INFORMATION STRUCTURE OF CLEFTS

IN SPOKEN ENGLISH

by

JENNIFER A. PIOTROWSKI

A THESIS

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Arts

September 2009

"Information Structure of Clefts in Spoken English," a thesis prepared by Jennifer A. Piotrowski in partial fulfillment of the requirements for the Master of Arts degree in the Department of Linguistics. This thesis has been approved and accepted by:

Dr. Doris L. Payne, Co-Advisor

Dr. Melissa A. Redford, Co-Advisor

Aug. 24, 2009
Date

Committee in Charge: Dr. Doris L. Payne
Dr. Melissa A. Redford

Accepted by:

Dean of the Graduate School

An Abstract of the Thesis of

Jennifer A. Piotrowski for the degree of Master of Arts

in the Department of Linguistics to be taken September 2009

Title: INFORMATION STRUCTURE OF CLEFTS IN SPOKEN ENGLISH

Approved: _____
Dr. Doris L. Payne, Co-Advisor

Approved: _____
Dr. Melissa A. Redford, Co-Advisor

Towards a more complete description of cleft constructions, this thesis comprises an investigation of the prosody, syntax, and information structure of IT clefts, REVERSE WH clefts, and existential THERE clefts in Spoken English. Cleft constructions were extracted from the Santa Barbara Corpus of Spoken American English on the basis of syntactic characteristics, and empirical methods were developed for evaluating clefts with respect to prosody and information structure factors. Native speaker-hearer judgments about cleft constructions in authentic spoken language were gathered to provide a basis for operational definitions of PROSODIC PROMINENCE, GIVENNESS, NEWNESS, CONTRASTIVENESS, and levels of contextual RELEVANCE. While cleft constructions have conventionally been discussed as contrastive focusing devices, the current study provides empirical evidence for a more complex view of clefts. Added to past corpus studies, this

thesis shows that English cleft constructions exhibit a broader range of subtypes and functions than captured by traditional accounts.

CURRICULUM VITAE

NAME OF AUTHOR: Jennifer A. Piotrowski

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon

University of Michigan, Ann Arbor, Michigan

Alpena Community College, Alpena, Michigan

DEGREES AWARDED:

Master of Arts, Linguistics, 2009, University of Oregon

Bachelor of Arts, Linguistics & English Language and Literature, 2003, University of Michigan

AREAS OF SPECIAL INTEREST:

Information Structure
Functions of Prosody
Corpus Linguistics
Language Acquisition

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, 2006-2009

Administrative Assistant, Literature And Teaching Ministries, 2005-2006

Composition Rater, University of Michigan's English Language Institute, 2004

Lecturer, Assumption University, 2003-2004

AWARDS AND HONORS:

Leon Culbertson Scholarship, 2008-2009

University Honors, University of Michigan, 2000-2003

James B. Angell Scholar, University of Michigan, 2002, 2003, 2004

Membership, The Honor Society of Phi Kappa Phi, 2002

Membership, Golden Key International Honour Society, 2001

ACKNOWLEDGMENTS

I gratefully acknowledge the support and constructive comments of Doris Payne and Melissa Redford in the preparation of this manuscript. Additionally, Eric Pederson and the Cognitive Linguistics Workgroup, as well as the Fall 2008 Discourse Analysis students, provided valuable input during the formative stages of this project. I also appreciate the many friends, neighbors and fellow graduate students who helped pilot the listening tasks for this study, and who gave an abundance of helpful feedback. Finally, I express sincere thanks to Niki Piotrowski, the extended Goad and Piotrowski families, and the Co-op Family Center for their invaluable support and care for Judah and Kazimir, without which I could not have completed this undertaking.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
1.1. Overview of the Thesis	1
1.2. Traditional Claims about Cleft Information Structure	1
1.3. Background: Cleft Constructions	2
1.3.1. Constructional Approach	2
1.3.2. Defining 'Cleft Construction'	3
1.3.3. Parts of a Cleft Construction	4
1.3.4. Disambiguation	5
1.3.5. Prosody and Defining Cleft Constructions	6
1.3.6. Summary	7
1.4. Background: Information Structure	7
1.4.1. What Is Information Structure?	7
1.4.2. An Illustration of Information Structure Coding	8
1.4.3. Conceptual Definitions	8
1.4.4. Speaker-hearer Concerns in the Current Study	10
1.4.5. Prosodic Coding of Information Structure	10
1.5. Originality of the Current Study	12
1.6. Structure of the Thesis	12
II. INFORMATION STRUCTURE AND CLEFT SUBTYPES	14
2.1. An Expanding View of Clefts	14
2.2. Lambrecht's Functional Typology of Cleft Constructions	15
2.3. Preliminary Predictions	19
2.4. Corpus Findings	20
2.4.1. Clefts in General	21
2.4.2. IT Clefts	22
2.4.3. REVERSE WH Clefts	24
2.4.4. Summary	26

Chapter Page

III. METHODOLOGY 28
3.1. Overview of Methodology 28
3.2. Data Selection 28
3.2.1. The Corpus 28
3.2.2. Syntactic Criteria for Target Constructions 29
3.2.3. Data Set Rationale 30
3.2.4. Use of Target Tokens in Tasks 31
3.3. Native Speaker-hearer Judgment Tasks 31
3.3.1. Subjects 32
3.3.2. Auditory Stimuli 32
3.3.3. Visual Stimuli 35
3.3.4. Procedure 36
3.4. Listener Judgments 37
3.4.1. Prosodic Prominence 37
3.4.2. Newness/Givenness 40
3.4.3. Contrastiveness 41
3.4.4. Context-based Relevance 42
3.5. Methods for Hypothesis-testing 44
3.6. Hypotheses and Predictions 45

IV. RESULTS 48
4.1. Overview of Results 48
4.2. Relative Frequency Results 48
4.3. Listener Judgments Results 49
4.3.1. Perceived Prosodic Prominence 49
4.3.2. Prosody and Information Structure 52
4.3.3. Newness/Givenness 53
4.3.4. Contrastiveness 56
4.3.5. Information Structure Patterns 57
4.3.6. Contextual Relevance 59
4.4. Summary of Results 64

Chapter Page

V. CONCLUSION 65
5.1. Summary of the Work 65
5.2. Various Motivations for Cleft Use 66
5.3. Speaker-hearer Issues 66
5.4. Methodological Improvements 68
5.4.1. Improvements to the Stimuli 68
5.4.2. Other Design Improvements 69
5.5. Concluding Remarks 69

APPENDICES 71
A. CORPUS SEARCHES PERFORMED 71
B. CLEFT CONSTRUCTIONS MEETING SEARCH CRITERIA 73
C. LISTENER JUDGMENT TASK SCRIPTS 79
D. TRAINING AND PRACTICE EXAMPLES 83

REFERENCES 90

LIST OF FIGURES

Figure Page

1. Prosody Judgment Example Screen 39
2. Given/New Judgment Example Screen with Clefted Element 41
3. Given/New Judgment Example Screen with Open Proposition 41
4. Contrastiveness Judgment Example Screen 42
5. Relevance Rating Example Screen with Clefted Element 44
6. Relevance Rating Example Screen with Open Proposition 44
7. Proportions for Prosodically Prominent Cleft Components 49
8. Proportions of Cleft Subtypes Exhibiting Prosodic Patterns 51
9. Proportions of Given/New/Indeterminate Clefted Elements 54
10. Proportions of Given/New/Indeterminate Open Propositions 55
11. Proportions of Clefted Elements Judged Contrastive 57
12. Relevance Levels of Clefted Elements 60
13. Relevance Levels of Open Propositions 61

LIST OF TABLES

Table		Page
1.	Skeletal Structures of Clefts under Study	4
2.	Terminology for Identifying Cleft Parts	5
3.	Focus Structure Predictions for Cleft Subtypes under Study	19
4.	Profiles of IT Clefts and REVERSE WH Clefts Based on Corpus Findings	27
5.	Structures of Subtypes of Clefts under Study, with Examples	29
6.	Predictions for Cleft Subtypes	47
7.	Token Counts for Prosodically Prominent Cleft Components	49
8.	Token Counts for Cleft Subtypes Exhibiting Prosodic Patterns	51
9.	Correlations among Prosodic Prominence and Information Structure	53
10.	Token Counts for New/Given/Indeterminate Cleft Components	55
11.	Token Counts for Contrastive Clefted Elements	56
12.	Token Counts for Newness Patterns of Cleft Components	58
13.	Contrastiveness of Given Clefted Elements	59
14.	Mean Relevance and Standard Deviations for Clefted Elements	62
15.	Mean Relevance and Standard Deviations for Open Propositions	63

CHAPTER I

INTRODUCTION

1.1. OVERVIEW OF THE THESIS. Towards a more complete description of the subtypes and functions of cleft constructions, the current study comprises an investigation of multiple subtypes of English clefts, with a particular interest in information structure-related properties. Past descriptions and analyses of cleft constructions are reviewed, and an original corpus investigation of IT clefts, REVERSE WH clefts, and existential THERE clefts in spoken American English is presented (see examples 1 a-c).

(1) a. It's an apartment that I want to rent. — IT cleft
b. An apartment is what I want to rent. — REVERSE WH cleft
c. There's an apartment that I want to rent. — THERE cleft

As part of this descriptive goal, a variant on the standard claim that cleft constructions function as syntactic focusing devices is evaluated.

A secondary goal of the thesis is the development, implementation, and evaluation of new empirical methods for characterizing clefts in authentic language data with respect to information structure related concepts. Results from native speaker-hearer judgments are used to operationalize NEWNESS, GIVENNESS, and CONTRASTIVENESS. RELEVANCE ratings are also considered as a possible avenue for investigating information structure in authentic data.

1.2. TRADITIONAL CLAIMS ABOUT CLEFT INFORMATION STRUCTURE. In addition to more general descriptive goals, the current study is designed to test previous claims about cleft constructions that have become fairly standard explanations of cleft function. The main claim is that clefts are syntactic devices for CONTRASTIVE FOCUS (Jespersen 1949, Chafe 1976, Givón 2001, and others). A weaker variant of this claim is that clefts clearly

separate the FOCUS or NEW INFORMATION from the PRESUPPOSITION or GIVEN/OLD INFORMATION in a proposition. These claims are elaborated in Lambrecht's (2001) analysis, which describes clefts as conventionally coding SPECIFICATIONAL ARGUMENT FOCUS, such that an open proposition is pragmatically presupposed, the specification of the value completing the open proposition is asserted, and the expression of the specified value (termed here CLEFTED ELEMENT) is the FOCUS DOMAIN. These claims are unpacked, further developed, and tested as hypotheses in the current study.

1.3. BACKGROUND: CLEFT CONSTRUCTIONS. While early discussions of cleft constructions (Jespersen 1949), as well as some more recent discussions of cleft constructions (e.g., Givón 2001), appear to consider only IT clefts under the heading of 'clefts', the current study employs a syntactic definition of CLEFT CONSTRUCTIONS that is far broader. Largely informed by Lambrecht (2001), this section of the thesis explicates the relevant characteristics of and assumptions about cleft constructions as they are understood here.

1.3.1. CONSTRUCTIONAL APPROACH. Of initial importance is the constructional approach to understanding cleft constructions that is adopted here. The current study proceeds from the view that the constructional meaning or function of cleft constructions relates to information structure, or the pragmatic construal of information based on a speaker's assumptions about the mental state of the hearer. The constructional meaning/function of clefts in general, and also of specific subtypes of clefts, is explored further in the sections and chapters that follow.

The current work does not assume that the meaning of a cleft construction is equal to the sum of its parts, nor is it invested in any claim that clefts can be derived and/or related to other constructions via transformation or from some more basic structure (e.g., Akmajian 1970, Higgins 1971, Hankamer 1974). Instead, it is assumed here that cleft constructions as a whole have an associated abstract constructional meaning and/or function, and that this constructional meaning need not necessarily be derived analytically from the sum of its parts (cf. Fillmore et al. 1988, Goldberg 1995, and others).

1.3.2. DEFINING 'CLEFT CONSTRUCTION'. Otto Jespersen was among the first to note some peculiar characteristics of what he termed CLEFT SENTENCES (1937). Coming from Jespersen is this notion that cleft constructions are characterized by the use of biclausal syntax for expressing a proposition that could be expressed using simpler syntax, without a change in truth conditions. Revising an earlier analysis of "relative clause adjuncts" (1927), Jespersen proposed that *it* and *is* can be understood as the "lesser subject and verb" in sentences like example 2 below, where *wife* and *decides* are the sentence's primary subject and verb.

(2) It is the wife that decides.

In a later work, Jespersen suggested that cleft sentences like example 1a and 2, 'serve to single out one particular element of the sentence and very often, by directing attention to it and bringing it, as it were, into focus, to mark a contrast' (1949). More recent analyses of cleft constructions have incorporated and expanded on these observations, and the notion that clefts are syntactic devices for marking CONTRASTIVE FOCUS persists as the dominant view (Chafe 1976, Givón 2001, and others). Variations on this view will be discussed further in Chapter II.

Expanding on Jespersen's discussion of cleft sentences, Knud Lambrecht provides the most complete and up-to-date framework for analyzing cleft constructions (2001). While Jespersen initially set out to describe only the kinds of constructions referred to here as IT clefts (a cleft subtype), the current study follows Lambrecht and others in considering a broader range of structures under the heading of CLEFT CONSTRUCTIONS.

To be clear, cleft constructions are understood here as a set of constructions, including a number of subtypes, characterized by the use of biclausal syntax to express a proposition that could grammatically be expressed using simpler syntax, but crucially without a change in truth conditions. To illustrate this property of cleft constructions, the same content as in examples 1a-c is expressed more simply in example 3, and example 2 is similarly simplified in example 4, with no change in the propositional semantics.

(3) I want to rent an apartment.

(4) The wife decides.

Clefts are also characterized by their SPECIFICATIONAL nature; that is, the property of specifying a value for a variable rather than directly predicating something about an argument (Declerck 1988, Lambrecht 2001)[1]. To illustrate, for examples 1a-c above, the expression of the variable, or open proposition, would be 'I want to rent something,' or 'I want to rent x,' and the specified value would be 'an apartment,' or x = 'an apartment'. For example 2 above, the open proposition would be 'someone decides', and the specified value for the variable would be 'the wife'. In the current study, clefts are assumed to have a specificational structure, but they are not assumed to have specificational focus, such that the specified value is necessarily new or contrastive and the open proposition is presupposed.

1.3.3 PARTS OF A CLEFT CONSTRUCTION. It was noted above that a cleft construction formally separates a proposition into two parts, termed here the CLEFTED ELEMENT and the OPEN PROPOSITION. In each of the cleft constructions illustrated in 1a-c, the clefted element is 'an apartment', and the open proposition is 'I want to rent something.' In example 2, 'the wife' is the clefted element, and 'someone decides' is the open proposition.

Table 1 summarizes the syntactic structures of the cleft subtypes under study here. The constituents expressing the clefted element and the open proposition are represented as X and Y, respectively, in the skeletal structures displayed in Table 1.

CLEFT SUBTYPE	STRUCTURE
IT	*It* BE X REL Y
REVERSE WH	X BE WH.REL Y
THERE	*There* BE X REL Y

TABLE 1. Skeletal structures of clefts under study

[1] While Declerck's (1988) and Lambrecht's (2001) use of *specificational* entails that the open proposition is presupposed, this is not the case here. The current use of *specificational* differs from previous uses in that it does not assume any particular focus structure.

In these structures, BE stands for some form of *to be* (e.g., *was, is, 's*), REL stands for a relativizer, such as *that* or *who*, and WH.REL stands for a so-called headless relativizer such as *what* or *who*. *It* and *There* as the initial elements in clefts are understood to be semantically empty, rather than fully referring, expressions.

A variety of terms have been used in the literature to discuss the two main parts of cleft constructions, and it will be useful to show how the terminology adopted here corresponds to terms used by others. A list of some other terms that have been used to identify these two parts of cleft constructions is provided in Table 2.

CLEFTED ELEMENT (CE)	OPEN PROPOSITION (OP)	
focus phrase	relative clause	Lambrecht 2001
focused part	presupposed part	Prince 1978
clefted constituent	cleft clause	Hedberg 1990, Weinert and Miller 1996
cleft head	cleft compliment	Delin (various)
highlighted element	relative clause	Collins (2004,2006)

TABLE 2. Terminology for identifying cleft parts

The terms used here, CLEFTED ELEMENT and OPEN PROPOSITION, are intended to be descriptive of the specificational structure of cleft constructions, but to remain neutral with regard to the information structure function of these elements. Thus, the formal partitioning of the variable (open proposition) and its specified value (clefted element) is recognized, but it is not assumed here that 'I want to rent something' is necessarily GIVEN or PRESUPPOSED (terms to be defined below), or that the specified value 'an apartment' is necessarily the ASSERTED part or the FOCUS DOMAIN. This is different from other accounts of cleft constructions, as is apparent from the function-denoting terminology that is sometimes used to identify these elements.

1.3.4. DISAMBIGUATION. Some possible sentences are ambiguous as to whether they qualify as clefts or not, when taken out of context. Thus, more than just skeletal structures are necessary for determining whether a particular utterance is a cleft or not. Lambrecht (2001: 493) provides the following example of a sentence that can be ambiguous between a cleft and non-cleft reading, renumbered here as example 5a-b.

(5) a. It's the country that suits her best.

b. The COUNTRY suits her best.

If the context for 5a is a discussion of whether the person in question has benefited from a recent move from a rural to urban area, then the sentence can have an IT-cleft reading. With the cleft-reading, the utterance retains its propositional meaning when simplified, as in example 5b. In another possible context for example 5a, the preceding discourse involves a discussion of the effect a particular country (e.g., *Mexico* or *Thailand*) has on the individual in question. In such a context, *it* in 5a can be understood as a fully referring anaphoric pronoun, likely co-referent with an antecedent specifying the particular country under discussion, and the relative clause could be interpreted as restrictive with respect to the referent of *the country*. Thus, the rephrasing in example 5b does not capture the propositional semantics of a predicate non-cleft reading. This illustrates how the simplification test (or 'canonical sentence' test, proposed in Lambrecht, 2001) can be used together with consideration of the discourse context to disambiguate genuine clefts from constructions that only superficially resemble clefts.

To further illustrate the potential ambiguity of structures with reference to clefthood, consider examples 1a and c. In example 1a, *it* could be a fully referring pronoun, with an antecedent reference to a particular building or place. Similarly, *there* in example 1c could be a fully referring spatial deictic reference, perhaps accompanied by a pointing gesture or preceded by an antecedent reference to a particular place. In these cases, the simplified expression in example 3 does not fully capture the propositional semantics of the look-alike constructions.

Thus, a simplification test, along with assessment of the context, is useful for disambiguating structures that look like clefts and determining whether a structure has a cleft-reading or not. Additionally, the presence or absence of a clear antecedent for the initial element in potential IT clefts and THERE clefts can also be considered in determining clefthood.

1.3.5. PROSODY AND DEFINING CLEFT CONSTRUCTIONS. Although prosody is acknowledged as an important aspect of any construction, the term CLEFT CONSTRUCTION

is used here to refer primarily to the syntactic structure of the construction; prosody is considered separately. This is important because particular information structure properties have commonly been attributed to cleft constructions, without regard to prosody. Interestingly, the term FOCUS has been used to refer both to constituents expressing clefted elements in cleft constructions (see above), and to constituents bearing pitch accent, or some form of prosodic prominence (e.g., Selkirk 1995), so there is an apparent assumption that prosodic prominence and cleft structure influence information structure similarly. Additionally, it has often been assumed, and in some cases claimed, that cleft constructions occur primarily or solely with prosodic prominence on the clefted element (Chafe 1976, Creider 1979, Givón 2001). Since disregard for varying prosodic patterns can lead to overly simplistic analyses, the prosodic patterns exhibited in cleft constructions are empirically investigated here, and *a priori* assumptions about the prosody of cleft constructions are avoided.

1.3.6. SUMMARY. To summarize the interpretation of the term CLEFT CONSTRUCTION in the current paper, clefts are seen here as a set of constructions encompassing a number of subtypes, defined by the following properties:

- Biclausal syntax expressing a proposition that could be expressed using simpler syntax, without a change in truth conditions
- A variable-expressing open proposition, adjoined to a specified value for the open variable (neither of which is definitively presupposed or asserted)

Additionally, the initial elements *it* and *there* in IT clefts and THERE clefts are seen as non-referring or semantically empty. This definition leaves open the possibility that various prosodic patterns and information structure properties could be associated with cleft constructions; this possibility is explored and discussed here.

