mid x)$? This last value is estimated from the proportion $n(pron, x)/n(x)$.

The results for the role-based method for determining syntactic prominence are shown in Table 4.2. Discourse referents introduced as subjects are 86% likely to have subsequent reference to them pronominalized compared to the 55%, 60%, and 65% likelihood for objects, obliques, and none ($\chi^2 = 25.6, p < 0.001$). Interestingly, estimates of the value of information of the various syntactic features here show a similar pattern in which the EIV of subjects is much higher than that of the other roles. In other words, it seems that knowing whether or not a particular discourse referent was realized as a subject or not is much more valuable (than knowing something about the other roles) for deciding whether subsequent reference to that referent is pronominalized or not. Both of these observations suggest a syntactic prominence hierarchy in which subject is more prominent than object and oblique, but in which object and oblique are (nearly)

Table 4.3. Corpus Results for Syntactic Prominence (Hierarchy-based Method)

x	$n(pron, x)$	$n(x)$	$p(pron \mid x)$	$EIV(x)$
1	121	141	0.86	0.031
2	43	61	0.70	0.004
3	17	20	0.85	0.002
4-5	9	16	0.57	0.009
none	34	53	0.64	0.014
			$EIV_{tot} =$	0.060

equally prominent. These results parallel the results of the psycholinguistic experiments in Chapter 3 where the data suggest a similar hierarchy.

Now let's consider how the results differ under the hierarchy-based method for determining syntactic prominence. Table 4.3 shows the results from this perspective. Here I show the results for the discourse referents introduced in the first (1), second (2), and third (3) most syntactically prominent positions in the respective utterance. I also show the results for the fourth and fifth positions (4-5) collapsed into one cell because the individual counts are so low. Then, finally is the none category representing discourse referents realized in non-argument positions.[7]

Here the pronominalization results are rather interesting. For discourse referents referred to in the subsequent utterance, approximately 86% of those introduced in the first most prominent position or the third most prominent position are subseqently referred to pronominally, while only 70% of those introduced in the second most prominent position are subsequently pronominalized ($\chi^2 = 17.2, p < 0.005$). The EIV values parallel those of the role-based method above suggesting that knowing that a discourse referent was realized in the syntactically most prominent position in an utterance is more valuable

[7]Interestingly, although there were 36 instances in the corpus of discourse referents with syntactic prominence values greater than 5 (i.e., in lower syntactic hierarchical positions), none of these participated in coreference across adjacent utterances.

than knowing it was realized in any lower position. The relatively high information value of discourse referents realized in non-argument positions (i.e., none; the syntactic prominence of referents is determined only for those realized in argument positions), may seem somewhat confusing at first, but is a reflection of the fact that the value of information figures are a measure of knowing whether *or not* a particular condition is true. In the present case, knowing whether a particular discourse referent was an argument (¬none) or was not an argument (none) is rather informative because if it was an argument, it was much more likely to become pronominalized. A similar pattern is observed for none items in the role-based method above. The patterns confirm what has long been thought: entities introduced in certain syntactic positions (i.e., adjunct, modifier) are much less prominent than those introduced in argument positions (see, e.g., McKoon et al., 1993).

The pronominalization results are somewhat mysterious and at present I do not have a very principled explanation for them. It is interesting, though, that Brown (1983) observed a similar pattern in a corpus investigation using a sample of Ian Fleming's 007 spy novel, "Dr. No". She measured the "persistence" of referents introduced in various syntactic positions, where persistence means the number of contiguous subsequent clauses (what I here call utterances) in which a particular referent is referred to again. She observed that indirect objects had an average persistence of 1.0 while direct objects had a persistence of 0.23 (i.e., indirect objects persisted longer). She also takes it as a puzzling result. In my study, perhaps an explanation can be found in the method used here to determine syntactic prominence. Looking closely at the 17 pronominalized cases of referents with a syntactic prominence value of 3, 8 of these were subjects appearing in such linguistic contexts as *John told Mary that Bill would go home.* Under

the role-based method, BILL has a syntactic prominence value of subject but under the hierarchical method has a value of 3. Of the remaining cases, 4 were objects and 5 were obliques. The results from the role-based method above show that subjects are pronominalized at a higher rate than the other roles, so this fact could have exaggerated the results here.

One final task in this section is to compare the two methods of determining syntactic prominence. The crucial point of comparison is in the information value of the two different methods. As shown in the tables above, the total EIV (i.e., sum total of right-hand column) for the role-based method is 0.101 while that for the hierarchical method is 0.060. Clearly, the role-based method provides considerably more information than the hierarchical method. Perhaps this is not surprising since the hierarchical method is more of a mathematical method based on a non-linguistic tree-search algorithm. On the other hand, the role-based method is more linguistic in nature, ranking elements in a manner directly related to certain relationships existing between predicates and arguments. Thus, there seems to be a closer connection between grammatical roles and pronominalization than between structural positions and pronominalization.

4.5.2. Semantic Role Information

Here I look at the contribution of semantic information to discourse salience from two perspectives: using the semantic role information provided by the FrameNet system and the semantic entailment information given by the Dowtian system of PROTO-roles.

4.5.2.1. FrameNet Semantic Roles. Table 4.4 shows the rate of occurrence of the different semantic features (here, frame elements) and subsequent pronominal reference.

Table 4.4. Corpus Results for Semantic Prominence (Frame Semantic Approach)

x	$n(pron, x)$	$n(x)$	$P(pron \mid x)$	$EIV(x)$
agent	38	43	0.88^{*}	0.010
cognizer	37	39	0.95^{**}	0.026
theme	24	36	0.67	0.006
experiencer	33	35	0.94^{**}	0.021
described	13	16	0.81	0.000
perceiver_agentive	14	15	0.93	0.008
phenomenon	11	13	0.85	0.001
perceiver_passive	10	11	0.91	0.004
goal	8	9	0.89	0.002

$\chi^2 significance: \quad {}^{*}p < 0.05 \quad {}^{**}p < 0.005$

In the corpus, 158 different frame elements occur. However, for many of these, the rate of occurrence is so low that little can be determined from them individually. I've therefore selected an arbitrary corpus frequency cut-off of $n(x) = 9$.

Before discussing this data, it is important to note two things. First, unlike the syntactic prominence data presented above in which only the most prominent occurrence was used to determine a referent's prominence value, the data here are initially presented where every role in which a discourse referent is realized is counted. Thus, in Table 4.4 there is overlap among the various semantic role counts. For example, for an utterance sequence like *John$_i$ wants PRO$_i$ to make something. He$_i$...* in which JOHN is both an experiencer (of *want*) and an agent (of *make*), the single discourse referent, JOHN, would be counted as having been pronominalized in both the agent and experiencer rows. This overlap means that a straightforward χ^2 analysis of the frame semantic approach across all the roles is not appropriate. Therefore, in the table above, I report the results of a χ^2 test for *each* semantic role using two-way tables pitting the respective role against pronominalization. That is, for the agent role, the table would be constructed as (agent,

not agent) vs. (pronominalized, not pronominalized). The EIV values also have the same overlap limitation, thus the EIVs for each semantic feature are reported, but a total EIV cannot be calculated as a simple summation of the individual EIVs. I will attempt to address these limitations below, but for the moment, I will discuss these results.

A quick glance at the data shows that the semantic information provided by FrameNet is a marginally better predictor than syntactic role information as to which discourse referents are more likely to be pronominalized. The posterior probabilities are mostly around 0.90 (e.g., 0.88 for agents, 0.95 for cognizers) as compared to 0.86 for subjects. This lends some further support (in addition to the psycholinguistic data) to my central hypothesis that semantic prominence may be a more accurate determiner of discourse salience than syntactic prominence. It is also interesting to note that the frame elements mainly responsible for this result all involve sentient participants. This result is underscored by the considerably higher EIV values for cognizer and experiencer. These results begin to suggest that perhaps the FrameNet element system merely reduces to the PROTO-role system, at least as a factor in determining semantic prominence.

A further interesting result arises when comparing the pronominalization results for themes and goals. While the results of the psycholinguistic experiments in Chapter 3 support a semantic hierarchy in which theme > goal, the corpus results here suggest an opposite ranking: goal arguments are more likely to be pronominalized (89%) than theme arguments (67%). However, I cannot express much confidence in this result because the results for goals depend on just 9 cases and therefore may be due to random variation. A larger corpus may be necessary to see if this trend is real.

One of the frames I had to add was the DESCRIPTION frame involving described and description elements. Many of the constructions that ended up in this group were sentences of the type *John is an angry man* in which there is really only one discourse referent. As such, one would expect that a subsequent sentence, in establishing coherence with the context, would be likely to regard this referent as salient. Yet, the data show that discourse referents with the described semantic feature are pronominalized only 81% (compared to the 89-95% range for more prominent roles) of the time. Thus, it seems that, the described role is rather low down on the semantic prominence hierarchy. Perhaps a different view of this is that these constructions are not used as highlighting constructions: They may add information to the discourse representation, but are not intended to establish coherence across utterances.

In order to evaluate the value of information from semantic prominence using the frame semantic approach, it is necessary to find some meaningful way to remove the overlap limitation discussed above and determine some single, discrete value which expresses the semantic prominence of a discourse referent. The approach I use here to do this is similar to that used with the syntactic prominence methods above: The semantic prominence value of a referent was determined by the one of its roles which is highest on a hierarchy of semantic roles. The difficulty here, of course, is determining this hierarchy. There is no existing ranking of the roles used in the FrameNet system. Furthermore, there is a very large number of roles. I opted, therefore, to arrange the roles into ranked groups by matching the FrameNet roles as best as possible to roles in existing thematic hierarchies proposed in the literature (e.g., Jackendoff, 1972, 1990; Speas, 1990). The hierarchy is as shown in (199).

Table 4.5. Corpus Results for Semantic Prominence (Revised Frame Semantic Approach)

x	$n(pron, x)$	$n(x)$	$P(pron \mid x)$	$EIV(x)$
1	67	88	0.86	0.013
2	70	75	0.93	0.045
3	18	30	0.60	0.012
4	22	31	0.71	0.002
5	7	12	0.58	0.005
6	8	12	0.62	0.004
7	32	52	0.62	0.019
			$EIV_{tot} =$	0.101

(199) **1:** agent, cause, content, deformer, donor, driver, employer, figure, focal_participant, grantor, ingestor, phenomenon, protagonist, self_mover

2: cognizer, communicator, experiencer, owner, perceiver_agentive, perceiver_passive, resident, responsible_party, speaker

3: emission, escapee, impactor, message, possession, projectile, stimulus, theme

4: action, addressee, affected, created_entity, effect, employee, evaluee, goal, impactee, item, object, patient, recipient, sleeper, undergoer, victim

5: direction, ground, intermediary, location, medium, path, sound_source, source

6: activity, entity, entity_1, event, eventuality, referent_1, referent_2, side_1

7: described, description, none, specifier

Using this hierarchy, the pronominalization results and EIVs were retabulated and are shown in Table 4.5.

Several observations can be made from this revised measure of semantic prominence. First there is significant relationship between this measure and pronominalization ($\chi^2 =$

$31.4, p < 0.001$). Looking at the role groups closely reveals that group 2 stands out in terms of its information value. The roles in this group are largely the sentient roles (e.g., cognizer, experiencer). This is somewhat surprising since the various thematic hierarchies proposed in the literature might lead one to predict that group 1 roles—consisting of the more agentive roles (e.g., agent, driver, protagonist)—should be greater determiners of discourse salience and thus pronominalization of subsequent reference. The fact that in this data set it is the sentient rather than agentive roles which better predict pronominalization provides further evidence that perhaps the FrameNet roles can be more efficiently expressed in terms of the Dowtian PROTO-role entailments with the PROTO-agent entailment of sentience as a central feature.

The second observation to be made is that the total EIV is the same as that of the syntactic prominence role-based method ($EIV_{tot} = 0.101$). This is a very interesting result suggesting that syntactic prominence and semantic prominence are equally informative for pronominalization. This parallels the results of the psycholinguistic experiments.

4.5.2.2. Dowtian PROTO-role Entailments. Using the same format as above, the data for the eight Dowtian PROTO-role entailments is presented in Table 4.6.

Before discussing the results, a couple of notes must be made. As with the frame semantic approach results above, the PROTO-role results have the same kind of overlap limitation: Each discourse referent may have more than one of the PROTO-role entailments. Therefore, the χ^2 results presented here are calculated for each entailment and are based on a two-way table pitting each respective entailment against pronominalization; thus, for sentience, [sentient, not sentient] vs. [pronominalized, not pronominalized]. The EIVs also have the same overlap limitation, so a total EIV is not shown here.

Table 4.6. Corpus Results for Semantic Prominence (PROTO-role Approach)

x	$n(pron, x)$	$n(x)$	$P(pron \mid x)$	$EIV(x)$
sentience	126	138	0.91^{**}	0.081
volition	92	103	0.89^{**}	0.037
cause event/c-o-s	22	28	0.79	0.000
movement	36	48	0.75	0.000
undergo c-o-s	29	37	0.78	0.000
causally-affected	18	22	0.82	0.001
incremental theme	21	28	0.75	0.000
stationary	23	30	0.77	0.000

$\chi^2 significance: \quad {}^{**}p < 0.005$

The results show that PROTO-agent entailments, on the whole, are better predictors of which discourse referents are likely to be pronominalized than PROTO-patient entailments. One might suggest that this is a rather meaningless result since, according to Dowty's theory, the argument with the most PROTO-agent entailments is selected into subject position and the results for syntactic prominence above already show that subjects are more likely to be pronominalized than objects or obliques. However, the pronominalization rate for subjects is only 86% while the rate for sentience-entailed arguments is 91%, a slight improvement. Indeed, while the overall pattern of results parallels the syntactic prominence results, the more granular approach used with the PROTO-role entailments turns out to be a better predictor of pronominalization.

