UCLA
UCLA Electronic Theses and Dissertations

Title
Subclausal Quotation Intonation

Permalink
https://escholarship.org/uc/item/8xg36556

Author
Sturman, Bethany Christine

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Subclausal Quotation Intonation

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

Bethany Christine Sturman

2022

ABSTRACT OF THE DISSERTATION

Subclausal Quotation Intonation

by

Bethany Christine Sturman

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2022

Professor Jessica L. Rett, Co-Chair

Professor Sun-Ah Jun, Co-Chair

While quotation has been a popular area of study among semanticists, there has been a dearth of work related to how people say quotations in spoken speech, in the absence of orthographic quotation marks. Using corpus methodology, I identify three key intonational features of subclausal quotation in spoken American English: the Emphatic Juncture, pitch range reset, and a quotation-final IP break. I then validate these three intonational features and determine which is most responsible for identifying the presence of subclausal quotation using a comprehension experiment. I then discuss how these three intonational features relate to semantic analyses of subclausal quotation (Cappelen and Lepore, 1997; Potts, 2007). Subclausal quotation intonation also poses challenges to the current understanding of intonational theory, particularly regarding the Strict Layer Hypothesis (Selkirk, 1986), through the discovery of evidence for embedded IPs in American English.

The dissertation of Bethany Christine Sturman is approved.

Byron Ahn

Dylan Bumford

Sun-Ah Jun, Committee Co-Chair

Jessica L. Rett, Committee Co-Chair

University of California, Los Angeles

2022

To my grandmother Elvira Garfield . . .

who was my biggest fan in my academic pursuits. I am so proud to carry on your feminist

dreams.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The acknowledgements I write here are woefully inadequate to convey the importance of the support I've received from my academic community, family, and friends throughout my graduate school career.

First, to my co-chairs, Sun-Ah and Jessica: thank you for your continual encouragement, support and challenge to go beyond what I thought I could. This dissertation would not have been finished without your mentorship and guidance. Sun-Ah, I will fondly remember marveling over sound files with you, particularly the ones that broke the model. Jessica, you helped me uncover a love of semantics, and for that I will be forever grateful. Thank you both for the care you showed for me as a person and the time you spent developing me into the scholar I am today.

To my committee members: Byron, thank you for serving as my external member. Our conversations were always stimulating and pointed my research in a good direction. Dylan, your questions were always pithy and insightful.

To Adam: from our UCLA office (The Contour Complex) to Oregon, you've seen me through so much of this process. Thanks for being my academic sounding board, R and stats expert, and MVP dog co-parent.

To my family: thank you for cheering me on even when you had no idea what I was doing or why. I love you all so much!

Last but certainly not least, to my partner Rebecca: you are the blessing I never thought I'd find. Thank you for loving me through the ugly parts of bringing this dissertation into being. Let's agree to not get any more advanced degrees.

2014   B.S. (Linguistics), Massachusetts Institute of Technology.

2016–2021   Teaching Assistant, Linguistics Department, UCLA.

2021   M.A. (Linguistics), UCLA, Los Angeles, California.

CHAPTER 1

Introduction

1.1 Setting the stage

Quotation is a phenomenon that has long piqued the interest of both semanticists and philosophers of language. There is something fundamentally metalinguistic about quotation; speakers use quotation to talk about language or, at a minimum, to draw attention to the language being used. The metalinguistic component of quotation has puzzled scholars seeking to analyze the meaning of quotation. The situation is further complicated by the fact that there are several different types of quotation.

Despite the complexity of and sustained interest in quotation, little has been said about how people actually say quotations in spoken speech. More precisely, there is a dearth of research on how people mark some content as a quotation in the midst of a larger utterance. This is the type of quotation referred to as subclausal quotation. In this dissertation, I explore how people mark subclausal quotation using intonation in spoken American English, ultimately identifying three key intonational features. I then explore how those key intonational features might be incorporated into existing semantic analyses of subclausal quotation.

1.2 Types of quotation

There are several different types of quotation that have been observed and analyzed. I will discuss three general types of quotation here: (pure) metalinguistic quotation, full clause quotation, and subclausal quotation.

1.2.1 Metalinguistic quotation

Some of the earliest work on quotation (Tarski, 1956; Quine, 1980) focuses on the metalinguistic aspect of quotation, noting the distinction between the use of an expression and the mention of it. A quoted expression functions to mention that particular linguistic expression itself rather than its denotation (i.e. its use). An example of a quotation being used purely for metalinguistic mention can be seen in 22:

(1)   a. "UCLA" is an acronym.

     b. "Linguistics" has eleven letters.

An important feature of this sort of metalinguistic quotation is its opacity. A substitution of a synonymous or co-referential expression does not preserve the truth value of the sentence, as shown in 23:

(2)   a. "UCLA" is an acronym. (*True*)

     b. "University of California, Los Angeles" is an acronym. (*False*)

This type of quotation has also been referred to as *pure quotation* because it is solely functioning to metalinguistically mention the quoted content rather than to use it. This contrasts it with other types of quotation where the quoted content is simultaneously used and mentioned.

1.2.2 Full clause quotation

As the name suggests, full clause quotation involves flanking entire clauses with quotation marks. Some examples of full clause quotation can be seen in 3 (from Partee (1973)).

(3)   a. The other day Tom said to me, "My grandfather was killed with a knife by a bachelor."

     b. The other day Tom said to me, "An unmarried man used a knife to cause the father of one of my parents to die."

In addition to the material being flanked by the quotation marks in written speech, full clause quotation is characterized by embedding under a verb of saying. The subject of the verb of saying is the source of the quoted material.

The synonymous content substitution between (a) and (b) in 3 does not result in the examples being synonymous. The contrast between (a) and (b) shows that like metalinguistic quotation, full clause quotation is sensitive to mention rather than simply the meaning of the quoted content (Partee, 1973).

1.2.3 Subclausal quotation

The type of quotation being investigated in this dissertation clearly contributes more than simply to mention the linguistic expression within the quotation. There are several sources of evidence that show subclausal quotation can function beyond pure mention. First of all, subclausal quotations can fill whatever position the quoted material does (Cappelen and Lepore, 1997). In other words, the content within the quotation is grammatically incorporated into the rest of the sentence as it would be if the quotation marks were absent. Several examples from the SQ corpus demonstrating this are shown below:

(4)   Romeo Mattison was "training" Goron Pezar's wife.   *SQ as V, from Barry*

(5)   We're throwing a "fundraiser" on your birthday.   *SQ as N, from Schitt's Creek*

(6)   It's almost offensive when we're criticized on how quote-unquote
      "terrible" healthcare is in this country.   *SQ as Adj, from C-SPAN*

(7)   Fanny and Freddie were found to "have been cooking the books." *SQ as VP, from*
      C-SPAN

The above examples show that subclausal quotations are not automatically collapsed into a single semantic type simply by virtue of being quotations. This is compelling evidence that subclausal quotation employs both the use and mention of the quoted expression (Davidson, 1979).

The complexity of subclausal quotation is not new to semanticists and philosophers

(see De Brabanter (2010) for an overview). In fact, it has been difficult for scholars in this area to even agree on a term for the phenomenon (or perhaps phenomena) at issue. It has been referred to as hybrid quotation (Recanati, 2001), mixed quotation (originating from mixed direct and indirect quotation (Partee, 1973)), incorporated quotation (Clark and Gerrig, 1990), double-duty quotation (García-Carpintero, 2003), impure quotation (Gómez-Torrente, 2003) or subclausal quotation (Potts, 2007). There has also been a distinction in the literature regarding scare quotes, distinguishing them from other types of subclausal quotation. For the purposes of this dissertation, I am referring to the phenomenon as subclausal quotation. This is an attempt to be theory-neutral, since the phenomenon of interest is quite literally a quotation that occurs on the subclausal level.

1.3   Quotation is more than quotation marks

Scholars who have studied quotation typically do so by discussing orthographic quotation (Davidson (1979), Cappelen and Lepore (1997), a.o.) or performance of quotation in coarse, high-level terms (Clark and Gerrig (1990), Potts (2007)).

But people use quotation in spoken and signed language, not just orthographically. This has been recognized by semanticists for quite some time, as exemplified in this quote from Partee (1973):

> "Quotation marks are an orthographic device with no direct analog in spoken language. [...] Perhaps the nearest thing spoken language has to a natural quotation device is an intonational one, namely, a pause before and after the quoted sentence, plus (imitation of (?)) the intonation the sentence would have in isolation."   (Partee (1973), p. 410)

In order to build a comprehensive theory of quotation, we first ought to understand how people say quotations (i.e. how speakers mark quotation). Speakers have a variety of strategies available to mark quotation, such as air quotes (produced by flexing the middle and index finger while saying the quoted material) and the lexical markers *quote* or *quote-*

4

unquote. But as I will show in this dissertation, speakers also reliably use intonation to mark quotation in spoken speech.

While one could hypothetically study all of the types of quotation outlined in Section 1.2 above, I have limited the scope of this dissertation to focus on subclausal quotation. There are a few reasons for choosing subclausal quotation rather than the other types. First, I was intrigued by the complexity of simultaneous use and mention. I also suspected that this type of quotation would most readily display intonational marking since the quoted material is a substring embedded within a larger sentence. With the potential for non-quoted material both before and after the quotation, there must be some method that speakers use to mark the start and the end of a subclausal quotation.

1.4 Understanding subclausal quotation intonation unearths important insights

Understanding how speakers use intonation to mark subclausal quotation is a boon to linguists in several ways. First, it paves the way for a richer semantic analysis of quotation that incorporates intonation. Incorporating intonation eliminates the need to rely on silent operators when extending a theory of quotation beyond orthography. Incorporating intonation also makes for a more fully compositional semantic theory. Speakers are using intonation as well as words and structure, so a fully compositional theory ought to include the meaning contributed by intonation.

Second, understanding how speakers use intonation to mark subclausal quotation can illuminate connections to other phenomena. In particular, the intonational feature used to mark the start of a subclausal quotation, known as the Emphatic Juncture, appears in other constructions, most notably Transparent Free Relatives (e.g. John mastered *what he called the sluve*.) (Smith, 2003; Grosu, 2003). The similarity in the intonational marking of subclausal quotation and Transparent Free Relatives suggests a possible connection in the semantics of these phenomena. I will explore this in more detail in Chapter 5.

Lastly, studying how speakers use intonation to mark subclausal quotation broadens our understanding of intonational theory. Subclausal quotation intonation challenges some key components of the Autosegmental-Metrical theory of English intonation undergirding the MAE_ToBI transcription system. First, the Emphatic Juncture challenges the conventional domain of downstep: it is not limited to two High tones within the same intermediate phrase. Second, the existence of embedded intonation phrases challenges the Strict Layer Hypothesis as it relates to prosodic structure. It is in fact a violable constraint rather than an inviolable principle. I will discuss both of these findings in greater detail in Chapter 3.

1.5 Methodology for understanding subclausal quotation intonation

1.5.1 Initial discovery: Subclausal quotation corpus (Chapter 2)

The first phase of the research was the assembly and analysis of the subclausal quotation corpus. In Chapter 2, I describe the methodology, assembly, analysis and results of this corpus. The intonation of the corpus tokens (N=90 tokens) was transcribed in the MAE_ToBI framework. Analyzing these transcriptions surfaced three key intonational features of subclausal quotation: an initial emphatic juncture at the start of the quotation, a pitch range reset on the quoted material, and an IP boundary at the end of the quotation. The presence of these features was remarkably consistent throughout the corpus.

Chapter 2 also includes a brief introduction to the MAE_ToBI transcription system (Beckman and Ayers-Elam, 1997; Beckman et al., 2005) for those who are less familiar with intonation.

1.5.2 Verification and refinement: Perception experiment (Chapter 4)

1.5.2.1 Motivation

Following the corpus analysis, I designed and ran an experiment isolating each of the three key intonational features. The goal was to learn more about how each feature

6

contributes to communicating the presence of subclausal quotation.

1.5.2.2 Methodology

In the experiment, participants heard an audio recording of a full sentence and were asked whether they thought the speaker used a gesture. The design of this question was to use the air quotes gesture as a proxy for orthographic quotation marks, which would signal the presence of a quotation. After this question, participants were asked how confident they were in their gesture answer on a 1-7 Likert scale. The "Yes Gesture" and "No Gesture" responses were introduced to the participants via a brief training at the start of the experiment.

The experiment consisted of a 1x5, within subject design. Of the five conditions, two were control conditions and three were experimental. The control conditions were floor and ceiling conditions, with either none of the intonational features or all three of the features of subclausal quotation intonation present, respectively. The three experimental conditions each isolated one feature of subclausal quotation to the exclusion of the other two. For example, the Emphatic Juncture condition used an emphatic juncture right before the beginning of the quoted material but did not employ a pitch range reset or an IP break at the end of the quoted material. The five conditions are shown in Table 4.1.

Condition	Left EJ	Pitch Range Reset	Right IP break
Neutral prosody (floor control)			
Emphatic Juncture (EJ)	X		
Pitch Range Reset (PRR)		X	
Right IP break (RIP)			X
SQ prosody (ceiling control)	X	X	X

Table 1.1: Each of the experiment conditions based on which intonational features were used. Note that in the Neutral prosody condition (first row), none of the intonational features that mark subclausal quotation were used.

1.5.2.3 Results

Of the three experimental conditions, the Emphatic Juncture condition had the highest rate of "yes" gesture responses, slightly higher than the other two experimental conditions. This result corresponds to participants detecting the presence of subclausal quotation at a slightly higher rate. This indicates that the Emphatic Juncture is the best of the three features at signaling the presence of subclausal quotation.

All three of the experimental conditions patterned more closely to subclausal quotation prosody control than to the neutral prosody control. In other words, each intonational feature served as a weaker signal than the full set of intonational features that subclausal quotation was present, but a signal to its presence nonetheless. This constitutes evidence that there is meaning present in intonational features below the tune level.

1.6 Overview of theoretical contributions

1.6.1 Intonation theory (Chapter 3)

The subclausal quotation corpus surfaced several interesting insights that challenge our current understanding of the theory of intonation in addition to its descriptive value. The two primary theoretical insights are the Emphatic Juncture and embedded IPs.

First reported in Sturman (2019) as part of work that set the stage for this dissertation, the Emphatic Juncture is a specialized type of Intonation Phrase boundary. The boundary tone sequence is a plateau, but it also includes an obligatory pause. In some cases, such as when it marks subclausal quotation, it is followed by a pitch range reset. But in other cases, such as in emphatic speech, downstepping is licensed across the juncture. This is a significant departure from the usual convention for the domain of downstep, which is that two high tones must be within the same intermediate phrase.

The second major finding related to the theory of intonation is the discovery of evidence for embedded intonation phrases (IPs) in English. In particular, a subclausal quota-

tion forms an IP which is nested inside a larger IP. The existence of embedded IPs is supported by three sources of evidence. First, splicing out the quotation leaves a continuous-sounding pitch track in many cases. The second and third forms of evidence come from predictions made by the possibility of embedding IPs. If the nested IP is embedded after the nuclear pitch accent of the larger IP, the content after the embedded IP will appear to be a headless IP since it has been stranded from its nuclear pitch accent. This prediction is born out in the corpus data. There is a second prediction that an embedded IP would be inserted in a location that splits a bitonal pitch accent. This was attested for L+H* in the corpus, though not for the other two bitonal pitch accents in English.