1.4. BACKGROUND: INFORMATION STRUCTURE

1.4.1. WHAT IS INFORMATION STRUCTURE? In the current study, the term INFORMATION STRUCTURE (Lambrecht 1994, due to Halliday 1967), or INFORMATION PACKAGING (Chafe 1976) is understood as a component of the linguistic system (i.e., the grammar) that systematically affects surface forms of utterances in interactive communication. It is proposed on the premise that interlocutors form mental

representations of the discourse as it progresses, and that speakers' linguistic choices are influenced by their sensitivity to not only their own mental models and communicative goals, but also to their hypotheses about their interlocutors' mental models of the discourse at hand. Information structure refers to the systematic coding choices, whether conscious or subconscious, that are influenced by speakers' communicative goals and their hypotheses about what hearers know and are thinking about at the time of utterance. Information structure is context-sensitive and is seen as influencing formal aspects of language such as article and pronoun use, placement of prosodic prominence, and potentially the use of particular constructions, such as clefts.

1.4.2. AN ILLUSTRATION OF INFORMATION STRUCTURE CODING. Pronoun use is among the least controversial examples that can be used to illustrate the influence of information structure on formal aspects of language. The speaker's assumptions about the hearer's knowledge and awareness clearly influence the choice between a lexical noun phrase and a pronoun. First, a definite pronoun cannot be felicitously used to express reference unless the speaker assumes the hearer has a mental representation for the unique referent of the pronominal expression. In other words, the referent must be assumed to be IDENTIFIABLE. Second, a definite pronoun cannot felicitously be used to express reference unless the speaker assumes the hearer is currently thinking about that referent, such that the mental representation for the referent is assumed to be ACTIVE in the hearer's mind. In contrast, if a cooperative speaker assumes that a referent is NON-ACTIVE at the time of utterance, he or she will likely use a full lexical noun phrase to express reference. Further, if the speaker assumes that the hearer does not yet have a stored mental representation for the intended referent, the expressed reference is likely to be coded as indefinite. Although pinpointing the direct mapping of information structure factors to surface forms seems elusive, even in the case of nominal reference in English, it is clear that information structure factors affect coding choices.

1.4.3. CONCEPTUAL DEFINITIONS. According to Lambrecht (1994), the basic elements of information structure are PRAGMATIC PRESUPPOSITION, PRAGMATIC ASSERTION, IDENTIFIABILITY, and ACTIVATION; the first two having to do with units of information, and the latter two having more to do with the status of referents in a hearer's knowledge

(identifiability) or consciousness (activation). Each of these concepts, as well as notions of FOCUS and CONTRASTIVENESS, is relevant to the current study of cleft constructions.

Based on Lambrecht 1994, PRAGMATIC PRESUPPOSITIONS are propositions that the speaker assumes the hearer knows or is ready to take for granted at the time of utterance, and PRAGMATIC ASSERTIONS are propositions that the speaker expects the hearer to know or take for granted as a result of hearing the utterance. Lambrecht's notion of FOCUS is built on these primitives of information structure; FOCUS is understood as the semantic component of a proposition (or utterance) whereby the pragmatic assertion differs from the presupposition. Since FOCUS is a semantic component, its syntactic expression is termed its FOCUS DOMAIN. This pragmatic, functional notion of FOCUS is adopted here, rather than a form-based definition (i.e., focus is not used to refer to, and is not assumed to directly correlate with, prosodic prominence or with a particular syntactic position).

Continuing in Lambrecht's framework, IDENTIFIABILITY relates to the speaker's assessment of whether the hearer has a stored mental representation for a referent as a unique entity, and ACTIVATION relates to the speaker's assessment of whether the hearer is consciously thinking about particular referents at the time of utterance. It is important to note that propositions can become referents, and can therefore be IDENTIFIABLE and ACTIVE.

The notion of FOCUS adopted here does not entail CONTRASTIVENESS. In fact, Lambrecht considers CONTRASTIVENESS an extragrammatical part of meaning that arises through implicature, rather than being expressed directly through formal coding in English. The current paper investigates CONTRASTIVENESS through hearer judgments, defining CONTRASTIVENESS as the degree to which interlocutors assume a referent is being talked about in opposition to something else, or as a member or part of a set of alternatives; a CONTRASTIVE referent is understood as a referent that is being talked about in opposition to something else, or as a member or part of a set of alternatives.

Some uses of information structure related terms in the current study were adapted for use with naïve native speaker-hearers in judgment tasks. These include CONTRASTIVENESS (as defined above), NEWNESS and GIVENNESS, whose definitions here are only loosely based on previous accounts (Lambrecht 1994, 2001; Chafe 1976). For

the purpose of the present study, NEWNESS is seen as a property associated with referents that are being brought up for the first time, and portions of utterances that express additional information. Thus, new referents or propositions are those that speakers assume hearers are not consciously thinking about at the time of utterance, new propositions are asserted propositions, and new information is focal information. Thus, NEW was used as a cover term for both NON-ACTIVE and ASSERTED. As the converse of newness, GIVENNESS is a property associated with referents and propositions that the speaker assumes the hearers already know and are thinking about. Thus, GIVEN was used as a cover term for ACTIVE and PRESUPPOSED; thus, NEW and GIVEN could be applied equally well to both referents and propositions. Operational definitions for these, as well as further explanation of their use with naïve hearers, will be presented in Chapter III.

1.4.4. SPEAKER-HEARER CONCERNS IN THE CURRENT STUDY. Although information structure has been described above as relating to the speaker's assumptions about the hearer's knowledge and consciousness at the time of utterance, the current study does not directly tap into the speaker's assumptions. Instead, the methodology developed for the current study involves two kinds of information gathered from participants simulating the hearer role: judgments about information structure of cleft components based on hearers' interpretations of the cleft constructions alone, and judgments about the perceived relevance of cleft components to their preceding contexts. Thus, hearer interpretations, rather than speaker assumptions, are the basis for much of the analysis presented here.[2]

1.4.5. PROSODIC CODING OF INFORMATION STRUCTURE. The importance of prosody in coding information structure is generally recognized (e.g., Chafe 1976, Lambrecht 1994, Givón 2001), and prosodic prominence has been said to affect information structure similarly to clefts. For example, some claim that a rise in pitch straightforwardly corresponds to 'new information' (e.g., Selkirk 1984), or to 'focus-marking' (e.g., Selkirk 1995). This may seem to be the case when considering decontextualized examples, as in

[2] The corpus studies reviewed in Chapter III are also apparently based on hearer interpretations and assumptions about the hearer's intentions (particularly those of the analyst/researcher). Methods for tapping into speakers' assumptions directly have yet to be developed.

6a (an IT cleft), 6b (a REVERSE WH cleft), and 6c (an apparently equivalent prosodically marked proposition).

(6) a. It's John who went to the store.
b. John is who went to the store.
c. JOHN went to the store.

While sentences like 6a and 6b have often been treated similarly (calling *John* 'focal'), Delin (1995) suggests that prosody and syntax have distinct functions in cleft constructions. Namely, she proposes that IT cleft syntax marks its relative clause content as *required* in the hearer's discourse model (regardless of whether it is assumed to be already known or active), while prosody serves as an indicator of whether the relative clause contains *shared knowledge* (i.e., information that speaker assumes is already present in the hearer's discourse model: identifiable and/or active). Thus, Delin proposes a strict division between logical presupposition and shared knowledge, which seems essentially a distinction between logical and pragmatic presupposition. To paraphrase Delin's proposal in the terminology of the current paper, Delin suggests that IT cleft syntax designates the open proposition as logically presupposed, and its prosody marks the open proposition as either pragmatically presupposed (with prosodic attenuation) or asserted (with prosodic prominence).

Taking a different approach, Lambrecht (1994, 2001), recognizes multiple pragmatic-marking functions of prosodic prominence, and thus suggests that clefts and prosody have some overlapping (focus-marking) functions. He sets forth the following guidelines for understanding prosodic prominence in English:

- An utterance must have at least one sentence accent/prosodic prominence to be informative
- Sentence accent/prosodic prominence signals the hearer to establish a focus or topic relation between a denotatum and a proposition

- Sentence accent/prosodic prominence generally marks the right boundary of a constituent expressing the focus domain or topic domain (General Phrasal Accent Principle, 1994:247)
- Following the General Phrasal Accent Principle, focus-coding prosodic prominence tends to fall on the last accentable constituent of the focus domain[3]
- A constituent is unaccented only if it expresses an entity or proposition that is both discourse-active and non-focal

The current paper approaches an analysis of prosody with Lambrecht's sentence accent principles in mind. Corpus findings about cleft prosodic patterns are reviewed in Chapter II, and the degree to which clefted elements and prosodic prominence correlate, as well as the influences of prosody on information structure judgments are investigated in the original corpus investigation reported here.

1.5. ORIGINALITY OF THE CURRENT STUDY. The current study of cleft constructions differs from past studies of clefts in several respects. First, it investigates cleft constructions in spoken American English, while past corpus studies of clefts have been based primarily on British, Scottish, and New Zealand varieties of English. Additionally, the current study represents a new approach to analyzing corpus data; while the linguists' analyses have been the sole source of information structure and other kinds of judgments about corpus data in past studies, the current study uses naïve native speaker-hearer judgments to describe information structure. Finally, the inclusion of existential THERE clefts among the investigated English cleft subtype is another way in which the current study differs from past corpus studies.

1.6. STRUCTURE OF THE THESIS. Chapter I has introduced the goals for the current investigation, and has laid a foundation for the thesis by providing the necessary background information about cleft constructions and information structure that form the basis for the thesis. Ways in which the current study differs from previous studies have also been discussed.

---

[3] A constituent is *accentable* if accent-placement there would not result in undesirable information structuring (Lambrecht 1994).

The goals laid out in the introduction are pursued in the remaining chapters. First, previous accounts of clefts and information structure are reviewed in Chapter II. The methodology for the original corpus investigation is described in detail in Chapter III, followed by the results of the investigation in Chapter IV. Finally, Chapter V concludes the thesis by discussing the implications of the original investigation, critiquing its methods, and summarizing what can currently be said about the extent to which English cleft constructions code contrastive focus and then suggesting some revisions to the traditional accounts of cleft constructions in English.

CHAPTER II

INFORMATION STRUCTURE AND CLEFT SUBTYPES

2.1. AN EXPANDING VIEW OF CLEFTS. The widespread claim that clefts are syntactic focusing devices likely stems from the centrality of IT clefts in traditional accounts of cleft function. When Jespersen coined the term *cleft sentence*, it was initially to describe only IT clefts (1937); this may have led to some assumptions about clefts as a group that are more appropriately considered specific to IT clefts as a subtype, or even to a subcategory of IT clefts. The fact that WH clefts are sometimes referred to PSEUDOCLEFTS contributes to the perceived centrality of IT clefts, as it suggests that WH clefts are somehow less-good examples of clefts. While IT clefts with specificational argument focus or contrastive focus remain central to discussions of clefts, other types of clefts, both in English and cross-linguistically, are gaining increasing recognition.

Lambrecht (2001) notes that while IT clefts and WH clefts feature most prominently in the literature, many more cleft types exist in English, as well as cross-linguistically. Lambrecht provides a framework for analyzing the information structure of cleft constructions, acknowledging multiple possibilities regarding their information structure and prosody. In addition, a number of corpus studies of clefts have provided insight into the information structure and prosodic patterns of several subtypes of cleft constructions in English (Calude 2008; Collins 2004, 2006; Delin 1989a, 1989b, 1990; Hedberg 1990; Oberlander and Delin 1996; Prince 1978; Weinert and Miller 1996, and others). With a particular interest in factors related to information structure (including prosody), this chapter reviews Lambrecht's framework and initial analysis of cleft information structure, as well as corpus-based findings about the similarities and differences between IT clefts and REVERSE WH clefts. As past corpus studies have not featured THERE clefts, only hypotheses for THERE clefts are reviewed here.

2.2. LAMBRECHT'S FUNCTIONAL TYPOLOGY OF CLEFT CONSTRUCTIONS. Lambrecht (2001) provides a framework for analyzing the information structure of cleft constructions, based on his theory of information structure and sentence form (1994). Lambrecht's use of PRAGMATIC PRESUPPOSITION and FOCUS were introduced in Chapter I, and they are employed and elaborated here with respect to cleft constructions.

Lambrecht (1994, 2001) identifies three major focus subtypes relative to structural domains: PREDICATE-FOCUS, ARGUMENT-FOCUS, and SENTENCE-FOCUS. Sentences with active and/or topical arguments, and whose focus domains are the expression of a predicate, are said to have predicate-focus; sentences with presupposed/active predicates and focal argument expressions are said to have argument-focus; and sentences with no presupposition and wholly new/asserted arguments and predicates are said to have sentence-focus. Of these three subtypes, Lambrecht identifies only argument-focus and sentence-focus as potentially characterizing cleft constructions. In discussing the cross-linguistic functional motivation for clefts, Lambrecht clearly states, 'Cleft constructions are focus-marking devices used to prevent unintended predicate-focus construal of a proposition' (2001: 489).

According to Lambrecht, the unmarked focus type in English is predicate-focus, to which he attributes a 'topic-comment' function.[4] This focus subtype would typically be expressed in a simple declarative form, with prosodic prominence on the clause-final constituent. This is exemplified in 7b, where the speaker is responding to the question in 7a, in which *caffeine* has clearly been established as a referent of current interest in the discourse (i.e., TOPICAL).[5]

(7) a. Does CAFFEINE bother you?
b. Caffeine makes me CRAZY.

In sentences with predicate-focus, the topicality of an argument (usually the syntactic subject) is presupposed, and the predication that is asserted about that topical argument

[4] The term UNMARKED is used here in a distributional sense, and also in the sense of 'pragmatically neutral'.

[5] Small capital letters indicate the locus of prosodic prominence.

(typically expressed in the verb phrase) is the focus domain. In other words, the referent of an argument is understood to be active and of current interest in the discourse, and a focal predication expresses new information about the given subject.

In Lambrecht's analysis, cleft constructions do not exhibit predicate-focus; instead, they are marked focus constructions that express either ARGUMENT-FOCUS or SENTENCE-FOCUS. That is, clefts can either have a presupposed open proposition and a focal argument expressed in the clefted element, or they can express a proposition that is entirely asserted. Thus, in Lambrecht's view, it would be surprising to find any cleft constructions with active and/or topical clefted elements and focal open propositions. This view is challenged to some degree in the current study, particularly in the case of REVERSE WH clefts.

As noted in Chapter I, associating cleft constructions with argument-focus is most compatible with traditional analyses of clefts as (contrastive) focusing devices. Following Declerck (1988), Lambrecht calls the function of argument-focus clefts *specificational*.[6] Lambrecht further subdivides specificational argument-focus clefts into those that are EXHAUSTIVE and NON-EXHAUSTIVE. Lambrecht claims that the best-known English clefts (IT, WH, and REVERSE WH clefts) are characterized as EXHAUSTIVE, such that the referent(s) expressed in the clefted element are interpreted as the full set of contextually relevant value(s) satisfying the presupposed open proposition. Exhaustive argument-focus clefts are exemplified in 8b and 8c, expressed as felicitous responses to 8a, where the open proposition is 'x is making [the second speaker] (act) crazy'.

(8) a. You're acting kind of crazy. Have you had too much sugar lately?
b. Actually it's CAFFEINE that's been making me crazy.
c. Actually, CAFFEINE is what's been making me crazy.
d. ?Yes, and it's also CAFFEINE that's been making me crazy.
e. ?Yes, and CAFFEINE is also what's been making me crazy.

[6] Thus, both Lambrecht's (2001) and Declerck's (1988) use of *specificational* entails that the open proposition, or predicate, is presupposed. This contrasts with the use of *specificational structure* in Chapter I of this thesis.

The exhaustive interpretation of 8b and 8c indicate that the value of x is 'only caffeine', and this accounts for the acceptability of 8b and 8c, in contrast with the infelicity of 8d and 8e, where 'caffeine' is not the only relevant value for the open variable x. If 8c seems somewhat acceptable, perhaps the exhaustive interpretation of IT clefts is implied, but can be contradicted, resulting in an anomalous, but not completely ungrammatical, utterance.

NON-EXHAUSTIVE ARGUMENT-FOCUS clefts, on the other hand, are interpretable as expressing one or more members of a set of possible values in the clefted element. Lambrecht notes that THERE clefts can have a non-exhaustive argument-focus information structure, as in example 9b (as a response to 9a), where x = '{sugar, caffeine...}'. Again, the exhaustive interpretation of IT-clefts and REVERSE WH clefts makes 9c and d less felicitous, if not ungrammatical.

(9) a. You're acting kind of crazy. Are you on something?
b. Well, there's CAFFEINE that's making me crazy, and also a lack of sleep.
c. ?Well, it's CAFFEINE that's making me crazy, and also a lack of sleep.
d. ?Well, CAFFEINE is what's been making me crazy, and also a lack of sleep.

A non-exhaustive interpretation of 9b seems most natural and acceptable if one imagines a non-final listing (i.e., rising) intonation on the clefted element.

While exhaustivity has little to do with newness or givenness, the presence or lack of implied exhaustivity likely contributes to the degree of contrastiveness expressed by a cleft construction. If contrastiveness is understood as a scalar property, its strength correlating negatively with the size of the relevant set, exhaustive argument-focus clefts would likely be seen as more contrastive than clefts with the potential for expressing non-exhaustive listing via the clefted element. Thus, THERE clefts can be expected to contain less contrastive clefted elements than IT clefts and REVERSE WH clefts, since THERE clefts can more readily express non-exhaustive listing.

The second marked focus subtype Lambrecht describes as available to cleft constructions is sentence-focus. In Lambrecht's terms, sentence-focus constructions have a presentational (introducing entities) or eventive (introducing situations) function. While Prince (1978), Hedberg (1990) and others have observed IT clefts with apparent sentence-focus (i.e., where the cleft components are 'all-new'; terming them *informative-presupposition IT clefts* or *comment-clause IT* clefts), Lambrecht suggests that IT clefts are primarily used for specificational argument-focus.[7] In contrast, among the examples Lambrecht provides for sentence-focus clefts is a THERE cleft with prosodic prominence in both cleft components (see example 10, which is Lambrecht's example 64a).

(10) There is a LINGUIST who wants to explain CLEFTS.

Example 10 could be uttered in a context where nothing is presupposed, where, 'a linguist wants to explain clefts' is entirely asserted, and thus the focus is the entire proposition.

In additional commentary regarding THERE clefts as sentence-focus clefts, Lambrecht suggests that the pragmatic function of THERE clefts is to mark the clefted element as newly introduced in the discourse. This suggestion is strongly supported by observations regarding THERE clefts in the SBCSAE, the vast majority of which have non-identifiable[8] clefted elements that are frequently coded as indefinite. This suggestion is also supported empirically in the current study (cf. Chapter IV).

---

[7] Lambrecht (2001) proposes that there is no need to posit two separate types of IT clefts. Instead, Lambrecht essentially claims that apparent sentence-focus IT clefts should be analyzed as constructions based on the argument-focus type of IT clefts, requiring pragmatic accommodation. That is, because IT clefts are prototypically argument focus constructions, when there is new information in the open proposition, the speaker assumes the hearer will readily accommodate that new information as if it were given.

[8] Each cleft token in the original corpus study (presented in Chapters III-V) was reviewed in context, and each was categorized regarding identifiability by the researcher, in consultation with Doris L. Payne. THERE clefts were overwhelmingly non-identifiable, while other types of clefts did not appear to be constrained in this way.

To summarize, Lambrecht discusses English IT clefts and REVERSE WH clefts as argument-focus constructions coding exhaustive specification of a variable. In contrast to IT clefts and REVERSE WH clefts, Lambrecht claims that existential THERE clefts may code either non-exhaustive specificational argument-focus or presentational-eventive sentence-focus. Thus, THERE cleft information structures stand out among the others. In Lambrecht's analysis, clefts are marked focus constructions, and are therefore not expected to exhibit predicate-focus.

2.3. PRELIMINARY PREDICTIONS. Focus structure predictions for cleft subtypes, based primarily on Lambrecht's framework (2001), are outlined in Table 3.

	PREDICATE-FOCUS	EXHAUSTIVE SPECIFICATIONAL ARGUMENT-FOCUS	NON-EXHAUSTIVE SPECIFICATIONAL ARGUMENT-FOCUS	PRESENTATIONAL/ EVENTIVE SENTENCE-FOCUS
IT clefts	NO	YES	NO	YES[9]
REVERSE WH clefts	NO	YES	NO	NO
THERE clefts	NO	NO	YES	YES

TABLE 3. Focus structure predictions for cleft subtypes under study

Unmarked predicate-focus is not expected, argument-focus is expected to some extent for all three cleft types, and sentence-focus is expected for some THERE clefts and some IT clefts (though Lambrecht does not clearly predict sentence-focus for IT clefts; see note 9).