It is interesting, though, that these results are largely driven by just two PROTO-agent entailments: sentience and volition. The sentience entailment even has the hightest EIV value of any feature in this study, syntactic or semantic. Apparently, discourse referents which bear the sentience (and possibly also volition) entailments are more salient in the discourse and thus influence pronominalization.

The data are also in contrast to some other data on default notions of focus and pronominal reference. Stevenson et al. (1994, 2000) and Stevenson (1999) argue for a model of discourse representation in which events can be seen as involving a sequence of states, and that entities associated with the end-state of an event are, by default, more focused than other entities. Hence, these entities are more accessible for subsequent pronominal reference. If this is so, then one would predict that in a corpus of naturally occurring descriptions of events, there would be a greater tendency toward pronominalization of discourse referents which are affected entities: In the Dowtian system, this would be entities for which undergo c-o-s and possibly causally-affected are entailed. However, the data in Table 4.6 above show that these discourse referents engender only a 78% and 82% rate, respectively, of pronominalization—in contrast to the 91%-89% rate for discourse referents with sentience and volition PROTO-agent entailments.

As with the frame semantic approach results above, here I present an attempt to transform the PROTO-role data into a single, discrete measure of semantic prominence. I used a fairly simple formula to calculate this value: the semantic prominence of a discourse referent was calculated as the total number of PROTO-agent entailments it bears minus the total number of PROTO-patient entailments it bears. Thus, the values range from +4 to −4. The Dowtian PROTO-role system can be seen as promoting some arguments with respect to PROTO-agent entailments while demoting other arguments with respect to PROTO-patient entailments. Thus, the syntactic prominence value results in a sort of ranking where discourse referents higher on the scale are more PROTO-agent-like and those lower on the scale are more PROTO-patient-like. Using this transformation,

Table 4.7. Corpus Results for Semantic Prominence (Revised PROTO-role Approach)

x	$n(pron, x)$	$n(x)$	$P(pron \mid x)$	$EIV(x)$
+4	2	3	0.67	0.000
+3	18	22	0.82	0.001
+2	58	61	0.95	0.045
+1	51	61	0.84	0.005
0	60	97	0.62	0.045
-1	27	35	0.77	0.000
-2	7	10	0.70	0.001
-3	1	1	1.00	0.002
-4	0	0	***	***
			$EIV_{tot} =$	0.098

the revised results for the PROTO-role approach are presented in Table 4.7. Note that there were no -4 cases in this corpus, so they are excluded from the calculations.

The results here seem again to parallel those of the frame semantic approach above: The relationship between this revised approach and pronominalization is significant ($\chi^2 = 26.9, p < 0.001$) and the EIV for semantic prominence under this revised PROTO-role approach is comparable to that of the revised FrameNet approach ($EIV_{tot} = 0.101$) as well as the role-based syntactic prominence method ($EIV_{tot} = 0.101$).

There are some interesting contrasts between the FrameNet approach and the PROTO-role approach which were observed during the mark-up process. In particular, one observation is that some of the verbs which FrameNet groups together into the same frame actually appear to have different PROTO-role entailments. For instance, consider the verbs *hurt* and *beat up* as in (200)-(201).

(200) a. John hurt Matt (with a stick).

b. A stick hurt Matt.

(201) a. John beat up Matt (with a stick).

b. #A stick beat up Matt.

Both of these verbs are assigned to the CAUSE_HARM frame in FrameNet. In (200a), John is assigned the agent role. However, in (200b), the non-agentive entity STICK takes on a role called cause. In (201a), JOHN similarly takes on the agent role. Thus within FrameNet, STICK may similarly take the cause role in (201b). But as the judgment shows, this interpretation is ruled it. FrameNet cannot account for this variation, except perhaps by stipulating this in the lexical entry for *beat up*. The PROTO-role system, on the other hand, captures the difference between these two fairly easily if we assume that *beat up* entails a volitional subject, but *hurt* does not. While John in (200a) may seem to act volitionally, this may be the result of inference from the fact that we assume that John is a sentient, volitional being and further from the fact that an instrument (stick) was used in the event. However, volition seems not to be entailed as can be seen by *John fell down and hurt Matt* which can be seen as an entirely non-volitional series of events. Facts such as these call into question how adequate the FrameNet system is at capturing some crucial generalizations while allowing for important distinctions among lexical items.

One final note about the revised PROTO-role results is as follows: Although the results are comparable to those of the other methods/approaches, I still treat them with some hesitation. This transformation technique fails to capture some possible correlations among the entailments. I have already noted that sentience and volition are not independent, yet this dependency is glossed over by the transformation. Future work

includes the develoment of an algorithm that would give a single semantic prominence value, but still somehow reflect these dependencies.

4.5.3. Syntactic and Semantic Prominence

The central question of my dissertation is whether or not semantic prominence *explains* syntactic prominence, or if they are independent, how they work together to determine discourse salience. The results presented above are somewhat ambivalent on this question: The prior probabilities ($P(pron \mid x)$ in the tables above) suggest that semantic prominence does a marginally better job than syntactic prominence of determining which referents are salient and thus more likely to be pronominalized. On the other hand, the EIV results suggest that syntactic role information and semantic role information are comparably valuable in determining subsequent pronominalization. While such comparisons are interesting and give some indication of the relative contribution of syntactic and semantic prominence, they do not indicate how syntactic prominence and semantic prominence may work *together* to determine discourse salience. In this section, I briefly consider this problem and discuss how to evaluate the interaction between syntactic and semantic prominence.

One straightforward procedure to test the interaction of syntactic prominence and semantic prominence is to assume a model of discourse salience in which one syntactic feature and one semantic feature on each discourse referent determine that referent's salience for subsequent pronominalization. For instance, the model may take the role-based syntactic prominence method and the revised frame semantic semantic prominence

approach and look at the EIV of referents realized with [subject, group 1 role] syntactic and semantic features, of referents realized with [subject, group 2 role] features, of referents realized with [subject, group 3 role] features, and so on through all the various combinations. For the role-based method and the frame semantic (revised) approach, that would mean $4 \times 7 = 28$ cells; for the role-based method and the (revised) PROTO-role approach, that would mean $4 \times 9 = 36$ cells; for the hierarchical method and the frame semantic (revised) approach, that would mean $5 \times 7 = 35$ cells; and for the hierachical method and the (revised) PROTO-role approach, that would mean $5 \times 9 = 45$ cells. For each of these pairings of syntactic and semantic information, a total EIV may be calculated which can then be compared to the EIVs reported above for syntactic and semantic prominence independently. If it could be shown that the EIV of syntactic and semantic prominence together is higher than that of either syntactic prominence or semantic prominence alone, that would constitute good evidence that syntactic prominence *and* semantic prominence contribute together in some fashion to the salience of discourse referents.

Results above already show that the role-based method is superior to the hierarchical method as a predictor of subsequent pronominal reference, so in the analyses which follow, I will only deal with these two out of the four pairings outlined above: the role-based method with the frame semantic approach and then the role-based method with the PROTO-role approach. For all three of these methods/approaches, the informativity was approximately $EIV = 0.100$. Therefore, in the current analysis, I will use this value as a baseline for comparison and refer to it as EIV_b.

Using the same basic format as in the preceding tables, the results for the role-based method and revised frame semantic approach are shown in Table 4.8. The primary and most interesting result is that the information value of syntactic and semantic prominence, $EIV_{tot} = 0.165$, exceeds the baseline, EIV_b. This is consistent with the results of the psycholinguistic experiments in Chapter 3 suggesting that *both* syntactic prominence and semantic prominence contribute to the salience of discourse referents. In short, the value of knowing some syntactic and some semantic information is higher for predicting subsequent pronominalization than knowing only one of these two pieces of information.

The results for the role-based method and revised PROTO-role approach are similar as shown in Table 4.9. The value of knowing the syntactic and semantic prominence information in this configuration is $EIV_{tot} = 0.141$, lower than that of the previous pairing, but still an improvement over the baseline, EIV_b.

4.5.3.1. Summary. The corpus results presented here strongly suggest that syntactic prominence and semantic prominence contribute comparably to the salience of discourse referents for subsequent pronominalization. Furthermore, the results also give quite clear evidence that syntactic and semantic prominence together provide greater information about subsequent pronominalization than either does alone. Towards this end, the role-based method of determining syntactic prominence and the frame semantic approach to determining semantic prominence appear to be the most informative match-up, yielding an overall value of $EIV_{tot} = 0.165$. One limitation on these results stems from the fact that the corpus is still relatively small and the analysis here is based on just 291 cases of interutterance coreference. In Chapter 5, I discuss some proposals for how a larger

Table 4.8. Corpus Results for Role-based Method and (Revised) Frame Semantic Approach

syntactic role (x)	semantic role group (y)	$n(pron, xy)$	$n(xy)$	$P(pron \mid xy)$	$EIV(xy)$
subject	1	56	65	0.86	0.011
subject	2	65	69	0.94	0.046
subject	3	11	19	0.58	0.009
subject	4	11	12	0.92	0.005
subject	5	2	2	1.00	0.003
subject	6	6	7	0.86	0.001
subject	7	7	9	0.78	0.000
object	1	1	2	0.50	0.002
object	2	0	1	0.00	0.007
object	3	5	7	0.71	0.000
object	4	5	9	0.56	0.005
object	5	2	3	0.67	0.000
object	6	2	5	0.40	0.008
object	7	2	4	0.50	0.003
oblique	1	3	3	1.00	0.004
oblique	2	2	2	1.00	0.003
oblique	3	0	0	***	***
oblique	4	6	9	0.67	0.001
oblique	5	2	6	0.33	0.013
oblique	6	0	0	***	***
oblique	7	2	5	0.40	0.008
none	1	7	8	0.88	0.001
none	2	3	3	1.00	0.004
none	3	2	4	0.50	0.003
none	4	0	1	0.00	0.007
none	5	1	1	1.00	0.001
none	6	0	1	0.00	0.007
none	7	21	34	0.62	0.011
				$EIV_{tot} =$	0.165

corpus might be designed to verify the current conclusions, while utilizing some design shortcuts in order to avoid some of the time-consuming labor of preparing such a corpus.

Table 4.9. Corpus Results for Role-based Method and (Revised) PROTO-role Approach

syntactic role (x)	semantic role group (y)	$n(pron, xy)$	$n(xy)$	$P(pron \mid xy)$	$EIV(xy)$
subject	+4	2	3	0.67	0.000
subject	+3	16	20	0.80	0.000
subject	+2	53	56	0.95	0.038
subject	+1	43	51	0.84	0.005
subject	0	26	30	0.87	0.005
subject	-1	14	18	0.78	0.000
subject	-2	3	4	0.75	0.000
subject	-3	1	1	1.00	0.001
subject	-4	0	0	***	***
object	+4	0	0	***	***
object	+3	0	0	***	***
object	+2	0	0	***	***
object	+1	3	4	0.75	0.000
object	0	9	20	0.45	0.026
object	-1	2	2	1.00	0.003
object	-2	3	4	0.75	0.000
object	-3	0	1	0.00	0.007
object	-4	0	0	***	***
oblique	+4	0	0	***	***
oblique	+3	0	0	***	***
oblique	+2	1	1	1.00	0.001
oblique	+1	1	2	0.50	0.002
oblique	0	5	10	0.50	0.009
oblique	-1	7	10	0.70	0.001
oblique	-2	1	2	0.50	0.002
oblique	-3	0	0	***	***
oblique	-4	0	0	***	***
none	+4	0	0	***	***
none	+3	2	2	1.00	0.003
none	+2	4	4	1.00	0.005
none	+1	4	4	1.00	0.005
none	0	20	37	0.54	0.027
none	-1	4	5	0.80	0.000
none	-2	0	0	***	***
none	-3	0	0	***	***
none	-4	0	0	***	***
				$EIV_{tot} =$	0.141

4.6. General Discussion

I begin the discussion section by reviewing some of the major findings of the corpus analysis. The results suggest some syntactic and semantic hierarchical rankings as shown in (202) and (203), respectively. More hierarchical relationships might be proposed, but because there are so few specific instances, such conclusions are not reliable.

(202) Syntactic Hierarchies

a. subject > { object, oblique }

b. {1, 3} > 2

(203) Semantic Hierarchies

a. sentience > non-sentience

b. {cognizer, experiencer, perceiver_agentive, perceiver_passive} > other roles

The syntactic hierarchies replicate the oft-observed phenomenon that discourse referents introduced as subjects are more salient than those introduced in other positions (cf., Hudson-D'Zmura and Tanenhaus, 1997; Mathews and Chodorow, 1988). However, there is little support for an often assumed ranking relationship between the lower members of the hierarchy: object and oblique. The psycholinguistic experiments in Chapter 3, showed a comparable result between the *spray/load* and *tough* construction stimuli: Those results were most easily explained by a hierarchy of syntactic roles in which the prominence difference between subjects and objects was greater than that between objects and obliques. The results from the corpus analysis here suggest the same conclusion: knowing that a discourse referent was a subject was more informative than learning that it was an object or oblique.

The suggested semantic hierarchies replicate thematic hierarchies that have been proposed in the literature (cf., Larson, 1988; Speas, 1990) with the interesting exception that agent-like roles in the FrameNet approach seem not to be as informative as sentience-entailed roles. The suggested prominence of the PROTO-agent entailment of sentience can be seen as replicating previous results showing the salience of animate entities (cf., Prat-Sala and Branigan, 1999).