1.6.2 Semantics (Chapter 5)

1.6.2.1 Integrating intonation into semantics

Based on the results of both the corpus and the experiment, I detail how the key intonational features externalize subclausal quotation in the framework of Potts (2007). First, the Emphatic Juncture serves the roles of signaling the presence of quotation and marking the beginning of the quoted material. In the framework of Potts (2007), the Emphatic Juncture is the intonational externalization of **quote-shift**.

The right IP boundary marks the end of the quoted string. In order for the meaning of a subclausal quotation to be calculated, the string must be clearly defined. As a large juncture, the IP boundary is often employed to signal large syntactic boundaries. As such, it is a natural candidate to mark the end of a quotation.

The pitch range reset does not have an obvious direct correlate in the semantic analysis the way the other two features do. Instead, it functions to increase the acoustic prominence of the quoted material. This helps the quoted material to stand out in relation to the surrounding content.

1.6.2.2 Communicative uses of subclausal quotation

The subclausal quotation corpus demonstrated that speakers employ subclausal quotation for a number of different communicative functions. I propose three different communicative uses for subclausal quotation: appealing to authority, speech reporting, and epistemic distancing. I analyze these three uses as being differentiated pragmatically. Following the Maxim of Relevance, the speaker employs a subclausal quotation because it is relevant that x said u, where x is the source of the quotation and u is the quoted content.

CHAPTER 2

Subclausal Quotation Corpus

2.1 Empirical motivation

In this chapter, I present the collection, analysis and results of a subclausal intonation corpus assembled from natural speech. This corpus provided a variety of both quantitative and qualitative research results that set the groundwork for a subsequent perception experiment (Chapter 4).

In addition, the corpus unearthed several interesting insights that spurred some proposed expansions and refinements to our current understanding of intonational theory and suggest promising avenues for future research. In Chapter 3, I expound on the results of the corpus through a theoretical analysis of the core intonational features of subclausal quotation and what these findings mean for the theory of intonation more broadly.

2.2 A brief description of the MAE_ToBI transcription system

The intonation of the subclausal quotation corpus was transcribed using the MAE_ToBI transcription system (Beckman and Ayers-Elam, 1997; Beckman et al., 2005). I describe the theoretical origins of the system in Chapter 3, but I am including a brief description of the system here to aid readers who are less familiar with MAE_ToBI. Then, in Section 2.2.3 I detail the ways in which my labeling in this dissertation deviates from the canonical labeling conventions (Beckman et al., 2005), which is intended to guide readers who are more familiar with labeling intonation.

The MAE_ToBI transcription system (short for Mainstream American English Tones

and Breaks Indices) includes three critical tiers[1] for labeling intonation: the Words tier, the Tones tier, and the Breaks tier. The separation of the Tones and Breaks tier corresponds to the fact that prosody is composed of both rhythm and tonal marking (Beckman et al., 2005). These tiers appear below the spectrogram of a given utterance, which includes a pitch track showing the speaker's fundamental frequency (f0) throughout the utterance. An example spectrogram with the Words, Tones and Breaks tiers is shown in Figure 2.1.
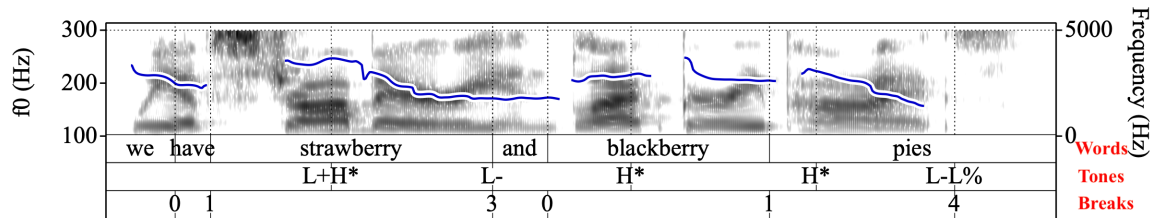
Figure 2.1: An example utterance (*We have strawberry and blackberry pies*) with the pitch track overlaid on the spectrogram. The three tiers (Words, Tones, and Breaks) are included below the spectrogram.

2.2.1 The Tones tier

The Tones tier marks the phonological tonal targets of an utterance, including pitch accents, phrase accents, and boundary tones. Pitch accents (e.g. L+H* and H* in Figure 2.1) mark words that are prominent relative to other words in the utterance. Pitch accents can be composed of one or two tonal targets. The starred tone of a pitch accent aligns to the stressed syllable of a word. Consider, for example, the L+H* pitch accent aligned with *strawberry* in Figure 2.1. The **H*** component of this pitch accent aligns to the stressed syllable of the word (i.e. *straw-*) and is realized as a pitch peak. The **L** component is realized as a trough before the stressed syllable, in this case on *have*. Pitch accents do not have a direct correlate on the Breaks tier.

Phrase accents are the second type of tone which is transcribed on the Tones tier. Phrase accents (L-, H- or !H-) mark the right edge of an intermediate phrase. The phrase accent is marked on the Tones tier after the final word of the intermediate phrase. This

---

[1]The Miscellaneous tier is not included here.

is not always where the phrase accent is realized, however. If there are words separating the final pitch accent of an intermediate phrase and the end of the intermediate phrase, the phrase accent is realized beginning on the word following the pitch accented word. Phrase accents are marked on the Breaks tier as *3*.

Boundary tones are the final type of tone on the Tones tier. Boundary tones (L% or H%) mark the end of an intonation phrase. The boundary tone is marked on the Tones tier after the final word of the intonation phrase. The boundary tone is realized on the final syllable of the intonation phrase. Boundary tones are marked on the Breaks tier as *4*.

2.2.2 The Breaks tier

The Breaks tier uses numerical indices (0-4) to mark the prosodic phrasing of an utterance. A break index is placed after every word, and the size of the index corresponds to the size of the break (also known as a juncture). The larger the break index, the larger the juncture. A phrase-medial word boundary is labeled as *1*. The *0* label is used when two orthographic words form a single prosodic word (Peperkamp, 1997), such as a determiner forming a prosodic word with a noun. The end of an intermediate phrase is labeled as *3*, and the end of an intonation phrase is labeled as *4*.

2.2.3 Deviations from standard MAE_ToBI

The labeling conventions I use in this dissertation deviate from the most recent iteration of MAE_ToBI in two respects. First, I do not employ the break index *2* to indicate mismatches between break size and the presence of a phrase accent. Instead, following the discussion at the 2004 ToBI workshop and taught in Sun-Ah Jun's Ling 111/211, I use *1m* to indicate a phrase accent is present but the size of the break is a word-level break rather than an intermediate phrase-level break. Relatedly, I use *3m* to indicate the break size corresponds to an intermediate phrase break but there is no phrase accent present. According to the MAE_ToBI transcription conventions, both of these situations would be labeled with the break index *2*.

Second, I do not strictly observe that the domain of downstep needs to be within a given intermediate phrase. Descriptively speaking, downstep is the process of a high tonal target being realized slightly lower than a previous high target. According to the MAE_ToBI conventions, that previous high target must be within the same intermediate phrase. As I will show with data involving the Emphatic Juncture, however, the conventional domain of downstep seems too strict. Instead, I label downstep based on a high target's height relative to the previous high target, regardless of whether that previous high target is part of the same intermediate phrase.

2.3 Methodology for corpus assembly

The subclausal quotation corpus is comprised of 90 tokens from a variety of sources. The tokens were gathered by the author and two undergraduate research assistants (Dylan Ross and Theo Chen) who were trained in transcribing intonation using the MAE_ToBI transcription system.

2.3.1 Sources for the corpus

In an attempt to simulate the gambit of settings in which natural language occurs, I did not enforce any exclusion criterion based on source, so long as the speaker was an American English speaker. Tokens were gathered from a wide variety of sources, from more formal registers such as National Public Radio (NPR) and TED Talks to instances from acting in shows like *The Office*, *Schitt's Creek*, *Barry*, and *The Good Place*, to C-SPAN callers and even YouTube videos that had searchable text transcripts. YouTube transcripts were found using the YouGlish search engine by searching for the lexical markers *quote* and *unquote*. As such, the corpus spans a variety of settings, styles and contexts. A full list of the sources and the number of tokens from each source is presented in Table 2.1.

Source	Number of tokens
C-SPAN	62
YouGlish	14
National Public Radio (NPR)	6
The Office	4
Schitt's Creek	2
The Good Place	1
Barry	1

Table 2.1: Sources used in the Subclausal Quotation Corpus and the number of tokens per source

2.3.2 Criteria for inclusion

There were three distinct criteria which constituted sufficient evidence for a given token to be added to the subclausal quotation corpus. This evidence is independent of intonation, being either lexical, orthographic, or gestural. The criteria for a token's inclusion in the corpus varied based on whether the source for the token was audio only or audio-visual. The lexical criterion was available for both source types. According to the lexical criterion, a token was considered an instance of subclausal quotation if the speaker used the lexical markers *quote*, *quote...unquote*, or *quote-unquote*. The orthographic criterion referenced the transcript of the token. If the recording included a transcript, the presence of orthographic quotation around a word or phrase was sufficient evidence for inclusion in the corpus. Lastly, the gestural criterion applied only if the source was audio-visual. The speaker employing a finger quotation gesture (i.e. air quotes) was taken as evidence that they were producing an instance of subclausal quotation.

Assuming that a core premise of this dissertation holds, namely that English speakers can and do mark instances of subclausal quotation using intonation, it seems possible that intonation itself might be a sufficient criterion. As such, one might contend that the set of criteria described above was overly conservative. In fact, this was an intentional design

15

choice. One purpose of this dissertation is to provide empirical evidence to support the intonational marking of subclausal quotation. Using intonation to identify instances of subclausal quotation for the corpus whose purpose is to identify the intonation used to mark subclausal quotation leads to an obvious issue of circularity. To avoid this pitfall, I selected only inclusion criteria that were independent of intonation. This is by no means meant to imply, however, that I believe these criteria would exhaustively select all instances of subclausal quotation.

2.3.3 Other metrics

There were two relatively minor additional criteria for exclusion. First, non-medial instances of subclausal quotation were excluded. This was to enable the observation of any intonational marking happening adjacent to the edges of the quotation (e.g. insertion of junctures, boundary tone selection, etc.) The other minor criterion of exclusion was the language. Any token not produced by an American English speaker was excluded (i.e. every utterance must be valid to annotate using the MAE_ToBI transcription system).

There are many other metrics one could use as inclusion or exclusion parameters for the corpus. For example, there are metrics related to the content of the utterance, syntactic position of the quotation, or even characteristics about the speaker. For most of these possible metrics, assembly of the corpus was agnostic to their status. One factor of note that was not a consideration for inclusion or exclusion was whether the subclausal quotation appeared after a verb of saying (e.g. *say, yell, whisper*, etc.). Verbs of saying often appear before direct quotation or when a sentence includes a mix of direct and indirect quotation (i.e. what some semanticists and philosophers have called *mixed quotation*). An example of mixed quotation (originally from Davidson (1979)) is shown in 8.

(8) Quine said that quotation has "a certain anomalous feature."

This particular sub-type of quotation has been a topic of interest for semanticists and philosophers of language (see De Brabanter (2010) for an overview), but as I will show in Section 2.6.1, it represents a subset of the observed uses of subclausal quotation. Since I

am interested in learning about the intonation of subclausal quotation more broadly than just in cases of mixed direct and indirect speech reports, I opted to not include the presence of a verb of saying as a criterion for inclusion or exclusion in the assembly of the corpus.

2.4 Summary of corpus findings

The intonation of the corpus tokens was analyzed in the MAE_ToBI framework (Beckman et al., 2005). As is standard for a MAE_ToBI analysis, the assessment of the intonation of the corpus was done at a phonological level. This means that the presence of an intonational feature was measured categorically or discretely (i.e. either present or absent) rather than gradiently or continuously, as would be typical in a more phonetically-driven analysis.

Analyzing the corpus revealed that American English speakers consistently utilize three key intonational features to mark subclausal quotation. First, the left edge of the quoted material is often marked with an Emphatic Juncture (Sturman, 2019). Speakers insert an emphatic juncture slightly less frequently, however, when they mark the start of the quotation using a lexical marker. Second, there is often a pitch range reset at the start of the quoted material, which results in the pitch accents associated with the quoted material being produced with higher **H** targets. Finally, the right edge of the quoted material is almost always marked with an Intonation Phrase (IP) boundary. I will now discuss each of the three key intonational features of subclausal quotation in greater detail.

2.4.1 Feature 1: Left edge emphatic juncture

Speakers often mark the start of a subclausal quotation through the insertion of an IP boundary. It is not just any ordinary IP boundary, however. In fact, speakers employ a highly specialized sub-type of IP boundary called the Emphatic Juncture in this location. To better illuminate the significance of the Emphatic Juncture, I will first describe the generic form of an IP boundary and then explain more about the Emphatic Juncture,

17

including what it looks like and where else it is found.

2.4.1.1 The generic form of the Intonation Phrase (IP) boundary

The Intonation Phrase (IP) is the largest prosodic phrase in English (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986). In the MAE_ToBI transcription system, the IP is labeled on the Breaks tier as *4*. (For comparison, a simple word boundary is labeled as *1* and an intermediate phrase boundary is labeled as *3*.) Each IP consists of one or more intermediate phrases (ip), which in turn are composed of words or syllables. Similar to syntactic structure, the hierarchical structure of English intonation can be depicted using a branching tree diagram. Figure 3.1 shows the maximal possible prosodic structure (including locations where recursion is possible) for an utterance on the left and the smallest possible structure on the right.

Figure 2.2: The hierarchical phrasal structure of English intonation according to the model put forth in Beckman and Pierrehumbert (1986). IP = Intonation Phrase, ip = intermediate phrase, T* = pitch accent, T- = phrase accent, T% = boundary tone. Figure from Sun-Ah Jun's Ling 111/211 class handout

The right edge of each IP is marked in two important ways that help the listener detect that the IP is ending. First, the final word, and in particular the final syllable, in an IP

is lengthened in comparison to the same word's duration when produce phrase-medially (Oller, 1973; Klatt, 1975; Wightman et al., 1992).

Second, the end of an IP is marked with a boundary tone, either H% or L%. The position of the boundary tone is marked with the generic label T% in Figure 3.1. The final intermediate phrase in a given IP also contributes a phrase accent (either H-, !H-, or L-and labeled as T- in Figure 3.1), which results in a sequence of two tones at the end of an IP. This final boundary tone sequence can surface as a fall (L-L%), a high rise (H-H%), a low rise (L-H%) or a plateau (H-L% or !H-L%). Following the boundary tone sequence, the speaker may optionally insert a pause.

In many cases, IP boundaries align with larger syntactic boundaries (Selkirk, 1996). IP boundary insertion can also be used to disambiguate between two possible syntactic structures, such as in the case of ambiguous attachment (Fodor, 1998). The following examples demonstrate how the location of an IP boundary can help disambiguate the attachment location of an adjunct phrase. The location of the IP boundary is marked with the % symbol.