Regarding predictions about the interaction of prosody and cleft constructions, focus domains are expected to bear prosodic prominence in English. However, due to the multiplicity of information structure functions that prosodic prominence can potentially indicate (in Lambrecht 1994 and 2001, focus domain, activation, reactivation and topic-ratification are among the functions that prosodic prominence may indicate), prosodic prominence is not expected to fall solely on the focus domain in all cases. Since clefts

[9] In Lambrecht's analysis, IT-cleft open propositions are at least *knowledge-presupposed*, though they may not be assumed to be active or topical. In other words, the speaker assumes the hearer either knows the open proposition or is ready to take it for granted via accommodation. Though this is something of a gray area for Lambrecht (2001), who wants to avoid positing multiple types of IT clefts based on differing information structure properties (as mentioned in note 7), Prince (1978) and others clearly report the occurrence of IT-clefts with all new cleft components.

are expected to always code either argument-focus or sentence-focus, and since clefted elements constitute all or part of the focus domain in each of these subtypes, prosodic prominence is expected to always fall on at least the clefted element in cleft constructions (though it may also fall elsewhere in the cleft construction).

In sum, based on Lambrecht's discussion, one would expect to find the following regarding the data analyzed in the current paper:

- IT clefts and REVERSE WH clefts code exhaustive specificational argument-focus; the clefted element is expected to be judged as new and/or contrastive, and the open proposition is expected to be judged given. IT clefts and REVERSE WH clefts are expected to have prosodically prominent clefted elements.
- Some THERE clefts code non-exhaustive specificational argument-focus; in these cases, the clefted element is expected to be judged as new and/or contrastive (but less decisively contrastive than IT clefts) and the open proposition is expected to be judged as given. These THERE clefts are expected to have prosodically prominent clefted elements.
- Some THERE clefts code presentational/eventive sentence-focus, where nothing is presupposed; both the clefted element and the open proposition are expected to be judged as new. For these THERE clefts, loci of prosodic prominence are expected in both cleft components.

Aspects of these predictions will be evaluated with respect to past corpus studies (in §2.4), as well as the original corpus study presented in Chapters III-V.

2.4. CORPUS FINDINGS. Prince (1978) first pointed out striking differences between cleft subtypes in her study of IT clefts and WH clefts. Prince's recognition of the distinctiveness of these two subtypes has been followed by numerous corpus studies of cleft constructions in authentic data, often discussing the observed differences among subtypes. A number of these corpus findings about cleft constructions are reviewed in this section. Some general conclusions about clefts are reviewed first, followed by particular findings about IT clefts and REVERSE WH clefts. Since previous corpus studies of clefts have not included THERE clefts as objects of study, these are not included in the

review of corpus findings (however, Lambrecht's 2001 observations about THERE clefts were discussed above).

2.4.1. CLEFTS IN GENERAL. The main finding from corpus studies about clefts as a general category is that they are a much more heterogeneous group than assumed in traditional accounts; cleft subtypes differ in terms of frequency, as well as prosodic, syntactic, semantic, and information structure properties (Prince 1978, Hedberg 1990, Delin and Oberlander 1995, Weinert and Miller 1996, etc.). REVERSE WH clefts are found to be the most frequent cleft type, by far (Calude 2008, Collins 2006, Weinert and Miller 1996), and IT clefts are found to be the least frequent in both speech and writing when compared with WH clefts and REVERSE WH clefts, considering a broad range of possible clefted elements (Weinert and Miller 1996, Collins 2006). Thus, the relative amount of attention in the literature devoted to IT clefts is not reflective of their frequency in the natural language. Additionally, the extrapolation of IT cleft properties to clefts as a general category is not warranted.

While differences among cleft sub-types abound, as will be demonstrated below, some characteristics appear to be shared among various cleft subtypes. Beyond defining features of clefts, all cleft types included in previous corpus studies have attested full noun phrases as clefted elements. Those corpus studies that have examined prosody have found that clefts, generally, rarely bear sole prosodic prominence on the clefted element, although prosodic patterns differ across subtypes (Weinert and Miller 1996, Delin 1990). Additionally, clefts are generally found to occur more frequently in speech than in writing (Breivik 1986).

As for what can be gathered about the information structure of clefts as a group, the varying methods and definitions employed in discussing information structure in various studies make it difficult to draw firm conclusions. Results regarding newness and givenness, in particular, depend largely on the breadth or narrowness of their applied definitions. Working with a narrow definition of 'given' (contextually explicit), Delin (1990) finds weak support, with counter-examples, for the hypothesis that clefts separate new information from given information, with new information as the clefted element and given information in the open proposition. Delin also concludes that all cleft

constructions must minimally express one unit of new information, whether in the clefted element or open proposition; however, the applicability of this conclusion for all clefts is weakened by Collins' (2004) claim that only a minority of REVERSE WH clefts express any new information. Collins is apparently applying a broader definition of 'given' than Delin.

Delin (1990) also points out that clefts, as a group, cannot present information of 'simple narrative consequence' in their relative clauses (i.e., as open propositions). This fact relates to the logical presupposition-conveying property of relative clauses: that is, the open proposition expressed in the relative clause, whether pragmatically presupposed or not, must be true in order for the truth or falsity of the cleft utterance as a whole to be evaluated. It may be that this logically presupposed nature enables the open proposition cleft component to readily express active and/or pragmatically presupposed (i.e., given) information. In Delin's view, presuppositionality is the key to cleft function, and the tendency for the clefted element to express new and/or focal information is an epiphenomenon of the two claims mentioned above, paraphrased here in terminology of the current study: (a) that all clefts minimally express new information in one cleft component, and (b) the tendency for logically presupposed open propositions to express information that is also pragmatically presupposed and/or active.

Findings about the particular syntactic, prosodic, information structure, and other related characteristics observed for IT clefts and REVERSE WH clefts are reviewed in the upcoming sections.

2.4.2. IT CLEFTS. Since IT clefts have long been central to discussions of cleft constructions, numerous corpus studies report observations about the syntactic, prosodic and information structure patterns of this cleft subtype (Collins 2006; Delin 1989a, 1990; Hedberg 1990; Weinert and Miller 1996; Prince 1978; among others). While full noun phrases are the most frequent clefted element for IT clefts, examples with prepositional phrases and adverbials as clefted elements are also attested (Prince 1978, Weinert and Miller 1996). Examples 11a and b from Prince (1978) are provided to illustrate clefted prepositional phrases and adverbials, respectively (Prince's examples 7a and 6a).

(11) a. It is AGAINST PARDONING THESE that many protest.

b. It was THEN that I became a young revolutionary.

IT clefts have been observed to occur most frequently with prosodic prominence on the clefted element (Collins 2006, Delin 1990, Weinert and Miller 1996). Of those with prominence on the clefted element, the majority bear prosodic prominence in both cleft components (Delin 1990, Weinert and Miller 1996). Though prosodic prominence is often assumed to indicate informativity, Collins (2006) finds that prosodic prominence in IT clefts does not always correlate straightforwardly with newness or givenness.

Prince (1978) first observed that IT clefts exhibit two major information structure patterns: first, the expected pattern with new and/or focal information in the clefted element and presupposed or given information in the open proposition (her STRESS-FOCUS type); second, with new and/or focal information in both the clefted element and the open proposition (her INFORMATIVE-PRESUPPOSITION type). One of the examples of stress-focus IT clefts that Prince provides is provided in abbreviated form in example 12a (Prince's 40a), and one of Prince's informative-presupposition examples is provided in 12b (Prince's 41a; ## signifies discourse-initial).

(12) a. ...When you last saw me with anyone, *it was Barbara I was with.*

b. ## It was just about 50 years ago that Henry Ford gave us the weekend.

Both Delin (1990) and Collins (2006) find both of these patterns in their corpora; however, there is, again, a quantitative discrepancy regarding the proportion of IT clefts associated with each of these information structure patterns: Collins finds that about half of IT clefts contain new information in the open proposition, while Delin finds that about ninety percent of IT clefts contain some new information in the open proposition, and 72 percent display a distinctive old before new (i.e., old information in the clefted element, and new information in the open proposition) information structure pattern. Collins (2006) reports that 27.8 percent of IT clefts had new information in both components, while 21.4 percent had new information in the open proposition, with a 'non-focal' clefted element ('focus' seems implicitly related to 'newness' for Collins). Methods for

determining such patterns in these studies remain somewhat unclear and likely vary greatly between the two studies.

Also of interest is Weinert and Miller's (1996) finding that IT clefts are the preferred cleft type for expressing contrast. They found that while less than 3 percent of REVERSE WH clefts and WH clefts expressed explicit contrast, a full 36 percent of IT clefts expressed explicit contrast. Their example of an explicitly contrast-expressing IT cleft is presented in example 13 (Weinert and Miller's example C39).

(13) We're we're after everything I mean not not not the phonetics because that's fairly well known anyway em *it's the SYNtax we're after.*

To summarize, a portion of IT clefts are found to pattern according to the traditional expectations: with specificational argument-focus and prosodic prominence on the clefted element. Also attested are apparent sentence-focus IT clefts (i.e., those with new information in both cleft components), and IT clefts with prosodic prominence on both cleft components.

2.4.3. REVERSE WH CLEFTS. As mentioned above, REVERSE WH clefts are apparently the most frequent cleft type in spoken English, usually occurring with a clefted demonstrative, most often *that* (Calude 2008; Collins 2004, 2006; Weinert and Miller 1996). Some examples of typical REVERSE WH clefts from the Santa Barbara Corpus of Spoken American English are presented in 14a-c, and an example of a less typical REVERSE WH cleft, with a clefted full noun phrase, is presented in 15.[10]

(14) a. S1: Llenar is to fill?
S2: Yeah.
S1: *That's what I thought.*

[10] The Talkbank (MacWhinney 2007) version Santa Barbara Corpus of Spoken American English (DuBois et al., 2000; Du Bois et al., 2003; DuBois and Englebretson, 2004; DuBois and Englebretson 2005) is the primary data source for the original investigation reported Chapter III-V.

b. I said, if you invite them up there for a party, they're gonna assume that they are staying with you, right? And she goes, no they're not, And I said, well *that's what they're gonna think.*

c. I'm sure a lot of women throw themselves at em, so *that's what they expect from women.*

(15) And so it was a lot cheaper than if she'd bought it… cause I only charged her, a buck, a buck a ton, which is what I paid for it. But then I had to pay a lot more to have it hauled in… *The hauling is what costs so much.*

Although the use of a pronoun is usually licensed only if the pronoun's referent is expected to be identifiable and active, Calude (2008) observes that it is often difficult to pinpoint a clear referent for *that* in REVERSE WH clefts (which she terms DEMONSTRATIVE CLEFTS to distinguish them from REVERSE WH clefts with other types of clefted elements). Calude and others note that the clefted demonstrative often refers anaphorically to a proposition or utterance in the discourse context, and that REVERSE WH clefts tend to serve a 'summative' function, bringing an episode in the discourse to a close (Collins 1991, 2004, 2006; Oberlander and Delin 1996; Weinert and Miller 1996). This discourse-deictic function of demonstrative clefts is exemplified in 14a and b, above, and possibly in 14c.

Though results for prosody and information structure of REVERSE WH clefts are somewhat mixed, they are consistently different from the patterns reported for other cleft types. The characteristic prosodic and information structure properties observed for REVERSE WH clefts may follow from the typically pronominal character of the clefted element.

Compared with other cleft types, clefted elements in REVERSE WH clefts are less likely to bear prosodic prominence (Collins 2006, Delin 1990, and Oberlander and Delin 1996 report that less than 10 percent bear nuclear accent on the clefted element; Weinert and Miller 1996 report a much higher figure: just under 40 percent). However, a majority of open propositions in REVERSE WH clefts are reported as bearing prosodic prominence (Collins 1991, Delin 1990, Oberlander and Delin 1996, Weinert and Miller 1996).

REVERSE WH clefts are also reported to occur under a single intonation contour most of the time (Collins 2006, Delin 1989a).

Probably due to numerous methodological factors, information structure results for REVERSE WH clefts are fairly mixed: Collins (2004, 2006) reports that the majority of REVERSE WH clefts have given information in both cleft components and that they are low in informativity, while Delin (1990) reports that a majority of REVERSE WH clefts have new information in their clefted elements. These findings are based on very different foundational definitions for terms like NEW and GIVEN. Taken together, the general finding about information structure of REVERSE WH clefts that sets them apart from other cleft types is that the open proposition is more likely than the clefted element to express new and/or focal information. Thus, the focus subtype that most closely fits the observations about REVERSE WH clefts is the predicate-focus type, which was not predicted for cleft constructions at all. Yet, if Collins (2004, 2006) is correct in his claim that REVERSE WH clefts tend to express given information in both cleft components, and that they are low in informativity, it is unclear which focus subtype they exhibit. Perhaps the assertion associated with some REVERSE WH clefts is on the level of discourse structuring, or speech act. If REVERSE WH clefts often serve a 'summative function', perhaps speakers use them to indirectly assert something like, 'Let's end this episode,' or, 'I have nothing more to say about this topic.' Calude (2008) suggests that demonstrative clefts are used as a means of taking the floor without causing an interlocutor to lose face. In any case, the findings from past corpus studies suggest that REVERSE WH clefts do not neatly fit into Lambrecht's functional taxonomy of clefts.

2.4.4. SUMMARY. Given the corpus findings reviewed above, IT clefts and REVERSE WH clefts appear to have very different profiles. IT clefts typically have clefted lexical noun phrases, prosodically prominent clefted elements and open propositions, and exhibit both argument-focus and sentence-focus; REVERSE WH clefts, on the other hand, usually have a clefted demonstrative pronoun that is not prosodically prominent, have a prosodically prominent open proposition, and appear to exhibit predicate-focus. Additionally, IT clefts have been described as the preferred cleft for expressing contrast, and REVERSE WH clefts have been said to frequently bring closure to episodes in

discourse. The varying syntactic, prosodic, information structure and other properties of the two cleft subtypes, based on the corpus studies reviewed above, are summarized in Table 4.

CLEFT TYPE	MOST FREQUENT CLEFTED CONSTITUENT	PROSODIC PROMINENCE IN CLEFTED ELEMENT?	PROSODIC PROMINENCE IN OPEN PROPOSITION?	INFORMATION STRUCTURE PATTERNS	OTHER CHARACTERISTICS
IT clefts	full lexical NP	in majority of tokens	in majority of tokens	2 main patterns: 1) new/focal CE, given/presupposed OP 2) new/focal CE, new/focal OP	-Preferred cleft for expressing contrast -Low frequency relative to REVERSE WH clefts
REVERSE WH clefts	*that*	in minority of tokens	in majority of tokens	-CE is usually active, sometimes topical -mixed results regarding information structure of OP; agreed upon finding is that OP is more likely to express new information than CE	-clefted *that/this* often discourse-deictic -often serves 'summative' function -High frequency relative to IT clefts

TABLE 4. Profiles of IT clefts and REVERSE WH clefts based on corpus findings; CE represents 'clefted element(s)' and OP represents 'open proposition(s)'

The predictions based on Lambrecht's framework for analyzing clefts are supported by these corpus findings to a limited extent. The prediction that some IT clefts would exhibit argument-focus (pattern 1, under 'Information structure patterns' in Table 4), and some IT clefts would exhibit sentence-focus (pattern 2, under 'Information structure patterns' in Table 4) is born out in the observations from past corpus studies. However, corpus findings about REVERSE WH clefts reveal an apparent gap in Lambrecht's framework for analyzing cleft information structure, as the predicate-focus subtype seems most applicable to REVERSE WH clefts.

CHAPTER III

METHODOLOGY

3.1. OVERVIEW OF METHODOLOGY. This chapter describes the methodology employed in an empirical investigation of several sub-types of cleft constructions as they occur in the Talkbank (MacWhinney 2007) version of the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000, Du Bois et al. 2003, Du Bois and Englebretson 2004, Du Bois and Englebretson 2005). Henceforth, the corpus will be referred to as the SBCSAE. A series of corpus searches were conducted, and empirical methods for gathering native speaker input about prosody and information structure of cleft constructions were developed and employed.

First, the corpus search and data selection processes are described. Then, several naïve speaker-hearer judgment tasks are described; general methods for the tasks are presented first, followed by task-specific methods for evaluating clefts with respect to prosodic prominence, newness/givenness, contrastiveness, and relevance to their preceding contexts. Next, methods for hypothesis-testing are made explicit, and hypotheses and predictions are summarized in a final section.

3.2. DATA SELECTION

3.2.1. THE CORPUS. The SBCSAE was chosen as the corpus for this study for a number of reasons. First, a spoken corpus was chosen, in part, because clefts have been found to occur more frequently in spoken English than in written English (e.g., Breivik 1986). Additionally, early searches of corpora with more formal and/or planned speech (such as radio/broadcast speech) did not yield high numbers of cleft constructions, perhaps because clefts are more likely to occur in highly interactive and spontaneous speaking contexts (such as many of those represented in the SBCSAE). A spoken corpus was also necessary for investigating the prosodic aspects of cleft constructions. The availability of audio files, in addition to searchable transcripts, for the Talkbank version

was also necessary for investigating the prosodic aspects of cleft constructions. The availability of audio files, in addition to searchable transcripts, for the Talkbank version of the SBCSAE made it amenable to prosodic analysis. Finally, as the SBCSAE represents speakers of varieties of American English from a wide variety of backgrounds, and in a reasonably broad range of contexts, it is likely a reasonably reliable representation of Spoken American English.

3.2.2. SYNTACTIC CRITERIA FOR TARGET CONSTRUCTIONS. A subset of cleft constructions was selected for study: only IT clefts, REVERSE WH clefts, and THERE clefts with lexical noun phrases as clefted elements,[11] and with overt relativizers (limited to *that, what,* and *who*) were extracted from the SBCSAE for use in the judgment tasks. Table 5 provides the skeletal structures (more constrained than those presented in Chapter I) of these three cleft subtypes, along with examples from the SBCSAE to illustrate each cleft type.

	STRUCTURE	SBCSAE EXAMPLE
IT cleft	IT BE $X_{NP.lex}$ $REL_{that/who}$ Y	*It was Barbara that has some seeds.*
REVERSE WH cleft	$X_{NP.lex}$ BE $REL_{what/who}$ Y	*That flood's what left that white line around the lake out there.*
THERE cleft	THERE BE $X_{NP.lex}$ $REL_{that/who}$ Y	*There're a lotta women out there who apparently don't believe in using condoms.*

TABLE 5. Structures of subtypes of clefts under study, with examples[12]

Searches for words, word-strings, and within-utterance combinations were performed on transcripts of the Talkbank version of the SBCSAE in .cha files, using CLAN language analysis software (Spektor 2008). A list of the attempted searches for the above

[11] For the purposes of the present study, the category LEXICAL NOUN PHRASES includes modified or quantified pronouns, such as the expressions: *different ones* and *three or four of 'em.* Thus, clefts with bare pronouns (with minimal lexical specification) as clefted elements, such as *this*, *that*, *him*, or *something*, are excluded from the analysis. Constructions with bare indefinite pronouns (*something, one, others,* etc.) were originally included in the hearer judgment tasks, but they are left out of the results and analysis, since bare definite pronouns were excluded altogether.

[12] In Table 5, a number of symbols are used: X represents the expression of the clefted element, Y represents the expression of the open proposition, REL stands for 'relativizer', NP stands for 'noun phrase', and LEX stands for 'lexical'.

structures is provided in Appendix A. For each search, CLAN generated a list of utterances in the SBCSAE meeting the search criteria. Each utterance in the search results was then compared to the relevant structural skeleton(s) in Table 5. When results matching one or more of the skeletal structures were found, the simplification test (cf. Chapter II) was implemented, along with context-analysis to distinguish clefts from look-alike (but non-cleft) constructions. A search result was included in the hearer judgment tasks (see below) and analysis if and only if it matched one or more skeletal structure(s) in Table 5 and could be straightforwardly expressed in a simplified declarative form, without changing the truth conditions of the utterance. To illustrate, the examples in Table 5 above were considered cleft tokens since they fit the prescribed structures, initial *it* and *there* in the IT cleft and THERE cleft, did not have clear antecedents, and they could be simplified, as in examples 16a-c below, without changing the propositional semantics.

(16) a. Barbara has some seeds.
b. That flood left that white line around the lake out there.
c. A lot of women out there apparently don't believe in using condoms.

There were a few additional reasons for excluding tokens from the judgment tasks and analysis. Clefts under the scope of negation (e.g., *It's not like there's...*) and clefts that were complements of *if* in conditionals (e.g., *If it's... then...*) were excluded. Finally, one IT cleft token was excluded because its clefted element, a proper name, was masked from the recording (to protect the referent's identity), making it unintelligible and unsuitable for the hearer judgment tasks.

3.2.3. DATA SET RATIONALE. The particular subset of cleft constructions under study was chosen for a number of reasons. For IT clefts, REVERSE WH clefts, and THERE clefts, the order of the clefted element (X) and the open proposition (Y) is kept constant across these three types, in contrast with WH clefts, which were excluded. Additionally, the data set was constrained to only those clefts with lexical noun phrases (including quantified or modified pronouns; see note 10) as clefted elements primarily because this is the main area of overlap among syntactic subtypes regarding possible clefted elements (cf. §2.4.1).