The evidence further suggests that syntactic prominence and semantic prominence make roughly comparable contributions to the salience of discourse referents and that together they are even more reliable predictors of salience than either factor is alone. Thus far, the results parallel those of the psycholinguistic experiments in Chapter 3, but a larger corpus will be necessary in the future to more accurately evaluate the interaction between syntactic and semantic prominence. However, the fact that parallel results from both the narrow, controlled linguistic environments used in the psycholinguistic experiments and from the broader, much-less controlled environments in the corpus suggests that the effects syntactic prominence and semantic prominence as well as the interaction between them is relatively robust.

As for practical concerns, the corpus results suggest that the role-based method is a better predictor of pronominalization patterns in the corpus than the hierarchy-based method of determining syntactic prominence. However, little difference was observed between the frame semantic approach and the PROTO-role approach by themselves to determining semantic prominence. Both semantic approaches were essentially consistent with each other in predicting patterns of pronominalization of discourse referents. But in conjunction with the role-based syntactic prominence method, the frame semantic

approach resulted in the highest informativity. On the other hand, the mark-up process did reveal some potential problems with the FrameNet system including the incomplete implementation of the hierarchical inheritance relationships among frames and frame elements; the presence of inconsistencies in the usage of lexical items within frames; and the fact that the most prominent roles are all roles entailing sentience suggesting that perhaps the FrameNet roles simply reduce to the Dowtian PROTO-role entailments.

However, the corpus itself is not without its limitations. This is intended as a pilot project designed to reveal how a larger, in-depth corpus-based study of syntactic and semantic prominence might be organized. As such, the present analysis is somewhat limited in its scope and several simplifying assumptions have been made such as the independence of the various syntactic and semantic features (that is, considering the features in isolation, not with respect to the features on other items in the same context). Anther crucial weaknesses is that I have not taken into account the nature of competing antecedents for pronominal reference and how their relative salience might affect a target antecedent, and along the same lines, I have not analyzed possible candidates in the same sentence as the pronoun (i.e., effects of the possibility of intra-sentential coreference). Finally, the mark-up was carried out wholly by myself: thus there is no inter-rater agreement. However, intra-rater agreement was promoted by frequent reference to earlier mark-up as discused in Section 4.3.2, above. Nonetheless, future corpus investigation of the effects of syntactic and semantic prominence must take these weaknesses into account.

4.7. Conclusion

The corpus analysis discussed in this chapter has shown that even in a relatively small-scale test, the influence of semantic prominence can be observed, and results suggest that it is comparable to syntactic prominence as a determiner of the salience of discourse referents. Results parallel those of the psycholinguistic experiments in showing that syntactic and semantic prominence interact with each other to provide more information than either does alone. The analysis makes some suggestions about how a more sophisticated corpus investigation might proceed. In the next chapter, I will discuss these possibilities as part of plan for continuing the investigation of the relative effects of syntactic and semantic prominence.

CHAPTER 5

Further Work

The final chapter of this dissertation is devoted to tying off some loose strands of the main argument as well as suggesting how some other loose strands may motivate further investigation and confirmation of the conclusions I have made. In addition, I will discuss some applications of the model of discourse salience I have presented for psycholinguistic models and computational implementations of discourse processing.

5.1. A Weighted Model of Discourse Salience

One of the recurring themes throughout this dissertation has been the observation that the ranking relationships among the discourse referents within the store appears to be gradient, with some referents much more prominent than others, and some that are nearly equally ranked. This gradience showed up in both the psycholinguistic experiments and the corpus analysis. In this section, I would like to propose one way in which the model of discourse salience outlined in Chapter 2 might be extended to account for this gradience. I will illustrate this by taking a selected part of the corpus results and applying it to the *tough*/non-*tough*-alternation from the psycholinguistic experiments.

During each new utterance, the context is updated as new discourse referents are added to the store. The prominence ranking relationships are determined among the items in the store with respect to the syntactic and semantic information in the current utterance (and also recency, as discussed in Section 4.3.3). These relationships are also

encoded in the store. In the graphical illustrations I have been using throughout this dissertation, these relationships have been shown by orderings stretching from more prominent to less prominent referents and operations over the store (i.e., filtering and salience) have proceeded with respect to these relationships.

The weighted model I propose here takes a slightly different approach and is essentially the same as that employed in Kennedy and Boguraev (1996) and Lappin and Leass (1994). I assume that the store contains a set of discourse referents, each of which has attached to it a list of prominence indices. The list is ordered with respect to type of prominence (e.g., syntactic prominence, semantic prominence, recency, etc.) and each index represents the ranking of that referent with respect to the respective prominence factor. The filtering operation remains the same, merely removing incompatible referents during each pronoun resolution process. However, the salience operation must be revised somewhat. I assume the task of the salience operation is to find the maximal referent (if any) by finding the referent with the greatest salience index, computed from the prominence indices. The exact algorithm is left to future study, but for illustrative purposes I will assume it is a simple summation of the prominence indices. Other possibilities include a weighted summation or possibly even some algorithm involving a cross-product of prominence factors. But this is purely speculative. The salience operation will return the most salient referent if the algorithm can determine one, or the empty set, $\emptyset$. This salience operation might even take advantage of some sort of stochastic mechanism to determine whether a particular candidate is likely more salient than another. In other words, two referents that are very close in their total salience, though not equal, might still be regarded as indistinguishable with respect to their salience.

Table 5.1. Syntactic and Semantic Information for Antecedents of Pronouns in Corpus

syntactic prominence	matrix subject	0.54
	not matrix subject	0.31
	none	0.15
semantic prominence	group 1	0.30
	group 2	0.31
	group 3	0.08
	group 4	0.10
	group 5	0.03
	group 6	0.04
	group 7	0.14

I will illustrate this model with a relatively simply model of discourse salience in which only two prominence factors contribute: syntactic prominence and semantic prominence. The table in Table 5.1 shows some corpus results answering the question, "Given a pronoun, what is the probability that the context discourse referent with which it is coreferent has some particular syntactic or semantic feature?" The syntactic prominence features and values are adapted from the hierarchical method,[1] and the semantic prominence values are from the revised frame semantic appoach.

It is important to note here that this data represents the speaker's perspective. That is, these are the sorts of patterns that speakers show when constructing coreference between adjacent utterances. In the present illustration, in which I'll be showing how a weighted model might be used in pronoun resolution processes, I am therefore assuming that hearers are sensitive to the speaker's perspective and therefore make judgments about pronoun resolution based on prior observations of speakers. This view seems to put the burden of speaker-hearer negotiation on the hearer. However, an alternative

[1]That is, refernts with a syntactic prominence hierarchical value of 1 which were realized as subjects were regarded as matrix subjects and all other arguments were regarded as not matrix subjects.

view might be to regard the hearer as providing the standard which the speaker strives to meet and that the observations above reflect speakers' attempts to structure coreference in ways that will be most economical for the hearer to process. See van Rooy (2003) for extended discussion of this issue.

Now, returning to the weighted model illustration, let's consider two fairly simple discourses as shown in (204) and (205). The syntactic and semantic information for the respective discourse referents in each context sentence (the (a) sentences) is also shown.

(204) a. John could easily push Matt away.

JOHN: matrix subject, agent MATT: not matrix subject, theme

b. He ...

(205) a. Matt$_i$ was easy for John to push $\emptyset_i$away.

JOHN: not matrix subject, agent MATT: matrix subject, theme

b. He ...

The illustration for (204) and (205) are shown in Figure 5.1 beginning with the filtered store (which is the same as the unfiltered store). Below each discourse referent is a list of prominence indices. For this example, the indices are simply the proportions shown in Table 5.1 as a percentage. Note that in these illustrations, the directed arcs are actually redundant, being derivable from the indices which are attached to each referent. The salience operation computes the net salience, *Sal*, of the discourse referents (by summing the prominence indices) and compares them to find the maximal referent. In the non-*tough* case, the most salient referent is JOHN ($Sal_{\text{JOHN}} = 84 > Sal_{\text{MATT}} = 39$), while in the *tough* case, there is no maximal referent ($Sal_{\text{JOHN}} = 61 > Sal_{\text{MATT}} = 62$). Thus, predictions turn out to be exactly as the psycholinguistic experiments predict: a

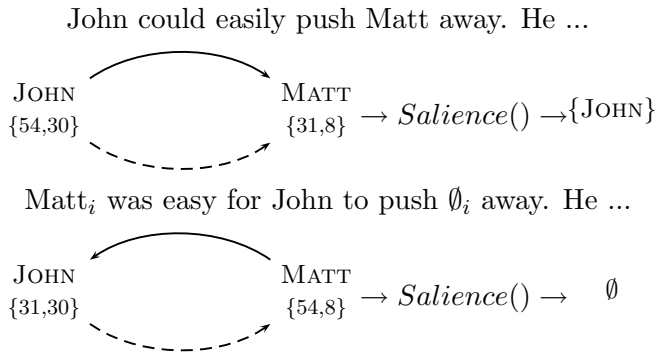


Figure 5.1. Illustration of a weighted illustration of the store for *tough*/non-*tough* items. Numerical lists below each discourse referent are prominence indices: syntactic prominence and semantic prominence, respectively. syntactic prominence relations are shown with solid lines and semantic prominence relations are shown with dashed lines.

preference for the subject-agent in the CONTROL condition, and no greater preference for either referent in the SPLIT condition.

This is still a very crude adaptation of the model and much more research is necessary to determine what sort of algorithm the salience operation should use to determine the maximal referent, as well as how to calculate and represent the prominence indices for each discourse referent. Much greater sophisticated computational procedures are certainly possible, but it may turn out that simple arithmetic operations are sufficient as in Kennedy and Boguraev (1996) and Lappin and Leass (1994).

5.2. FrameNet versus PROTO-roles

One of the stated goals of the research reported in this dissertation has been to evaluate the effectiveness of two different syntax-semantics linking approaches—frame semantics (Fillmore, 1968, 1976) as represented by the FrameNet system (Baker et al., 1998) and the PROTO-role entailment system (Dowty, 1991)—toward determining the

salience of discourse entities for subsequent pronominal reference. The results are somewhat ambivalent, not showing either system to have more explanatory power for the present purposes (though in conjunction with syntactic prominence, the FrameNet system was somewhat more informative than the PROTO-role system; see Section 4.5.3). However, the results suggest that the systems do have some strengths and weaknesses which may be useful to recognize in future investigation. In this section, I will review each of the systems in some detail.

5.2.1. FrameNet

The FrameNet system is rapidly-growing and as of May, 2005 has a lexical database covering 8,900 lexical units based on the annotation of over 135,000 sentences. A large number of ongoing projects are using FrameNet in both linguistic investigations as well as practical applications, and the theoretical foundation of FrameNet is being used to build similar databases in other languages. On-line tools allow users to browse the database via the internet. In particular a graphical tool may be used to visually explore the frame inheritance hierachy (although it is not yet possible to view frame element inheritance).

That said, the present research reveals some (current) practical inadequacies in FrameNet. The primary difficulty which I faced in using FrameNet in the present investigation was with the frame elements (i.e., semantic roles). In the version of FrameNet employed in my investigation (Version 2.0), I counted more than 500 unique element labels (e.g., agent, patient, etc.) in the database. In frame semantic theory, these roles should be derived via frame inheritance from some very small set of primitives (see Fillmore, 1968) contained in some abstract semantic frames. In practice,

however, these derivational relationships are not fully implemented. For instance, the CHANGE_OF_LEADERSHIP frame covers such lexical items as *elect* and *appoint* as in (206). One of the elements of this frame is selector which is the entity directly responsible for bringing about the change of leadership, syntactically, most often realized in subject position. Intuitively, the selector seems to be an agentive role, acting volitionally and causing a change-of-state in another entity. Thus, it would be reasonable for selector to be somehow derived from the agent role of a more abstract frame, say INTENTIONALLY_AFFECT. However, in this version of FrameNet, CHANGE_OF_LEADERSHIP is a base frame, not derived from any other.

(206) a. The prime minister appointed a new cabinet.

b. The people elected a new president.

It appears that this problem is not getting better, and may be getting worse. FrameNet now has over 3,000 frame elements and only about half are tied into the inheritance hierarchy (C. Baker, personal communication, March 15, 2005). Admittedly, accomplishing this is no small task, and some difficult decisions must be made. For instance, the SIMILARITY frame (e.g., *John is the spitting image of his father*) assigns ENTITY_1 and ENTITY_2 to the two entities being compared but what these two roles derive from is a tough question.

In the present research, the large number of frame element labels made for a somewhat unwieldy dataset and a large number of labels for which there were only one or two instances in the entire corpus. In order to complete some parts of the corpus analysis, I had to construct my own inheritance hierarchy of sorts, organizing labels into discrete groups.

These shortcomings notwithstanding, the FrameNet system was somewhat more informative than the PROTO-role system in the corpus analysis when taken in conjunction with the role-based syntactic prominence information. However, the results were somewhat unusual suggesting that agent-like roles (e.g., agent, driver) were not as informative as sentient roles (e.g., experiencer, cognizer). Paralleling that information was also the fact that given a coreference link, the probability that the antecedent discourse referent was a sentient role was slightly higher than the probability that it was an agent-like role. This is an interesting result because it contrasts with many proposed thematic hierarchies which place agent above experiencer. As such, it may be that thematic hierarchies and prominence hierarchies cannot be assumed to be the same. An interesting result which deserves further exploration.