(9)   The artist drew the man with a pen.

(10)   The artist drew % the man with a pen.

(11)   The artist drew the man % with a pen.

In 9, there are two possible interpretations: either the artist drew the man who had a pen, or the artist drew a particular man and used a pen to draw him. The insertion of an IP after the verb in 10 gives rise to the former meaning. In contrast, inserting an IP between the object and the adjunct phrase results in the latter meaning (Jun, 2003).

2.4.1.2 Specific features of the Emphatic Juncture

The Emphatic Juncture as initially described in Sturman (2019) consists of three core features which position it as a sub-type of the generic IP boundary. Acoustically, an emphatic juncture is characterized by: (1) IP-level final lengthening on the segments preced-

ing the juncture, (2) a pitch plateau boundary tone (labeled as H-L% or !H-L%), and (3) an obligatory pause following the boundary tone. This pause is intentionally planned by the speaker rather than a disfluency. The emphatic juncture is labeled on the Breaks tier using *4e*.

Unlike other types of IP boundaries, which typically align with larger syntactic boundaries (Selkirk, 1996), emphatic junctures often interrupt even the tightest constituency relationships. For example, an emphatic juncture can be inserted between a determiner and a noun phrase (such as in Figure 5.3) or between negation and an adverb (such as in Figure 3.5).
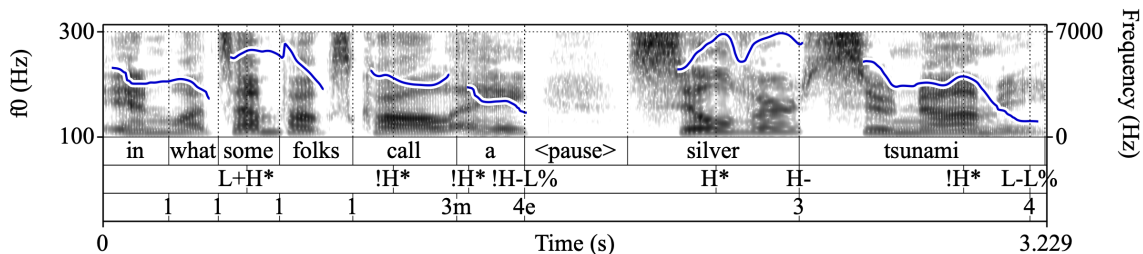
Figure 2.3: Pitch track of *in what some folks call a % silver tsunami.* This example shows an emphatic juncture (marked with a box in the figure) interrupting a highly local syntactic relation, between a determiner and noun phrase. This example also shows the use of an emphatic juncture in a transparent free relative. (Data from National Public Radio)
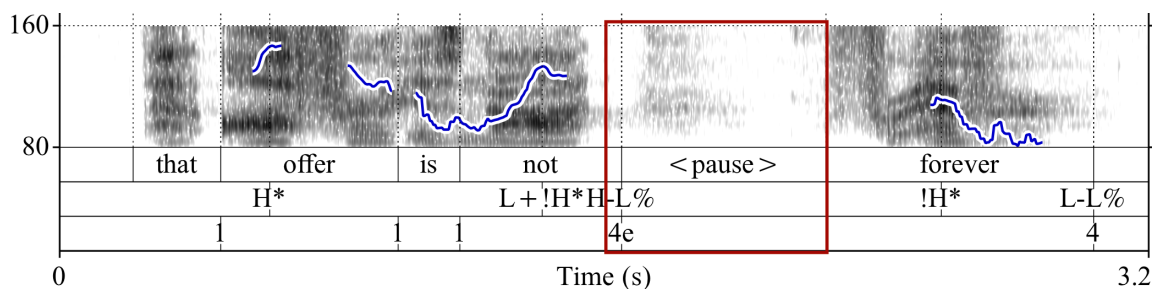
Figure 2.4: Pitch track of *That offer is not % forever.* This example shows an emphatic juncture (marked with a box in the figure) interrupting a highly local syntactic relation, between an adverb and negation, as well as the use of emphatic junctures in oratorical speech styles such as sermon speech. (Data from a Baptist sermon).

Emphatic junctures regularly appear in transparent free relatives. Transparent free relatives (Wilder, 1998; Grosu, 2003) are a special type of free relative that embeds a syntactic constituent (usually a DP) within a free relative. As can be seen in 35, they consist of a non-sortal use of *what*, an optional source, an embedding verb that selects an equative structure or small clause, and the pivot (bolded in the following examples). The pivot is the content embedded within the TFR that is being attributed to the TFR source.

(12)   Subject source TFR: The pitcher mastered <u>what he calls **the slurve**</u>.

        No-source TFR: The pitcher mastered <u>what is called **the slurve**</u>.

Crucially, speakers often insert an emphatic juncture directly preceding the noun within the pivot (Sturman, 2019). An example of a TFR that includes an emphatic juncture can be seen in Figure 5.3.

Emphatic junctures are also very commonly found in particularly emphatic speech (Sturman, 2019). In some cases, speakers may even insert emphatic junctures within words, such as in Figure 3.4. Additionally, certain styles of oratorical speech regularly include emphatic junctures. For example, the sentence shown in Figure 3.5 is from a Baptist sermon, a speaking style that employs high rate of emphatic junctures to emphasize key points of the message. In this particular example, the preacher was making an important point (i.e. that the offer of forgiveness he had been preaching on would not be available forever). Using an emphatic juncture is a strategy to signal a particular moment as noteworthy in a quiet way, with a pause and a lesser acoustic intensity than previous louder and faster-paced parts of the sermon.

2.4.2   Feature 2: Pitch range reset

The second key feature of subclausal quotation intonation is the pitch range reset. Speakers reset their pitch range at the beginning of the quotation, meaning that there is no longer any drop in the height of **H** targets due to declination or downstepping. According to MAE_ToBI labelling guidelines, pitch range resets may optionally occur at any intermediate phrase (ip) break. In the case of quotation, however, this reset could often also be
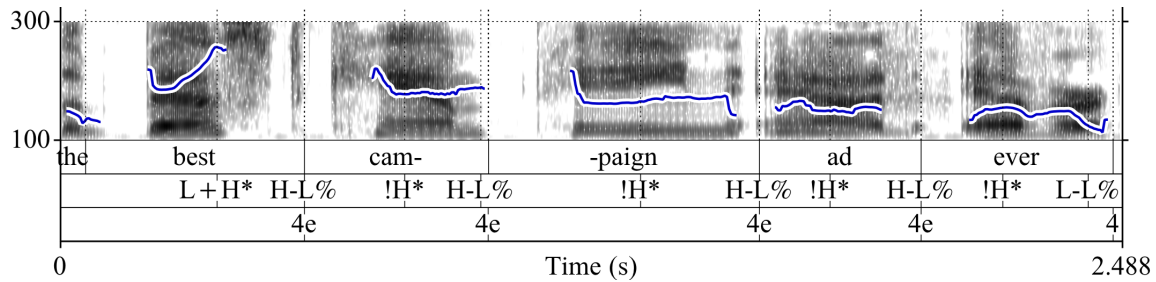
Figure 2.5: Pitch track showing emphatic junctures appearing after each content word and even within a word. Transcript: *The best % cam- % paign % ad % ever.* (data from National Public Radio)

characterized as a pitch range expansion. The set of figures below demonstrate an instance where the speaker expands the pitch range for the quoted material (Figure 5.1) as well as a token where the pitch is reset but not expanded in comparison to the rest of the utterance (Figure 2.7). To complete the contrastive triplet, Figure 2.8 shows an utterance that does not include a pitch range reset or expansion on the quoted material.
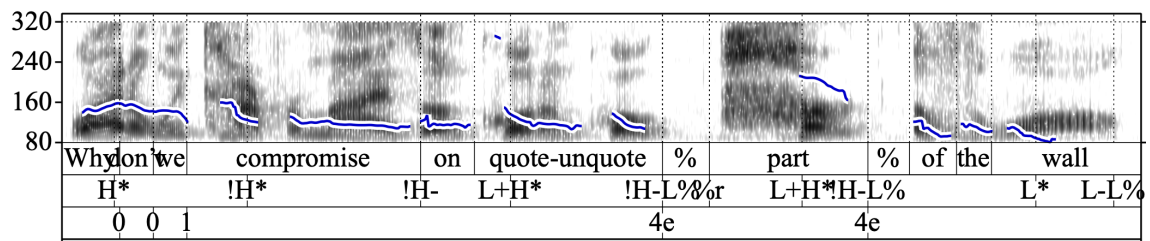
Figure 2.6: Pitch track showing an example of pitch range expansion on the quoted material. Transcript: *Why don't we compromise on quote-unquote % part % of the wall* (data from C-SPAN)

Resetting the pitch range is a means for the speaker to increase the acoustic prominence of the content of the quotation (Ladd, 1990). Moreover, Ladd has proposed a relation between pitch reset and information structure (Ladd, 1988). I suspect that in the case of subclausal quotation, the pitch range reset is also serving as an indicator that the material within the quotation is distinct from the preceding content at an information structural level. In particular, the subclausal quotation content has a source other than the speaker. I leave further exploration, testing and verification of this speculation to future research.
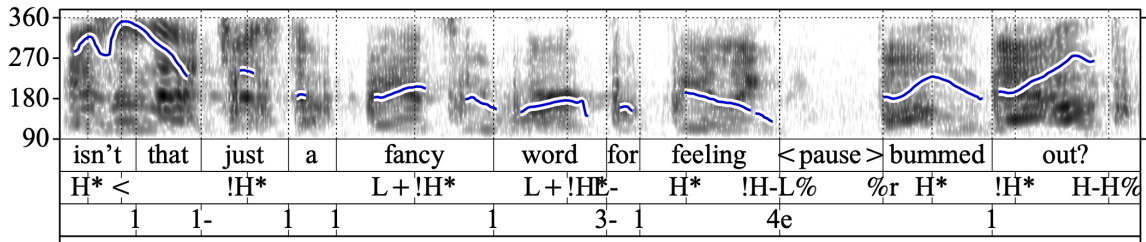
Figure 2.7: Pitch track of '*Isn't that just a fancy word for feeling % "bummed out?"*' (data from "The Office" S3E20 "Safety Training") Example demonstrating pitch range reset without pitch range expansion.
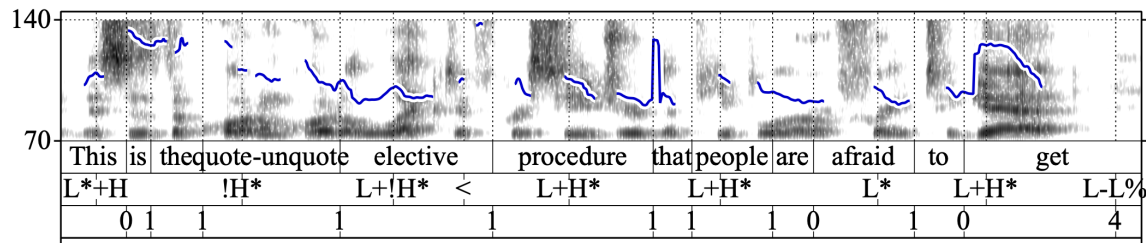
Figure 2.8: Pitch track of '*This is the quote-unquote "elective procedure" that people are afraid to get.*' (data from C-SPAN) Example of a subclausal quotation without a pitch range reset or pitch range expansion.

Interestingly, following the end of the quotation, the speaker often returns to their pre-quotation pitch register. This is surprising because one of the major theoretical points of Pierrehumbert's autosegmental-metrical model of intonation is that there are no non-local dependencies in intonation. (Pierrehumbert, 1980). If non-local dependencies are indeed non-existent in intonation, it would be entirely accidental that the speaker would return to their pre-quotation pitch register once the quotation has finished.

In many cases, however, the material within the quotation can be spliced out to leave a relatively continuous-sounding pitch track. When low-pass filtered, the spliced version sounds like a felicitous English contour. The pair of spectrograms in Fig. 2.9 demonstrates the original sentence, with subclausal quotation around *disastrous design*, above and the spliced version below.

I will expound on this finding about the continuity of the pitch track preceding and following the quotation and its implications in greater detail in the following chapter.
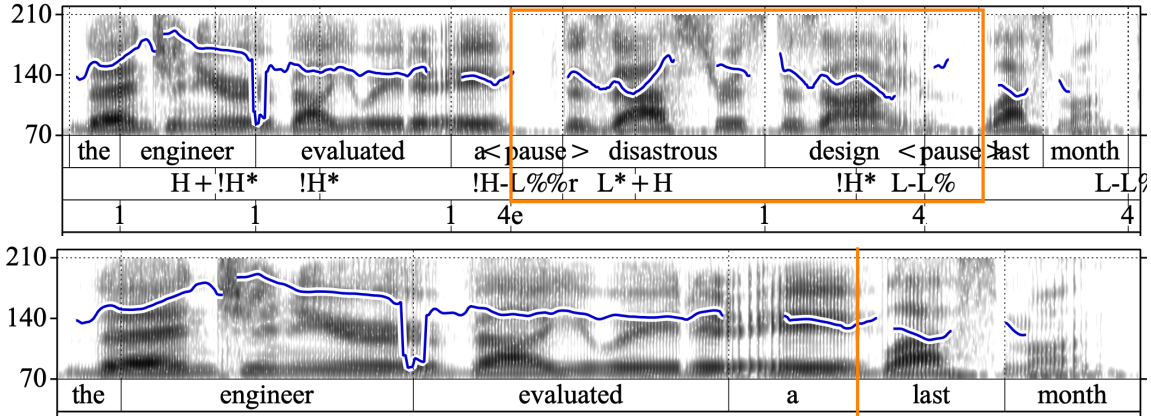
Figure 2.9: Pitch tracks of *The engineer evaluated a (% "disastrous design" %) last month.* (data from SQ Exp.) A pair of sentences demonstrating pitch track continuity before and after the quotation. The original sentence is above and the spliced sentence without the object is below.

2.4.3 Feature 3: Right edge IP boundary

The final intonational feature of subclausal quotation is an IP boundary at the right edge of the quoted material. The presence of this break is most apparent in sentence medial instances of subclausal quotation, which eliminates the boundary being confounded with sentence final boundary tones. The boundary tone sequence associated with the IP break at the end of the quotation is typically L-L% or L-H%. An example of each can be seen in Figs. 2.10 and 2.11 below.
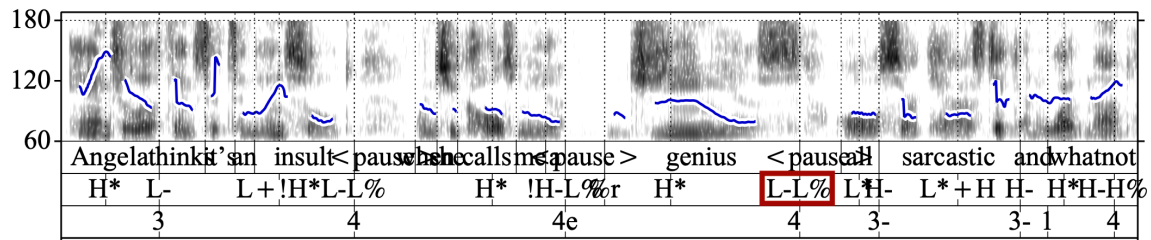
Figure 2.10: Pitch track of '*Angela thinks it's an insult when she calls me a % "genius" % all sarcastic and whatnot?*' (data from "The Accountants" webisode of "The Office") This example shows an IP break at the end of the quotation with an L-L% boundary tone sequence.