These three cleft types were also chosen because each type as salient for the discussion of clefts in at least one specific way. First, IT clefts were selected for study because of their centrality to past discussions of clefts in linguistics, despite their relative infrequency. REVERSE WH clefts, on the other hand, have been found to be the most frequent cleft type, and have also been found to differ significantly in informativity and discourse function from other cleft types, as reviewed in Chapter II. Finally, THERE clefts were selected for study primarily because of their lack of representation in the literature on English clefts (although similar constructions in other languages have uncontroversially been described as cleft constructions, as pointed out in Lambrecht 2001). The fact that THERE clefts have not been explored in past corpus studies on clefts, along with their relative frequency in the SBCSAE, made them particularly ripe for investigation in the current study.

3.2.4. USE OF TARGET TOKENS IN TASKS[13]. As a result of the processes described above, 31 target THERE cleft tokens, 13 target IT cleft tokens, and 9 target REVERSE WH cleft tokens were used in the cleft-based hearer judgment tasks described below (for judgments relating to newness/givenness and constrastiveness, and prosodic emphasis). A complete list of these target tokens is provided in Appendix B. Fewer tokens (30 target THERE clefts, 11 target IT clefts, and 9 target REVERSE WH clefts) were used in the context-based relevance rating task, due to some problems relating to discourse content.[14] Due to the selection methods, the proportions of cleft subtypes used in the judgment tasks roughly represent the relative distributions of these subtypes in natural interactional language, with THERE clefts being much more frequent than either IT clefts or REVERSE WH clefts.

3.3. NATIVE SPEAKER-HEARER JUDGMENT TASKS. Judgments made by participants simulating the hearer role informed the main prosodic and information structure analyses.

---

[13] A complete list of the cleft tokens included in the analysis is available in Appendix B.
[14] During the piloting phase of the project, volunteers provided input and feedback regarding many aspects of the task. Multiple volunteers reported feeling off-put and less cooperative because of the high number of clips containing strongly religious language. Thus, three of the most strongly religious contexts were removed from the tasks, including one THERE cleft and two IT clefts.

For all tasks, subjects were presented with auditory stimuli (authentic speech segments) and accompanying visual (orthographic) stimuli. All stimuli related in some way to target cleft constructions occurring in spoken discourse. In four different tasks, subjects were instructed to listen to the speech segments, and then to indicate their perceptions about how the orthographic stimuli relate to the speech segments by pressing keys or clicking buttons on a computer screen. The four different task types correspond to judgments about four different aspects of the constructions, including locus/loci of prosodic prominence, newness/givenness of cleft components (i.e., clefted elements and open propositions), contrastiveness of the clefted element, and perceived relevance of cleft components to preceding contexts.

3.3.1. SUBJECTS. Subjects were undergraduates at the University of Oregon enrolled in a psychology course. Subjects participated to partially satisfy a course requirement. All participants were adult native English speakers. No other selection criteria were used in recruiting participants.

3.3.2. AUDITORY STIMULI. All auditory stimuli, including non-cleft practice and training examples and cleft-related test stimuli, were extracted from SBCSAE media files (available through Talkbank, MacWhinney 2007). Two different types of auditory stimuli were presented in the various perception tasks: CLEFT UTTERANCES, and CLEFT CONTEXTS.

Cleft utterances were the speech segments containing only the target constructions (see Appendix B for a complete list). Some guidelines followed in extracting cleft utterances from the speech stream must be noted. First, if the cleft construction was embedded (see example 17a), then the portion preceding the cleft in the utterance was excluded from the segment used as a stimulus. Additionally, if the cleft construction was preceded by a conjunction (see examples 17b and 17c), the conjunction was excluded from the segment used as a stimulus.

(17) a. Well I think [it was Barbara that has some seeds].
b. But [that flood's what left that white line around the lake out there].

c. Cause [there're a lotta women out there who apparently don't believe in using condoms].

Thus, in examples 17a-c, respectively, *Well I think*, *But*, and *Cause* were not included in the cleft utterance audio files used as stimuli; only the bracketed portions of 17a-c were included. The endpoint of the extracted cleft utterance was also determined on syntactic grounds; a cleft utterance segment ended where its final relative clause ended. In this way, subjects presented with cleft utterances as auditory stimuli heard only the cleft construction, extended to its apparent syntactic boundaries, and not necessarily the full compound or complex of clauses within which the cleft construction occurred.

Cleft contexts, the second type of auditory stimuli, consisted of a portion of the discourse context preceding the target construction; cleft contexts led up to, but did not include the target constructions. Cleft context audio segments varied in length from 6 seconds to nearly 2 minutes (averaging about 30 seconds). Each cleft context began at a seemingly natural starting point for the episode or section of discourse leading up to the target construction, and ended with the last complete phrase before the target construction. To illustrate, examples 18a-c are transcriptions[15] of the contexts presented for the IT cleft, REVERSE WH cleft, and THERE cleft, respectively, that were presented in Table 5, as well as in example 17a-c.

(18) a. Angel: Can I grow some basil? From seed?
Sam: Yes, that's how, I don't have any this year, but I've grown it other years.
Angel: Well--
Sam: There's no problem. It will not take any frost. Soon as the--
Angel: Yeah.
Sam: A little bit of frost, it's gone.
Angel: I learned that, one time.
Sam: I learned it just went down.

[15] Modified from the original transcripts to more closely resemble conventional spelling and punctuation, for readability. There remain some differences between conventional punctuation and that used here: commas (,) indicate minor intonation unit boundaries, periods (.) indicate major intonation boundaries, and question marks (?) indicate phrase-final rising intonation.

Angel: The hard way. Hm? Uh huh.
Sam: Yes?

b. What happened that year, is they had a very heavy snow pack in the Rocky Mountains, and that was followed by rain and hot weather, which created a flooding condition. We had more water coming down the Colorado river system, than they could put through the power plant. Every river in the Colorado river system was full to overflowing that year. On July second, water overtopped our spillways out here. Six days later, it was going over four and a half feet high, and that went on for sixty-six days. At the end of that time, there was no damage to the dam, or to the power plant, but further on downstream where people had built in a flood plain, they did have thirty million dollars worth of damage, but no loss of life.

c. Miles: I didn't know this, but apparently in Brazil, they have a very very high AIDS infection rate.
Harol: Really?
Jamie: Oh.
Pete: Oh.
Jamie: Probably.
Miles: It's supposed to be very high,
Pete: Mm.
Harol: You mean like-
Jamie: Very unmonitored too.
Harol: higher than the United States, higher than around here?
Miles: That's what I've heard.
Pete: Hm.
Jamie: I wouldn't be surprised.
Harol: Have you heard these figures, that like, um, it's something like forty percent of males in the Bay Area are supposed to be infected?
Miles: Well, last July
Pete: Oh really?
Miles: This is what-
Jamie: Of homosexual males, or of males.
Pete: Of all males.
Harol: Of males in general.
Miles: This infectious disease woman-
Harol: OR males under thirty I think it is.
Miles: At San Francisco General-
Pete: Mhm.
Miles: She said that, this doesn't seem like it can be true, but she said that, ninety percent of gay men, are HIV positive, and fifty percent of all males, are HIV positive.
Harol: Yeah.

Miles: That's what she said.
Harol: Fifty perc- It was some- Like, half or more than half of, and it was fairly young men, but, were, had been exposed to HIV virus. It wasn't necessarily that they were infected.
Miles: Well, if you're HIV positive, it's the same difference.
Harol: Mhm.
Miles: Since they feel that, sooner or later you'll come down with the actual disease.
Pete: Mhm.
Jamie: Yeah.
Miles: But that's what she said. Now I don't know if she meant the Bay area or San Francisco, but those are some ferocious numbers, if one out of--
Pete: Yeah.
Jamie: That's horrible.
Miles: --two guys, you meet.
Harol: Yeah.

The above transcriptions, in 18a-c, serve to illustrate some of the variety in the contexts presented as stimuli for the relevance rating task. Similarly to 18a-c, the full set of stimuli varied in terms of subject matter, length, degree of interactivity, and, to a lesser extent, level of formality (most contexts were towards the informal end of the scale).

3.3.3. VISUAL STIMULI. All visual (orthographic) stimuli represented components of the target cleft constructions: either the clefted element, the open proposition, or both were presented orthographically. Two main types of visual stimuli were presented orthographically on the computer screen: TRANSCRIBED CLEFT COMPONENTS and PARAPHRASED CLEFT COMPONENTS.

Transcribed components reflected the exact words uttered to express the clefted elements (CE) and open propositions (OP), as closely as possible. Only the parts of the cleft corresponding to X and Y in the skeletal structures (Table 5) were included as transcribed cleft components. The underlined portions of the IT cleft, REVERSE WH cleft, and THERE cleft examples in Table 5 represent the transcribed components of these tokens. Transcribed cleft components were presented as orthographic stimuli in the prosody judgment task and in the contrastiveness judgment task.

Paraphrased cleft components, presented as orthographic stimuli in the given/new judgment and relevance-rating tasks, were meant to capture the semantic content, or the

ideas, contained in the cleft components, so that hearers could make judgments about whether concepts were likely to be new in the discourse or were likely given in the preceding contexts. For the IT cleft, REVERSE WH cleft, and THERE cleft examples in Table 5, the paraphrased clefted elements are listed in 19a, 20a, and 21a; the paraphrased open propositions are listed in 19b, 20b, and 21b.

(19) a. someone named Barbara
b. someone having some seeds

(20) a. a flood
b. something leaving a white line around a lake

(21) a. a lot of women
b. some people not believing in using condoms

Paraphrased clefted elements were presented as indefinite noun phrases (see examples 19a, 20a, and 21a), and paraphrased open propositions were nominalized (as in examples 19b, 20b, and 21b). Thus, paraphrased cleft components did not necessarily represent the exact words used in the cleft (as is evident in examples 19-21 above).

3.3.4. PROCEDURE. For all tasks, initial scripted instructions, along with any necessary supplemental help from the experimenter, were delivered orally (see Appendix C). For some tasks, definitions were provided to guide participants in their judgments. For cleft-based information structure judgments (givenness/newness judgments and contrastiveness judgments), the researcher verbally acknowledged that subjects were making guesses about the newness or contrastiveness of part of the utterance with respect to their larger contexts, since they did not have access to the context for the clip, but explained that they were making educated guesses, since they were native speakers of English and could probably imagine likely speaking contexts for the clips. Along with verbal explanations, training sessions (with feedback from the researcher) and/or practice sessions (done independently) using non-cleft stimuli were completed for each task.

Training sessions were provided primarily to clarify instructions, to illustrate less obvious aspects of the definitions provided for new, given, and contrastive, and to help

make the participants comfortable with the task. Practice sessions (completed without researcher feedback) were provided primarily to (further) familiarize participants with the tasks. After training and/or practice, task instructions were reiterated in a condensed form on the opening screen for the main (i.e., cleft-related) part of each task.

All tasks were presented and executed as subject-directed sessions, using Praat Multiple Forced Choice experiment files and accompanying sound files in .wav format. During each session, subjects used a mouse to click buttons in order to progress through the task. While navigating through a task, participants saw phrases like, "Click to continue," on the screen, where appropriate. Task-specific instructions and processes are further explicated as necessary in the task-specific sections below, and in Appendix D.

For the most part, tasks were conducted with a between-subjects design; however, for logistical reasons, some subjects completed multiple tasks. Specifically, all twenty subjects who made new/given judgments, and five of the subjects who made contrastiveness judgments also completed the prosodic prominence judgment task. Measures were taken to counterbalance any task-order effect: half of those judging newness and prosody judged prosody first (five judged prosody, then CE newness; five judged prosody, then OP newness), and half judged newness first (five judged CE newness, then prosody; five judged OP newness, then prosody). Similarly, five of the ten participants who judged contrastiveness judged prosody first, and the other five judged only contrastiveness. As an aside, t-tests were carried out to detect any task-order effect. From these, the only significant effect obtained was that of judging prosody first on judging newness of the clefted element; for THERE clefts, those who judged prosody first were less likely to judge the clefted element as new. Though this effect was significant, it was quantitative rather than qualitative: the overall pattern of differences among the cleft sub-types remained the same across groups. Thus, the results reported in Chapter IV collapse the differently ordered groups for each task.

3.4. LISTENER JUDGMENTS

3.4.1. PROSODIC PROMINENCE. As discussed in Chapter I, there has been a lingering assumption in the literature that prosodic prominence and cleft structure influence information structure similarly (i.e., clefts have been called 'syntactic focusing devices'

and pitch prominence has been termed 'prosodic focus'). Following this, it has often been assumed that clefted elements are prosodically prominent. Based on past corpus studies (including those reviewed in Chapter II), however, cleft constructions in authentic spoken data appear to exhibit a greater variety of prosodic patterns than accounted for in descriptions of clefts that claim or assume an equivalence between prosodic and syntactic focus. As explained in Chapter I, the current study does not assume that syntactic clefting and prosodic prominence equivalently code the information structure of clefts; instead, the notion that clefted elements are prosodically prominent is tested as a hypothesis, and prosodic prominence is explored as a possible correlate to the information structure characteristics under study (especially newness and contrastiveness).

In the current study's prosodic analysis, perceived prosodic prominence was assessed empirically through native speaker-hearer judgments. As past information structure studies (including those reviewed in Chapter II) have typically defined prosodic prominence in terms of pitch, the use of hearer judgments for evaluating prosodic prominence calls for some justification. First, it is likely that in addition to pitch changes, other acoustic cues, such as intensity, duration, and pausing, contribute to the coding and perception of prosodic prominence. Vainio and Järvikivi (2007) report, for example, that syntactic cues play a role in both production and perception of prosodic prominence in Finnish. With a goal of empirically reflecting the various cues to prosodic prominence that interlocutors produce and attend to, prosodic prominence in the current study is based on holistic judgments made by multiple naïve native English speaker-hearers, about what sounds emphasized in cleft constructions. An additional reason for using speaker judgment rather than instrumental measurements to analyze prosody was the ability of native speaker-hearers to make judgments despite noisiness of sound files. That is, acoustic information in the available sound files was obscured, at times, by overlapping speech and/or background noise, and subjects simulating the hearer role were able to screen out obscuring factors when acoustic analysis software could not. Thus, this task was intended to provide data for an empirically-based analysis of where prosodic prominence is perceived as occurring in the cleft constructions under study.

In a forced-choice judgment task, 25 adult native English speakers simulating the hearer role made judgments about what sounded emphasized in SBCSAE cleft tokens presented aurally. EMPHASIZED was not directly defined for participants; it was demonstrated in a training session, through feedback regarding non-cleft examples with (more or less) clear pitch-accents. After receiving guidance during the training session, participants were encouraged to make their own subjective judgments about what sounded emphasized. Participants also completed a non-cleft practice session for this task.

In the main part of the task, subjects heard cleft utterances and saw both transcribed cleft components on the screen, with the clefted element labeled A and the open proposition labeled B. They were then asked to make a forced choice from among the following: no emphasis, emphasis in A only, emphasis in B only, emphasis in both A and B, and emphasis somewhere else. Test stimuli for the prosodic prominence judgment task, including 31 target THERE clefts, 13 target IT clefts, and 9 target REVERSE WH clefts, were presented in a randomized order.

Figure 1 shows an example of the screens participants saw upon hearing cleft utterance clips, with transcribed cleft components from the THERE cleft exemplified in Table 5.

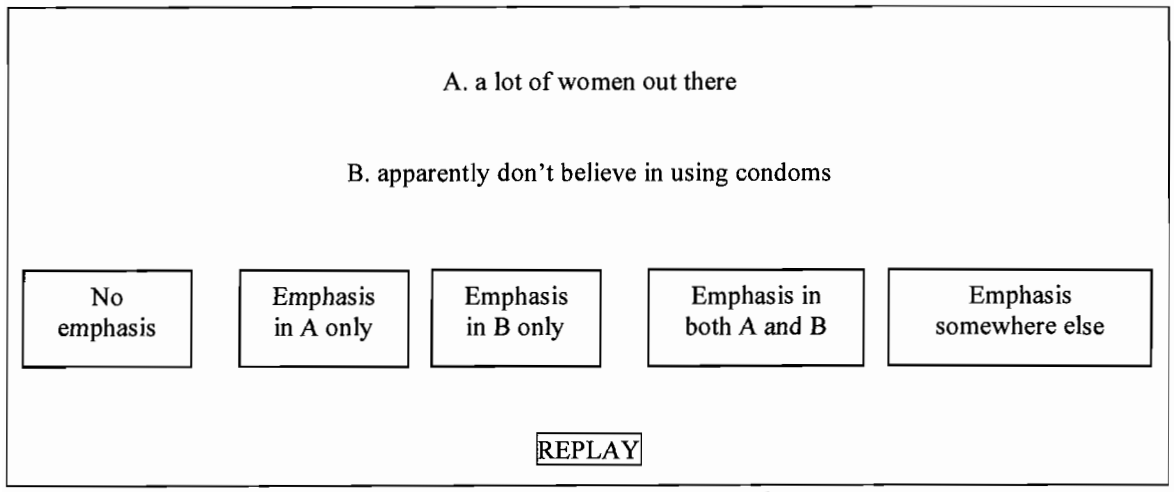

FIGURE 1. Prosody judgment example screen

3.4.2. NEWNESS/GIVENNESS. As described in Chapter I, information structure coding is assumed to reflect speakers' hypotheses about hearers' knowledge and awareness of referents and propositions at the time of utterance. If information structure coding is conventionalized (i.e., if the same form-meaning correspondences are used widely among speakers), and/or if information structure coding informative (i.e., if the coding adds to the knowledge of the speakers in some way, including meta-linguistically), then native speaker-listeners are likely able to detect the encoded information structure characteristics of referents and/or propositions in utterances. The cleft-based information structure tasks (this task and the contrastiveness judgment task) are designed on the assumptions that information structure is both conventionalized and informative to interlocutors.

The new/given judgment task was developed to test hypotheses about newness and givenness of cleft components by gaining insight into how the hearer interprets the structure of cleft subtypes as signals of new/given information status. Toward this aim, participants were asked to make judgments about the newness/givenness of cleft components after hearing only cleft utterances out of context, which inherently included both the syntactic and prosodic structural characteristics of the cleft tokens.

In a forced-choice judgment task, twenty adult native English speakers heard cleft utterances and determined whether paraphrased cleft components, presented visually, were likely to be new or given in their larger speaking contexts. Ten subjects judged the newness/givenness of only clefted elements (CE group), and a different group of ten subjects judged the newness/givenness of only open propositions (OP group). NEW was defined for participants as, 'being brought up for the first time, or adding information to the discussion.' GIVEN was defined as, 'already under discussion, or part of what the speaker and hearers already know or already consider relevant to the discussion.' Training examples and contexts were also provided to illustrate these concepts for participants. Test stimuli in the newness/givenness judgment task, including 31 target THERE clefts, 13 target IT clefts, and 9 target REVERSE WH clefts, were presented in a randomized order.

An example of the screens that participants in the CE group saw upon hearing each cleft utterance clip for this task is presented in Figure 2 (for the THERE cleft in Table 5), and Figure 3 provides an example of the screens that participants in the OP group saw (also for the THERE cleft in Table 5).

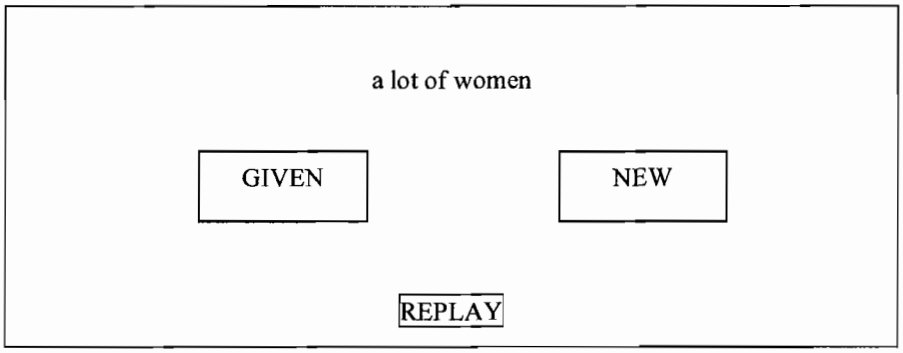

FIGURE 2. Given/new judgment example screen with clefted element

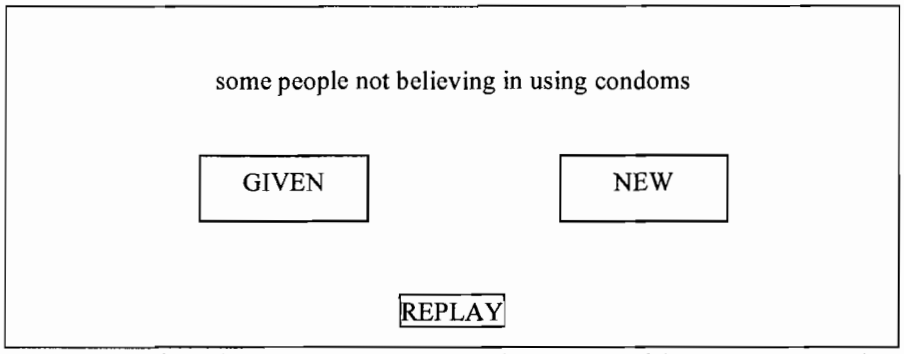

FIGURE 3. Given/new judgment example screen with open proposition

3.4.3. CONTRASTIVENESS. As with the new/given judgments, it was assumed that if contrastiveness is conventionally and/or informatively coded, listeners can likely detect syntactic and prosodic cues to contrastiveness of referents and/or propositions in utterances, even without being exposed to the relevant context. Thus, contrastiveness assessments were based on judgments made by listeners who heard cleft utterances only. In this task, only clefted elements were evaluated with respect to contrastiveness, in order to test the hypothesis that cleft constructions code contrastive focus, with contrastive clefted elements.