5.2.2. PROTO-role entailments

In the present research, the Dowtian PROTO-role entailment system of predicate-argument linking performed nearly as well as the FrameNet system, showing a relatively high informativity in the corpus analysis for the pronominalization of referring expressions. In particular, though, this informativity was highest for the sentience entailment. In fact, the informativity of sentience alone, was higher than any other single syntactic or semantic feature. This parallels the results of the FrameNet informativity results, suggesting that perhaps the FrameNet system of elements merely reduces to the PROTO-role entailments. This is a point I'd like to emphasize. It would be very interesting if once the FrameNet inheritance hierachy is fully implemented, the set of primitives from which the

elements are derived is essentially the same as the Dowtian PROTO-role entailments. So far, the evidence points in this direction.

In the literature, I have been able to find some researchers who make use of the PROTO-role system in order to explain certain linguistic observations (e.g., Filip et al., 2001; Hamilton, 1994). However, I have not yet found any evaluation of the theory in any broad linguistic context. However, the corpus described in Chapter 4 now provides a basis for such a test across a fairly wide range of linguistic environments. It is not directly relevant to the question at hand and it is beyond the scope of this chapter to explore this deeply, but here I present some preliminary results from such an investigation.

The basic question I am dealing with here is how well the argument-selection principle (see Section 2.7) correctly predicts which argument will end up in which syntactic position. Recall that the argument with the most PROTO-agent entailments is selected into subject position, the (remaining) argument with the most PROTO-patient entailments is selected into object position, and any remaining arguments are selected into oblique positions. In the corpus, there are 331 verb types which bear on this question (321 other verb types trivially verify the PROTO-role system because there is only one argument, or because none of the arguments carry any PROTO-role entailments). Of these 331 verbs, 7.8% or 26 verb tokens displayed an argument ordering which did not follow the argument selection principle. For example, this included the verb *emit* as in (207).

(207) The sun emits ultraviolet radiation.

SUN: stationary RADIATION: movement

The PROTO-agent entailment of movement applies to RADIATION but it is realized in object position. Conversely, the PROTO-patient entailment of stationary applies to SUN but it is realized in subject position.

The PROTO-role system was therefore $100\% - 7.8\% = 92.2\%$ accurate. I'm not sure what should be an acceptable standard to claim the success of the PROTO-role system, but I expect 95% would be a reasonable cut-off point (cf., $\alpha = 0.05$). As such, the PROTO-role system here is only marginally acceptable. However, these conclusions should be treated as very tentative: Currently there is no mark-up in the corpus to distinguish passive from active constructions. In addition, my analysis used a fairly simple technique which did not take into account preposed arguments (I merely extracted the <np> siblings of each <verb> node, in left-to-right order). Thus, under a more careful analysis, the success rate is almost sure to improve.

5.3. Psycholinguistic Experimentation

In this section, I would like to discuss some possibilities for examining the relative influence of syntactic and semantic prominence further in psycholinguistic experimentation. Taking at face value the results of the experiments presented in Chapter 3 suggesting that syntactic and semantic prominence contribute to discourse salience in a complementary fashion, the results are best explained by a revised syntactic prominence hierarchy in which objects and obliques are not ranked much differently from each other. If this is the case, then the hypothesis should be testable using passive constructions (initially rejected because of a possible structural confound). Thus, for the active-passive

alternates shown in (208), If syntactic and semantic prominence indeed work in complementary fashion as the results suggest, then there should be no advantage for either CARPENTER or WOOD in the passive context.

(208) a. The carpenter$_i$ cut the wood$_j$ in half. He$_i$/It$_j$ ...

b. The wood$_j$ was cut in half by the carpenter$_i$. He$_i$/It$_j$ ...

A similar experiment might also be performed using *psych*-verbs as shown in (209). In a semantic-role approach, the experiencer role is regarded as higher on the thematic hierarchy than the stimulus role (cf., Jackendoff, 1972; Larson, 1988; Speas, 1990, inter alia) and in a PROTO-role approach, sentience-entailed entities are more prominent than non-sentience-entailed entities. Thus, under either approach, it is predicted that VIOLINIST and AUDIENCE in (209b) should be regarded as comparably salient.

(209) a. The audience$_i$ admired the violinist$_j$. They$_i$/It$_j$ ...

b. The violinist$_j$ amazed the audience$_i$. They$_i$/It$_j$ ...

The results of the psycholinguistic experiments using the repeated-name penalty design suggests that the penalties really only show up clearly when a pronoun can be interpreted unambiguously. Thus, the experimental stimuli in this experiment could be slightly revised to the form shown in (210)-(211) in order to verify this hypothesis.

(210) a. John sprayed the paint$_i$ on the walls$_j$. It$_i$/They$_j$ ...

b. John sprayed the walls$_j$ with the paint$_i$. It$_i$/They$_j$ ...

(211) a. Jane$_i$ easily hit Matt$_j$. He$_i$/She$_j$ ...

b. Matt$_j$ was easy for Jane$_i$ to hit $\emptyset_j$. He$_i$/She$_j$ ...

With the discourse referents properly disambiguated, then the repeated-name and definite-description penalties should be clearer. That is, alternating the pronouns with

definite descriptions in (210) or names in (211) should result in larger penalties (and no advantages).

If the participants' behavior observed in the psycholinguistic experiments discussed above are determined by fundamental cognitive strategies in discourse processing, as opposed to merely reflexive responses to the experimental tasks, then the effects should be observable in other experimental paradigms. One such candidate method would be an oral discourse completion task. For instance in Arnold (1998a), participants listened to an incomplete story up to and including a test sentence. Then, participants spoke a continuation of the story. Her experiments successfully observed clear differences in participants' focusing choices depending on whether the context sentence was a cleft construction (*What the burglar heard was the neighbor's dog*) or a non-cleft construction (*The burglar heard the neighbor's dog*). This experimental procedure could also be readily used in the investigation of the effects of syntactic and semantic prominence by using the stimuli already available from the previous experiments. Thus, for the *tough* stimuli shown in (212), the prediction would be that whichever discourse referent in the context a participant sees as most salient will be referred to in the participant's oral continuation.

(212) a. John and Matt fought each other in a boxing match. It was twelve rounds long. John easily hit Matt in the final round of the match.

b. John and Matt fought each other in a boxing match. It was twelve rounds long. Matt was easy for John to hit in the final round of the match.

5.4. Corpus Investigation

The corpus investigation described in Chapter 4 successfully showed an effect of both syntactic and semantic prominence in the salience of discourse referents. However, the analysis is somewhat limited due to the relatively small size of the corpus. In particular, I had to assume that the various features (i.e., grammatical functions, semantic roles and entailments) are independent of each other. For instance, I assumed that for a given discourse referent realized as an agent, the contribution that semantic prominence makes to the salience of that referent is constant, regardless of whether it appeared in a context in which there was also a patient, or a goal, or a theme, and so on. While a fairly robust effect of semantic prominence was observed in this analysis, it might be the case that a more detailed corpus analysis might show that these effects are context-sensitive (or, perhaps more precisely, verb-frame sensitive). The reason for making the independence assumption is because a more detailed analysis will require a much larger corpus. Here I'd like to propose how such a corpus might be obtained, but with some practical simplifications.

Crucially, the corpus needs several pieces of information to parallel the corpus used in the present investigation. This includes a clause-level parse with (at least) part-of-speech tags for noun phrases (including pronouns) and verbs, syntactic role information, semantic role/PROTO-role entailment information, and coreference information. Fortunately, there are some tools that can make the large-scale mark-up of this information much easier.

Currently, there are a number of reasonably accurate part-of-speech taggers and syntactic parsers from which the requisite information and structure can be obtained:

the CLAWS tagger (Garside, 1987; Garside and Smith, 1997), the Penn Supertagger (Joshi and Srinivas, 1994; Srinivas, 1997), the slot-grammar parser (Lappin and McCord, 1990), and the link parser (Sleator and Temperley, 1991) among others. Syntactic role information could be obtained through a simple algorithm of assigning subject role to the last pre-verbal noun phrase, object role to the first post-verbal noun phrase, and oblique to any other noun phrase in a clause.

The information required to determine semantic prominence is perhaps one of the more difficult pieces of information to acquire. However, one possibility is to use a statistical algorithm proposed by Gildea and Jurafsky (2002) designed to determine FrameNet semantic roles for noun phrases in text. The algorithm has an accuracy rate of 82%, so some post-application hand-checking will still be necessary. However, automating the mark-up process will reduce the total time considerably.

The PROTO-role mark-up, on the other hand, is perhaps the most difficult of all. I can find no published report of any procedure for the automatic determination of PROTO-role entailments within a text. This could partly be the result of the fact that there doesn't seem to be any available database of verb lexemes with their PROTO-role entailments spelled out. However, the corpus analysis I've described in Chapter 4 has produced just such a database. While it still lacks something in breadth (there are 654 verbs with their entailments listed), holes may be filled in as necessary by applying the linguistic tests described in Chapter 2 to systematically determine the PROTO-role entailments for other verbs. As such, the database and entailment tests may serve as the foundation for the development of a system for automatic mark-up of thematic PROTO-role information in texts.

Finally, the last piece of information necessary for the proposed extended corpus investigation is coreference information. This process may also be automated to some success. Pronominal coreference can be determined by using several quite successful algorithms including those of Kennedy and Boguraev (1996), Lappin and Leass (1994), and Mitkov (1994, 1996) although each of these require varying degrees of other information (e.g., full syntactic parse, domain-specific semantic information) to do their work.

Of course, these practical shortcuts will surely introduce some noise into the results. However, they will make the process of investigating the relative effects of syntactic and semantic prominence more feasible.

5.5. Implications and Applications

The results suggest that some of the salience effects that have been attributed to syntactic prominence are actually due to semantic prominence. This has various implications for different investigative programs and practical applications. In this section I will briefly discuss some of these.

Most pronoun resolution algorithms depend on some procedure for determing the salience of entities in the current context, and most (in fact, all the algorithms I've ever seen) use some notion of syntactic prominence. While some use semantic information of some sort (e.g., animacy of entities), none use semantic role or entailment information as a factor. This study suggests that this information may provide more accurate estimates of discourse salience.

The results also suggest that the repeated-name penalty experimental method (Gordon et al., 1993), a commonly used on-line psycholinguistic experimental technique, may

be interestingly limited in its applicability. In particular, penalties are difficult to observe when a context has more than one discourse referent compatible with a subsequent pronoun.

The results have shown that the salience of discourse referents can be seen as being influenced by the semantic information (thematic roles or entailments) imposed on the arguments of a verb. In my test of this, I looked only at the information imparted by the verb, ignoring how the verb phrase might be modified. But perhaps such additional information, especially if it places constraints on the noun phrase arguments of a verb, also influences semantic prominence. For example, the verb *bite* as in *The dog bit my finger* does not assign any PROTO-role entailments to its object because it does not have to have undergone any change-of-state. However, modification of *bite* changes this: *The dog bit off my finger*. In this case, it is clear that the object has undergone a change-of-state and is also causally-affected. Similarly, adjectival modification of noun phrases can also place certain entailments on the corresponding discourse referents. For instance, adjectives describing some mental state (e.g., *happy*, *frustrated*, *angry*) can be seen as entailing sentience in the referents of the noun phrases they modify. It could be the case that discourse salience is further determined by these entailments.

5.6. Conclusion

In this chapter, I have proposed an extended version of the discourse salience model which allows for weighted prominence relationships. In the remainder of this chapter I have proposed some further work to continue to investigate the relative effects of syntactic and semantic prominence and also some implications of the present results for psycholinguistic and computational models of discourse processing.

To sum up, the main conclusion of this dissertation is that the salience of discourse referents is determined at least partly by the semantic roles or entailments imposed on the referent by the verb for which it serves as an argument. This effect was relatively robust and consistent in each of the psycholinguistic experiments and corpus analysis . Furthermore, the results were explainable in terms of either a frame semantics approach or a Dowtian PROTO-role approach to determining the semantic prominence of referents. However, between two methods used to determine syntactic prominence—a role-based method and a hierarchial method—the role-based method was found to be more accurate and consistent across the two investigative paradigms.

References

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106:748–765.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24:65–87.

Arnold, J. (1998a). Marking salience: The similarity of topic and focus. Manuscript submitted for publication.

Arnold, J. (1998b). *Reference Form and Discourse Patterns*. PhD thesis, Stanford University.

Arnold, J., Eisenband, J., Brown-Schmidt, S., and Trueswell, J. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76:B13–B26.

Badecker, W. and Straub, K. (1994). Evidence that binding principles participate in a constraint satisfaction process. In *7th CUNY Conference on Human Sentence Processing*, New York, NY.

Baker, C., Fillmore, C., and Lowe, J. (1998). The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*.

Baker, M. (1997). Thematic roles and syntactic structure. In Haegeman, L., editor, *Elements of Grammar: Handbook in Generative Syntax*, pages 73–138. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Beaver, D. (2003). The optimization of discourse anaphora. *Linguistics and Philosophy*, 27:3–56.

Blutner, R. (1998). Lexical pragmatics. *Journal of Semantics*, 15:115–162.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17:189–216.

Boas, H. (2003). Towards a lexical-constructional account of the locative alternation. In *Proceedings of the 2001 Western Conference on Linguistics*, volume 13, pages 27–42.

Bock, J. K. and Miller, C. (1991). Broken agreement. *Cognitive Psychology*, 23:45–93.

Boland, J., Acker, M., and Wagner, L. (1998). The use of gender features in the resolution of pronominal anaphora. Cognitive Science Technical Report 17, The Ohio State University Center for Cognitive Science.