Some utterances, like the examples in Figs. 2.10 and 2.11, included a notable pause

24
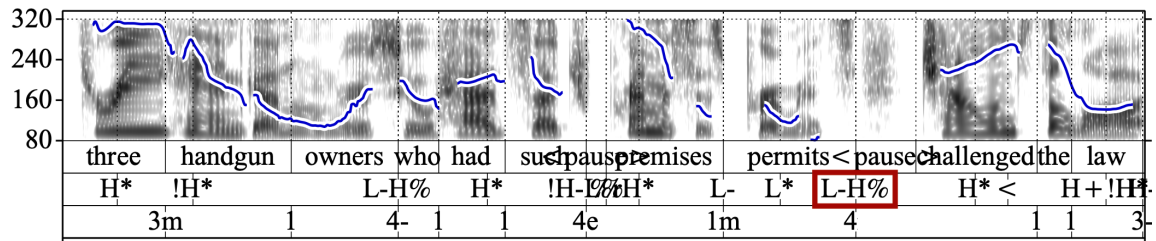
Figure 2.11: Pitch track of '*Three handgun owners with such % "premises permits" % challenged the law...*' (data from National Public Radio) This example shows an IP break at the end of the quotation with an L-H% boundary tone sequence.

after the subclausal quotation, similar to the emphatic juncture at the start of the quotation. This pause seems to be optional, however, as other utterances did not include a quotation-final pause. An example omitting the pause is shown in Fig. 2.12 below.
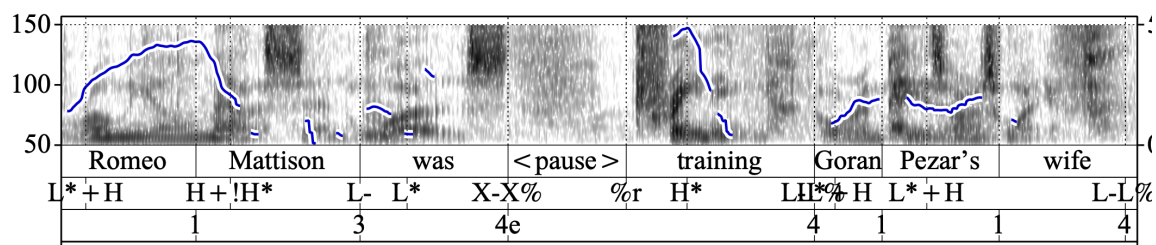
Figure 2.12: Pitch track of '*Romeo Mattison was % "training" % Goran Pezar's wife.*' (data from "Barry" S1E3: "Make the Unsafe Choice") This example shows an IP break at the end of the quotation with no notable pause.

It should be noted that while the pause is optional, an IP break was almost always present at the end of the quotation. Recall from Section 2.4.1.1 that for typical IP boundaries, pauses are optional. This contrasts with the Emphatic Juncture, where a pause is a required feature.

2.5 Quantitative results from the corpus

Looking at the corpus from a quantitative perspective illuminates several different insights about the data and the phenomenon. Of the 90 total tokens in the corpus, 70 of those tokens utilized the lexical markers *quote* or *quote-unquote* to mark an instance of

subclausal quotation. The mean number of words in each subclausal quotation was 1.98, with a maximum of 7 words.

The occurrence of the three key intonational features described in the previous section was remarkably consistent throughout the corpus. Overall, tokens included an emphatic juncture preceding the quoted material 76% of the time, a pitch range reset 80% of the time, and an intonation phrase break at the end of the quoted material 89% of the time. A summary of the percentage of tokens that included each of the three features is shown in Table 5.1 below.

	EJ	Pitch Range Reset	Right Edge IP
No lexical marker (N=20)	90%	85%	90%
Lexical marker (N=70)	73%	79%	89%
Overall (N=90)	76%	80%	89%

Table 2.2: Occurrence rates for each of the key intonational features found in the Subclausal Quotation Corpus, separated by row based on whether the token included a lexical marker (*quote* or *quote-unquote*) as well as the overall rates.

Only one token in the corpus did not utilize any of the three key intonational features. The speaker did, however, mark the start of the quotation with *quote-unquote*. This token is shown in Figure 2.13.
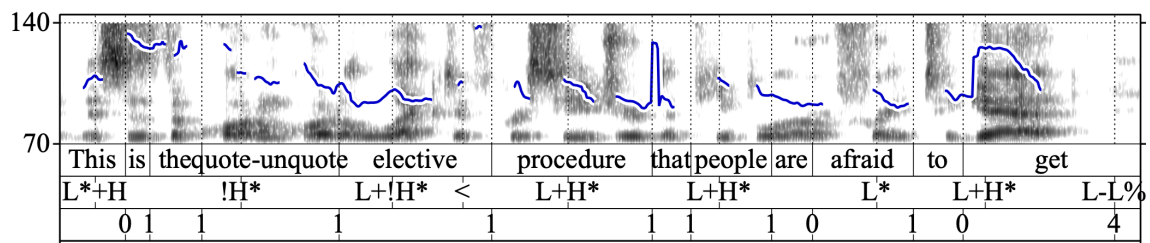
Figure 2.13: Pitch track of '*This is the quote-unquote "elective procedure" that people are afraid to get.*' (data from C-SPAN) This example does not use any of the three key intonational features found throughout the rest of the corpus.

Perhaps not surprisingly, when the subclausal quotation was signaled without using the lexical marker *quote*, the occurrence rate of all three intonational features was very

high (at least 85% for all three features). Interestingly, when considering the tokens that *did* use a lexical marker at the beginning of the quotation, the occurrence rate dropped, but only for two out of the three intonational features. Speakers' usage of the right edge IP boundary at the end of the quotation remained virtually unchanged.

I propose that this is due to the need to mark the end of the quoted material and is directly related to how speakers use the lexical markers *quote, unquote* and *quote-unquote*. To foreshadow one semantic function of the intonational features of subclausal quotation, in order for part of an utterance to be successfully interpreted as quotation, the listener must determine what material belongs inside of the quotation. In other words, the substring of the utterance functioning as a quotation must have a defined start and end. The right boundary tone serves to signal the end of the quoted material. It is interesting that even when a lexical marker is used to indicate the beginning of a quotation, the other lexical bookend *unquote* is rarely placed at the end of the quoted material. Speakers opt to use the right boundary tone to mark the end of the quotation instead.

In addition to observing the rate of right boundary tone usage, we can also examine the types of boundary tones (i.e. boundary tone sequences) speakers used. Boundary tone usage rates by tone sequence type are shown in Table 2.3. The most frequently used right boundary tone type was L-L%, followed by L-H%.

	L-L%	L-H%	H-L% or !H-L%	H-H%
No lexical marker (N=20)	60%	27%	13%	0
Lexical market (N=70)	47%	30%	7%	6%
Overall (N=90)	58%	29%	9%	4%

Table 2.3: Occurrence rates of boundary tone sequences found in the Subclausal Quotation Corpus, separated by row based on whether the token included a lexical marker (*quote* or *quote-unquote*) as well as the overall rates.

2.6 Qualitative observations from the corpus

In addition to the quantitative results, the corpus surfaced a few very interesting qualitative findings. These observations are less technical, but such a wealth of natural data is worth mining for insights beyond what is well-established enough within theoretical linguistics to confidently measure and analyze. I will discuss observations about the content of words within the quotation, explore differences in the use of lexical markers to indicate quotation, and finish by presenting a few examples of how subclausal quotation intonation seems to manifest differently in different varieties of American English.

2.6.1 Observations about quoted content

One of the first questions that comes to mind when studying subclausal quotation relates to what kind of words people put in quotes. Unfortunately, just as with trying to come up with comprehensive (dictionary-style) meanings for words, this question is much easier to ask than it is to answer. Despite this difficultly, I will make a few high-level, non-exhaustive observations about content of subclausal quotations based on the corpus.

There are three broad categories of subclausal quotation content. Two are relatively straightforward but the third is more nebulous. I should note before preceding further that these categories are for descriptive purposes only. I am not attempting to claim any kind of theoretical difference between the three, and some tokens may be well-described by more than one type.

2.6.1.1 Neologisms and uncommon terms

The first category is used to flag neologisms or uncommon terms (e.g. a term an expert knows but non-experts have probably never heard). As an example, consider the token in 13, collected from C-SPAN:

(13)   The ones who were really into quote-unquote "juuling" could often do a pod in a day.

I speculate that in some cases, people are quoting another speaker's neologism or expert term, and in other cases, the speaker is highlighting an unfamiliar word by putting it in quotations.

2.6.1.2   Verbatim speech reports

The second category of content can be characterized as using quotation to explicitly report what someone said. This is the kind of quotation that is commonplace in news reporting, courtrooms, and other places where people are concerned with replicating the exact words someone used as accurately as possible. An example of this kind of quotation, spoken during Donald Trump's impeachment hearing and broadcast via NPR, can be seen in 33

(14)   President Trump directed us to quote "talk with Rudy."

In such a high stakes setting, it is important for the speaker to not only accurately portray what was said but also sidestep any chance of having illicit activities be attributed to them by mistake.

2.6.1.3   Controversial content

The third type of quotation content can be broadly construed as controversial, something that when used might be characterized by some as provocative or offensive. Based on the corpus data, speakers seem to often employ subclausal quotation to use less savory words in a more palatable way. Quotation seems to diffuse some of the provocative or offensive tone brought by the same content used without quotation. An example of this type of quotation content is shown in 15.

(15)   More and more people are quote-unquote "home-grown Americans" who don't
        come from other countries.

There seem to be some potentially interesting pragmatic differences between the three

types of content that appeared within subclausal quotation. I will revisit these categories and consider them further in Chapter 5 when I discuss the semantics of subclausal quotation.

2.6.2 Lexical marker usage

Early in the data collection process, I was intrigued by the ways speakers used lexical markers to indicate quotation. In particular, I was fascinated by the use of *quote... unquote* (with unquote occurring at the end of the quotation) versus *quote-unquote* occurring as a compound at the start of the quotation. While *quote-unquote* suffices to mark the start of a quotation, it is less informative than using *quote* at the beginning of the quote and *unquote* separately at the end of the quotation. Despite this, *unquote* seemed to be used primarily by news reporters and sometimes politicians when delivering a verbatim speech report.

This leads me to wonder if using *unquote* to mark the end of a quotation is actually somewhat prescriptive, perhaps a strategy that is taught to news reporters for their particular speech style in order to lexically delineate both the start and the end of a quotation. Perhaps this has to do with the prevalence and importance of transcripts in these settings. Using both lexical markers ensures that both the start and the end of the quotation will be clearly marked in the transcript. It is unclear how well the intonational features described here transition to orthography such as in a transcript, particularly if the transcription is being done by a computer program.

Unfortunately, I am unable to comment on intonational patterns, and in particular right boundary tone usage, when a speaker uses *unquote* at the end of the quotation. I found only one token like this that employed *unquote* at the end of the quotation but not at the end of the utterance. In that example, the speaker used all three of the intonational features discussed in this chapter, including an IP boundary after the final quoted word but before *unquote*. Further investigation of more tokens like this is needed before making any claims about how the usage of intonational features when quotation-final *unquote* is used.

2.6.3 Variation

Finally, I would be remiss if I failed to mention a few tokens from the corpus that suggest differences in how different varieties of American English prosodically mark subclausal quotation. For example, consider the token from African American English in Figure 2.14. This token shows several interesting characteristics that contrast on the surface to the key intonational markers of subclausal quotation discussed in this chapter. First, there are two strikingly high L+H* accents on *looks* and *really*. On the quoted material (*betraying... values*), however, there appears to be a pitch range compression happening rather than a pitch range expansion.
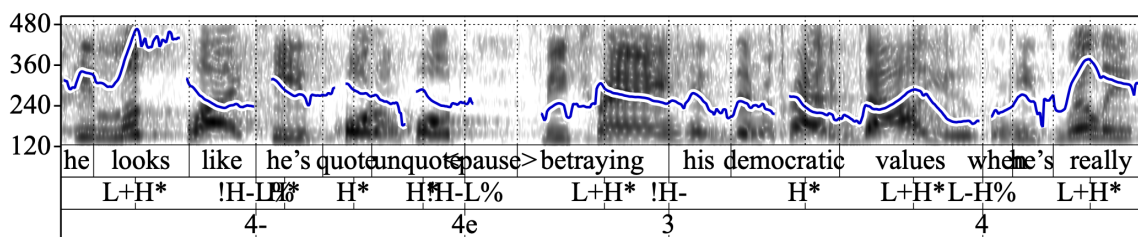
Figure 2.14: Pitch track of '*He looks like he's quote-unquote % "betraying his democratic values" % when he's really...*' (data from YouGlish) This example from a Black speaker uses two L+H* accents with very high peaks both before (*looks*) and after (*really*) the quotation. The pitch range on the quotation itself is compressed rather than expanded.

These two observations taken together with a pair of findings from previous research suggest something quite interesting might be happening. First, speakers of AAE typically employ L+H* pitch accents more frequently than in Standard American English (Holliday, 2016). Secondly, cross-linguistic analysis of prosodic systems shows that the goal of focus prosody is to violate the rules of language-specific default prosody (Jun, 2019). Taken in the context of these two theoretical findings, the compressed pitch range on the quoted material might actually be intentionally violating the default prosody of AAE to contrast the quoted content and thereby draw attention to it. This seems to be an interesting direction for future research.

Accounting for variation has traditionally been a struggle among theoretical linguists, particularly when it comes to intonation. Holliday (2021) outlines many of the challenges,

31

one of which being the limits of MAE_ToBI for exploring variation. When it comes to characterizing speech beyond the scope of what the designers considered "Mainstream American English," MAE_ToBI is too phonological to systematically note differences in how category labels are realized (i.e. phonetic-level differences). In addition to these methodological limitations, there were no quotas surrounding speaker identity, resulting in a dearth of AAE tokens. As such, I am unable to make more than the few preliminary observations above that may prove useful in guiding future research.

CHAPTER 3

Intonation Theory

3.1 Introduction

In this chapter I will expound on the ways the Emphatic Juncture as well as other intonational observations relating to subclausal quotation discussed in Chapter 2 challenge and broaden our understanding of intonational theory. I will begin by briefly overviewing the history of the study of intonation, culminating in the development of the ToBI framework. I will then demonstrate how the Emphatic Juncture poses problems for the current iteration of the MAE_ToBI model. I will then present evidence for embedded Intonation Phrases (IPs) in English.