In a forced choice judgment task, ten adult native English speakers heard cleft utterances and were asked to think about the larger speaking context of the clip and judge whether transcribed clefted elements were contrastive or non-contrastive in their larger contexts. CONTRASTIVE was defined for participants as, 'being talked about in opposition to something else, or as a member or part of a set of alternatives.' NON-CONTRASTIVE was explicitly defined as 'not contrastive; being talked about in a more neutral way.' The given definition of contrastive was illustrated through written, non-cleft examples of contrastive elements in their immediate contexts, and participants were further familiarized with the task and definitions via a non-cleft training session with feedback. Test stimuli in the contrastiveness judgment task, including 31 target THERE clefts, 13 target IT clefts, and 9 target REVERSE WH clefts, were presented in a randomized order.

Figure 4 provides an example of the judgment screens for this task (again, for the THERE cleft from Table 5).

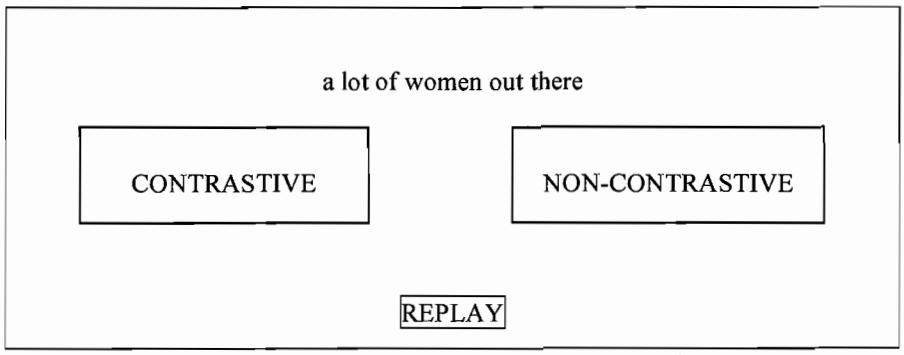

FIGURE 4. Contrastiveness judgment example screen

3.4.4. CONTEXT-BASED RELEVANCE. Though RELEVANCE is not generally considered an element of information structure, it is seen here as a potential correlate of information structure characteristics such as newness, activeness/givenness and topicality. In this task, participants rated the relevance of paraphrased cleft components (presented visually) to contexts preceding, but not including, cleft utterances (as described in §3.3). These context-based relevance ratings are taken as a reflection of the degree to which listeners incorporated the rated concepts (in this case, paraphrased cleft components) into

their mental models of the discourse at a given point (in this case, at the time of cleft utterance).

If relevance ratings reflect the degree to which referents or propositions are incorporated into interlocutors' mental models of the discourse at hand, a rating of 'extremely relevant' would likely be assigned for a referent or proposition that is solidly incorporated into the discourse model—likely a given (active or presupposed) referent or proposition, possibly even a topical referent or proposition. On the other hand, a rating of 'not at all relevant' would likely indicate that a referent or proposition is not yet incorporated in the discourse model, possibly because it is new (non-active and/or not presupposed). In other words, low relevance is assumed to correspond with newness, and high relevance is assumed to correspond with activeness/givenness, and possibly even with topicality.

As noted in Chapter I, hearers' perceived relevance ratings in this task reflect listeners' interpretations of contexts, rather than directly reflecting speakers' information structure coding. With this in mind, this task was developed in order to investigate the context-dependent nature of cleft information structure, and to provide insight into whether/which cleft component(s) tend to be included and/or salient in a hearer's mental discourse model at the time the cleft is uttered.

Twenty adult native English speakers were asked to rate the relevance of target cleft components to their preceding discourse contexts: ten participants were asked to rate the relevance of clefted elements to pre-cleft contexts (CE group), and a separate group of ten participants rated the relevance of open propositions to pre-cleft contexts (OP group). Subjects in the relevance-rating task heard each cleft context only once, and cleft components were rated on a scale of 1 to 7, where 1 indicated 'not at all relevant' and 7 indicated 'extremely relevant'. RELEVANCE was not defined explicitly, and although participants completed a practice session before proceeding to the main task, no feedback was provided beyond basic clarification of how to perform the task. Test stimuli, including 30 THERE clefts, 11 IT clefts, and 9 REVERSE WH clefts, were presented in a randomized order.

Figure 5 provides an example of the relevance rating screens presented to the CE group, and Figure 6 provides an example of the rating screens for the OP group (both corresponding to the THERE cleft in Table 5).

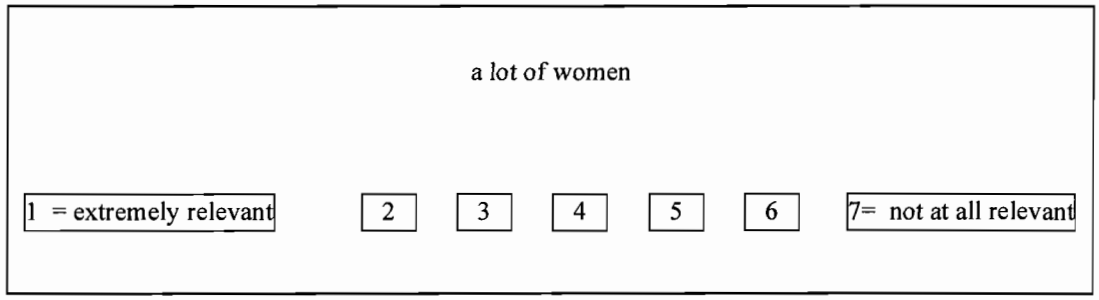

FIGURE 5. Relevance rating example screen with clefted element

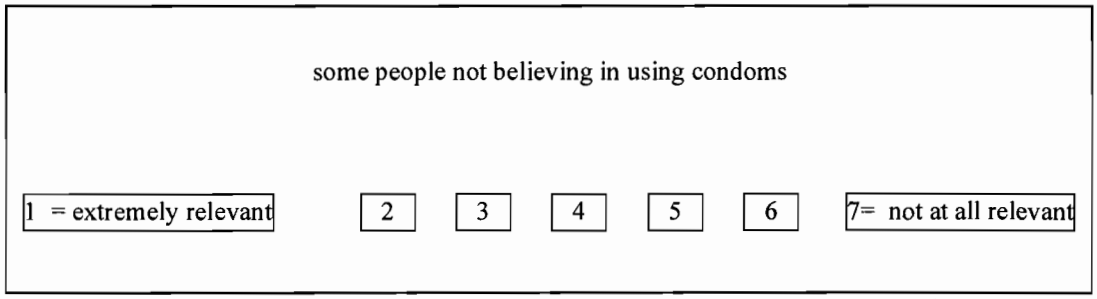

FIGURE 6. Relevance rating example screen with open proposition

3.5. METHODS FOR HYPOTHESIS-TESTING. Operational definitions were developed in order to use the gathered listener judgments to test hypotheses and predictions about prosodic patterns, focus structure and associated newness/givenness and contrastiveness patterns, and contextual relevance of cleft components (hypotheses and predictions are explicated in § 3.6 below). Specifically, a cleft component (i.e., a clefted element or an open proposition for a given token) was categorized as:

PROSODICALLY PROMINENT if at least 60 percent of participants judged it 'emphasized'. If a cleft component for a given token did not meet this threshold, it was categorized as NON-PROMINENT with respect to prosody.

NEW if at least 60 percent of participants judged it 'new'.

GIVEN if at least 60 percent of participants judged it 'given'.

CONTRASTIVE if at least 60 percent of participants judged it 'contrastive'. If a cleft component for a given token did not meet this threshold, it was categorized as NON-CONTRASTIVE.

MINIMALLY RELEVANT if its mean relevance rating was 2 or below.

MODERATELY RELEVANT if its mean relevance rating was between 2 and 6.

HIGHLY RELEVANT if its mean relevance rating was 6 or above.

The above operational definitions of PROSODICALLY PROMINENT, GIVEN, NEW, CONTRASTIVE, MINIMALLY RELEVANT, MODERATELY RELEVANT, and HIGHLY RELEVANT are implemented in the results reported in Chapter IV.

3.6. HYPOTHESES AND PREDICTIONS. The tasks described above were designed to test aspects of the traditional claims about clefts (introduced in Chapter I), to test some claims about prosody and information structure (based on Lambrecht's theoretical framework, also introduced in Chapter I), and also to test some claims about distinctive characteristics of cleft subtypes that stem from a combination of Lambrecht's analysis of cleft constructions (described in Chapter I) and from past corpus findings (reviewed in Chapter II). Since traditional claims about clefts are based primarily on fabricated examples of IT clefts, which have been central in the literature on clefts in the past (as noted in the introductory chapters), hypotheses related to traditional claims are most appropriately understood as hypotheses about IT clefts, although they have been stated as general hypotheses about cleft constructions. The second set of hypotheses relate to the function of prosody in cleft constructions, and the third set of hypotheses are intended to account for the distinctive and/or characteristic functions of the different cleft subtypes under study.

Several hypotheses that relate to traditional claims and/or assumptions are presented in 22 through 24:

(22) Clefted elements are prosodically prominent.

(23) Cleft constructions separate new from given information by presenting new information in the clefted element, and given information in the open proposition.

(24) Cleft constructions are contrastive focusing devices with contrastive clefted elements.

The prediction that clefted elements are prosodically prominent follows from hypothesis 22. From hypothesis 23, it is predicted that clefted elements are consistently judged 'new', and open propositions are consistently judged 'given'. Based on hypothesis 24, one would expect that clefted elements are consistently judged 'contrastive'.

Regarding the prosody of cleft constructions, the following hypotheses are of interest:

(25) a. Prosodic prominence codes newness.
b. Prosodic prominence codes contrastiveness.

Based on the hypotheses in 25, newness and contrastiveness are expected to correlate with prosodic prominence. Additionally, contextual relevance is expected to correlate negatively with prosodic prominence, due to the assumed association between newness and minimal contextual relevance.

Finally, aspects of Lambrecht's framework for analyzing clefts (cf. Chapter I) and findings from past corpus studies (cf. Chapter II) were synthesized to form hypotheses regarding the distinctive functions of each cleft subtype under study. These are intended as hypotheses about (part of) the constructional meaning of THERE clefts, IT clefts, and REVERSE WH clefts. These hypothesized distinctive functions are provided in 26:

(26) a. THERE clefts function to introduce new referents as clefted elements.
b. IT clefts function to express contrastive focus, via contrastive clefted elements.
c. REVERSE WH clefts function to express predicate-focus, via a given/topical clefted element, and a new open proposition

Predictions related to the hypotheses in 26 are summarized in Table 6.

THERE clefts	IT clefts	REVERSE WH clefts
-New, minimally relevant, prosodically prominent CE -CE may be contrastive	-Contrastive CE -CE may be given or new, may be minimally to highly relevant	-Given, highly relevant CE -New, prosodically prominent open proposition

TABLE 6. Predictions for cleft subtypes; CE represents 'clefted element'

The results of the procedures described here will be presented and discussed in Chapter IV with reference to the above hypotheses and predictions.

CHAPTER IV

RESULTS

4.1. OVERVIEW OF RESULTS. In the sections that follow, the results of the corpus searches and native speaker-hearer judgment tasks are presented, and implications for a more complete description of English clefts are discussed. First, corpus search results are reviewed and relative frequencies are discussed. Then, listener judgment results are presented and are related to the hypotheses and predictions outlined in the final section of Chapter III, including hypotheses based on traditional claims, hypotheses relating prosody to information structure, and hypotheses about the distinguishing characteristics of the cleft subtypes under study. Results for perceived prosodic prominence and prosody-information structure correlations are presented; after that, results for newness/givenness, contrastiveness, and contextual relevance are presented and discussed. The results are then summarized in a final section.

4.2. RELATIVE FREQUENCY RESULTS. As reported in Chapter III, SBCSAE searches for IT clefts, REVERSE WH clefts, and THERE clefts meeting specific syntactic criteria (specified in Chapter III) yielded 31 tokens of target THERE clefts, 14 tokens of target IT clefts, and 9 tokens of target REVERSE WH clefts. Among the target constructions, THERE clefts were far more frequent in the SBCSAE than either REVERSE WH clefts or IT clefts.[16] It is important to note this distribution would likely be different if all clefted noun phrases had been considered, such that REVERSE WH clefts with clefted demonstrative pronouns (*that* and *this*) would likely be the most frequent type, as they are in past corpus studies. As in past corpus studies, however, IT clefts are relatively infrequent in the SBCSAE. In

[16] Despite their relative frequency, THERE clefts are all but ignored in the literature on English clefts. Given the space in the literature that has been devoted to other, less frequent, cleft types, THERE clefts merit additional attention in future cleft studies.

keeping with past corpus studies, these relative frequency results call into question the centrality of IT clefts to discussions of cleft constructions.

4.3. LISTENER JUDGMENTS RESULTS

4.3.1. PERCEIVED PROSODIC PROMINENCE. Perceived prominence results for clefted elements and open propositions are presented first, followed by results for the various possible combinations of prominent and non-prominent cleft components. Overall, clefted elements (CE) were perceived as prominent in nearly twice as many cleft constructions as open propositions (OP).

Table 7 provides the token counts of prominent clefted elements and open propositions, for each cleft type, employing the operational definition of prominence provided in Chapter III (cleft components were categorized as prominent if at least 60 percent of participants judged them emphasized). Figure 7 depicts the proportions of each cleft subtype with prominent clefted elements and open propositions.

	TOKENS IN TASK	TOKENS WITH PROMINENT CE	TOKENS WITH PROMINENT OP
THERE clefts	31	16	11
IT clefts	13	12	4
REVERSE WH clefts	9	5	2
Total	53	33	17

TABLE 7. Tokens counts for prosodically prominent cleft components

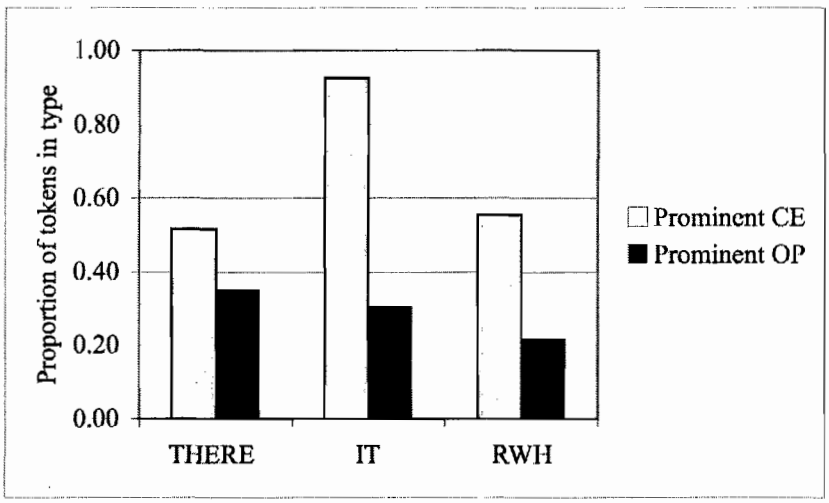

FIGURE 7. Proportions for prosodically prominent cleft components

As shown in Figure 7, clefted element prominence was strongest among IT clefts, for which 92 percent of tokens had prominent clefted elements. For the other two subtypes, slightly over half of the tokens had clefted elements perceived as prosodically prominent. This higher perceived prominence of clefted elements is consistent with traditional expectations about cleft constructions, and with the hypothesis (in 25) that clefted elements are prosodically prominent received limited support.

Nevertheless, the proportion of clefted elements judged as prominent is lower than what one might traditionally expect, and the proportions differ across cleft subtypes. There were some prosodically prominent open propositions for each subtype: 35.5 percent of THERE clefts, 30.8 percent of IT clefts, and 22.2 percent of REVERSE WH clefts had open propositions judged as prominent. Though this result would not be expected according to the traditional claim that open propositions express only given or presupposed information, the finding that some open propositions were perceived as prominent is not inconsistent with the hypotheses in Chapter III, and it is consistent with findings in past corpus studies. Additionally, though REVERSE WH clefts were the only subtype specifically hypothesized to occur with a new, therefore possibly prominent, open proposition, REVERSE WH clefts actually exhibited the lowest proportion of prominent open propositions (cf. hypotheses 25a and 26c in Chapter III).

Taken together, the present study and past corpus studies suggest that although cleft constructions are often perceived as having prosodically prominent clefted elements, a complete description of cleft constructions must account for the multiple prosodic patterns exhibited by cleft constructions.

In addition to looking at isolated cleft components, four possible prosodic patterns were identified, with various combinations of prominence and non-prominence of cleft components. The possibilities for prosodic type include: prominence in both the clefted element and the relative clause (+ +), prominence in only the clefted element (+ -), prominence in only the relative clause (- +), and prominence in neither the clefted element, nor the relative clause (- -). Prosodic types based on emphasis judgment results

were assigned for each of the 53 cleft stimuli, and the most frequent pattern was the + - pattern, followed by the - - pattern, then the + + pattern, and finally the - + pattern.

Token counts for each prosodic type are presented in Table 8, and proportions of tokens in each type exhibiting these patterns are presented in Figure 8.

	TOKENS IN TASK	+ -	- -	+ +	- +
THERE clefts	31	10	10	6	5
IT clefts	13	8	1	4	0
REVERSE WH clefts	9	4	3	1	1
Total	53	22	14	11	6

TABLE 8. Token counts for cleft subtypes exhibiting prosodic patterns

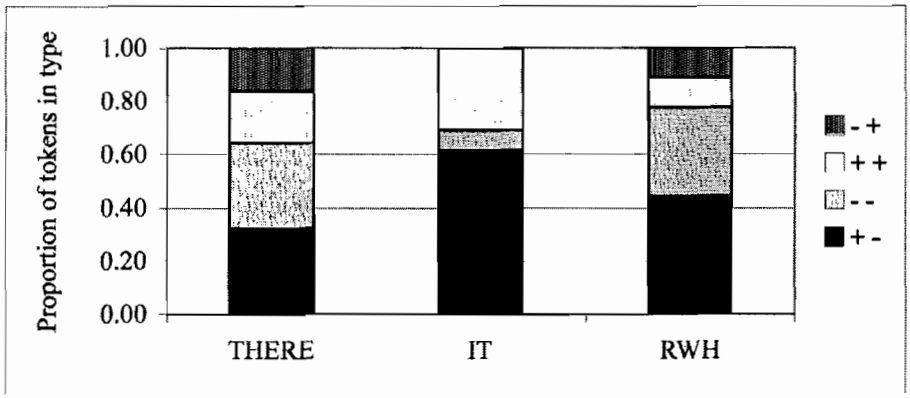

FIGURE 8. Proportions of cleft subtypes exhibiting prosodic patterns

Supporting traditional claims about clefts, some tokens in each cleft subtype were judged as having only prosodically prominent clefted elements. The + - pattern was particularly common among the IT clefts in the task (61.5 percent), and least common among THERE clefts (32.3 percent). Though this perceived pattern is expected according to the traditional descriptions of clefts, it is inconsistent with past corpus findings that sole prominence on the clefted element is rare. This discrepancy, between the current results and past results, may be due to a number of factors, including varying syntactic constraints on the data set, as well as difference between holistically perceived prominence and specifically pitch-related prominence.

Though many clefts had perceived prominence only in the clefted element, this was not always the majority pattern. A lack of prominence in either cleft component, and prominence in both components were also fairly common patterns. For THERE clefts and REVERSE WH clefts, about a third of tokens in each type lacked perceived prominence on either cleft component, and 1 IT cleft had no perceived prominence. A lack of prosodic prominence was not predicted for any cleft type, and it is likely an artifact of the methodology employed here. Some IT clefts (30.8 percent) and THERE clefts (19.4 percent) had perceived prominence in both components, which would be the expected pattern for sentence-focus clefts with new information in both cleft components (given hypothesis 25a in Chapter III). Very few cleft tokens had sole perceived prominence on the open proposition, including a single REVERSE WH cleft, the only cleft type for which this pattern was predicted.