Bouma, G. (2003). Doing Dutch pronouns automatically in Optimality Theory. In *Proceedings of the EACL 2003 Workshop on The Computational Treatment of Anaphora*, Budapest.

Brennan, S., Friedman, M., and Pollard, C. (1987). A centering approach to pronouns. In *Proceedings, 25th Annual Meeting of the Association for Computational Linguistics*, pages 155–162.

Brown, C. (1983). Topic continuity in written English narrative. In Givón (1983a), pages 313–342.

Büring, D. (1995). *The 59th Street Bridge Accent: On the Meaning of Topic and Focus*. PhD thesis, Universität Tübingen.

Carlson, G. and Tanenhaus, M. (1988). Thematic roles and language comprehension. In Wilkins (1988), pages 263–288.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In Li (1976), pages 25–55.

Chambers, C. and Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language*, 39:593–608.

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

Chomsky, N. (1977). On *Wh*-movement. In Culicover, P., Wasow, T., and Akmajian, A., editors, *Formal Syntax*, pages 71–132. Academic Press, New York.

Chomsky, N. (1981). *Lectures on government and binding*. Foris, Dordrecht.

Christianson, K., Hollingworth, A., Halliwell, J., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42:368–407.

Clark, H. and Haviland, S. (1977). Comprehension and the given-new contract. In Freedle, R., editor, *Discourse Production and Comprehension*, pages 1–40. Ablex, Norwood, NJ.

Clark, H. and Sengul, C. (1979). In search of referents for nouns and pronouns. *Memory and Cognition*, 7:35–41.

Cornish, F. (1996). 'antecedentless' anaphors: deixis, anaphora, or what? some evidence from English and French. *Journal of Linguistics*, 32:19–41.

Cote, S. (1997). Ranking forward-looking centers. In Walker et al. (1997), pages 55–70.

Crawley, R., Stevenson, R., and Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19(4):245–264.

Di Eugenio, B. (1997). Centering in italian. In Walker et al. (1997), pages 115–137.

Diestel, R. (2000). *Graph Theory*. Springer-Verlag, New York.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67:547–619.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Ferreira, F. and Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25:348–368.

Fiengo, R. and May, R. (1994). *Indices and Identity*. MIT Press, Cambridge, MA. Linguistic Inquiry Monograph Twenty-Four.

Filip, H., Tanenhaus, M., Carlson, G., Allopenna, P., and Blatt, J. (2001). Reduced relatives judged hard require constraint-based analyses. In Stevenson, S. and Merlo, P., editors, *Lexical Representations in Sentence Processing*. John Benjamins, Amsterdam.

Fillmore, C. (1968). The case for case. In Bach, E. and Harms, R., editors, *Universals in Linguistic Theory*, pages 1–90. Holt, Rhinehart and Winston, New York.

Fillmore, C. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, pages 20–32.

Fillmore, C. (1977). Scenes-and-frames semantics. In Zampolli, A., editor, *Fundamental Studies in Computer Science*, pages 55–81. North Holland Publishing, Amsterdam.

Firbas, J. (1964). On defining the theme in functional sentence analysis. *Travaux Linguistiques de Prague*, 1:267–280.

Firbas, J. (1966). Non-thematic subjects in contemporary English. *Travaux Linguistiques de Prague*, 2:229–236.

Garside, R. (1987). The CLAWS word-tagging system. In Garside, R., Leech, G., and Sampson, G., editors, *The Computational Analysis of English: A Corpus-based Approach*, pages 30–41. Longman, London.

Garside, R. and Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In Garside, R., Leech, G., and McEnery, A., editors, *Corpus Annotation: Linguistic Information from Computer Text Corpora*, pages 102–121. Longman, London.

Gernsbacher, M. A. and Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27:699–717.

Geurts, B. (1999). *Presupposition and Pronouns*. Elsevier, Oxford.

Gildea, D. and Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28:245–288.

Givón, T., editor (1983a). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamins, Amsterdam.

Givón, T. (1983b). Topic continuity in discourse: An introduction. In Givón (1983a), pages 1–42.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, IL.

Goldberg, A. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics*, 13:327–356.

Gordon, P. and Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language*, 34:216–231.

Gordon, P., Grosz, B., and Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17:311–347.

Gordon, P. and Hendrick, R. (1997a). Intuitive knowledge of linguistic co-reference. *Cognition*, 62:325–370.

Gordon, P. and Hendrick, R. (1997b). The representation and processing of coreference in discourse. *Cognitive Science*, 22:389–424.

Gordon, P., Hendrick, R., Ledoux, K., and Yang, C. L. (1999). Processing of reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes*, 14:353–379.

Grice, H. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics 3: Speech Acts*. Academic Press, New York.

Grimshaw, J. (1990). *Argument Structure*. MIT Press, Cambridge, MA.

Grosz, B. (1981). Focusing and description in natural language dialogues. In Joshi, A., Sag, I., and Webber, B., editors, *Proceedings of the Workshop on Computational Aspects of Linguistic Structure and Discourse*, pages 84–105. Cambridge University Press, Cambridge.

Grosz, B., Joshi, A., and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21:203–225.

Grosz, B. and Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204.

Gruber, J. (1965). *Studies in Lexical Relations*. PhD thesis, Massachusetts Institute of Technology.

Guindon, R. (1985). Anaphora resolution: Short-term memory and focusing. In *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics*, pages 218–227.

Gundel, J. (1974). *The Role of Topic and Comment in Linguistic Theory*. PhD thesis, University of Texas, Austin.

Gundel, J., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions. *Language*, 69:274–307.

Halliday, M. and Hasan, R. (1976). *Cohesion in English*. Longman, London.

Hamilton, R. (1994). Double object dative and the possession condition in l2 english relativization: A proto-role entailment analysis. In *Proceedings of the Eastern States Conference on Linguistics (ESCOL)*, pages 140–151, University of South Carolina.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst.

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In Bauerle, R., Schwarze, C., and Von Stechow, A., editors, *Meaning, Use, and Interpretation*, pages 164–189. DeGruyter, Berlin.

Hirst, G. (1981). *Anaphora in Natural Language Understanding: A Survey*. Springer-Verlag, Berlin.

Hobbs, J. (1978). Resolving pronoun references. *Lingua*, 44:311–338.

Hobbs, J. (1979). Coherence and coreference. *Cognitive Science*, 3:67–90.

Hobbs, J. and Kehler, A. (1997). A theory of parallelism and the case of VP ellipsis. In *Proceedings of the 35^{th} Annual Meeting of the Association for Computational Linguistics and 8^{th} Conference of the European Chapter for the Association of Computational Linguistics (ACL-97/EACL-97)*, Madrid, Spain.

Hornstein, N. (2001). *Move! A Minimalist Theory of Construal*. Blackwell Publishers.

Hudson-D'Zmura, S. and Tanenhaus, M. (1997). Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In Walker et al. (1997), pages 199–226.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Project Gutenberg, www.gutenberg.org.

Hymes, D. (1974). *Foundations of Sociolinguistics: An Ethnographic Approach*. University of Pennsylvania Press, Philadelphia.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, MA.

Jackendoff, R. (1987). Status of thematic relations in linguistic theory. *Linguistic Inquiry*, 18:369–411.

Jackendoff, R. (1990). *Semantic Structures*. MIT Press, Cambridge, Massachusetts.

Johnson, C., Fillmore, C., Petruck, M., Baker, C., Ellsworth, M., Ruppenhofer, J., and Wood, E. (2002). Framenet: Theory and practice. unpublished manuscript.

Joshi, A. and Srinivas, B. (1994). Disambiguation of super parts of speech (or supertags): Almost parsing. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING '94)*, pages 154–160, Kyoto, Japan.

Kameyama, M. (1985). *Zero Anaphora: The Case of Japanese*. PhD thesis, Stanford University.

Kameyama, M. (1986). A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 200–206, New York.

Kameyama, M. (1995). Indefeasible semantics and defeasible pragmatics. In Kanazawa, M., Piñón, C., and de Swart, H., editors, *Quantifiers, Deduction, and Context*, pages 111–138. CSLI Publications, Stanford.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J., Janssen, T., and Stokhof, M., editors, *Formal Methods in the Study of Language*, pages 277–322. Mathematisch Centrum, Amsterdam.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer Academic, Dordrecht.

Karttunen, L. (1976). Discourse referents. In McCawley, J., editor, *Syntax and Semantics, Vol. 7: Notes from the Linguistic Underground*, pages 363–385. Academic Press, New York.

Kehler, A. (1995). *Interpreting Cohesive Forms in the Context of Discourse Inference*. PhD thesis, Harvard University.

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford University, CA.

Kennedy, C. and Boguraev, B. (1996). Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING '96)*, pages 113–118, Copenhagen, Denmark.

Kibble, R. (2001). A reformulation of rule 2 of Centering Theory. *Computational Linguistics*, 27:579–587.

Kibble, R. and Power, R. (2004). Optimizing referential coherence in text generation. *Computational Linguistics*, 30:401–416.

Kim, B. (1996). Predication in tough-constructions. In *The Proceedings of the Fourteenth West Coast Conference on Formal Linguistics*, pages 271–286, Stanford, CA. Center Study Language & Information.

Kintsch, W. and van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85:363–394.

Krahmer, E., van Erk, S., and Verleg, A. (2003). Graph-based generation of referring expressions. *Computational Linguistics*, 29:53–72.

Kruijff-Korbayová, I. and Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*, 12:249–259.

Lambrecht, K. (1994). *Information Structure and Sentence Form : Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, UK.

Lappin, S. and Leass, H. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20:535–561.

Lappin, S. and McCord, M. (1990). A syntactic filter on pronominal anaphora for slot grammar. In *Proceedings, 28th Annual Meeting of the Association for Computational Linguistics*, pages 135–142.

Larson, R. (1988). On the double object construction. *Linguistic Inquiry*, 19:335–392.

Lasnik, H. and Fiengo, R. (1974). Complement object deletion. *Linguistic Inquiry*, 5:535–571.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago, IL.

Levin, B. and Rappaport, M. (1986). The formation of adjectival passive. *Linguistic Inquiry*, 17:623–661.

Li, C., editor (1976). *Subject and topic*. Academic Press, New York.

Li, C. and Thompson, S. (1976). Subject and topic: A new typology of language. In Li (1976), pages 457–489.

MacDonald, M. and MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29:469–492.

Mann, W. and Thompson, S. (1987). Rhetorical structure theory: A theory of text organization. Technical Report RS-87-190, Information Sciences Institute.

Marantz, A. (1984). *On the Nature of Grammatical Relations*. MIT Press, Cambridge, MA.

Mathews, A. and Chodorow, M. (1988). Pronoun resolution in two-clause sentences: Effects of ambiguity, antecedent location, and depth of embedding. *Journal of Memory and Language*, 27:245–260.

McEnery, T. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh University Press, Edinburgh.

McKoon, G., Ratcliff, R., Ward, G., and Sproat, R. (1993). Syntactic prominence effects on discourse processes. *Journal of Memory and Language*, 32:593–607.

Miller, G. (1995). Wordnet: A lexical database for english. *Communications of the ACM*, 38:39–41.

Minsky, M. (1975). A framework for representing knowledge. In Winston, P., editor, *The Psychology of Computer Vision*, pages 211–277. McGraw-Hill, New York.

Mitkov, R. (1994). An integrated model for anaphora resolution. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, pages 1170–1176, Kyoto, Japan.

Mitkov, R. (1995). Anaphora resolution in natural language processing and machine translation. Working Paper. Saarbrücken: IAI.

Mitkov, R. (1996). Anaphora resolution: a combination of linguistic and statistical approaches. In *Proceedings of the Discourse Anaphora and Anaphor Resolution (DAARC'96)*, pages 76–85, Lancaster, UK.

Mitkov, R. (1999). Anaphora resolution: The state of the art. Unpublished Manuscript.

Nasukawa, T. (1994). Robust method of pronoun resolution using full-text information. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, pages 1157–1163, Kyoto, Japan.

Nicol, J. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18:5–19.

Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press, Edinburgh.

"Odds and Ends" (2003, December 4). Reuters.

Ostler, N. (1979). *Case-Linking: A Theory of Case and Verb Diathesis Applied to Classical Sanskrit*. PhD thesis, Massachusetts Institute of Technology.

Palmer, M., Dahl, D., Schiffman, R., Hirschman, L., Linebarger, M., and Dowding, J. (1986). Recovering implicit information. In *Proceedings of the 24th Annual Conference of the Association for Computational Linguistics*, pages 10–19.

Parsons, T. (1990). *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press, Cambridge, Massachusetts.

Pinker, S. (1989). *Learnability and Cognition*. MIT Press, Cambridge, MA.

Poesio, M., Stevenson, R., Di Eugenio, B., and Hitzeman, J. (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30:309–364.

Prat-Sala, M. and Branigan, H. (1999). Discourse constraints on syntactic processing in langauge production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, 42:168–182.

Prince, E. (1981). Toward a taxonomy of given-new information. In Cole, P., editor, *Radical Pragmatics*, pages 223–254. Academic Press, New York.

Prince, E. (1986). On the syntactic marking of presupposed open propositions. In *Papers from the Parasession on Pragmatics and Grammatical Theory, 22nd Regional Meeting, Chicago Linguistic Society*, pages 208–222, Chicago, IL.

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In Mann, W. and Thompson, S., editors, *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, pages 295–325. John Benjamins, Amsterdam.

Rappaport, M., Laughren, M., and Levin, B. (1987). Levels of lexical representation. Technical report, MIT. Lexicon Project Working Papers 20.

Rappaport, M. and Levin, B. (1988). What to do with θ-roles. In Wilkins (1988), pages 7–36.