3.2 Background: The origins of the Autosegmental-Metrical framework and MAE_ToBI

The goal of this section is to outline the research in intonation that eventually led to the development of the Autosegmental-Metrical theoretical framework for intonation and the corresponding MAE_ToBI transcription system that followed. It is not meant to be a comprehensive history of the study of intonation as a whole, but rather a highlight reel of how this particular framework came about.

3.2.1 Early analyses of intonation: Phonetic or phonological?

Historically, researchers have characterized intonation from either a phonetic or phonological perspective. There was good reason for this turf war. Parts of intonation seem more

phonetic in nature, best studied by analyzing gradient acoustic cues like f0 that are constantly in motion throughout an utterance (Lieberman, 1967). Unfortunately, conclusive findings using these methods tended to be elusive.

On the other hand, those approaching intonation from a phonological lens treated intonation in terms of small number of categorically distinct elements, such as pitch phonemes or nuclear tones. In order for phonological categories to be useful, however, there must be agreement about what the categories *are*, much less how distinctions in them are defined. Early on, categories were based on the researchers' impressions and had little evidence to garner agreement across the field.

There were two primary traditions: the American tradition and the British tradition. The American tradition analyzed intonation as four phonemic levels (e.g. Pike (1945), Wells (1945)). They viewed pitch contours (i.e. movement between phonemic pitch levels) as contrastive. Unfortunately, this leads to an overgeneration of contours that are phonologically contrastive. Mathematically, this theory predicts 12 possible bitonal contours. It seems highly unlikely, however, that a /41/ fall is meaningfully different from a /31/ fall, or that a /14/ rise is meaningfully different from a /13/ rise.

Rather than splitting a speaker's range into phonemic bands, the British tradition (Crystal, 1969; O'Connor and Arnold, 1961) defined intonation in terms of three varieties of phonological tones: static (i.e. High, Low), kinetic (i.e. rising, falling), or complex (i.e. rising-falling, rising-falling-rising). In contrast to the fixed status of a tone in the American-style phonemic levels, the British system interprets each tone relative to what has preceded it. In other words, a tonal level can only be perceived relative to other tones.

The British school and the American school provide two contrasting perspectives on how to think about tonal units. The American school analyzed tone in terms of *levels*, whereas the British school analyzed tone in terms of *configurations*. The confluence of these two approaches paved the way for an autosegmental approach to intonation.

3.2.2 Intonation as autosegments

Analyzing intonation in an autosegmental framework began with Bruce (1977) and remains the predominant theoretical framework for intonational analyses. Bruce bridged the gap between pitch levels and pitch configurations by reducing the number of tonal levels to two, High and Low, and positing that pitch accents can be configurational (i.e. two or more pitch targets can comprise a single pitch accent). Bruce also reduced the division between the phonetic and phonological approaches to intonation by positing that several phonetic realizations may be possible for a single phonological tone.

In practice, an autosegmental analysis of intonation views the pitch movements of an utterance as a series of discrete, independent phonological events (i.e. pitch accents and boundary tones) that have phonetic targets. The speaker modulates f0 to hit these pitch targets and smoothly interpolates between them.

3.2.3 The Autosegmental-Metrical theory of intonation

Janet Pierrehumbert's seminal dissertation (Pierrehumbert, 1980), in conjunction with subsequent work by Beckman and Pierrehumbert (1986), set the stage for the language-specific ToBI models of intonation in use today. The models proposed in Pierrehumbert (1980) and Beckman and Pierrehumbert (1986), however, are specifically designed to handle American English intonation.

3.2.3.1 Phrasal structure of intonation

The primary, overarching hypothesis of Pierrehumbert's approach is that intonation has a phonological organization and a predictable structure. As with syntax and even other areas of phonology, the intonation of an utterance is composed of hierarchical phrasal structure. The hierarchical phrasal structure of English intonation proposed in Beckman and Pierrehumbert (1986) (building off of Pierrehumbert (1980)) is shown in Figure 3.1. The highest level of structure is the Intonation Phrase (IP). Each intonation phrase is composed

of one or more intermediate phrases (ip). Each intermediate phrase contains one or more words. Within any given intermediate phrase, there must be one or more pitch accents associated with word(s) within that ip.
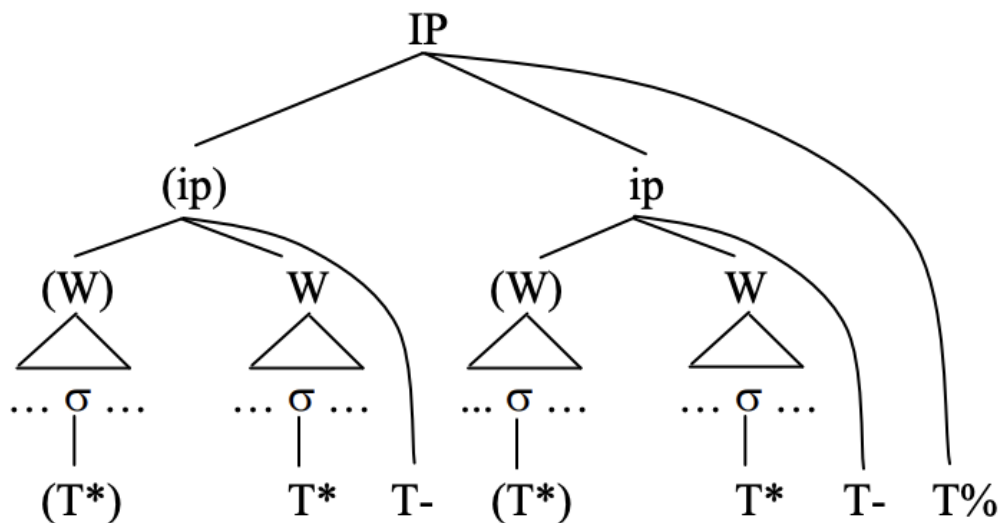
Figure 3.1: The hierarchical phrasal structure of English intonation according to the model put forth in Beckman and Pierrehumbert (1986). IP = Intonation Phrase, ip = intermediate phrase, T* = pitch accent, T- = phrase accent, T% = boundary tone. Figure from Sun-Ah Jun's Ling 111/211 class handout

As with other levels of phonological structure (e.g. syllables, feet), it has been proposed that the phrasal structure of intonation obeys the Strict Layer Hypothesis (Selkirk, 1986). Broadly speaking, the Strict Layer Hypothesis (Selkirk, 1984; Nespor and Vogel, 1986) stipulates two parameters: (1) A given non-terminal unit contains one or more elements of the category immediately below it; (2) If not the root node, a unit for a given level must be exhaustively dominated by an element from the level directly above it. The Strict Layer Hypothesis was widely accepted as it pertains to intonational structure (Shattuck-Hufnagel and Turk (1996); Jun (1998) though c.f. Ladd (1986)). Based on data from the subclausal quotation corpus, however, I will outline some evidence that seems to counter the Strict Layer Hypothesis (in its strongest form) in Section 3.4 later in this chapter.

3.2.3.2 Tonal inventory

Following Bruce (1977), the entirety of the tonal inventory is built using just two primitive tones: High and Low. There are three types of phonological tones in the framework: pitch accents, phrase accents, and boundary tones.

Phrase accents and boundary tones are single-tone targets that mark the edges of intermediate phrases and intonation phrases, respectively. Due to the Strict Layer Hypothesis, the final syllable of an IP is also the final syllable of an ip. When an overlap like this occurs, the IP-final syllable could carry both syllables or one of the tones could be deleted. In English, tones corresponding to phrase edges are realized sequentially (i.e. a phrase accent followed by a boundary tone) rather than only the tone from the highest phrase being realized (i.e. the boundary tone overrides the phrase accent), such as in Bengali or Korean.

Pitch accents are used to mark words as prominent within phrases. I will describe the prominence marking and alignment of pitch accents in greater detail in the following section. In the framework described in Beckman and Pierrehumbert (1986), there are six pitch accents: two single-tone pitch accents and four bitonal pitch accents.

3.2.3.3 Prominence and alignment

At the word level, English marks particular syllables as more prominent than others using stress. Stress is specified lexically and is realized with a host of phonetic features including intensity, duration, and pitch. Even though every word in English has a stressed syllable, however, not every word is prominent at the phrasal level. The words that receive phrasal prominence are not fully predictable, though we know it is dependent on speaker intention and information structure constraints (Bolinger, 1972; Calhoun, 2007).

The "metrical" component of the Autosegmental-Metrical framework indicates that there are rules for how autosegmental tones and tunes associate with the text, and those rules center around prominence. Pierrehumbert and Beckman and Pierrehumbert represent the prominence of a stressed syllable compared to an unstressed syllable within a word or within a phrase through the lens of metrical phonology (Liberman, 1975; Liberman and

Pierrehumbert, 1984). Intonational prominence is treated as part of the stress continuum. Pitch-accented syllables have a higher degree of stress than word-level stress. The highest level of intonational prominence (and the highest level on the stress continuum) is the final pitch accent in any intermediate phrase, known as the nuclear pitch accent. The phrase accent is realized beginning on the word following the nuclear pitch accented word until the end of the intermediate phrase.

3.2.4 Development of the MAE_ToBI model

3.2.4.1 Origins of the labeling conventions

The Autosegmental-Metric framework for English intonation proposed by Pierrehumbert and refined in Beckman and Pierrehumbert effectively opened up a new research area as more linguists began to explore new frontiers in intonation. Despite the theoretical innovations, however, intonation researchers lacked a standardized system or set of conventions for labeling intonation in English. Beginning in 1991, several prosody researchers from both academia and industry came together to establish conventions and guidelines for intonation labeling (Silverman et al., 1992). Born out of this collaboration was the first iteration of the ToBI (Tone and Break Indices) labeling system. The goal of this system was to achieve high levels of inter-transcriber reliability using (primarily) phonological labels.

The first iteration of the MAE_ToBI labeling system was built using a sample set of 25 utterances labeled by twenty transcribers (Silverman et al., 1992). As the system gained more widespread usage by more transcribers labeling more types of speech data, researchers recognized the need for further revision to the labeling conventions. Over the course of several workshops, intonation researchers collaborated again to introduce some new labels and refinements to the system (Beckman and Ayers-Elam, 1997; Beckman et al., 2005). While certain researchers or labs may have their own conventions beyond this update, there have been no official changes to the labeling standards since then.

3.2.4.2 Tonal inventory of contemporary MAE_ToBI

The tonal inventory of MAE_ToBI consists of five pitch accents which mark words as prominent (L*, H*, L*+H, L+H*, and H+!H*), two phrase accents that mark the end of ips (L- and H-), and two boundary tones which mark the end of IPs (H% and L%). With the exception of the boundary tones, each of the High tonal targets also have the potential to be downstepped, which is marked with an exclamation point before the H.

3.3 The Emphatic Juncture: A challenge to intonational theory

The MAE_ToBI model provides a helpful, consistent standard for labeling intonation. It has its limits, however, and testing those limits often surfaces new and important insights for intonational theory. I believe this is particularly true when considering more complex syntactic constructions and more spontaneous speech than what was used in developing the early Autosegmental-Metrical frameworks for English intonation. Subclausal quotation, and particularly the naturally occurring speech found in the corpus described in Chapter 2, is an ideal domain to test the limits.

The most significant finding that is difficult to accurately label in the most recent MAE_ToBI convention (Beckman and Ayers-Elam, 1997; Beckman et al., 2005) is the Emphatic Juncture. Recall that the Emphatic Juncture is one of the key intonational features of subclausal quotation, used to mark the beginning of the quoted material.

3.3.1 Acoustic markers of the Emphatic Juncture

To recapitulate the brief description given in Chapter 2, the Emphatic Juncture is a large juncture that is marked by IP-level final lengthening, a plateau boundary tone sequence (H-L% or !H-L%), and an obligatory pause. The descriptions of these features in the following sections draws on earlier work on the Emphatic Juncture (Sturman, 2019).

3.3.1.1 Final lengthening

The Emphatic Juncture's status as an IP-type boundary is clear from significant final lengthening of the word directly preceding the juncture. Consider, for example, the word *an* in Figures 3.2 and **??**, uttered by the same female speaker. These tokens were elicited as part of an early, exploratory production task that involved participants reading sentences that included subclausal quotation. Because these tokens were produced as part of an experiment with a counterbalanced design, there are no tokens to act as a precise minimal pair that vary only in the presence of an emphatic juncture. These two utterances form a near minimal pair, however, given that the word of interest occurs in the same syntactic position and with very similar surrounding phonetic segments on either side of the potential juncture location ([d] _ V).

In Fig. 3.2, *an* is produced IP-medially, with a duration of 160 ms. In Fig. **??**, *an* is produced directly before an emphatic juncture, with a duration of 439 ms. In both figures, the realization of *an* is highlighted with a box.
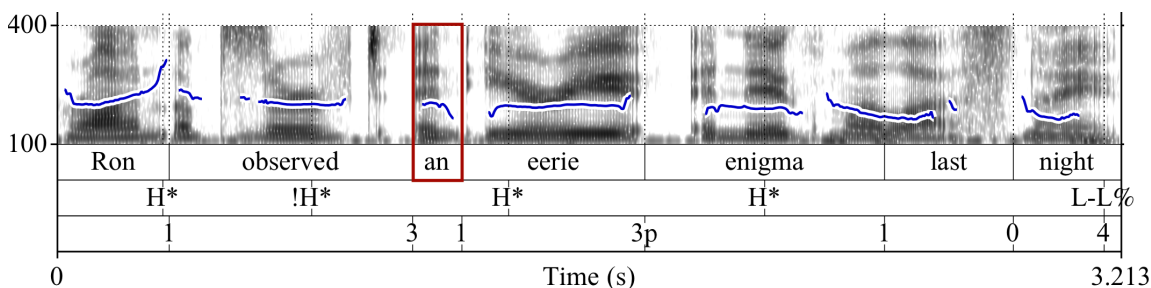
Figure 3.2: *Ron observed an eerie enigma last night.* Duration of *an* = 160 ms (Data from a reading production task)

3.3.1.2 Obligatory pause

The Emphatic Juncture's phonetic status as an IP boundary is also supported by the presence of a notable pause. Although a pause is optional as part of a canonical IP boundary, the pause is obligatory in the case of the Emphatic Juncture. These pauses can be rather lengthy, sometimes >600 ms. The pause in Figure **??** between *an* and *alarming*, for
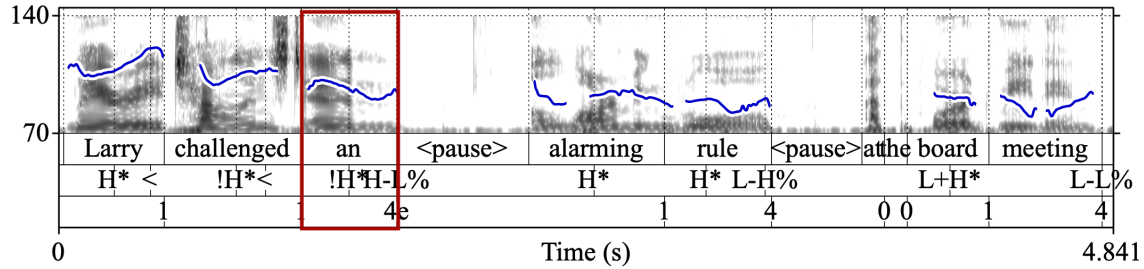
Figure 3.3: *Larry challenged an "alarming rule" at the board meeting.* Duration of *an* = 439 ms (Data from a reading production task)

example, is 606 ms. However, this pause is not due to a slowdown in phonological planning or some other disfluency, as would typically be marked on the MAE_ToBI Breaks tier with the *2p* or *3p* label. In fact, it is intentionally placed by the speaker to signal the use of an emphatic juncture. Some phonetic evidence for this intentionality includes the continuity of the pitch track across the juncture, as can be clearly seen in Figure 3.3.