To summarize the results for prominence patterns of cleft constructions, the dominant pattern for all subtypes is the pattern that has been typically assumed true for clefts: that of sole prominence on the clefted element. Still, the fact that there are clefts exhibiting each of the other prosodic types is evidence that more than one prosodic pattern must be considered in a complete description of cleft constructions.

4.3.2. PROSODY AND INFORMATION STRUCTURE. Because findings about prosody alone may indicate various information structure characteristics, results regarding relationships between prosody results and information structure characteristics are explored in this section. Only clefted elements were included in this portion of the analysis.

Proportions of subjects judging each clefted element as emphasized were compared with proportions of subjects judging clefted elements as new, with proportions of subjects judging clefted elements as contrastive, and with mean relevance ratings for clefted elements. In looking for relationships between prosodic prominence of clefted elements and the information structure factors of newness/givenness, contrastiveness, and relevance, only weak correlations were found.

Table 9 displays the correlations among prosodic prominence judgments and information structure judgments.

	PERCEIVED PROMINENCE		
	THERE CLEFTS	IT CLEFTS	REVERSE WH CLEFTS
Newness	-0.002	0.358	0.037
Contrastiveness	-0.047	0.269	-0.191
Relevance	-0.459	0.177	0.118

TABLE 9. Correlations among prosodic prominence judgments and information structure judgments for clefted elements

Despite weak correlations, two relationships are of note. First, a negative correlation is apparent between perceived prosodic prominence and relevance for THERE clefts, in accordance with expectations (assuming a connection between newness and low perceived relevance; cf. hypothesis 25a). Second, both newness and contrastiveness correlate positively with perceived prosodic prominence for IT clefts, also as expected (cf. hypothesis 25a-b). Overall, these results suggest that information structure judgments of cleft constructions in the present study are not strongly influenced by prosodic patterns; however, when there is a relationship, it is in accordance with expectations based on hypothesis 25.

4.3.3. NEWNESS/GIVENNESS. The most striking result regarding the information structure of cleft constructions was the difference in perceived newness/givenness of clefted elements for THERE clefts on the one hand, and for IT clefts and REVERSE WH clefts on the other hand. Clefted elements in THERE clefts were consistently perceived as new, while clefted elements in the other cleft types were consistently perceived as given.[17]The perceived newness of clefted elements in THERE clefts strongly supports hypothesis 26a and Lambrecht's (2001) claim that the existential THERE BE component of these clefts functions to introduce new referents.

[17] As noted in Chapter III, NEW tokens of cleft components are those judged 'new' by at least 60 percent of participants, and GIVEN tokens of cleft components are those judged 'given' by at least 60 percent of participants. INDETERMINATE tokens were those for which between 40 percent and 60 percent judged the component as *new* or *given*, in a binary judgment task.

Figure 9 shows the proportions of cleft subtypes with new, given and indeterminate clefted elements.

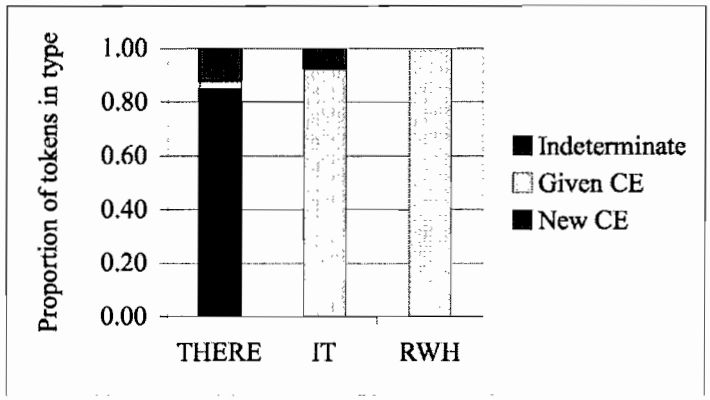

FIGURE 9. Proportions of given/new/indeterminate clefted elements (CE)

The consistent perception of clefted elements in REVERSE WH clefts as given supports hypothesis 26c, and is in agreement with past corpus analyses of this cleft type, for which the majority of clefted elements were found to be active and sometimes topical. This was in spite of the fact that, unlike in other studies, REVERSE WH clefts with clefted demonstrative pronouns were excluded. Clefted elements in IT clefts were also perceived as given more often than new. This result was surprising, as it may seem to challenge an analysis of clefted elements in IT clefts as focal. It is important to note, however, that constituents bearing contrastive focus need not be new in the discourse context; although clefted elements of IT clefts and REVERSE WH clefts were perceived as given, this does not necessarily indicate that they are non-focal. For further comment on this point, see the results for information structure patterns (§4.3.5).

Similar, but weaker, patterns emerged for open propositions of the target constructions. Open propositions in THERE clefts were perceived as new more often than given, and open propositions were perceived primarily as given for IT clefts and REVERSE WH clefts. Thus, both cleft components tend to be perceived as new for THERE clefts, while both cleft components tend to be perceived as given for the other two subtype.

Figure 10 depicts the proportions of new, given, and indeterminate open proposition tokens in each subtype.

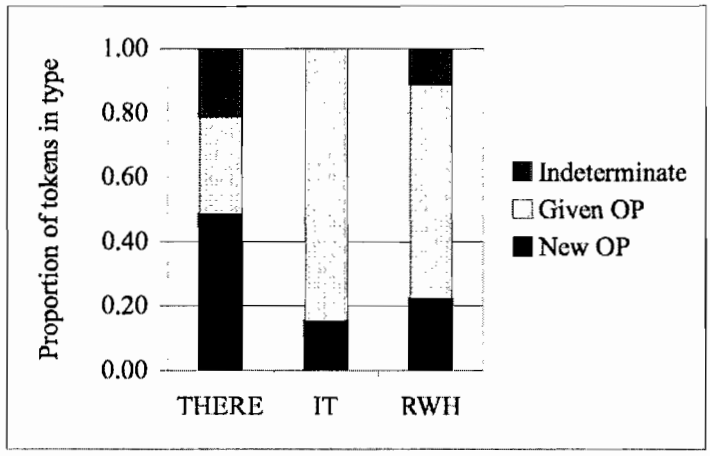

FIGURE 10. Proportions of given/new/indeterminate open propositions (OP)

Particularly for IT clefts and REVERSE WH clefts, these results support, to a limited extent, the traditional claim that open propositions in clefts express given, old, or presupposed information. It is interesting to note, however, that open propositions were less consistently perceived as given than clefted elements for IT clefts and REVERSE WH clefts. Thus, in agreement with Delin (1990), these results challenge the hypothesis (in 23) that clefts function to clearly separate new from given information. Additionally, the finding that some open propositions in each subtype were perceived as new is evidence that open propositions in clefts are not necessarily pragmatically/contextually presupposed.

Table 10 displays the token counts for clefted elements and open propositions perceived as new, given, and indeterminate.

		CLEFTED ELEMENTS (CE)			OPEN PROPOSITIONS (OP)		
	TOKENS IN TASK	NEW	GIVEN	INDET.	NEW OP	GIVEN OP	INDET.
THERE clefts	31	26	1	4	15	10	6
IT clefts	13	0	12	1	2	11	0
REVERSE WH clefts	9	0	9	0	2	6	1
Total	53	26	22	5	19	27	7

TABLE 10. Token counts for new/given/indeterminate cleft components

To summarize, newness/givenness judgments were considerably different for THERE clefts, compared with the other cleft types, particularly for clefted elements. For both types of cleft components, THERE clefts were perceived as new more consistently than the other cleft types. The combined results for all target constructions present a mixed profile for clefts regarding patterns in givenness and newness of cleft components. Generally, open propositions were perceived as given somewhat more frequently than were clefted elements, weakly supporting a traditional view of cleft information structure. Lumping all the cleft types together, however, obscures important differences among subtypes.

The analysis of givenness and newness of cleft components provides only a partial picture of the information structure of cleft components. In order to provide a more complete picture, perceived contrastiveness of clefted elements is investigated in the next section.

4.3.4. CONTRASTIVENESS. A majority of clefted elements was perceived as contrastive for all cleft subtypes under study. These results indicate partial support for the hypothesis that cleft constructions serve as contrastive focusing devices (cf. hypothesis 24 in Chapter III). The strength of support for claims about the contrastiveness of clefts, however, is limited by the presence of apparently non-contrastive clefted elements in each subtype: over 30 percent of tokens in each subtype were non-contrastive, according to the operationalized definition employed here.

Table 11 displays the token counts of clefted elements perceived to be contrastive by at least 60 percent of participants.

	TOKENS IN TASK	CONTRASTIVE TOKENS
THERE clefts	31	19
IT clefts	13	9
REVERSE WH clefts	9	6
Total	53	34

TABLE 11. Token counts for contrastive clefted elements

Figure 11 displays the proportions of contrastive tokens in each cleft subtype.

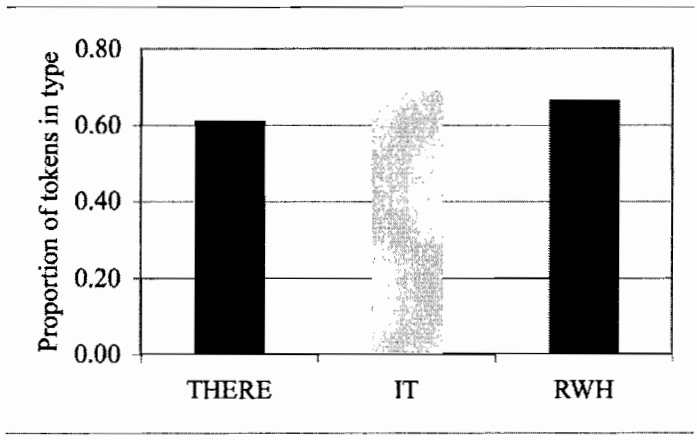

FIGURE 11. Proportions of clefted elements judged contrastive

The slightly stronger contrastive result for IT clefts is consistent with hypothesis 26a, and with Weinert and Miller's (1996) finding that IT clefts are the preferred cleft type for expressing contrast. However, IT clefts were found to be only slightly more (proportionally) contrastive than the other cleft types, and an appreciable proportion of IT clefts (about 31 percent) was not perceived as contrastive.

The finding that a relatively large proportion of clefts were not perceived as contrastive is somewhat surprising, especially given the broad definition of contrastive that participants were trained to employ ('being talked about in opposition to something else, or as a member or part of a set of alternatives'). This suggests that expression of contrast is only one among multiple information structure-related functions of cleft constructions, and that the specificational character of clefts (i.e., the expression of an open variable in the open proposition and its specification in the clefted element, as described in Chapter I) may not be salient to the naïve native English speaker-hearer.

4.3.5. INFORMATION STRUCTURE PATTERNS. In addition to assessing the newness/givenness and contrastiveness of isolated cleft components, clefts tokens were divided into subtypes, based on various combinations of information structure characteristics cleft components. First, cleft tokens were categorized as either NEW-NEW (with a new clefted element and a new open proposition), NEW-GIVEN (with a new clefted

element and a given open proposition), GIVEN-NEW (with a given clefted element and a new open proposition, or GIVEN-GIVEN (with a given clefted element and a given open proposition). Tokens with an indeterminate clefted element and/or an indeterminate open proposition were categorized as 'indeterminate' for the purpose of this initial pattern analysis. Table 12 displays the token counts in each of these categories, for each cleft subtype.

	NEW-NEW	NEW-GIVEN	GIVEN-NEW	GIVEN-GIVEN	INDETERMINATE
THERE	12	8	0	1	10
IT	0	0	2	10	1
RWH	0	0	2	6	1
Total	12	8	4	17	12

TABLE 12. Token counts for newness patterns of cleft components

As seen in Table 12, THERE clefts exhibit some new-new clefts (expected for sentence-focus clefts), and some new-given clefts (expected for argument-focus clefts). These patterns are in agreement with predictions based on Lambrecht's preliminary description of English THERE clefts, and also with hypothesis 26a (Chapter III), which states that THERE clefts function to introduce new referents as clefted elements.

The strong given-given pattern observed for the IT clefts and REVERSE WH clefts is somewhat surprising at first glance; it seems to suggest a lack of new information, and therefore a lack of focus, if focus is understood simply as the part of the utterance expressing new information. However, focus is better understood as a broader category subsuming both the expression of new information and the expression of contrastive information, and so the given-given pattern need not indicate a lack of focus (as noted above in §4.3.3).

With the above concept of focus in mind, the contrastiveness of clefts with given clefted elements was assessed to determine whether they expressed contrastive focus. While none of the given-new clefts were judged contrastive, the majority of given-given clefts were perceived as having contrastive clefted elements.

Table 13 displays the token counts for contrastive given-new and given-given clefts, along with the total given-new and given-given tokens in each cleft subtype.

	GIVEN-NEW		GIVEN-GIVEN	
	TOTAL	CONTRASTIVE	TOTAL	CONTRASTIVE
THERE	0	0	1	1
IT	2	0	10	8
RWH	2	0	6	5
Total	4	0	17	14

TABLE 13. Contrastiveness of given clefted elements

As most of the given-given IT clefts and REVERSE WH clefts were perceived as contrastive, it seems as though these clefted elements are, in fact, functioning as focus domains. These results provide some support for hypothesis 24 (which states that clefts are contrastive focusing devices, according to the traditional claim), and also substantiate the view that focus is divisible into separate subtypes based on newness and contrastiveness.

4.3.6. CONTEXTUAL RELEVANCE. As described in Chapter III, relevance ratings of cleft components to their preceding contexts were given on a scale of 1 to 7, with 1 indicating 'not at all relevant' and 7 indicating 'extremely relevant', and levels of relevance were assigned for clefted elements and open propositions based on mean relevance ratings. To remind the reader of the levels, cleft components were considered minimally relevant to their preceding contexts if they received a mean relevance rating of 2 or lower; cleft components were considered moderately relevant if they received a mean relevance rating between 2 and 6; cleft components were considered highly relevant if they received a mean relevance rating of 6 or above.

Relevance results were most striking for REVERSE WH clefts, of which the largest proportion of clefted elements were perceived as highly contextually relevant. At the other end of the rating scale, no clefted elements of REVERSE WH clefts were found to be minimally relevant. These findings indicate that the referents of clefted elements in REVERSE WH clefts are likely active, if not topical in their discourse contexts, as would be expected if REVERSE WH clefts express predicate-focus (cf. hypothesis 26c). In comparison, clefted elements of THERE clefts and IT clefts were seen as less relevant to their preceding contexts, and therefore probably less active, and less likely to be

established topics in their discourse contexts. Clefted elements of THERE clefts had the fewest highly relevant clefted elements, as would be expected if THERE clefts introduce new referents in their clefted elements (cf. hypothesis 26a).

Proportions of clefted elements for each cleft subtype falling into categories based on the three relevance levels (a mid range, and two extremes) are displayed in Figure 12.

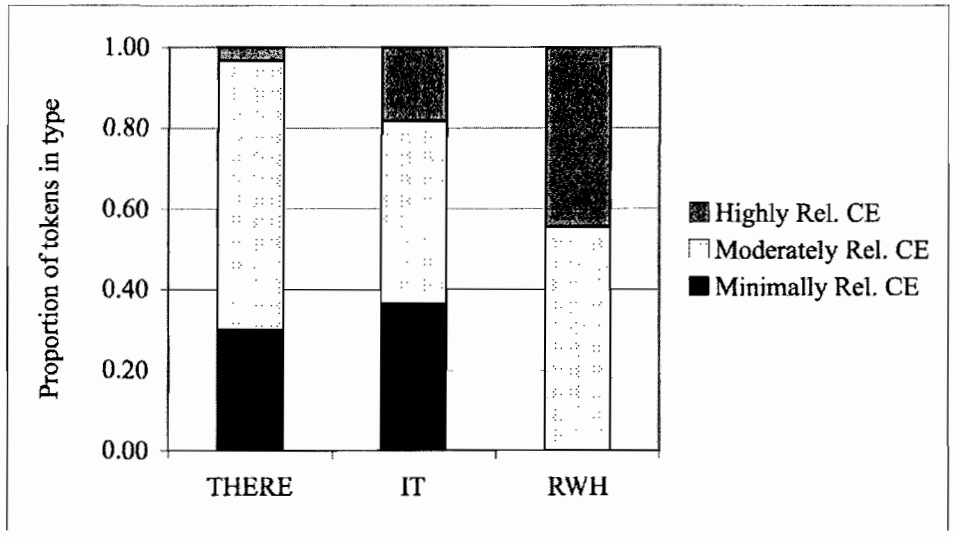

FIGURE 12. Relevance levels of clefted elements (CE)

Though differences among cleft subtypes are less striking for relevance ratings of open propositions, some interesting patterns emerged for these as well. While clefted elements of REVERSE WH clefts tended to be rated as more relevant to their preceding contexts than those of other clefts, open propositions of REVERSE WH clefts were rated as less relevant than those of other clefts. If REVERSE WH clefts tend to have active or even topical clefted elements and less active (newer, more focal) open propositions, this suggests that their information structure is more like unmarked predicate focus than that of either IT clefts or THERE clefts as expected from hypothesis 26c. Thus, REVERSE WH clefts with clefted lexical noun phrases appear to pattern much like the more frequent demonstrative REVERSE WH clefts in terms of information structure (cf. past corpus studies reviewed in Chapter II).

Figure 13 displays the proportions of open propositions at each relevance level for THERE clefts, IT clefts, and REVERSE WH clefts.

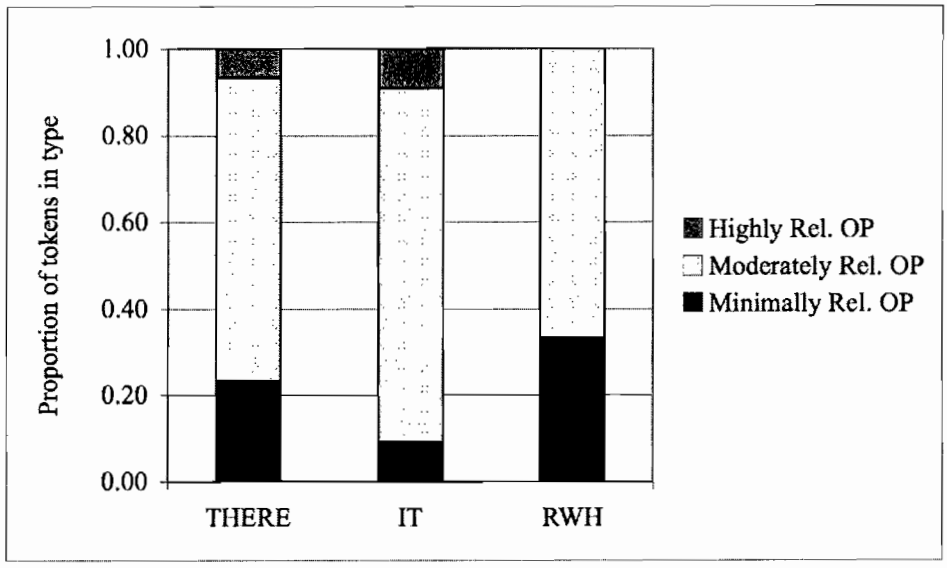

FIGURE 13. Relevance levels of open propositions (OP)

Probably due in part to multiple interpretations of the term RELEVANCE (which was not defined in the task instructions), there was much variation in the relevance ratings for most stimulus sets. For THERE clefts and IT clefts, clefted elements with lower mean relevance ratings tended to be more consistently rated than clefted elements with higher relevance ratings. For THERE clefts, mean ratings correlate positively with standard deviations (at 0.543); for IT clefts, mean ratings correlate positively with standard deviations (at 0.408). However, this is not the case for REVERSE WH clefts. For clefted elements of REVERSE WH clefts, there was a strong negative correlation between mean relevance ratings and standard deviations (at -0.867).

Table 14 displays the mean relevance ratings for clefted elements, along with their standard deviations, ordered from lowest to highest standard deviation.