Rose, R. (2002). Discourse salience and movement constructions. In *15th CUNY Conference on Human Sentence Processing*, New York, NY.

Schank, R. and Abelson, R., editors (1977). *Scripts, Plans, Goals, and Understanding*. Laurence Erlbaum Associates, Hillsdale, NJ.

Schank, R. and Rieger, C. (1974). Inference and the computer understanding of natural language. *Artifical Intelligence*, 5:375–412.

Sgall, P. (1967). Functional sentence perspective in a generative description. *Prague Studies in Mathematical Linguistics*, 2:203–225.

Shibatani, M. (1976a). The grammar of causative constructions: A conspectus. In Shibatani (1976b), pages 1–40.

Shibatani, M., editor (1976b). *Syntax and Semantics 6: The Grammar of Causative Constructions*. Academic Press, New York.

Sidner, C. (1981). Focusing for interpretation of pronouns. *American Journal of Computational Linguistics*, 7(4):217–231.

Sleator, D. and Temperley, D. (1991). Parsing English with a link grammar. Computer Science Technical Report CMU-CS-91-196, Carnegie Mellon University.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23(3):197–229.

Smyth, R. and Chambers, C. (1996). Parallelism effects on pronoun resolution in discourse contexts. In *Proceedings of the Discourse Anaphora and Anaphor Resolution (DAARC'96)*, pages 268–280, Lancaster, UK.

Speas, M. (1990). *Phrase Structure in Natural Language*. Kluwer Academic Publishers, Dordrecht, Netherlands.

Srinivas, B. (1997). *Complexity of Lexical Descriptions and its Relevance to Partial Parsing*. PhD thesis, University of Pennsylvania.

Stevenson, R. (1999). The production of referring expressions: Who gets referred to and how? In *Proceeding of the ESSLLI Workshop on NP Generation*.

Stevenson, R., Crawley, R., and Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9:519–548.

Stevenson, R., Knott, A., Oberlander, J., and McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15(3):225–262.

Stevenson, R., Nelson, A., and Stenning, K. (1995). The role of parallelism in strategies of pronoun comprehension. *Language and Speech*, 38(4):393–418.

Strube, M. and Hahn, U. (1996). Functional centering. In *Proceedings of the 34th Conference of Association for Computational Linguistics*, pages 270–277, Santa Cruz, CA.

Talmy, L. (1976). Semantic causative types. In Shibatani (1976b), pages 43–116.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In *Language Typology and Syntactic Description, Vol. III: Grammatical Categories and the Lexicon*, pages 57–194. Cambridge University Press, Cambridge.

Taranto, G. (2002). An event-structure analysis of causative and passive *GET*. Unpublished Manuscript.

Tenny, C. (1994). *Aspectual Roles and the Syntax-Semantics Interface*. Kluwer Academic Publishers, Dordrecht, Netherlands.

Tetreault, J. (2001). A corpus-based evaluation of centering theory and pronoun resolution. *Computational Linguistics*, 27:507–520.

Turan, U. (1995). *Null vs. Overt Subjects in Turkish Discourse: A Centering Analysis*. PhD thesis, University of Pennsylvania.

Vallduví, E. (1990). *The Informational Component*. PhD thesis, University of Pennsylvania.

van Eijck, J. and Kamp, H. (1997). Representing discourse in context. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 179–237. Elsevier Science, Amsterdam.

van Rooy, R. (2003). Relevance and bidirectional OT. In Blutner, R. and Zeevat, H., editors, *Pragmatics in Optimality Theory*, pages 173–210. Palgrave Macmillan, Oxford.

Vonk, W., Hustinx, L. G., and Simons, W. H. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes*, 7:301–333.

Walker, M., Joshi, A., and Prince, E., editors (1997). *Centering Theory in Discourse*. Clarendon Press, Oxford.

Ward, G. (1985). *The Semantics and Pragmatics of Preposing*. PhD thesis, University of Pennsylvania.

Weaver, W. and Shannon, C. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois.

Wilkins, W., editor (1988). *Syntax and Semantics 21: Thematic Relations*. Academic Press, New York.

Yule, G. (1979). Pragmatically controlled anaphora. *Lingua*, 49:127–135.

APPENDIX

Stimulus Items Used in Psycholinguistic Experiments

1. Items Using *Spray/Load* Verbs

(1) a. Patrick, a hotel custodian, gathered some trash and found a dumpster outside the hotel.
b. He was cleaning up after a big party.
c. [He/Patrick] loaded the trash into the dumpster.
c'. [He/Patrick] loaded the dumpster with the trash.
d. [It/The trash] became soggy and mushy after a night's rain.
d'. [It/The dumpster] was rolled into an alley later that evening.
e. The clean-up effort continued all night long.

(2) a. While visiting the seashore, Frank hurriedly pulled a camera and some film out of his bag.
b. He wanted to get a picture of the sun disappearing over the horizon.
c. [He/Frank] loaded the film into the camera.
c'. [He/Frank] loaded the camera with the film.
d. [It/The film] became partially exposed and thus useless.
d'. [It/The camera] was somewhat hard for a novice to operate.
e. Unfortunately, he missed the chance to get a good picture.

(3) a. Bill drove a truck to a vacant lot and found some trash there.
b. He decided to clean up the lot.
c. [He/Bill] loaded the trash onto the truck.
c'. [He/Bill] loaded the truck with the trash.
d. [It/The trash] consisted mostly of cardboard boxes.
d'. [It/The truck] was much harder to drive afterward.
e. The neighbors were very grateful for all his work.

(4) a. Linda borrowed a wheelbarrow and some fertilizer from her neighbor.
b. She began to replant the garden in front of her house from scratch.
c. [She/Linda] loaded the fertilizer onto the wheelbarrow.
c'. [She/Linda] loaded the wheelbarrow with the fertilizer.
d. [It/The fertilizer] was then spread over the whole new garden bed.

d'. [It/The wheelbarrow] became difficult to maneuver around the yard.
e. The garden took many hours to finish.

(5) a. Jack went outside his mountain cabin where there was a bin and some firewood.
b. He was spending the day getting ready for the coming winter
c. [He/Jack] loaded the firewood into the bin.
c'. [He/Jack] loaded the bin with the firewood.
d. [It/The firewood] was then burned little by little over the winter.
d'. [It/The bin] closed just enough to keep out the rain and snow.
e. The winter proved to be one of the coldest ever.

(6) a. Alex went out onto the rear porch where there was a refrigerator and some beer.
b. He was helping to get ready for a frat party.
c. [He/Alex] loaded the beer into the refrigerator.
c'. [He/Alex] loaded the refrigerator with the beer.
d. [It/The beer] was eventually drunk by all the partygoers.
d'. [It/The refrigerator] was subsequently set to a supercool setting.
e. The fraternity ended up getting in trouble for underage drinking, though.

(7) a. Jane, who is a confectionary expert, put a briefcase and some candy on her desk one day.
b. Then she began preparing to go to a confectioner's business convention.
c. [She/Jane] packed the candy into the briefcase.
c'. [She/Jane] packed the briefcase with the candy.
d. [It/The candy] didn't melt in spite of the hot weather.
d'. [It/The briefcase] snapped shut afterward with no trouble.
e. The convention turned out to be a great success.

(8) a. Roger pulled up a chair and sat down in front of a chest and some bedding.
b. He was doing some spring cleaning.
c. [He/Roger] packed the bedding into the chest.
c'. [He/Roger] packed the chest with the bedding.
d. [It/The bedding] became pretty wrinkled in a few weeks.
d'. [It/The chest] was then closed to keep out the moths.
e. Roger didn't like cleaning, but it was a job that had to be done.

(9) a. Mark carried a bucket and some cement to his backyard.
b. He was getting ready to put up a brick wall.
c. [He/Mark] packed the cement into the bucket.
c'. [He/Mark] packed the bucket with the cement.

d. [It/The cement] dried in a very short time.
d'. [It/The bucket] was rusty and almost broke.
e. Building the brick wall took much longer than had been expected.

(10) a. Amy set out a cooler and some soda pop on the counter.
b. She was getting ready for a drive through the Mojave Desert.
c. [She/Amy] packed the soda pop into the cooler.
c'. [She/Amy] packed the cooler with the soda pop.
d. [It/The soda pop] soon went flat though because of the heat.
d'. [It/The cooler] also contained a lunch and one grapefruit.
e. Fortunately, the drive only took one day.

(11) a. Mary bought a basket and spent a couple of hours preparing a lunch.
b. She was planning to take her family on a picnic that day.
c. [She/Mary] packed the lunch into the basket.
c'. [She/Mary] packed the basket with the lunch.
d. [It/The lunch] got partially eaten on the way to the park.
d'. [It/The basket] got too full and the bottom began to break.
e. The picnic still turned out to be a lot fun for the whole family.

(12) a. Judy went out to her greenhouse where there was a crate and some fruit.
b. She started getting ready to go to the farmer's market that day.
c. [She/Judy] packed the fruit into the crate.
c'. [She/Judy] packed the crate with the fruit.
d. [It/The fruit] was ripe and had a delicious aroma.
d'. [It/The crate] was not old but had many splinters.
e. Judy made some good money at the market that day.

(13) a. Martha was in her kitchen one chilly morning and took out a muffin and some butter.
b. She wanted to make some breakfast for herself.
c. [She/Martha] spread the butter on the muffin.
c'. [She/Martha] spread the muffin with the butter.
d. [It/The butter] melted quickly and dripped onto the floor.
d'. [It/The muffin] sat on the plate steaming in the cold air.
e. Her breakfast also included some bacon and eggs.

(14) a. Howard went to the beach with a surfboard and some wax.
b. He had heard that the waves were really good.
c. [He/Howard] spread the wax on the surfboard.
c'. [He/Howard] spread the surfboard with the wax.
d. [It/The wax] was very soft and easy to apply.

d'. [It/The surfboard] looked like a glistening rocket.
e. The weather that day turned out to be awesome.

(15) a. Sally mixed up a bucket of plaster and sat down in front of a wall in her old house.
b. She had decided that she would make some renovations.
c. [She/Sally] spread the plaster on the wall.
c'. [She/Sally] spread the wall with the plaster.
d. [It/The plaster] dried slowly but felt almost like rock.
d'. [It/The wall] was quite big and took most of the day.
e. The renovation took many months, but it was worth it.

(16) a. Tom bought a sponge cake and some whipped cream at the supermarket.
b. He decided to make something special for his wife's birthday.
c. [He/Tom] spread the whipped cream on the sponge cake.
c'. [He/Tom] spread the sponge cake with the whipped cream.
d. [It/The whipped cream] melted quickly in the afternoon heat.
d'. [It/The sponge cake] was too soft and broke apart easily.
e. Fortunately, his wife was very forgiving.

(17) a. Marty went into a workshop in his garage where he had put a tabletop and some varnish.
b. He had decided to try his hand at finishing some furniture.
c. [He/Marty] spread the varnish on the tabletop.
c'. [He/Marty] spread the tabletop with the varnish.
d. [It/The varnish] soon dried clear as glass and became waterproof.
d'. [It/The tabletop] was then bolted onto three legs made previously.
e. The finished product became a main fixture in his dining room.

(18) a. Nancy went into the cafeteria and picked up a muffin and some jam.
b. She wanted to eat a small mid-afternoon snack.
c. [She/Nancy] spread the jam on the muffin.
c'. [She/Nancy] spread the muffin with the jam.
d. [It/The jam] oozed down over the whole plate.
d'. [It/The muffin] was hot and broke apart easily.
e. She enjoyed her snack very much and washed it down with some herbal tea.

(19) a. Katy brought a doll and some cotton to her sewing class.
b. She was in the process of trying to make a gift for her niece.
c. [She/Katy] stuffed the cotton into the doll.
c'. [She/Katy] stuffed the doll with the cotton.
d. [It/The cotton] got all clumped together in one segment.

d'. [It/The doll] ended up resembling a plump grandmother.
e. Her niece was very happy to receive a handmade gift.

(20) a. Florence set out an album and some memorabilia on her dining room table.
b. She had just gotten back from a trip to the British Isles.
c. [She/Florence] stuffed the memorabilia into the album.
c'. [She/Florence] stuffed the album with the memorabilia.
d. [It/The memorabilia] had been collected from six different places.
d'. [It/The album] couldn't even be closed completely afterward.
e. The trip was one of the best things that ever happened to her.

(21) a. Fred got out a pita bread and some tuna salad.
b. He started to make himself a nice healthy lunch.
c. [He/Fred] stuffed the tuna salad into the pita bread.
c'. [He/Fred] stuffed the pita bread with the tuna salad.
d. [It/The potato salad] was too runny and made a mess on the floor.
d'. [It/The pita bread] was too dry and broke open easily, however.
e. Fortunately, there was also a frozen dinner in the freezer.

(22) a. Janet put a turkey and some rice on the counter in her kitchen.
b. She started getting ready for her family's thanksgiving dinner.
c. [She/Janet] stuffed the rice into the turkey.
c'. [She/Janet] stuffed the turkey with the rice.
d. [It/The rice] became a delicious and moist stuffing.
d'. [It/The turkey] was plump and meaty and cooked nicely.
e. The dinner turned out very delicious and everyone had second helpings.

(23) a. Raymond got a rucksack out of his closet and some laundry out of his clothes hamper.
b. He was getting ready for a weekend trip to his parents' house.
c. [He/Raymond] stuffed the laundry into the rucksack.
c'. [He/Raymond] stuffed the rucksack with the laundry.
d. [It/The laundry] came out later all wrinkled and messy.
d'. [It/The rucksack] was thin and broke apart at the seams.
e. The weekend turned out to be really relaxing in the end.