3.3.2 Use of the Emphatic Juncture beyond subclausal quotation

Subclausal quotation intonation is not the only place I have observed instances of the Emphatic Juncture. Aside from marking the beginning of a quotation, the Emphatic Juncture can be used to highlight the material following the juncture as prominent, a strategy particularly common in certain speech styles. An example of this function is shown in Figure 3.4. In this utterance, the speaker places an emphatic juncture after every syllable besides *the*, even within a single word (*campaign*), and each of these syllables is pitch-accented.

In addition to this kind of emphatic marking, some speech styles use the Emphatic Juncture for rhetorical effect. By placing a pause, the speaker can signal that something important is coming, a way to vary the speech cadence to draw listeners back in to heed an important point. An example of this type of emphatic juncture usage can be seen in Figure 3.5.

41
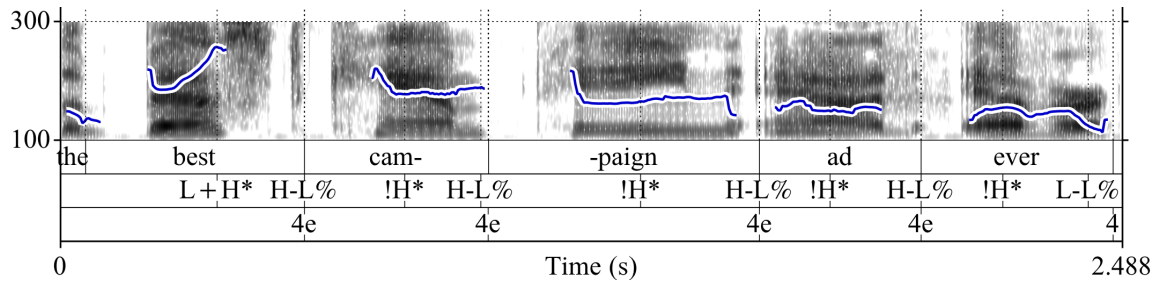
Figure 3.4: Pitch track showing emphatic junctures appearing after each content word and even within a word. Transcript: *The best % cam- % paign % ad % ever.* (data from National Public Radio)
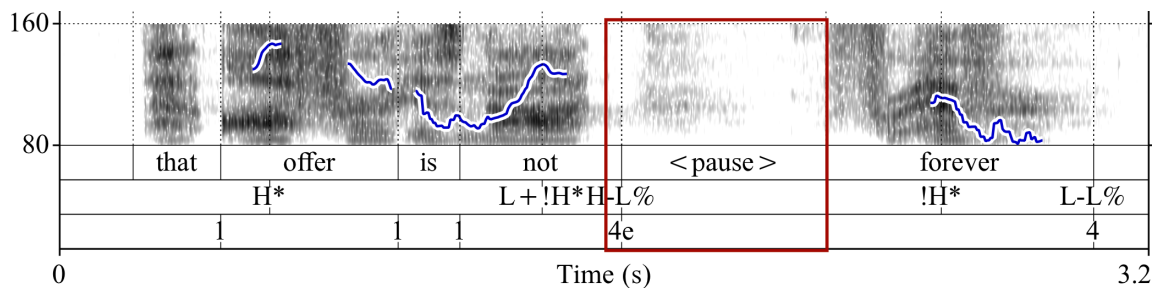
Figure 3.5: Pitch track (including an emphatic juncture marked with a box) of *That offer is not % forever.* This example shows the use of an emphatic juncture in oratorical speech styles such as sermon speech (data from a Baptist sermon).

3.3.3 The Emphatic Juncture challenges the domain of downstep

In English, downstepping refers to a high target being realized lower relative to a previous high target than cannot be explained by declination (e.g. !H* in MAE_ToBI), resulting in a lowering of the upper bound of the pitch range. Subsequent high targets will thus be realized no higher than the pitch of the !H* target unless the pitch range is reset at an Intermediate Phrase break. Though it is an optional process, downstepping often occurs when the information is predictable or backgrounded (Pierrehumbert and Hirschberg, 1990). The MAE_ToBI model defines the Intermediate Phrase as the domain of downstep (following Liberman and Pierrehumbert (1984)). This means that downstepping can only occur between two high targets that are within the same intermediate phrase, but not across an intermediate phrase boundary. As a reminder, the way I label downstep in this dissertation does not follow this convention. Rather, I label downstep as a relation between high

tones regardless of the prosodic boundaries that occur between the high tones.

Interestingly, despite the unmistakable junctures in Figure 3.4, all of the pitch accents after *best* are clearly being downstepped relative to the previous pitch accent. Given that there is ample evidence that the Emphatic Juncture marks an IP boundary, this is a clear challenge to the traditionally defined domain of downstep.

3.4 Embedded Intonation Phrases

3.4.1 Evidence for embedded IPs from subclausal quotation

Recall from Section 3.2.3.1 that the Strict Layer Hypothesis has been widely accepted in the development of intonational theory. Broadly speaking, the Strict Layer Hypothesis (Selkirk, 1984; Nespor and Vogel, 1986) stipulates two parameters: (1) A given non-terminal unit contains one or more elements of the category immediately below it; (2) If not the root node, a unit for a given level must be exhaustively dominated by an element from the level directly above it.

Given the Strict Layer Hypothesis, one of the most compelling intonational findings from the corpus is evidence that suggests the existence of embedded IPs in English. In many cases the material within the quotation can be spliced out to leave a continuous-sounding pitch track. In fact, when I played the spliced (bottom) sentence in Figure 3.6 for another colleague who works on intonation, they were surprised to learn that it was not the original sound file (Adam Royer, p.c.). They were even more surprised when they heard the significant pitch excursion that happened during the spliced material.

It should be noted that similar observations have been made regarding pitch in parentheticals. Kutik et al. (1983) note that parentheticals involve a drop in f0 on the parenthetical material and then a sharp rise in f0 upon returning to the main clause. Future research could attempt a similar splicing exercise for parentheticals or appositives. If the findings of Kutik et al. produce the same sort of results that splicing subclausal quotation has here, it would be additional evidence for the existence of embedded IPs in English.

43
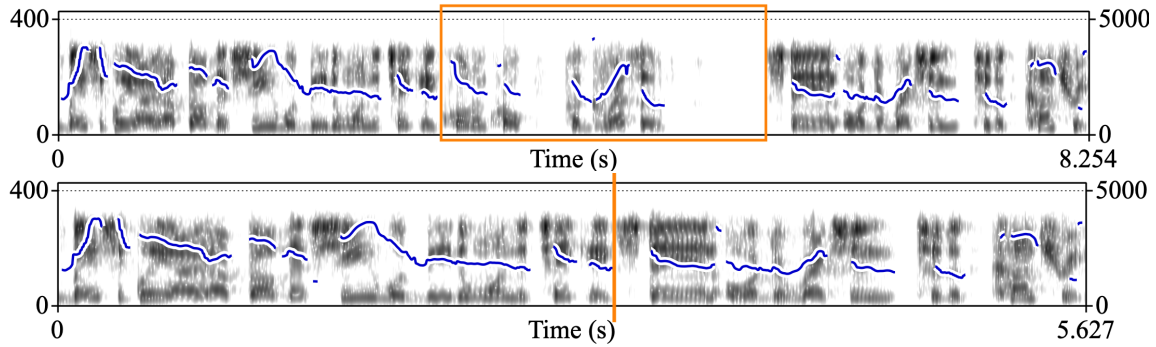
Figure 3.6: Pitch tracks of *Basically, I thought that she would be the one to take a (%* *quote unquote "progressive" %) stand or direction to the country.* (data from C-SPAN) A pair of sentences demonstrating pitch track continuity before and after the quotation. The spliced material is marked with a box in the original sentence and the location of the splice is marked with a line in the sentence below.

3.4.2 Predictions of embedded IPs

The existence of several of these continuous-sounding spliced pitch tracks is unexpected in itself. Contemplating the existence of embedded IPs, however, leads to a couple of additional predictions that also seem difficult to fathom.

3.4.2.1 Prediction 1: (Seemingly) headless intermediate phrases

First, embedding an IP opens up the possibility of a seemingly-headless intermediate phrase or intonation phrase. If the nested IP is inserted after the nuclear pitch accent but before the end of the intermediate phrase, the material following the nested IP would appear to be totally unaccented, or headless. An example of just this sort of prosodic stranding appeared in the subclausal quotation corpus, shown in Figure 3.7 here. In this utterance, the second *is* is separated from the rest of its natural intermediate phrase or intonation phrase by the insertion of the subclausal quotation directly preceding it.

There are a few different ways that the prosodic structure resulting from this sort of IP insertion could be analyzed. For reference, a depiction of the prosodic structure of the original utterance is shown in Figure 3.8.

44
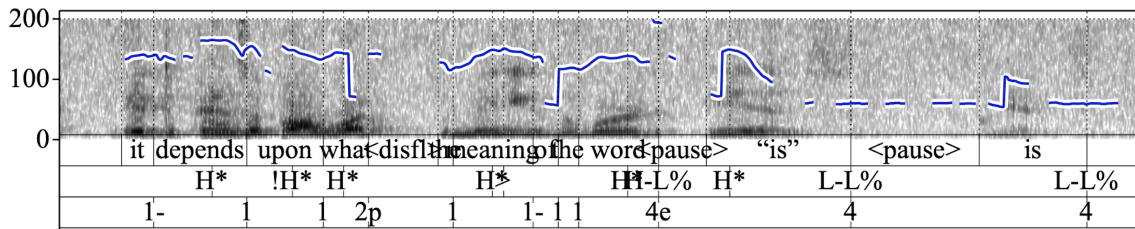
Figure 3.7: Pitch track of *It depends on what the meaning of the word % "is" % is.* This example shows the existence of a (seemingly) headless intermediate phrase following the subclausal quotation (data from the Bill Clinton deposition).
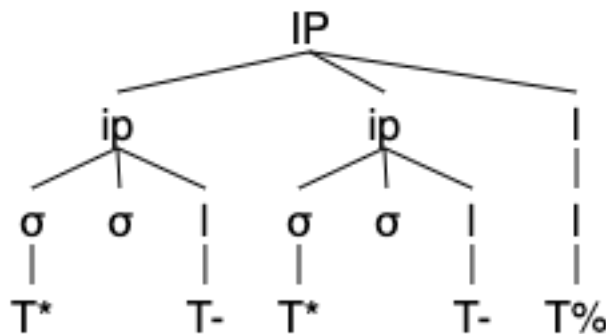
Figure 3.8: The prosodic structure of the (abstract) original utterance. IP = Intonation Phrase, ip = intermediate phrase, T* = pitch accent, T- = phrase accent, T% = boundary tone

The most basic option is that the inserted IP does not alter the pre-insertion prosodic structure (aka Basic Insertion Option, shown in Figure 3.9).

The material following the inserted IP is stranded from its intermediate phrase of origin, meaning the segmental material as well as the phrase accent associated with the original ip are realized following the inserted IP. The interruption of the original intermediate phrase is marked with the insertion of an emphatic juncture. This would mean, however, that although the Emphatic Juncture is as large a break as the junctures associated with the end of intonation phrases, the Emphatic Juncture itself does not necessarily mark the end of an IP.

Another option is that the material following the inserted IP is parsed into a headless intonation phrase (aka IP Layering Option, shown in Figure 3.10). This scenario involves
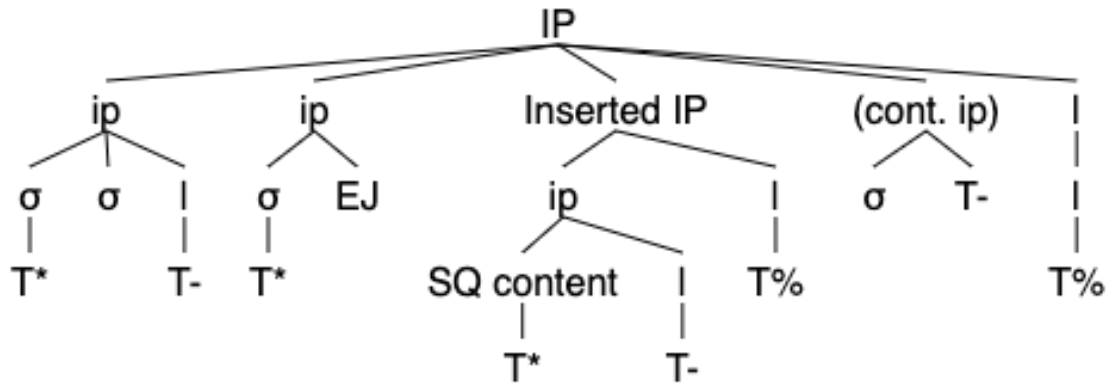
Figure 3.9: The Basic Insertion Option, which does not involve the addition of additional structure and results in stranding material from the original intermediate phrase that is realized at the end of the utterance.

the insertion of additional IP layers, which in turn accounts for the use of the Emphatic Juncture. The IP Layering Option corresponds with the IP layering suggested as part of Ladd's Compound Prosodic Domain (Ladd, 1988).
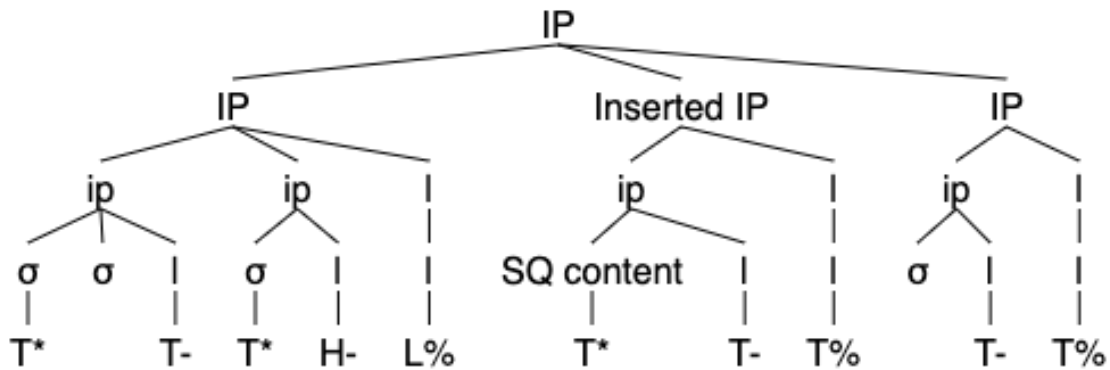
Figure 3.10: The IP Layering Option involves the insertion of additional IPs, including an IP after the subclausal quotation that is headless.