THERE CLEFTS		IT CLEFTS		REVERSE WH CLEFTS	
CE MEAN RELEVANCE	CE S.D.	CE MEAN RELEVANCE	CE S.D.	CE MEAN RELEVANCE	CE S.D.
1.0	0.000	1.1	0.316	3.2	2.348
1.1	0.316	1.3	0.483	3.8	1.398
1.3	0.483	1.3	0.483	3.9	1.729
1.9	1.729	1.9	0.738	4.9	1.853
1.9	0.994	2.4	1.897	5.0	2.000
2.0	1.054	2.4	1.776	6.1	0.994
2.0	1.054	2.4	1.350	6.4	0.966
2.0	1.563	2.7	1.160	6.8	0.422
2.0	1.247	5.4	1.838	7.0	0.000
2.3	1.337	6.4	1.265		
2.6	1.955	6.6	0.966		
2.7	1.636				
2.8	1.229				
3.3	1.059				
4.0	1.333				
4.2	1.619				
4.3	1.703				
4.3	1.418				
4.3	1.889				
4.4	1.506				
4.6	1.838				
4.7	1.252				
4.9	1.729				
5.0	1.491				
5.1	1.524				
5.3	1.337				
5.4	1.430				
5.5	1.716				
5.8	1.989				
6.2	0.919				

TABLE 14. Mean relevance ratings and standard deviations (among ten participants) for clefted elements, ordered from lowest to highest mean relevance rating

Relevance ratings for open propositions were even less consistent than relevance ratings for clefted elements. The correlations among mean ratings and standard deviations for open propositions were weakly positive for all cleft types (0.400 for THERE clefts, 0.384 for IT clefts, and 0.437 for REVERSE WH clefts). Thus, open propositions were more consistently rated if they were judged less relevant to their preceding contexts.

Table 15 shows mean relevance ratings and standard deviations for open propositions.

THERE CLEFTS		IT CLEFTS		REVERSE WH CLEFTS	
OP MEAN RELEVANCE	CE S.D.	OP MEAN RELEVANCE	CE S.D.	OP MEAN RELEVANCE	CE S.D.
1.1	0.316	1.5	0.972	1.4	0.699
1.4	0.699	2.7	1.337	1.7	0.483
1.4	0.699	2.8	0.789	1.8	1.398
1.5	0.972	2.9	1.595	2.7	2.058
1.7	0.949	3.7	1.636	3.4	1.897
1.8	1.033	3.9	2.424	4.8	1.751
2.0	1.491	4.0	2.749	5.0	1.764
2.5	1.581	4.7	2.058	5.1	1.370
2.5	1.780	4.9	2.079	5.4	1.174
2.6	1.265	5.1	2.025		
2.8	1.932	6.1	1.101		
2.9	1.792				
2.9	2.183				
3.1	1.287				
3.5	0.972				
3.8	2.658				
4.0	2.211				
4.1	1.663				
4.3	1.567				
4.3	2.359				
4.5	1.780				
4.8	1.814				
4.9	1.912				
5.0	1.633				
5.1	1.595				
5.1	1.853				
5.5	1.581				
5.7	1.059				
6.0	1.054				
6.3	1.337				

TABLE 15. Mean relevance ratings and standard deviations (among ten participants) for open propositions, ordered from lowest to highest mean relevance rating

In addition to varying interpretations of the task, this variation may be evidence that the undoubtedly dramatic variation in the background knowledge that individual hearers have and bring to a discourse context affects their judgments about what is relevant to a conversation. This also likely reflects differences in how individual speakers construct their mental models of the discourse.

4.4. SUMMARY OF RESULTS. Predictions based on traditional accounts of cleft constructions outlined at the end of Chapter III (cf. hypotheses 22-24), were supported to some extent, in that 1) clefted elements tended to be perceived as prosodically prominent for all cleft types, 2) clefted elements tended to be perceived as new and/or contrastive, 3) open propositions were perceived as given more consistently than contrastive, and 4) the dominant discernable patterns for IT clefts and REVERSE WH clefts were contrastive clefted elements with given open propositions.

Expected correlations between prosodic prominence and newness, contrastive and relevance were only weakly attested.

The hypothesized distinctive functions of THERE clefts, IT clefts, and REVERSE WH clefts were each supported, but to different extents. As expected, clefted elements of THERE clefts were consistently perceived as expressing new information, strongly supporting the hypothesized function of THERE clefts (cf. hypothesis 26a). IT clefts were perceived as contrastive only slightly more consistently than other cleft types, weakly supporting hypothesis 26b. Finally, while the newness/givenness results for REVERSE WH clefts do not evidence a unique predicate-focus function for this cleft subtype, the perceived relevance results do suggest that clefted elements in REVERSE WH clefts tend to be seen as highly relevant to their preceding discourse contexts, supporting the notion that REVERSE WH clefts have given, possibly topical, clefted elements (cf. hypothesis 26c and previous corpus studies reviewed in Chapter II). Thus, while the hypothesized characteristics partially distinguish THERE clefts, IT clefts, and REVERSE WH clefts from one another, the strongest distinction is between the expression of new clefted elements in THERE clefts, and the lack of this characteristic among IT clefts and REVERSE WH clefts.

CHAPTER V

CONCLUSION

5.1. SUMMARY OF THE WORK. This thesis has been an investigation of naïve native speaker-hearer judgments about THERE clefts, IT clefts, and REVERSE WH clefts in authentic Spoken American English. The goals of the project included testing traditional claims about clefts and developing and implementing methods for empirically describing the information structure of cleft constructions. These goals, as well as the nature of cleft constructions and information structure were introduced in Chapter I. In Chapter II, traditional claims about clefts were further developed with respect to Lambrecht's (2001) framework for analyzing clefts, and past corpus studies about the target constructions were reviewed. The methodology for the original investigation was described in Chapter III, and the results of the investigation were presented in Chapter IV.

The principal results of the current study suggest the following tendencies with regard to the information structure of the cleft subtypes under study:

- Clefted elements in THERE clefts strongly tend to be interpreted as new, and are sometimes interpreted as contrastive; open propositions in THERE clefts may be interpreted as new or given
- Clefted elements in IT clefts tend to be interpreted as given and contrastive; open propositions in IT clefts tend to be perceived as given
- Clefted elements in REVERSE WH clefts tend to be interpreted as given and contrastive, and possibly also topical—or at least highly relevant to the preceding discourse context; open propositions in REVERSE WH clefts tend to be interpreted as given

Prosodic prominence was also explored, both in relation to information structure and cleft components. This investigation yielded weak evidence for positive correlations between perceived prosodic prominence and newness, and between prosodic prominence

and contrastiveness. In support of standard claims, clefted elements were perceived as prosodically prominent more consistently than open propositions, for all cleft subtypes under study.

The remaining sections of this chapter conclude the thesis by further discussing interpretations and some theoretical implications of the investigation, by critiquing the methodology, and by suggesting some revisions to traditional descriptions of cleft constructions.

5.2. VARIOUS MOTIVATIONS FOR CLEFT USE. While the information structure characteristics summarized above (in §5.1) are apparent in the results of the study, they are only tendencies: no categorical patterns emerged. These tendencies, or gradient patterns, support the notion that information structure is probabilistic, rather than rule-based. Two explanations are suggested here regarding possible sources for information structure variation in production of cleft constructions. First, the choice to code a message in the form of a cleft construction may be motivated by processing needs. During discourse, a participant may know that he/she wants to make a comment about a particular referent in the conversation, but may need additional time to construct the predication. In such a case, use of a cleft construction allows the participant to first mention the referent/argument in the clefted element, and then provides the participant with more processing time than the simplified version would to construct the predication (which would likely be new) and express it in a relative clause. Secondly, cleft constructions (particularly THERE clefts) may be seen as ideal for expressing sentence-focus (where both an argument and a predication are new), since they allowing a speaker to distribute a highly informative proposition over multiple clauses. Thus, processing needs, informativity, newness and contrastiveness may all motivate the use of cleft constructions at different times.

5.3. SPEAKER-HEARER ISSUES. As noted in the opening chapter, the original investigation reported here does not directly describe the information structure of cleft constructions, since it describes hearers' interpretations rather than speakers' intentions. If communication is effective, hearers' interpretations should reflect the speakers' intentions to some degree, but the correlation is certainly imperfect. For example, the

speaker may not be making correct hypotheses about what the hearer knows or is aware of at the time of utterance.

In the current study, there was certainly a mismatch between the speakers' hypotheses at the time of utterance and the experimental subject-hearers' knowledge and awareness because the two were dislocated in time and space. This was especially obvious in the information structure judgments, since cleft utterances were completely decontextualized, but it may also have been problem for the relevance rating task, since the contexts subjects heard were restricted in many ways: they did not have the visual input, the background knowledge, or contextual knowledge about the setting, level of formality, familiarity among participants, etc., that the intended audience would have had. In addition, the kinds of judgments that the experimental subject-hearers were making in these tasks required them to consciously process the information structure, thereby making explicit a task that is implicit under normal circumstances. Explicit processing may have rendered the tasks in the current study more difficult, thereby affecting the consistency of the results.

Though the above issues are problems for the ecological validity of the investigation, the removal of contexts allows us to better understand what is actually communicated to the hearer about the information structure in the speech stream itself. In fact, the potential mismatch between speakers' hypotheses and hearers' actual states is arguably one of the reasons for linguistic coding of information structure. Information structure, whether syntactically or prosodically encoded in the speech stream, should give interlocutors continuous input about what speakers think hearers know or are aware of, allowing hearers to continuously monitor how well they are tracking conversation, or how closely their model of the message matches the intended message.

In any particular utterance, a speaker's coding choices reflect his/her unambiguous intentions; ambiguity arises only when hearers are interpreting the signal. Since the results reported here are based on hearers' interpretations rather than speakers' intentions, they may reflect ambiguity that was not present for the speakers who uttered the clefts. This is an additional possible source of inconsistency in the results.

5.4. METHODOLOGICAL IMPROVEMENTS. Information structure analyses in linguistics have often been based on the judgments of individual researchers, each influenced by his or her own biases and/or agenda. The current study has been conducted with a goal of lessening this problem by capitalizing on the conventionality of language, both in use and interpretation. Toward this aim, the use of naïve native speaker-hearer judgments provides the foundation for the information structure analysis presented in this thesis. This aspect of the methodology is considered its most prominent strength, in the interest of producing realistic (rather than idealized) analyses of authentic language data. As is often the case, however, this strength is also a source of problems for the interpretation of the current study. This section highlights some of the limitations of the methods employed in the current study, and provides some suggestions for future research.

5.4.1. IMPROVEMENTS TO THE STIMULI. First, the fact that the SBCSAE, from which the auditory stimuli were extracted, includes data from speakers from a variety of regions of the country, and from quite varied contexts may be seen as both an asset and a problem for the current study. The variety is an asset because it allows for reasonably good representation of 'American English', but it is not clear whether the use and function of clefts is the same or different across the many varieties of American English represented in the clip. Investigating dialect variation is beyond the scope of this thesis, but would make a potentially interesting topic for future studies of cleft constructions, and should be investigated for a more complete description of clefts in American English.

Also, the varying number of clefts tokens presented as stimuli for each subtype is a weakness of the current study, particularly the low number of IT clefts and REVERSE WH clefts. Future studies employing similar methods would do well to compare similar quantities of the each cleft type under study, and to gather additional tokens (depending on the goals of the study).

An additional design problem regarding stimuli in the current study is the lack of non-cleft stimuli with apparent unmarked predicate focus, as these are hypothesized to have marked argument focus or sentence focus (Lambrecht 1994, 2001). Since participants were only exposed to cleft stimuli in the main part of the experiment, their judgments of

cleft component newness and contrastiveness may have been less consistent than they would have been otherwise, due to a resistance to answering the same way every time. By presenting both cleft and non-cleft stimuli, future studies would be able to compare perceived information structure of clefts with non-clefts, and would likely gain in terms of consistency of results for target cleft constructions as well.

Additionally, a considerable measure of subjectivity was involved in deciding how much of the preceding context to include in the auditory stimuli for the relevance rating task. Based on pilot studies, it seems that varying the amount of context affects relevance judgments; some pilot participants reported that they felt what was said at the beginning of the clip strongly influenced their relevance judgments. This problem may be difficult to avoid, but this apparent effect suggests that varying the starting point in a similar task might help to expand our understanding of the way that speakers construct mental models of the discourse.

5.4.2. OTHER DESIGN IMPROVEMENTS. Several other characteristics of the design could be improved in future research. First, the fact the cleft stimuli represented multiple varieties of English in combination with the homogeneous group of experimental participants (i.e., university students in a western state) presents a potential problem for the ecological validity of the study. Since dialect differences in cleft use have not been ruled out, future studies might do well to control for dialect of both speakers and hearers. Additionally, task-related terminology was not narrowly defined for all tasks. It has been noted that the term RELEVANCE is somewhat polysemous, and therefore could have been interpreted to mean something like 'important' or more loosely 'connected' to the context. Clarification of potentially ambiguous task-related terms would likely promote consistency in responses. Other aspects of the training procedures, such as the quality and quantity of training and practice examples, could also be improved.

5.5. CONCLUDING REMARKS. The primary goal of this thesis has been the evaluation of the traditional claims that clefts are contrastive focusing devices and that clefts separate the new/asserted/focal information from the given/presupposed information. Both of these claims are supported in the current study, but to a limited extent. First, it has been noted that clefts are defined by their specificational structure, and as such, they

can be considered inherently contrastive. On the other hand, the contrastiveness of cleft tokens in the current study was apparently not always salient to naïve native speaker-hearer subjects, which casts doubt on the claim that expressing contrastiveness is a central pragmatic function of clefts. Second, clefted elements were typically perceived as focal in terms of contrastiveness and/or newness, and open propositions were perceived as given more often than not. Challenging the notion that clefts separate new/asserted/focal information from given/presupposed information, however, is the attestation of both new-new and given-given clefts in the SBCSAE. The original corpus study reported here provides only partial support of these traditional claims, favoring a more complex view of clefts as a broader category, subsuming several formally and functionally distinct subtypes.

Thus, while the traditional accounts are not wholly rejected here, the current study indicates that these traditional accounts of cleft constructions have been overly simplistic. Different cleft subtypes do, in fact, exhibit differing information structure patterns (as summarized above in §5.1), the clearest of which is the consistent newness of clefted elements in THERE clefts. Based on the current study, and on past corpus studies, English cleft constructions must account for multiple subtypes of clefts and distinctions among cleft subtypes, including their potential for sentence-focus and predicate-focus information structures. Furthermore, it is recommended that explanations for these differences be sought with reference to informativity and processing considerations, in addition to information structure characteristics.

APPENDIX A

CORPUS SEARCHES PERFORMED

SEARCH STRING ENTERED IN CLAN		TRANSCRIPT REPRESENTATION
FOR IT CLEFTS:	combo +sit^is^*^that	it is … that
	combo +sit^is^*^who	it is … who
	combo +sit^was^*^that	it was … that
	combo +sit^was^*^who	it was … who
	combo +s"it's"^*^that	it's … that
	combo +s"it's"^*^who	it's … who
FOR REVERSE WH CLEFTS:		
	combo +sis^who	is who
	combo +sis^what	is what
	combo +s"*'s"^who	's who
	combo +s"*'s"^what	's what
	combo +sare^who	are who
	combo +sare^what	are what
	combo +swere^who	were who
	combo +swere^ what	were what
FOR THERE CLEFTS:	combo +sthere^is^*^that	there is … that
	combo +sthere^is^*^who	there is … who
	combo +s"there's"^*^that	there's … that
	combo +s"there's"^*^who	there's … who
	combo +sthere^are^*^that	there are … that
	combo +sthere^are^*^who	there are … who

combo +s“there’re”^*^that	therere … that
combo +s“there’re”^*^who	there are … who
combo +sthere^was^*^that	there was… that
combo +sthere^was^*^who	there was… who
combo +sthere^were^*^that	there were … that
combo +sthere^were^who	there were … who

APPENDIX B

CLEFT CONSTRUCTIONS MEETING SEARCH CRITERIA

ELEMENTS OF EACH ITEM:

Preceding part of utterance, if any [cleft construction] following part of utterance, if any

Simplification

Explanatory note, if necessary

SPECIAL USE OF SYMBOLS:

(.) pause

(..) long pause

/ intonation unit boundary

? phrase final rising intonation

NOTE ABOUT TRANSCRIPTION:

Transcriptions as they appears here are somewhat modified from the original version, for readability. Explanatory notes regarding modifications are provided where necessary. All instances of intonation boundaries, pauses, and phrase-final rising are from the original version.

THERE clefts:

(1) [There's a white line / that you go by].

You go by a white line.

(2) [There's this girl / that's workin' with him / for the summer] ?

This girl is workin' with him for the summer.

(3) Cause [there're a lotta women out there who (.) apparently don't believe in using condoms.]

A lot of women out there apparently don't believe in using condoms.

(4) [There's like one lemon left on this tree that I can reach].

I can reach like one lemon left on this tree.

(5) This is (.) a raging beauracracy / (.) and [there's nothing that I can do.]

I can do nothing.

(6) [There are seats right here / (.) um (.) / that are for the audience].

(The/Some) seats right here are for the audience.

(7) And then [there's a woman who sits here who's the clerk].

A woman who sits here is the clerk.

(8) And [there's a woman who sits here th- who's the court reporter.]

A woman who sits here is the court reporter.

(9) [There's manners of- of speaking that I was very outspoken about / at this meeting].

I was very outspoken about (some) manners of speaking at this meeting.

(10) [There is a town right around here / that is- / still has a zocalo- / that's built around a zocalo].

A town right around here is built around a zocalo.

(11) [There's one deck here / I think / that's set up for that].

I think one deck here is set up for that.

(12) [There's usually two little arrows that will light up].

Usually, two little arrows will light up.

(13) It seems / just like / whatever we think / (.) we can (.) make it / (.) but [there's one technology that's um / (.) gonna overtake that] / and that's DNA research.

One technology is gonna overtake that.

(14) Well now uh / you know / [there are experiences of ecstasy / th- uh / that can be induced by a transformation of consciousness].

Experiences of ecstasy can be induced by a transformation of consciousness.

(15) [There are two young people that I've talked to about the movie].

I've talked to two young people about the movie.

(16) You see / [there is nothing that (.) can come to us].

Nothing can come to us.

(17) I know [there's stuff that she- / (.) that you won't tell me].

You won't tell me stuff.

(18) [There's (.) this scientist / that came up with irrefutable proof] / (.) that / there are / (..) there are / there (.) they found two planets circulating around (..) a sun.

This scientist came up with irrefutable proof.

(19) [So there's some chemical reaction that has to take place.]

Some chemical reaction has to take place.

(20) There are different ones / that kinda go all the way up.

Different ones kinda go all the way up.

(21) But also / [there was an- another motive / that we'll get to in just a second].

We'll get to another motive in just a second.

(22) Cause [there's these poets that (.) I like].

I like these poets.

(23) And then [there's other kids who go / nanananazabata].

Other kids go nanananazabata.

(24) Basically there's / (..) [there's a disease called (..) Newcastle's Disease / that / is (.) is very (..) contagious apparently].

A disease called Newcastle's Disease is very contagious apparently.

(25) No [there are cops there that bust 'em].

Cops there bust 'em.

(26) There are two things that stabilize your knee.

Two things stabilize your knee.

(27) And [there's guys that've been doing that four or five years].

(Some) guys have been doing that four or five years.

(28) [There are three or four of 'em that said / why did we ever refuse] / if we had known.

Three or four of 'em said, "Why did we ever refuse?"

(29) [There was a tree / that burned all night] / and never- / and (.) and wasn't consumed by the flame.

A tree burned all night.

(30) Because [there's a light on the top / that tells you if they're on duty or off-duty / and if they have a fare or not].

A light on the top tells you if they're on duty or off-duty and if they have a fare or not.

(31) [There are (..) a couple of other story-telling events (..) that I'd just like to share with you].

I'd just like to share a couple of other story-telling events with you.

IT CLEFTS:

(1) See [it's little rules like that / (.) that I'm not gonna remember].

I'm not gonna remember little rules like that.

(2) I was so glad that he opened up on this disclosure bit / because [it was Burns that's been objecting to our having it].

Burns has been objecting to our having it.

(3) Well I think [it was Barbara that (..) has some seeds].

Barbara has some seeds.

(4) You mean because (.) if (..) the people don't vote / (.) [it'll be (.) more of a (.) monarchy / (..) that has control].

More of a monarchy will have control.

(5) [It's the dual one that doesn't].

The dual one doesn't.

(6) So [it's God's love for us / that's poured out in our hearts / by the Holy Spirit given to us].

God's love for us is poured out in our hearts by the Holy Spirit given to us.

(7) And it's the assurance that your God loves you / (.) that he seeks you / (.) that he's with you / that he comes to you / [it's this assurance / that draws you to repentance / and draws you to faith / and draws you to himself].
This assurance draws you to repentance and draws you to faith and draws you to himself.

(8) Salvation is of the Lord / and uh / [it's the Lord's love / who quickens your love].
The Lord's love quickens your love.

(9) [It was a big hood that went over the top of this thing.]
A big hood went over the top of this thing.
Note: Original transcript has 'there was a big hood…' but after listening numerous times, the utterance sounds more to the present author like a token of, 'It was a big hood…'

(10) But I thought [it was his dad who was / in the hospital].
His dad was in the hospital.

(11) But [it's always the federal workers that should take the cut].
Always, the federal workers should take the cut.

(12) And [it was a group of us that went].
A group of us went.

(13) I guess [it probably was / (..) a dozen in our group that went].
A dozen in our group went.