(24) a. Walter took a wallet and some money out of his dresser.
b. He was getting ready for a night on the town.
c. [He/Walter] stuffed the money into the wallet.
c'. [He/Walter] stuffed the wallet with the money.
d. [It/The money] totalled more than a thousand dollars.

d'. [It/The wallet] was too full and wouldn't fold closed.
e. He had a great evening, but he had to ask his boss for a raise.

(25) a. Joseph looked at a bowl and some plastic wrap on the table in his kitchen.
b. He was trying to put away some leftovers from a small party.
c. [He/Joseph] draped the plastic wrap over the bowl.
c'. [He/Joseph] draped the bowl with the plastic wrap.
d. [It/The plastic wrap] didn't stick right and clung together horribly.
d'. [It/The bowl] suddenly slipped off the counter and shattered.
e. The leftovers ended up getting thrown away, unfortunately.

(26) a. Karen walked under a large open tent where there was a statue and a sheet.
b. She was preparing for the commemoration of a new memorial.
c. [She/Karen] draped the sheet over the statue.
c'. [She/Karen] draped the statue with the sheet.
d. [It/The sheet] almost blew away because of some high winds.
d'. [It/The statue] remained covered until the opening ceremony.
e. The new memorial was well-received by the public.

(27) a. Wendy, a hotel employee in room service, prepared a napkin and a teapot for one order.
b. She wanted to make sure to keep the order hot while delivering it.
c. [She/Wendy] draped the napkin over the teapot.
c'. [She/Wendy] draped the teapot with the napkin.
d. [It/The napkin] got stained and had to be sent out for dry cleaning.
d'. [It/The teapot] still cooled off very quickly in the drafty hallway.
e. Eventually, the hotel decided to invest in some electric warming platters.

(28) a. Marta approached a coffin at a burial ceremony and picked up a flag.
b. She was mourning the death of a fallen soldier.
c. [She/Marta] draped the flag over the coffin.
c'. [She/Marta] draped the coffin with the flag.
d. [It/The flag] blew away ominously in an unusual gust of wind.
d'. [It/The coffin] was rather heavy and was carried by eight men.
e. The ceremony was an emotional time for everyone there.

(29) a. Harry walked into the living room of his mansion where there was a mantle and some holly.
b. He loved to decorate his huge mansion during the holidays.
c. [He/Harry] draped the holly over the mantle.
c'. [He/Harry] draped the mantle with the holly.
d. [It/The holly] even hung down in front of the unused fireplace.

d'. [It/The mantle] was then used to support a large nativity scene.
e. The mansion emanated holiday spirit from every room.

(30) a. Lester, a restaurant worker, went into the kitchen and got some spaghetti and some sauce.
b. He started to prepare a platter for the Italian lunch buffet.
c. [He/Lester] heaped the sauce on the spaghetti.
c'. [He/Lester] heaped the spaghetti with the sauce.
d. [It/The sauce] was quite watery and dripped off the platter.
d'. [It/The spaghetti] soon became soft and limp on the big platter,
e. The platter looked quite unappetizing and few peopled tried it.

(31) a. Irene, a high school student, touched a pimple on her face and then took out some acne cream.
b. She was desperately searching for any method to clear up her skin.
c. [She/Irene] heaped the acne cream on the pimple.
c'. [She/Irene] heaped the pimple with the acne cream.
d. [It/The acne cream] dried up quickly and began to work at once.
d'. [It/The pimple] remained red and sensitive for a long time.
e. Keeping one's skin clear is a challenge for every teenager.

(32) a. Brad walked into his company conference room where there was a table and some paperwork.
b. He was trying to get organized after a long leave of absence.
c. [He/Brad] heaped the paperwork on the table.
c'. [He/Brad] heaped the table with the paperwork.
d. [It/The paperwork] eventually got straightened into a few piles.
d'. [It/The table] became completely covered right to the edges.
e. It took a couple of weeks, but Brad finally got caught up on his work.

(33) a. Molly got a hot dog and some relish at the baseball park.
b. She decided she wanted to get the full baseball experience.
c. [She/Molly] heaped the relish on the hot dog.
c'. [She/Molly] heaped the hot dog with the relish.
d. [It/The relish] was too runny and left a mess on the ground.
d'. [It/The hot dog] was undercooked and didn't taste that great.
e. The baseball game was otherwise a lot fun and the home team won.

(34) a. Rupert pulled a cart into the slums and collected some garbage from the streeets.
b. He was working on a neighborhood renovation project.
c. [He/Rupert] heaped the garbage onto the cart.

c'. [He/Rupert] heaped the cart with the garbage.
d. [It/The garbage] gradually settled and became like a mass of mud.
d'. [It/The cart] flipped over and broke under the extreme weight.
e. The neighborhood took many months to see much improvement.

(35) a. Doris stood in front of her cooking class with a lamb chop and some curry powder.
b. She had decided to teach her class a little about Indian cuisine.
c. [She/Doris] sprinkled the curry powder on the lamb chop.
c'. [She/Doris] sprinkled the lamb chop with the curry powder.
d. [It/The curry powder] soaked in quickly and gave a succulent flavor.
d'. [It/The lamb chop] grilled quickly giving off a very tasty aroma.
e. The class learned very quickly and the meal turned out to be delicious.

(36) a. Ellen carried a plant and some water into her private greenhouse.
b. She was trying to create a new garden bed.
c. [She/Ellen] sprinkled the water on the plant.
c'. [She/Ellen] sprinkled the plant with the water.
d. [It/The water] evaporated slowly because of the high humidity.
d'. [It/The plant] grew several inches during the next three days.
e. The new garden grew quickly and made Ellen very happy.

(37) a. Edwin placed a cookie and some sugar on the counter in front of him.
b. He was spending a lot of time this year preparing for the holidays.
c. [He/Edwin] sprinkled the sugar on the cookie.
c'. [He/Edwin] sprinkled the cookie with the sugar.
d. [It/The sugar] turned into a nice glaze in the warm oven.
d'. [It/The cookie] still needed some more festive decoration.
e. After much work, the decorations were all admired by everyone.

(38) a. Nora bought some charcoal and some lighter fluid before her outdoor dinner party.
b. She then had quite a lot of trouble starting the barbecue.
c. [She/Nora] sprinkled the lighter fluid on the charcoal.
c'. [She/Nora] sprinkled the charcoal with the lighter fluid.
d. [It/The lighter fluid] evaporated quickly and couldn't be easily lit.
d'. [It/The charcoal] mysteriously broke up into many little pieces.
e. The outdoor party had to be changed into an indoor party.

(39) a. Richard brought a canvas and some paint to his art class.
b. He had decided to experiment by making some abstract art.
c. [He/Richard] sprinkled the paint on the canvas.

c'. [He/Richard] sprinkled the canvas with the paint.
d. [It/The paint] formed tiny droplets in strange artistic shapes.
d'. [It/The canvas] was hung on a wall for some students to examine.
e. The other students were also motivated to try abstract expression.

(40) a. Louise, a cattle rancher, pulled a wagon into a barn and brought some hay down from the loft.
b. She had the responsibility of feeding the animals daily.
c. [She/Louise] piled the hay onto the wagon.
c'. [She/Louise] piled the wagon with the hay.
d. [It/The hay] was soon devoured by the cattle out in the fields.
d'. [It/The wagon] creaked loudly as it was pulled out to the fields.
e. The animals always seemed to know when food was coming.

(41) a. Anna took out a French roll and laid out some salami on the counter.
b. She was preparing some tea sandwiches for an afternoon party.
c. [She/Anna] piled the salami on the French roll.
c'. [She/Anna] piled the French roll with the salami.
d. [It/The salami] didn't have the usual spicy aroma.
d'. [It/The French roll] was smashed flat under the weight.
e. Unfortunately, the sandwiches turned out to be not very popular.

(42) a. Daniel, a waiter, was busy with a tray and some food in the kitchen.
b. He was getting ready to make a delivery to some customers.
c. [He/Daniel] piled the food onto the tray.
c'. [He/Daniel] piled the tray with the food.
d. [It/The food] was not all eaten because there was too much.
d'. [It/The tray] had a slight crack and broke into two pieces.
e. The customers were not very happy and left only a modest tip.

(43) a. Scott, a pilot, examined an airplane and some medicine at the airfield.
b. He was helping out in an emergency project for international aid.
c. [He/Scott] piled the medicine into the airplane.
c'. [He/Scott] piled the airplane with the medicine.
d. [It/The medicine] was used to help contain the spread of HIV and AIDS.
d'. [It/The airplane] took off to the neediest parts of equatorial Africa.
e. Scott tried to help with these projects as often as he could.

(44) a. Philip looked on top of his dresser which had a frisbee and some loose change on it.
b. He decided to started cleaning up his messy dorm room.
c. [He/Philip] piled the loose change onto the frisbee.

c'. [He/Philip] piled the frisbee with the loose change.
d. [It/The loose change] was eventually used to buy some coke and pizza.
d'. [It/The frisbee] didn't fly again until after final exams ended.
e. Although it took Philip several hours, his room eventually became clean again.

(45) a. Betty, a new bakery employee, was given a doughnut and some custard.
b. She was told to practice the important task of filling.
c. [She/Betty] injected the custard into the doughnut.
c'. [She/Betty] injected the doughnut with the custard.
d. [It/The custard] squirted all over and made a mess on the floor.
d'. [It/The doughnut] bloated up and became like a huge round orange.
e. The bakery had a good reputation and expected better from the employees.

(46) a. Robert, a veterinarian, took out an apple and some medicine.
b. He was trying to cure a sick horse.
c. [He/Robert] injected the medicine into the apple.
c'. [He/Robert] injected the apple with the medicine.
d. [It/The medicine] became a little less potent but was still strong.
d'. [It/The apple] remained crisp and red and looked very delicious.
e. Robert successfully cured the horse in just a matter of days.

(47) a. Matthew, on the day before Thanksgiving, set out a turkey and some beef broth.
b. He had decided to try a new way of preparing Thanksgiving dinner.
c. [He/Matthew] injected the beef broth into the turkey.
c'. [He/Matthew] injected the turkey with the beef broth.
d. [It/The beef broth] squirted all over and made a mess on the floor.
d'. [It/The turkey] baked up nice and tender in a clay baking dish.
e. The holidays are a fun time for experimenting with cuisine.

(48) a. Gary sneaked into his school's locker room and took out a football and some helium.
b. He wanted to give himself a small edge in that day's game.
c. [He/Gary] injected the helium into the football.
c'. [He/Gary] injected the football with the helium.
d. [It/The helium] mixed with air and ultimately had no effect.
d'. [It/The football] was rather old and burst apart at the seams.
e. In the end, Gary's team was able to win without cheating.

(49) a. Jill works in the information technology department.
b. One day she had a very big maintenance job.

c. [She/Jill] loaded a new operating system onto a hard drive.
c'. [She/Jill] loaded a hard drive with a new operating system.
d. [It/The operating system] had to be deleted soon after.
d'. [It/The hard drive] started to smoke on the next reboot.
e. The job ended up taking several days to complete.

(50) a. Frank was selling some of his old clothing at a flea market.
b. He prepared a lot of his old leather goods for the sale.
c. [He/Frank] rubbed some polish on a jacket.
c'. [He/Frank] rubbed a jacket with some polish.
d. [It/The polish] was a little old and difficult to apply.
d'. [It/The jacket] then sold for more than two hundred dollars.
e. The flea market saw a lot of money change hands that day.

(51) a. Judy was getting herself ready one day for a long road trip.
b. She tried to figure out how to make her car look a little nicer.
c. [She/Judy] rubbed some wax on a fender.
c'. [She/Judy] rubbed a fender with some wax.
d. [It/The wax] was easy to apply with a soft sponge.
d'. [It/The fender] now looked as good as new.
e. The car shined up nicely and looked great afterward.

(52) a. Howard was getting ready to go to the beach.
b. He had heard that the waves were really good.
c. [He/Howard] rubbed some wax on a surfboard.
c'. [He/Howard] rubbed a surfboard with some wax.
d. [It/The wax] was very soft and easy to apply.
d'. [It/The surfboard] looked like a glistening rocket.
e. The weather that day turned out to be awesome.

(53) a. Janet bought a new gas grill and decided to try it out one night.
b. She was getting everything ready to cook.
c. [She/Janet] rubbed some olive oil on a steak.
c'. [She/Janet] rubbed a steak with some olive oil.
d. [It/The olive oil] was poured from a newly opened bottle.
d'. [It/The steak] then cooked up really nicely.
e. The grill worked fine and was so easy to clean up afterward.

(54) a. Robert helped a friend who was moving his family across town.
b. He was very tired at the end of the day.
c. [He/Robert] rubbed some camphor oil on a sore shoulder.
c'. [He/Robert] rubbed a sore shoulder with some camphor oil.

d. [It/The camphor oil] unfortunately caused an allergic skin reaction.
d'. [It/The sore shoulder] soon felt much better.
e. The next day was still very difficult for him.

(55) a. Gary took a class about bicycle maintenance and repair.
b. He started practicing with some of the equipment.
c. [He/Gary] injected some air into a tire.
c'. [He/Gary] injected a tire with some air.
d. [It/The air] made a hissing sound while escaping.
d'. [It/The tire] burst into many pieces.
e. The teacher was fortunately very patient and helpful.

(56) a. Audrey worked at a pharmaceutical laboratory.
b. She began a new experiment one day.
c. [She/Audrey] injected some serum into a rat.
c'. [She/Audrey] injected a rat with some serum.
d. [It/The serum] was also tested on a monkey and a rabbit.
d'. [It/The rat] soon fell fast asleep in the cage.
e. The project was deemed a success and they moved on to phase two.