Ladd's proposal was initially put forth to account for differences in declination slopes[1] in multi-clause sentences based on the structural grouping of the clauses. For example, if a sentence contains three clauses arranged in the form of A and B or C, there are two

---

[1]Declination refers to the gradual decrease in pitch over the course of an utterance. Declination slope refers to the rate of this decrease.

groupings available that affect the truth conditions of the sentence. One reading involves the disjunction scoping higher (i.e. (A and B) or C), and the other involves the conjunction scoping higher (i.e. A and (B or C)). In both cases, each of A, B and C is produced as its own intonation phrase. If the disjunction scopes higher, the declination slope of the entire first disjunct matches the declination slope of the second disjunct. Likewise, if the conjunction scopes higher, the declination slope of the first conjunct matches the declination slope of the entire second conjunct. To account for declination slopes corresponding to the larger conjuncts rather than to each intonation phrase, Ladd proposed that each of the intonation phrases that make up the complex conjunct are layered under a larger intonation phrase. This sort of IP layering is what I am proposing as the IP Layering Option for analyzing the prosodic structure of subclausal quotation.

Given that there is precedent for layered IPs in the literature for an unrelated phenomenon and given that it does not necessitate redefining the Emphatic Juncture to be an IP-sized boundary that does not mark the end of an IP, I will proceed with the IP Layering Option as the analysis of the prosodic structure of subclausal quotation. This analysis involves the use of headless IPs and ips.

Thus, the prediction of seemingly-headless intermediate phrases is attested. English is not the first language that has been found to have (seemingly) headless intermediate phrases. Royer and Jun (2019) found that Tatar has headless intermediate phrases which occur in the absence of any phrasal embedding.

3.4.2.2 Prediction 2: Split bitonal pitch accents

The second prediction of embedded IPs is the possibility of a split bitonal pitch accent. This would occur if the nested IP is inserted between the first and second pitch targets of a bitonal pitch accent. Splicing out the nested IP would thus restore a typical bitonal pitch accent. The assembly of the subclausal quotation corpus surfaced an example of this type of bitonal splitting, shown in Figure 3.11. The top is the original and the bottom splices out *quote-unquote* to restore the continuity of the bitonal pitch accent.
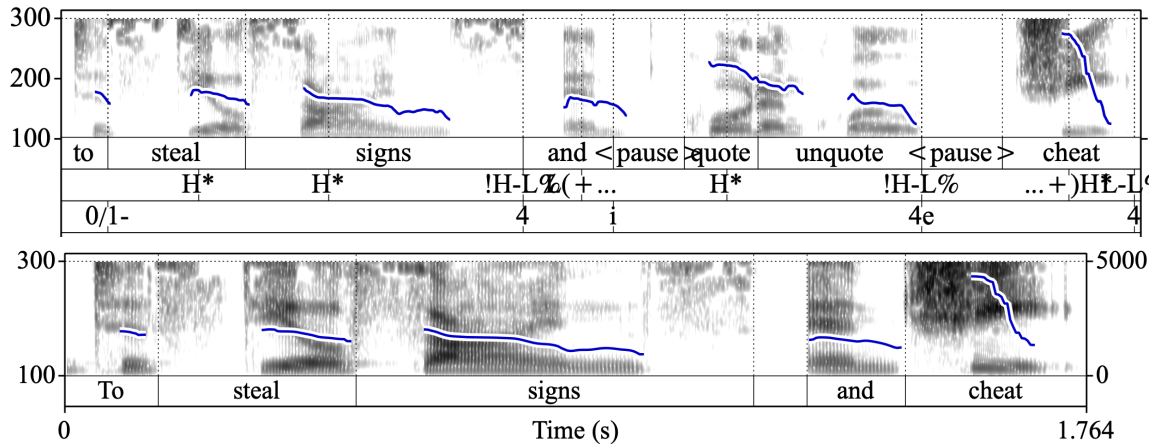
Figure 3.11: Pitch tracks of *To steal signs and (% quote unquote) cheat.* (data from the podcast *Sports? with Katie Nolan*) A pair of sentences demonstrating how an embedded IP can interrupt a bitonal pitch accent. The original sentence is above and the spliced sentence with an intact L+H* is below.

These are only a few examples, but the two predictions that come as a consequence of the possibility of embedded IPs are both attested in a relatively small corpus (N=90 tokens). Thus, the preliminary evidence for embedded IPs in English is promising.

3.4.3   Discussion

The evidence for embedded IPs presented here poses a challenge to the Strict Layer Hypothesis discussed in Section 3.2.3.1. Despite this, I propose that only the strongest version of the Strict Layer Hypothesis is challenged by these data in a meaningful way. A slightly weakened version of the Strict Layer Hypothesis as a violable constraint appears to be perfectly viable and in fact warranted.

Both the Basic Insertion Option and the IP Layering Option discussed in Section 3.4.2.1 involve violations of the Strict Layer Hypothesis. The Basic Insertion Option challenges the tenet of the Strict Layer Hypothesis that a unit for a given level must be exhaustively dominated by an element from the level directly above it (Selkirk, 1984). Both the Basic Insertion Option and the IP Layering Option violate the other tenet of the Strict Layer Hypothesis, that a given non-terminal unit contains one or more elements of

the category immediately below it. Rather, both analysis options involve an IP dominating another IP.

Rather than being an inviolable theoretical principle, I propose that the Strict Layer Hypothesis ought to be treated as a violable constraint. The Strict Layer Hypothesis was proposed and generally accepted for the theory of intonation because under many circumstances, perhaps even most circumstances, languages seem to obey it. Even in the data presented here, the violations of the Strict Layer Hypothesis only involve issues related to one layer of prosodic structure and only with one embedded IP. This contrasts with other types of linguistic embedding, where infinite recursion is (theoretically) possible.

3.4.4 Future research

Since English has three bitonal pitch accents and only one was found to be disrupted in the subclausal quotation corpus, future research may surface instances of the other two being split by embedded material. A more in-depth phonetic analysis would also bolster the initial observation of the continuity of the pitch track when the subclausal quotation is spliced out. Future research could also explore the possibility of embedded IPs in other syntactic phenomena, such as parentheticals or appositives.

The concept of embedded IPs also has several possible implications for speech planning as well as prosodic perception and processing. It would be very interesting to investigate potential predictions of embedded IPs using the methodologies of psycholinguistics.

Finally, if IP embedding is allowed at all, it naturally raises questions regarding how many layers this embedding process can produce. The examples here are limited to one level of IP embedding. Is this a hard limit? Is there a principled theoretical reason for a limit, or is infinite recursion a possibility? While all interesting concepts to consider, I leave the exploration of them to future research.

CHAPTER 4

Subclausal Quotation Intonation Experiment

4.1   Purpose of the experiment

The purpose of the subclausal quotation corpus described in Chapter 2 was to determine how people produce subclausal quotation in spontaneous speech. The corpus data showed that English speakers use three key intonational features to mark subclausal quotation: an emphatic juncture at the start of the quotation, a pitch range reset on the quoted material, and an IP boundary at the end of the quotation.

What the corpus data are not capable of answering, however, is which of these three features are most important to conveying the presence of subclausal quotation. Are any (or all) of the three necessary? Is one sufficient? In other words, which of these features might be doing the heavy lifting semantically? Answering these questions will provide insights to shape the incorporation of intonation into the semantic analysis of subclausal quotation.

4.2   Experiment methodology

4.2.1   Design

The experiment consisted of a 1x5, within subject design. Of the five conditions, two were control conditions and three were experimental. The control conditions were floor and ceiling conditions, with either none of the intonational features or all three of the features of subclausal quotation intonation present, respectively. The three experimental conditions each isolated one feature of subclausal quotation to the exclusion of the other

two. For example, the Emphatic Juncture condition used an emphatic juncture right before the beginning of the quoted material but did not employ a pitch range reset or an IP break at the end of the quoted material. The five conditions are shown in Table 4.1.

Condition	Left EJ	Pitch Range Reset	Right IP break
Neutral prosody (floor control)			
Emphatic Juncture (EJ)	X		
Pitch Range Reset (PRR)		X	
Right IP break (RIP)			X
SQ prosody (ceiling control)	X	X	X

Table 4.1: Each of the experiment conditions based on which intonational features were used. Note that in the Neutral prosody condition (first row), none of the intonational features were used.

The experiment was conducted online using the PennController for IBEX. Given the within subject design, each participant saw each item in one of five conditions. This resulted in five counterbalanced lists where trials were randomized.

4.2.2 Task

4.2.2.1 Designing the task for critical and filler trials

The design of the experimental task was carefully considered. First and foremost, I wanted the task that participants completed to be able to link back to the overarching research question, to identify which of the three intonational features most readily indicates the presence of subclausal quotation, without exposing the purpose of the experiment to participants. I considered several different options for how to ask participants about whether they perceived an instance of subclausal quotation in the stimulus.

Perhaps the most straightforward task would be to ask participants whether they heard any quotation in the sentence. The drawback to this direct approach, however, is that it would have precluded the use of any non-quotation fillers without modifying the basic task.

Having a different task that was specific to the critical trials would have risked making the purpose of the experiment highly transparent.

Another option would have been to ask participants to choose which of two contexts best fits the sentence, one of which would be a better context for subclausal quotation. Right away, though, several potential pitfalls and challenges with this approach appear. An explicitly given experimental context must be pithy and concise. Introducing another speaker making a speech act that the quotation is referencing or who uses expert language is difficult to do in one to two easy-to-understand sentences. Another drawback to this methodology is that the quality of the results hinges on the quality of the contexts the experimenter can come up with. To complicate things further, there are several descriptive usage patterns of subclausal quotation content that I discussed in Chapter 2, Section 6.1. Constructing contexts that concisely capture all of these descriptive categories would be particularly challenging. This methodology would involve two degrees of abstraction: first, participants may vary in whether they think a context is appropriate to license subclausal quotation and second, if they think the intonation used matches the appropriate context.

Yet another option would have been to ask a more abstract question derived from a proposed meaning of subclausal quotation. For example, the task could inquire whether participants thought that part of the sentence had been produced by another speaker at some time before the current utterance. Similar to the forced choice between contexts, however, this methodology would tie the quality and longevity of the experimental results to the quality of the theory of subclausal quotation used. Rather than tether the results to a theoretical abstraction about the meaning of subclausal quotation, I opted to ask participants a question that would get at whether subclausal quotation was present without having to make any theoretical assumptions about its meaning. Additionally, like asking directly about whether a quotation is present, this methodology would make it more difficult to use fillers to mask the target items.

Given the drawbacks to the various alternatives, I decided to ask participants about whether the speaker used a gesture. In written language, subclausal quotation is denoted

using quotation marks. Speakers have the option of visually porting the orthographic quotation marks over to spoken language using the air quotes gesture. By asking about the presence of gesture, I was also able to incorporate fillers into the experiment that did not include quotation. The gesture question may have made it more difficult for participants to consciously recognize quotation or make an overt connection to the airquotes gesture. In my opinion, however, this is not detrimental to the design. Answering based on linguistic intuition without thinking about a particular construction still signals that something linguistically interesting is happening. As a linguist interpreting the results in conjunction with the corpus results, a participant's intuition that a gesture is present is a strong indicator that the participant is detecting a quotation even if they are not consciously aware of what they are identifying.

Thus, the final design of the experimental task consisted of three parts for both critical and filler trials. First, the participant heard an audio recording of an English sentence. Then the participant was asked a Yes/No forced choice question, "Did the speaker use a gesture?" After answering this, the participant was asked to indicate how confident they were in their answer on a 1-7 Likert scale, with 1 being not at all confident and 7 being extremely confident. Participants answered both of these questions for each item. A screenshot of the experimental task is shown in Figure 4.1.

4.2.2.2 Introductory training

Before beginning the body of the experiment, participants completed two training items that were of the same form as the experimental trials. The two items demonstrated the control conditions for the experiment, the same sentence produced with either neutral prosody or with all intonational cues for subclausal quotation. The introductory item in its two versions is shown in 16 and 17. The %EJ symbol indicates an emphatic juncture, the *%r* notation indicates a pitch range reset, and the % symbol indicates an IP break. After completing each training item, participants were given explicit feedback about whether they should have answered yes or no for that item.

Did the person use a gesture?

Yes   No

'F'   'J'

How confident are you in your answer?

Not Confident   Completely
at all   Confident
1   2 3 4 5 6 7
○   ○ ○ ○ ○   ◉

Figure 4.1: Screenshot demonstrating the experimental task. The Yes/No question (top) was displayed immediately after the audio stimulus finished playing, and the confidence question appeared after the participant answered the above question.

(16)   Pierre took his kids to the *%EJ %r* local park % to enjoy the afternoon sunshine.

(17)   Pierre took his kids to the local park to enjoy the afternoon sunshine.

There were two reasons for including explicit training items at the beginning of the experiment. First, it gave the participants a chance to be familiarized with the task design and what they would have to do before reaching the critical items. The second reason has to do with the methodological design of the experiment. Specifically, the design of the experiment was not to determine how well participants can detect the presence of quotation apropos of nothing (although I do think that would be interesting to investigate in future research). Rather, the experiment was designed as a novel, loose variation to the classic phonetic categorical perception (consonant categorization) paradigm (Repp, 1984). For each of the experimental conditions, I was seeking to determine if a single intonational feature (without the other two) was sufficient for participants to classify the intonation

54

of the stimulus as subclausal quotation intonation. This classification was done via the participant selecting whether the speaker used a gesture (presumably air quotes).

The training items provided an explicit opportunity for the participant to hear a canonical subclausal quotation followed by explicit feedback that that sort of sentence ought to be categorized as yes in the experiment. In other words, the participants were being instructed on how to categorize the control conditions. Establishing the sort of stimulus that was expected to elicit a "yes" gesture response gives participants an idea of the category that they ought to consult to determine if a given stimulus should be categorized as "yes" gesture or "no" gesture.

4.2.3   Items

The items for the experiment were audio recordings produced by a native English speaker trained in ToBI production (the author). The recordings were made using a head-mounted microphone in a quiet room. A full list of all target and filler items for the experiment is available in Appendix A.

4.2.3.1   Target items

The experiment had 30 total target items, each of which had five versions corresponding to the five conditions of the experiment (shown in Table 4.1 above). Each target item was designed to (optionally) contain a subclausal quotation in object position which was one or two words in length. An example item quintuplet is shown in 18:

(18)   a. The decorator gave the plastic veneer a coat of paint.   *Neutral prosody control*

b. The decorator gave the *%EJ* plastic veneer a coat of paint.   *EJ condition*

c. The decorator gave the *%r* plastic veneer a coat of paint.   *PRR condition*

d. The decorator gave the plastic veneer *%* a coat of paint.   *Right IP condition*

e. The decorator gave the *%EJ %r*plastic veneer *%* a coat of paint.   *SQ control*

The content of the subclausal quotation was a point of careful consideration. First of

55

all, each item had to be felicitous both with and without quotation. The observations from the corpus (Chapter 2, Section 6.1) showed that subclausal quotations can typically be categorized in (at least) one of three ways: (1) the use of expert or highly precise language, (2) a (near) verbatim speech report, or (3) borrowing another speaker's language with some level of epistemic distancing. The third use is the least straightforward and will be explored in greater detail in Chapter 5. Because of this complexity and the difficulty constructing an experimental context that would license any kind of speech report, I decided to focus the content of the subclausal quotations on phrases that could be considered expert or highly precise language.