REVERSE WH CLEFTS:

(1) [Farrier's what they're called.]
They're called farrier(s).

(2) I guess [all I can't figure out is / what the square root of negative two thir- (.) th- (.) two thirds is].
The square root of negative two thirds is all I can't figure out.

(3) We tell the world / (..) that [democracy (.) is what (.) we practice].
We practice democracy.

(4) [The close relationship is what makes it so different].

The close relationship makes it so different.

(5) Cause [the track's what it- (.) what's important].

That track is important.

(6) See [the entitlement is what turned the West Side / into a damn desert (.) of oasis and poverty].

The entitlement turned the West Side into a damn desert of oasis and poverty.

(7) But [the designs you see / is what that Italian family saw / and interpretated from Southwest Indian blankets and pottery].

That Italian family saw and interpretated the designs you see from Southwest Indian blankets and pottery.

(8) But [that flood's what left that white line around the lake out there].

That flood left that white line around the lake out there.

(9) [The hauling is what costs so much].

The hauling costs so much.

APPENDIX C

LISTENER JUDGMENT TASK SCRIPTS

PROSODY JUDGMENT TASK SCRIPT:

PRE-TRAINING. During this session, you will hear audio clips of someone talking. These come from a variety of contexts. Some may be difficult to hear. You should try not to listen for the meaning of the clip, but for what sounds emphasized. With each clip, you will be presented with two parts of the sentence you heard in the clip, labeled A and B. Your task is to indicate which of the given parts sounds emphasized. For the training session, you can click on buttons to indicate that A sounds emphasized, B sounds emphasized, or BOTH sound emphasized. We will go through some training examples together.

POST-TRAINING. Now, you'll complete a practice session on your own, and then you can ask any questions you might have and go on to the main part of the task. In the practice, and also in the main part, there are a few additional buttons that we didn't have in the training session. You'll have one button over here (point to far left of screen): NO EMPHASIS, that you can choose if you think nothing sounds emphasized, and you'll have one button over here (point to far right of screen): EMPHASIS SOMEWHERE ELSE that you can you choose if something sounds emphasized, but it does not appear in A or B. You may play each example up to five times, in case you feel you need to hear it again.

NEW/GIVEN JUDGMENT TASK SCRIPT:

PRE-TRAINING. During this session, you will hear short audio clips of someone talking. These segments come from a variety of contexts. Some may be difficult to hear. With each audio clip, you will be presented with a concept from the clip. Your task is to indicate whether the concept is likely to be GIVEN or NEW in the larger speaking context that the clip comes from. By GIVEN, I mean that it is probably already under discussion,

or part of what the speaker and hearers already know or already consider relevant to the discussion. By NEW, I mean that it is probably being brought up for the first time, or it is adding information to the discussion. We'll do a training session, and I'll show you some examples to illustrate given and new concepts.

POST-TRAINING. Just like in the training, you will see two buttons labeled with your response options: given or new. Unlike the training, you will only rate one concept for each sound clip. You may replay each segment up to five times, if you feel you need to hear it again.

CONTRASTIVENESS JUDGMENT TASK SCRIPT:

PRE-TRAINING. During this session, you will hear audio clips of someone talking. These segments come from a variety of contexts. Some may be difficult to hear. With each audio clip, you will be presented with a part of the clip (a word or multiple words), in print on the screen. Your task is to indicate whether anything in the part of the clip that you see is CONTRASTIVE or if the part that you see is NON-CONTRASTIVE in the larger context that the clip comes from. By CONTRASTIVE, I mean that it is being talked about in opposition to something else, or as a member or part of a set of alternatives. This is probably a broader definition for CONTRASTIVE than what you might intuitively think of, because of the second part: I consider something contrastive not only if it is being talked about in opposition to something else, but also if it is talked about as a member or part of a set of alternatives. Here are some examples of elements that are contrastive in their contexts. I have underlined some of the things that I think are contrastive, and I've paraphrased the contrasts below each example; that's the starred part you see. Please look these over and let me know if you have any questions about why I indicated that something was contrastive (show participant these examples on paper):

EXAMPLE 1:
Speaker 1: I love fruit. I think my favorite fruits are <u>apples</u> and <u>peaches</u>.
Speaker2: I love fruit too, but I think <u>my</u> favorites are <u>mangos</u>.

*Mangos / apples, peaches / other possible favorite fruits
*Speaker 1 / Speaker 2

EXAMPLE 2:
Speaker 1: Michael really wants to go to Disney Land this summer.
Speaker 2: Seriously? I thought he said he didn't want to go there.

*Michael wants to go / doesn't want to go

EXAMPLE 3:
Speaker: There was a big group of hikers on this trip together. They had been hiking for hours in the mountains. Most of them were going strong, but one of them was really lagging behind.

*Most of the hikers / one of the hikers
*going strong / lagging behind

EXAMPLE 4:
Speaker 1: What do you think about this painting?
Speaker 2: I think it's really ugly. I would never put it on my wall.
Speaker 1: Hmm. I actually kind of like it.

*negative / positive reaction towards a painting
*Speaker 1 / Speaker 2

During the session, you will hear a short clip, and you will not have the context available to you. You will see one or more words from the clip on the screen. You will not have to decide which part is contrastive, you will only have to say whether anything in the portion of the clip that you see on the screen is CONTRASTIVE or not. So, you are making a guess, because you do not have the context, but it is an educated guess, since you are a native speaker of English. To clarify, you are judging whether the part that you see is contrastive in the larger speaking context of the clip. If you think something in the sound clip is contrastive, but the contrastive part does not appear on the screen, then you should choose NON-CONTRASTIVE. I will go through a training session with you, to make sure that you are comfortable with the task. You may ask questions at any time, but hopefully I can answer most of your questions during the practice session. After that, you will start the main part of the experiment. You may play each example up to five times if you feel you need to hear them again.

RELEVANCE RATING TASK SCRIPT:

During this session, you will hear sound clips with one or more people talking. These sound clips are of varying lengths (a few sentences up to two minutes), and they are taken from a variety of contexts. Some may be difficult to hear because of the recording quality, or movement near the microphones, but please try to pay attention to the speaking. With each segment, you will be presented with a concept to rate. Your task is to rate how relevant the concept is to the sound clip. Some concepts may not seem relevant at all, while others may seem extremely relevant. You are asked to rate relevance on a scale of 1 to 7, with 1 indicating NOT AT ALL RELEVANT, 7 indicating EXTREMELY RELEVANT. You can indicate your response by clicking on a button or pressing a number key. We will start with some practice examples. You'll hear each clip only once.

APPENDIX D

TRAINING AND PRACTICE EXAMPLES

Full example sets are provided below for each task, in a specific format. For each item in example sets, the first line represents the auditory stimulus used; the second line indicates the orthographic stimulus used, and remaining lines vary, depending on the task and are therefore labeled and/or explained at the beginning of the relevant section. Underlining indicates locus of prosodic prominence, based on pitch-prominence (as assessed by the researcher).

PROSODY JUDGMENT TRAINING:

(1) I hate being eighty, until I thought of the alternative.

A. I hate being eighty B. I thought of the alternative

Intended answer: Emphasis in A only

(2) I had never heard of her.

A. I had never B. heard of her

Intended answer: Emphasis in A only

(3) And I feel like I'm in a spaceship.

A. I feel B. I'm in a spaceship

Intended answer: Emphasis in B only

(4) I thought we were gonna wear out the buttons on the phone.

A. we were gonna wear out B. the buttons on the phone

Intended answer: Emphasis in B only

(5) People are the same everywhere.

A. people B. the same everywhere

Intended answer: Emphasis in both A and B

(6) And then I came back into Italy and we went into the invasion of South France.

A. I came back into Italy B. the invasion of South France

Intended answer: Emphasis in both A and B

PROSODY JUDGMENT PRACTICE:

(1) Look at you with the uh little armies down here.

A. look at you B. little armies down here

Intended answer: Emphasis in A only

(2) This is your favorite kind of cake, Mom.

A. your favorite B. kind of cake

Intended answer: Emphasis in A only

(3) Those are two different words.

A. those are B. two different words

Intended answer: Emphasis in B only

(4) I've been sleeping about ten hours every night.

A. I've been sleeping B. ten hours every night

Intended answer: Emphasis in B only

(5) I was constructed inside of some woman's womb.

A. I was constructed B. inside of some woman's womb

Intended answer: Emphasis in both A and B

(6) I didn't even see the fence.

A. I didn't even B. see the fence

Intended answer: Emphasis in both A and B

(7) He didn't quite know what to do.

A. he didn't B. know what to do

Intended answer: Emphasis somewhere else

NEW/GIVEN JUDGMENT TRAINING:

Note: The second and third lines each represent orthographic stimuli for the different groups judging clefted elements (CE-group) and open propositions (OP-group). The

training stimuli were adapted for each group so that the training stimuli would be more like the test stimuli (i.e., longer phrases for the OP group). Orthographic stimulus pairs in the second and third lines for each item are formatted as NEW / GIVEN, as subjects were trained on one intended given item and one intended new item for each training example.

(1) *I hate being eighty, until I thought of the alternative.*

CE-group: hating something / being eighty

OP-group: the (quoted) speaker hating something / being eighty

Transcription of preceding context shown to subjects: "He turned eighty, and someone came to him and said, 'What do you think about that?' And he said…" (regarding President Nixon)

(2) *I didn't even see the fence.*

CE-group: a fence / seeing something

OP-group: a fence / the speaker not seeing something

Transcription of preceding context provided: "When we came back from the trip, we found he'd built the fence. And I said to Debby…"

(3) *And I feel like I'm in a spaceship.*

CE-group: a space ship / the speaker's feelings or experience / a spaceship

OP-group: being in a spaceship / the speaker feeling or experiencing something

Explanation given: This is probably a conversation where the hearer understands hat this speaker is talking about her feelings or experience, and maybe she has been talking about her feelings for a while already; this is the first time she has brought up the *spaceship* metaphor.

(4) *That's an excellent idea.*

CE-group: something being excellent / a particular idea

OP-group: something being excellent / a particular idea

Explanation given: Someone has just brought up an idea, and this speaker is commenting that he thinks the idea is *excellent.* Beyond that, in the context of

this sentence, this speaker has actually just said, "THAT'S a good idea," so even going from 'good' to 'excellent' is adding some information.

CONTRASTIVENESS JUDGMENT TRAINING:

(1) *This is <u>your</u> favorite kind of cake, Mom.*

Your favorite

Intended answer: Contrastive

Commentary: 'your favorite' versus 'my favorite'

(2) *<u>People</u> are the same <u>everywhere</u>.*

people

Intended answer: Non-contrastive

Commentary: 'People' sounds like it's people talked about as a general category. One can imagine the speaker going on to talk about different sub-categories contrastively (e.g., people from Africa, people from Asia, people from Europe…), but this sounds more like a general category.

(3) *<u>I</u> had never heard of her.*

I

Intended answer: contrastive

Commentary: One can imagine that the speaker is reporting a conversation where everyone seemed to know of a certain person, but the speaker had never heard of *her*. So, *I* may be contrastive with some relevant *others*.

(4) *I <u>hate</u> being eighty, until I thought of the alternative.*

I

Intended answer: Non-contrastive

Commentary: Something in the example sounds contrastive: being eighty versus implied alternative of being dead, but the part on the screen does not seem contrastive. This example helps clarify the task of rating the part on the screen, not the whole sentence, with respect to the larger speaking context.

(5) *And then I went back into Italy, and we went into the invasion of South France.*

Italy

Intended answer: Contrastive

Commentary: 'Italy' is a member of relevant set of places, an illustration of the less intuitive second part of the *contrastive* definition.

(6) *I thought I was gonna wear out the buttons on the phone.*

was gonna

Intended answer: Non-contrastive

Commentary: There's nothing that really sounds contrastive about this example; it doesn't sound like the speaker is talking about future versus past or non-future, which would be a possible contrast for 'was gonna'.

RELEVANCE RATING TRAINING:

Note: In the examples below, everything in italics is part of the auditory stimuli. The two lines following the auditory stimuli represent the orthographic stimuli. As with the given/new judgment training examples, different orthographic stimuli were given for the different groups judging clefted elements (CE-group) and open propositions (OP-group). Orthographic stimuli are labeled accordingly.

(1) *Speaker 1: What does that have to do with heaven and hell in the book?*

Speaker 2: Well, I'm just sort of reiterating. I could read you some.

Speaker 1: No.

Speaker 2: I mean is that allowed?

Speaker 1: No, I don't wanna hear anything out of a book with a chapter called 'Heaven and Hell.'

Speaker 2: You don't.

Speaker 1: No.

Speaker 2: Mmkay. Well then let's talk about our vacation.

Speaker 1: I'm gonna be closed-minded about it.

Speaker 2: Oh dear. That's hell.

Speaker 1: Well, I didn't like the book the way I, the minute I looked at it.

Speaker 2: You didn't?

Speaker 1: No.

Speaker 2: That's 'cause you—

Speaker 1: It's because I have my own ideas about it I guess, that I'm, pretty comfortable with.

Speaker 2: Oh.

CE-Group: a book about heaven and hell

OP-Group: the guy not liking something

Note: Orthographic stimuli for this training example were intended to be highly relevant.

(2) *Speaker 1: We're all tired.*

Speaker 2: That must be it.

Speaker 1: Yeah, simple, simple explanation—

Speaker 3: I mean when Mary tells me to get sleep over the weekend, you know I need to get sleep over the weekend.

Speaker 1: That's good you're getting a good rest. I have been getting about ten hours.

Speaker 3: I know, I've been sleeping about ten hours, every night. And I'm still just like— But I think I'm over it faster than I would be.

Speaker 1: I was gonna ask the doctor, I'm like, what is wrong with me that I am sleeping so much?

CE-Group: someone named Kevin

OP-Group: someone being pregnant

Note: Orthographic stimuli for this training example were intended to be minimally relevant.

(3) *Speaker 1: Let's see {faded out name} got me a nice shirt. Umm. What color. Uh, all these, plain cotton? It's blue with white stripes?*

Speaker 2: Oh, that's pretty.

Speaker 1: Hundred percent cotton, and uh, they're comfortable as hell, and she likes the color and so on.

Speaker 2: Yeah, well you look good in blue anyway.

Speaker 1: Well, that's what everybody says. In fact, I stopped wearing it to work because the uh, some of the ladies, uh, were commenting on my shirt, so I stopped wearing that one, I figure it's none of their damn business.

Speaker 2: They're just being nice. What do you think, they're putting a make on you or something?

Speaker 1: No. But I don't wanna give 'em any reason to, so (laugh).

CE-Group: a blue shirt

OP-Group: the woman buying something

Note: CE-Group orthographic stimulus was intended to be highly relevant, and OP-Group orthographic stimulus was intended to be of low relevance.

(4) *So, I um, decided and I went to see him. And I entered his very simple little room, I looked around it. And I remember there were very coarse curtains and the, sun shown through them, made them like gold. And I stood in his room, and he was very nice gentlemanly man. He didn't quite know what to do, so he came, and he stood opposite me, and looked at me. And, I saw a painting on his wall.*

CE-Group: a glass of water

OP-Group: the speaker seeing something in a room

Note: CE-Group orthographic stimulus was intended to be of low relevance, and OP-Group orthographic stimulus was intended to highly relevant.

REFERENCES

AKMAJIAN, ADRIAN. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1.146-168.

BREIVIK, LEIV EGIL. 1986. Some remarks on cleft sentences in present-day English. *Linguistics across Historical and Geographical Boundaries, Volume II*, ed. by Kastovsky, Deiter and Aleksander Szwedek, 815-826. Berlin: Mouton de Gruyter.

BOERSMA, PAUL AND DAVID WEENINK. 2008. *Praat: doing phonetics by computer, Version 5.0.36*. Amsterdam: Institute of Phonetic Sciences.

CALUDE, ANDREEA. 2008. Demonstrative clefts and double cleft constructions in spontaneous spoken English. *Studia Linguistica* 62.78-118.

CHAFE, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. *Subject and Topic,* ed. by Li, Charles N., 25-56. New York: Academic Press.

COLLINS, PETER. 1991. *Cleft and Pseudo-cleft Constructions in English.* London: Routledge.

COLLINS, PETER. 2004. Reversed what-clefts in English: Information structure and discourse function. *Australian Review of Applied Linguistics* 27.63-74.

COLLINS, PETER. 2006. *It*-clefts and *wh*-clefts: Prosody and pragmatics. *Journal of Pragmatics* 38.1706-1720.

CREIDER, CHET A. 1979. On the explanation of transformations. *Discourse and Syntax.* ed. by Givón, Talmy, 3-21. New York: Academic Press.

DECLERCK, RENAAT. 1988. *Studies on Copular Sentences, Clefts, and Pseudo-clefts.* Leuven: University Press; Dordrecht.

DELIN, JUDY. 1989a. *Cleft Constructions in Discourse.* Ph.D. Dissertation, Centre for Cognitive Science, University of Edinburgh.

DELIN, JUDY. 1989b. *The Focus Structure of* It-*Clefts*. Centre for Cognitive Science, University of Edinburgh.

DELIN, JUDY. 1990. A multi-level account of cleft constructions in discourse. *Proceedings of the 13th Conference on Computational Linguistics*. 2.83-88.

DELIN, JUDY. 1995. Presupposition and shared knowledge in *it*-clefts. *Language and Cognitive Processes* 10.97-120.

DELIN, JUDY AND JON OBERLANDER.1995. Syntactic constraints on discourse structure: the case of *it*-clefts. *Linguistics* 33.465-500.

DU BOIS, JOHN W., CHAFE, WALLACE L. MEYER, CHARLES, AND THOMPSON, SANDRA A. 2000. *Santa Barbara corpus of spoken American English, Part 1.* Philadelphia: Linguistic Data Consortium.

DU BOIS, JOHN W., CHAFE, WALLACE L. MEYER, CHARLES, THOMPSON, SANDRA A., AND MARTY, NII. 2003. *Santa Barbara Corpus of Spoken American English, Part 2.* Philadelphia: Linguistic Data Consortium.

DU BOIS, JOHN W. AND ROBERT ENGLEBRETSON. 2004. *Santa Barbara Corpus of Spoken American English, Part 3.* Philadelphia: Linguistic Data Consortium.

DU BOIS, JOHN W. AND ROBERT ENGLEBRETSON. 2005. *Santa Barbara Corpus of Spoken American English, Part 4.* Philadelphia: Linguistic Data Consortium.

FILLMORE, CHARLES, PAUL KAY, AND MARY O'CONNOR. 1988. Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language* 64. 501-538.

GIVÓN, TALMY. 2001. *Syntax, Volume II.* Amsterdam: John Benjamins Publishing Co.

GOLDBERG, ADELE. 1995. *A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

HALLIDAY, MICHAEL A. K. 1967. Notes on transitivity and theme in English, Part II. *Journal of Linguistics* 3.199-244.

HANKAMER, JORGE. 1974. On the non-cyclic nature of wh-clefting. *Chicago Linguistics Society* 10.221-233.

HEDBERG, NANCY. 1990. *Discourse Pragmatics and Cleft Sentences in English.* University of Minnesota dissertation.

HIGGINS, F. ROGER. *The Pseudo-cleft Construction in English.* MIT dissertation.

JESPERSEN, OTTO. 1927. *A Modern English Grammar, Volume III.* London: Allen and Unwin.

JESPERSEN, OTTO. 1937. *Analytic Syntax*. London: Allen and Unwin.

JESPERSEN, OTTO. 1949. *A Modern English Grammer, Volume VII.* London: Allen and Unwin.

LAMBRECHT, KNUD. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

LAMBRECHT, KNUD. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39.463-516.

MACWHINNEY, BRIAN. 2007. The Talkbank Project. In J. C. Beal, K. P. Corrigan & H. L. Moisl (Eds.), *Creating and Digitizing Language Corpora: Synchronic Databases, Vol.1*. Houndmills: Palgrave-Macmillan.

OBERLANDER, JON AND JUDY DELIN. 1996. The function and interpretation of Reverse *wh*-clefts in spoken discourse. *Language and Speech* 39.185-227.

PRINCE, ELLEN. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54.883-906.

SELKIRK, ELISABETH O. 1984. *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, MA: MIT Press.

SELKIRK, ELISABETH O. 1995. Sentence prosody: Intonation, stress, and phrasing. *The Handbook of Phonological Theory,* ed. by J. Goldsmith, 550-569. Oxford: Blackwell.

SPEKTOR, LEONID. (2008). *CLAN: Computerized Language Analysis, Version 05.* Pittsburgh: Carnegie Mellon University.

VAINIO, MARTTI AND JUHANI JÄRVIKIVI. 2007. Focus in production: Tonal shape, intensity and word order. *Journal of the Acoustical Society of Americal* 121.EL55-EL61.

WEINERT, REGINA AND JIM MILLER. 1996. Cleft constructions in spoken language. *Journal of Pragmatics* 25.173-206.