(57) a. Lynn worked at a veterinary clinic.
b. She was asked one day to do an unenviable task.
c. [She/Lynn] injected some poison into a dog.
c'. [She/Lynn] injected a dog with some poison.
d. [It/The poison] worked very quickly and apparently painlessly.
d'. [It/The dog] whimpered for a few seconds and then became quiet.
e. The clinic often struggled with how to do this work.

2. Items Using *Tough*/non-*Tough* Constructions

(1) a. Marty and Darren were playing king-of-the-hill with their friends.
b. Everyone tried to push the person at the top of the hill off.
c. Marty could hardly move Darren during the game.
c'. Darren was hard for Marty to move during the game.
d. [He/Marty] finally became frustrated and started to bawl.
d'. [He/Darren] couldn't stop bragging when the game was over.
e. Still, everybody had a lot of fun.

(2) a. Stephanie and Marjorie served in the same platoon in the Marines.
b. One day, their commander ordered the platoon to secure a village.
c. Stephanie could hardly command Marjorie during the military operation.
c'. Marjorie was hard for Stephanie to command during the military operation.
d. [She/Stephanie] resolved to be more authoritative in the future.
d'. [She/Marjorie] was finally court-martialed for insubordination.
e. The operation turned out to be a miserable failure.

(3) a. John and Matt fought each other in a boxing match.
b. It was twelve rounds long.
c. John could hardly hit Matt in the final round of the match.
c'. Matt was hard for John to hit in the final round of the match.
d. [He/John] still managed to land a knockout punch, though.
d'. [He/Matt] completed the round without getting hit at all.
e. The judges had no trouble deciding the winner.

(4) a. Carl and Will played "flag tag" with some friends from school.
b. The object is to capture members of the other team by pulling a cloth out of their pocket.
c. Carl could hardly capture Will during the game.
c'. Will was hard for Carl to capture during the game.
d. [He/Carl] eventually became quite frustrated and then quit.
d'. [He/Will] successfully resisted with some clever maneuvers.
e. The game ended fairly quickly.

(5) a. Jenny and Vicki both entered a foreign language speech competition.
b. Each participant had to make a five-minute speech.
c. Jenny could hardly hear Vicki during the speeches.
c'. Vicki was hard for Jenny to hear during the speeches.
d. [She/Jenny] then listened very closely but it was no use.
d'. [She/Vicki] still got a loud applause at the end, though.
e. Unfortunately, though, neither of them won any prizes that day.

(6) a. Kathy and Tessa went to summer camp together.
b. The campers all slept in rustic log cabins.
c. Kathy easily woke up Tessa in the mornings.
c'. Tessa was a cinch for Kathy to wake up in the mornings.
d. [She/Kathy] eventually woke up everybody in the cabin.
d'. [She/Tessa] promptly yawned once and then sat upright.
e. If nothing else, every morning was always interesting in their cabin.

(7) a. Andrea and Carolyn decided to play a game of chess one day.
b. Chess games can often be very long, consisting of hundreds of moves.
c. Andrea easily outsmarted Carolyn after only a few chess moves.
c'. Carolyn was a cinch for Andrea to outsmart after only a few chess moves.
d. [She/Andrea] then did a little victory dance around the table.
d'. [She/Carolyn] became very upset over losing the game so easily.
e. Afterward, they decided to return to the old standard–checkers.

(8) a. Peter and William liked to go to magic shows.
b. The magicians always performed such amazing and mysterious illusions.
c. Peter easily tricked William after many magic shows.
c'. William was a cinch for Peter to trick after many magic shows.
d. [He/Peter] also fooled several grownups who happened to watch.
d'. [He/William] typically just sat there, confused and dumbfounded.
e. Eventually, though, they both became amateur magicians.

(9) a. Ted and Jim went to the company picnic this weekend.
b. There were three games: the piggy-back race, the three-legged race and the water-balloon fight.
c. Ted easily carried Jim in the piggy-back race.
c'. Jim was a cinch for Ted to carry in the piggy-back race.
d. [He/Ted] didn't stumble at all in spite of the excess weight.
d'. [He/Jim] held on tight with two hands to prevent falling off.
e. Their coworkers knew they would win the race easily.

(10) a. Greg and Max went to the same elementary school.
b. The children had two recess periods every day–one indoor and one outdoor.
c. Greg easily irritated Max during the afternoon recess.
c'. Max was a cinch for Greg to irritate during the afternoon recess.
d. [He/Greg] then usually got in trouble with the teachers.
d'. [He/Max] sometimes fought back with a vengance, though.
e. The playground was a pretty tough place during recess time.

(11) a. Steve and Jeff were doing their math homework late at night.
b. There were many very difficult problems.
c. Steve easily helped Jeff on many of the math homework problems.
c'. Jeff was a cinch for Steve to help on many of the math homework problems.
d. [He/Steve] expected to get something in return, though.
d'. [He/Jeff] decided to pay for a deluxe pizza in return.
e. They finally finished the assignment at about 3 a.m.

(12) a. Liz and Jane were playing hide-and-seek late one afternoon.
b. A little before six-o'clock, the sun started to set.
c. Liz could barely find Jane as it started to get dark.
c'. Jane was difficult for Liz to find as it started to get dark.
d. [She/Liz] continued to search until after the sun went down.
d'. [She/Jane] remained hidden until it was just too dark to see.
e. Afterward, they both had a big laugh and went home.

(13) a. Ted and Ray started playing tag at school one day.
b. Nine of their friends also joined the game.
c. Ted could hardly catch Ray during the game of tag.
c'. Ray was difficult for Ted to catch during the game of tag.
d. [He/Ted] was still 'it' when the end-of-recess bell rang.
d'. [He/Ray] eventually got tired and was tagged by somebody.
e. The children always enjoyed playing tag very much.

(14) a. Mark and John were roommates at college.
b. One night, a fire alarm went off in the dormitory.
c. Mark could hardly wake up John during the fire alarm.
c'. John was difficult for Mark to wake up during the fire alarm.
d. [He/Mark] then tried crashing a pair of cymbals together.
d'. [He/John] continued to sleep until the alarm had stopped.
e. After the alarm stopped, everyone went back to bed.

(15) a. Stan and James were both doctors at the city hospital.
b. One day, a patient with severe chest injuries came in.
c. Stan assisted James with great difficulty during the medical operation.
c'. James was difficult for Stan to assist during the medical operation.
d. [He/Stan] tried to be as helpful as possible, anyway.
d'. [He/James] even refused help from the chief physician.
e. The operation was a success, but they never worked together again.

(16) a. Barry and Martin worked together at a large company.
b. One day, the company got a new Canon copy machine with many

complex functions.
c. Barry trained Martin with much difficulty after the new copy machine was installed.
c'. Martin was difficult for Barry to train after the new copy machine was installed.
d. [He/Barry] also tried to teach the rest of the staff but found it impossible.
d'. [He/Martin] finally learned the functions and then sought to return the favor.
e. The new copy machine was apparently a little bit over-rated.

(17) a. Judy and Laura were neighbors in a small midwestern town.
b. Most of the townspeople were not very hospitable toward each other.
c. Judy delightedly visited Laura despite the bad feelings among all the neighbors.
c'. Laura was fun for Judy to visit despite the bad feelings among all the neighbors.
d. [She/Judy] even stayed for the family dinner on occasion.
d'. [She/Laura] eventually became a model hostess in the town.
e. They got together for tea practically every day.

(18) a. Jennifer and Allison were in the same class in their elementary school.
b. One day the whole school went on a field trip to a nature museum by bus.
c. Jennifer delightedly teased Allison on the bus ride to school.
c'. Allison was fun for Jennifer to tease on the bus ride to school.
d. [She/Jennifer] even dared to say some really spiteful things.
d'. [She/Allison] got tired of being picked on and became angry.
e. However, the teacher had seen it all and finally told everyone to be quiet.

(19) a. Marilyn and Jennifer both took part in a talent show at school.
b. There were eight comedy acts, five musicians, as well as two dramatic performances in the show.
c. Marilyn joyfully watched Jennifer during the talent show.
c'. Jennifer was fun for Marilyn to watch during the talent show.
d. [She/Marilyn] was even the first to applaud afterward.
d'. [She/Jennifer] had the audience in stitches at the end.
e. The show turned out to be a great fundraiser for the school.

(20) a. Kevin and Larry worked together in the same office and were close friends.
b. One night they worked so late, the trains and buses had stopped runnning.
c. Kevin delightedly had Larry as a guest.
c'. Larry was fun for Kevin to have as a guest.
d. [He/Kevin] turned out to be a really hospitable host.

d'. [He/Larry] turned out to be a very considerate guest.
e. They went back to work the next day fully refreshed.

(21) a. Carrie and Julie went to the same junior high school.
b. At their school, talented older students tutored younger students who needed help.
c. Carrie delightedly tutored Julie in math.
c'. Julie was fun for Carrie to tutor in math.
d. [She/Carrie] even explained some good test-taking strategies.
d'. [She/Julie] improved enough to get a B-plus in the semester.
e. Their teachers considered them perfect examples of the success of the tutoring program.

(22) a. Rachel and Lucy went to a masquerade party at school.
b. Their dates each picked them up at home and drove them to the party.
c. Rachel easily recognized Lucy at the school masquerade party.
c'. Lucy was easy for Rachel to recognize at the school masquerade party.
d. [She/Rachel] even figured out the principal in the ape suit.
d'. [She/Lucy] was easily identified by everyone at the party.
e. The party was a big hit and everyone raved about the costumes long afterward.

(23) a. Barney and Joseph were playing a war game in the forest.
b. The red team members were supposed to hunt for the blue team members.
c. Barney easily evaded Joseph throughout the red-blue war game.
c'. Joseph was easy for Barney to evade throughout the red-blue war game.
d. [He/Barney] was finally caught by the red commander, though.
d'. [He/Joseph] captured many of the blue team soldiers, though.
e. In the end, the red team won the games.

(24) a. Sue and Meg were good friends throughout most of their adult lives.
b. They helped each other out through good times and bad.
c. Sue easily supported Meg during the financially lean times.
c'. Meg was easy for Sue to support during the financially lean times.
d. [She/Sue] also helped out several other friends in town.
d'. [She/Meg] gladly returned the favor as soon as possible.
e. Their friendship easily survived these turmoils and grew ever stronger.

(25) a. Nelson and William were swimming in the sea one windy day.
b. Suddenly, a strong undertow started pulling many swimmers out into deep water.
c. Nelson easily rescued William during the emergency.

c'. William was easy for Nelson to rescue during the emergency.
d. [He/Nelson] was later given a special community award for bravery.
d'. [He/William] was forever grateful and didn't forget the experience.
e. Fortunately, everyone got out of the water safely that day.

(26) a. Mark and Dan were bitter rivals on their wrestling team at school.
b. At the beginning of one practice, the coach carelessly assigned new practice pairs.
c. Mark easily hurt Dan during the practice.
c'. Dan was easy for Mark to hurt during the practice.
d. [He/Mark] was then told never to use a choke-hold again.
d'. [He/Dan] ended up needing a trip to the emergency room.
e. The coach should have been more careful in assigning partners.

(27) a. Nancy and Susan ran in the state championship track meet this weekend.
b. The meet included 100-meter, 500-meter, and 1000-meter races.
c. Nancy easily beat Susan in the 100-meter race.
c'. Susan was easy for Nancy to beat in the 100-meter race.
d. [She/Nancy] became the state champ for the second year.
d'. [She/Susan] was frustrated and dejected after the race.
e. The state championship meet is always highly competitive every year.

(28) a. Gary and Tony were rivals at high school–in sports, academics, and everything.
b. At the senior awards ceremony, the valedictorian was announced.
c. Gary congratulated Tony with great difficulty after the ceremony.
c'. Tony was tough for Gary to congratulate after the ceremony.
d. [He/Gary] managed to mumble a brief congratulatory word.
d'. [He/Tony] accepted the congratulatory remark graciously.
e. Their rivalry ended when they went to different colleges.

(29) a. Robert and David were brothers.
b. One evening, their parents went out to see a concert.
c. Robert babysat David with much difficulty.
c'. David was tough for Robert to babysit.
d. [He/Robert] decided not to have any children in the future.
d'. [He/David] eventually became quite rebellious and naughty.
e. Sure enough, when their parents came home, the boys were fighting.

(30) a. Rhonda and Tara are sisters who spent the day alone last Saturday.
b. Their mother couldn't find a babysitter in time.
c. Rhonda could hardly control Tara.

c'. Tara was tough for Rhonda to control.
d. [She/Rhonda] eventually threatened to use corporal punishment.
d'. [She/Tara] wouldn't even respond to any corporal punishment.
e. Taking care of children is more challenging than one might think.

(31) a. Tony and Harry were at a bar one night.
b. Suddenly a huge brawl broke out in the parking lot among some of the customers.
c. Tony could hardly punch Harry during the fight.
c'. Harry was tough for Tony to punch during the fight.
d. [He/Tony] missed once and struck a waitress instead.
d'. [He/Harry] was able to sidestep nearly every attempt.
e. The police finally came and arrested everybody.

(32) a. Heidi and Lisa were on their high school karate team.
b. Their coach decided to have challenge matches before the district finals.
c. Heidi could hardly throw Lisa during the match.
c'. Lisa was tough for Heidi to throw during the match.
d. [She/Heidi] lost the match by an unusually wide margin.
d'. [She/Lisa] fought off every attack in the whole match.
e. The rest of the team was really surprised at the result.