As with any experiment, it is crucial to the integrity of the experiment to minimize any variation within an item outside of the designed manipulation. To accomplish this, the two control conditions were recorded with careful attention to pitch accent type and placement. In order to control within each item as much as possible, the three experimental conditions were derived from splicing and combining the two control conditions. All of the sound file manipulations for the experiment were done in Praat (Boersma and Weenink, 2022).

For the emphatic juncture condition, the content from the start of the utterance through the emphatic juncture was spliced in from the subclausal quotation recording and the remainder of the content originated from the neutral prosody recording.

For the pitch range reset condition, the beginning and end of the utterance originated from the neutral prosody recording. The quoted material was spliced in from the quotation recording, but the final syllable of the quotation originated from the neutral recording to remove the lengthening and pitch excursion associated with the quotation-final IP.

For the right IP condition, the final syllable of the quotation was spliced in from the quotation recording and the rest of the utterance was from the neutral prosody recording.

An example quintuplet of the splicing for each condition is shown in Figure 4.2. The blue boxes mark material that originated from the subclausal quotation recording and the orange boxes mark material that originated from the neutral prosody recording.
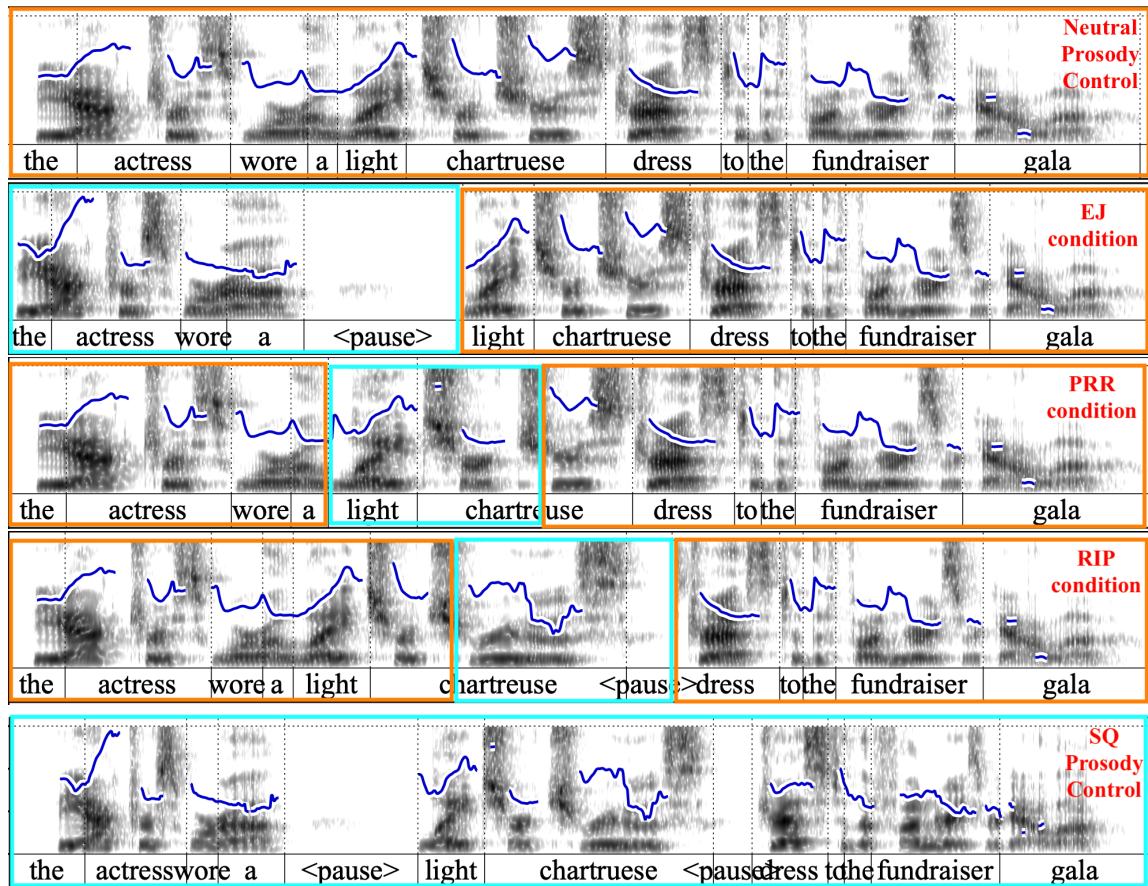
Figure 4.2: An example item quintuplet (*The actress wore a "light chartreuse" dress to the fundraiser gala* showing the source of the recording content for each condition. Orange boxes correspond to the neutral prosody control and cyan boxes correspond to the subclausal quotation prosody control. Order of conditions from top to bottom: Neutral prosody control, EJ condition, PRR condition, RIP condition, SQ prosody control

4.2.3.2   Filler items

In addition to the 30 target items, there were 24 filler items. The fillers in the experiment were sentences that used *there* or *that*. Here are a few examples of filler items:

(19)   Anna put that book in her backpack.

(20)   That's the intersection where the accident happened.

(21)   The hunter hid there hoping to spot a deer.

The reason for including these words in the filler items is that they can easily be ac-

companied by a pointing gesture. This is particularly the case if the speaker puts contrastive/narrow focus prosody on *that* or *there* (Roustan and Dohen, 2010).

Each filler item was designed so that it would be felicitous but not necessary to interpret *there* or *that* as accompanied by a pointing gesture. Having this gesture possibility be available to the participant is what allowed these sentences to be suitable fillers without having to change the response question (i.e. *Did the speaker use a gesture?*).

All of the fillers were produced with broad focus prosody. In an earlier iteration of the experiment, the fillers utilized both broad and contrastive focus prosodic patterns. Unfortunately, this seemed to confused what constituted a "yes" response from participants. Pilot results showed that participants were struggling to categorize the control conditions for subclausal quotation and were often selecting "yes" for all fillers regardless of prosody, suggesting that they were tuning in more to the lexical content (e.g. *that* or *there*) than to the prosody. In the final version of the experiment, therefore, only the broad focus prosody was used for the filler items. With this modification, the filler items patterned more closely to the neutral prosody control condition (more "no" gesture responses).

4.2.4 Participants

The participants for the experiment were 91 UCLA undergraduates who self-identified as native speakers of English. Participants were recruited via the UCLA Psychology Department subject pool and were awarded course credit for their participation. Participants were instructed to complete the experiment on a laptop or desktop computer rather than a mobile device and to use headphones.

4.2.5 Post-experiment debriefing

After the experiment, the participants were prompted to complete an optional debriefing questionnaire. The purpose of the debriefing was to give the participants a chance to express anything about the experiment that might surface some qualitative insights about the task or the experiment more broadly. The participants were asked three open-ended

questions:

1. *What did you think of the experiment?*

2. *Did you notice any patterns?*

3. *Did you use any strategies to answer the questions?*

Through these open-ended responses, participants demonstrated varying degrees of awareness about the purpose of the experiment and their strategies. Many participants gave responses that confirmed they were engaging well with the task. A sample of responses along with the corresponding question number are shown in Table 4.2.

Q	Participant Response
1	"It was quick and simple but because each sentence was unique it kept me attentive."
1	"Well thought out and user friendly."
1	"I was very confused as to what I was doing and needed a better explanation before I started."
2	"Some sentences were read normally some seemed to have hesitation between words"
2	"Just use in tonality. The people had higher pitches at certain vowels."
2	"When words were exaggerated, there was more of a chance that hand movements were used."
3	"I just tried to think of what I would mean if I said that sentence with the same tone infl[e]ctions and rising or falling sounds."
3	"I tried to imagine myself saying the sentence and what I would be doing with my body."
3	"I used the indicated pattern stated in the previous question."

Table 4.2: A selection of participant responses to the post-experiment debriefing questions

Broadly speaking, participants reported listening for pauses or certain words being emphasized. Some participants reported more intuitive responses without an explicit strategy in mind. A few participants reported being attuned to lexical indicators of gesture like *there* and *that* from the filler items. Overall, the debriefing responses confirm that participants were engaged in the task and were noticing the features related to subclausal quotation intonation when listening to the stimuli.

4.3 Results

4.3.1 Data processing

The cleaning and analysis of the results were done in R (R Core Team, 2019) using the Tidyverse package set (Wickham et al., 2019). In addition to working with the raw results, a weighted response score was calculated for each response a participant gave for a given trial. This weighted score was calculated by converting the Likert confidence scale from 1 through 7 to 0 through 6, then multiplying the new rating by either positive or negative 1, corresponding to a yes or no answer respectively. For example, if a participant gave a "yes" response and a confidence rating of 4, this would be converted to a weighted response score of 3. If a participant gave a "no" response and a confidence rating of 6, this would be converted to a weighted response score of -5.

The reason for turning the Likert scale into 0 through 6 was to discount responses where the participant said they were not at all confident in their answer. These "0" responses collapse the gesture responses into one, since we are not interested in their gesture response if they are not at all confident in the answer they selected. The weighted score is used to combine the gesture and confidence responses into a single measure that could be modeled easily.

In the following sections, I will report the raw data from the gesture responses (Section 4.3.2) and the confidence ratings (Section 4.3.3). I will then turn to the weighted gesture response score (Section 4.3.4 and statistical modeling of the results (Section 4.3.5).

4.3.2 Gesture responses

The graph in Figure 4.3 below shows the counts of yes and no responses across our 2 control conditions (i.e., "Neutral Prosody" and "Subclausal Quotation Prosody") and our three experimental conditions.

From these results, the control conditions mirror each other in the expected direction. The Neutral Prosody condition (which had none of the three key intonational features of
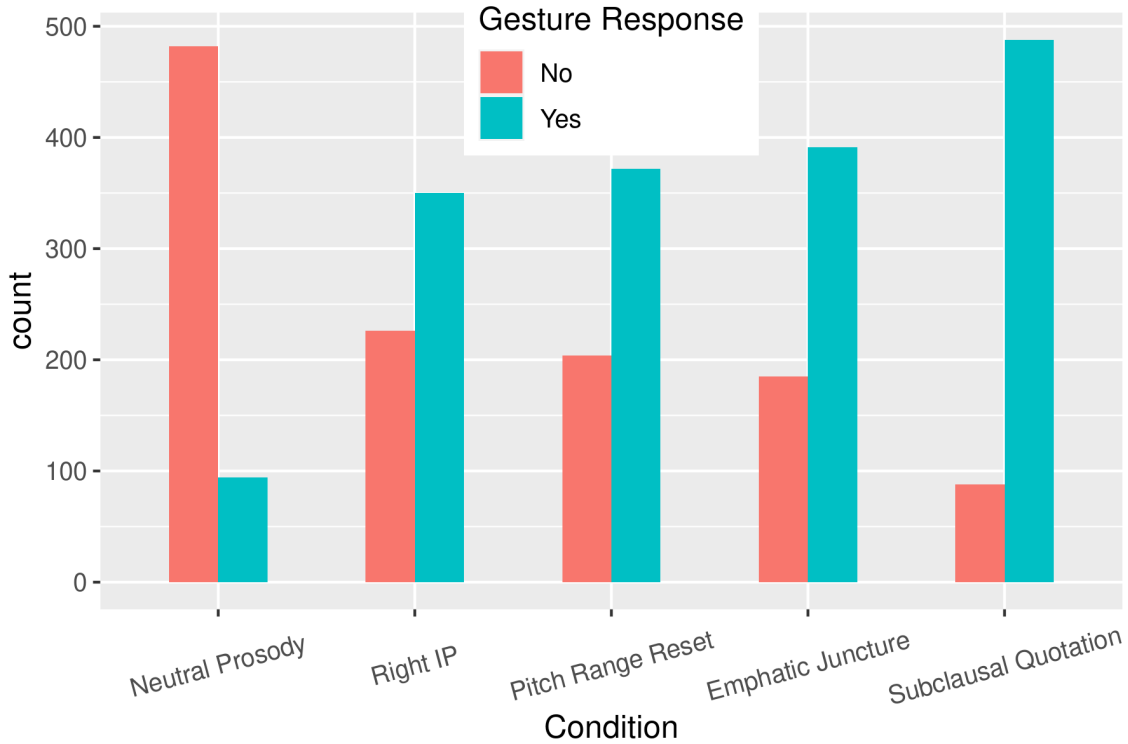
Figure 4.3: Total counts of gesture responses (*yes* or *no*) by condition.

subclausal quotation) received 482 "no" responses (83.7%) out of a total 576 trials. On the other end of the spectrum, the Subclausal Quotation Prosody condition (which had all three key intonational features) received 488 "yes" responses (84.7%). We take this to indicate that our control conditions formed the benchmark for the floor and ceiling for responses as they were designed to do. When evaluating the control conditions, it is important to keep in mind that each participant only saw 6 items per condition. As such, if a participant categorized 5 out of the 6 items the same way (i.e. "no" in the Neutral Prosody condition or "yes" in the Subclausal Quotation Prosody condition), that would give a response rate of about 85%, which is what we see here. It is reasonable, if not expected, that experiment participants do not perform perfectly in a task.

Numerically, the experimental conditions are roughly the same, with a higher proportion of "yes" responses than "no" responses. In aggregate, participants answered "yes" for the experimental conditions 64.4% of the time (1113 out of 1728 trials). The full numerical breakdown of gesture responses by condition is shown in Table 4.3.

Condition	No Responses	Yes Responses
No Quotation (Control)	83.7% (482)	16.3% (94)
Right IP	39.2% (226)	60.8% (350)
Pitch Range Reset	35.4% (204)	64.6% (372)
Emphatic Juncture	32.1% (185)	67.9% (391)
SQ (Control)	15.3% (88)	84.7% (488)

Table 4.3: Participants' gesture response rates by condition. Raw counts of No and Yes responses are listed in parentheses

4.3.3 Confidence ratings

The confidence ratings that participants gave for their gesture responses are shown in Figure 4.4 below.

Once again, the confidence responses for the control conditions pattern as expected. Participants exhibited high confidence when choosing the expected response (5.54 when answering "no" in the No Quotation condition and 5.85 when answering "yes" in the Sub-clausal Quotation condition). Although participants were slightly less confident overall in the experimental conditions, within the experimental condition they were consistently more confident when answering "yes" than when answering "no." Of the three experimental conditions, however, there was the least difference in the distribution of the confidence ratings participants assigned to "no" versus "yes" in the Right IP condition.

4.3.4 Weighted gesture responses

In addition viewing the gesture responses and the confidence ratings, we can get a high-level sense of how participants categorized the stimuli in each condition by examining the weighted gesture responses. Recall that the weighted gesture responses allow us to collapse the binary gesture response and the Likert scale confidence rating into a single metric. This is done by shifting the Likert responses from 1-7 to 0-6 and then multiplying by 1 for "yes" responses or -1 for "no" responses. The mean weighted response for each
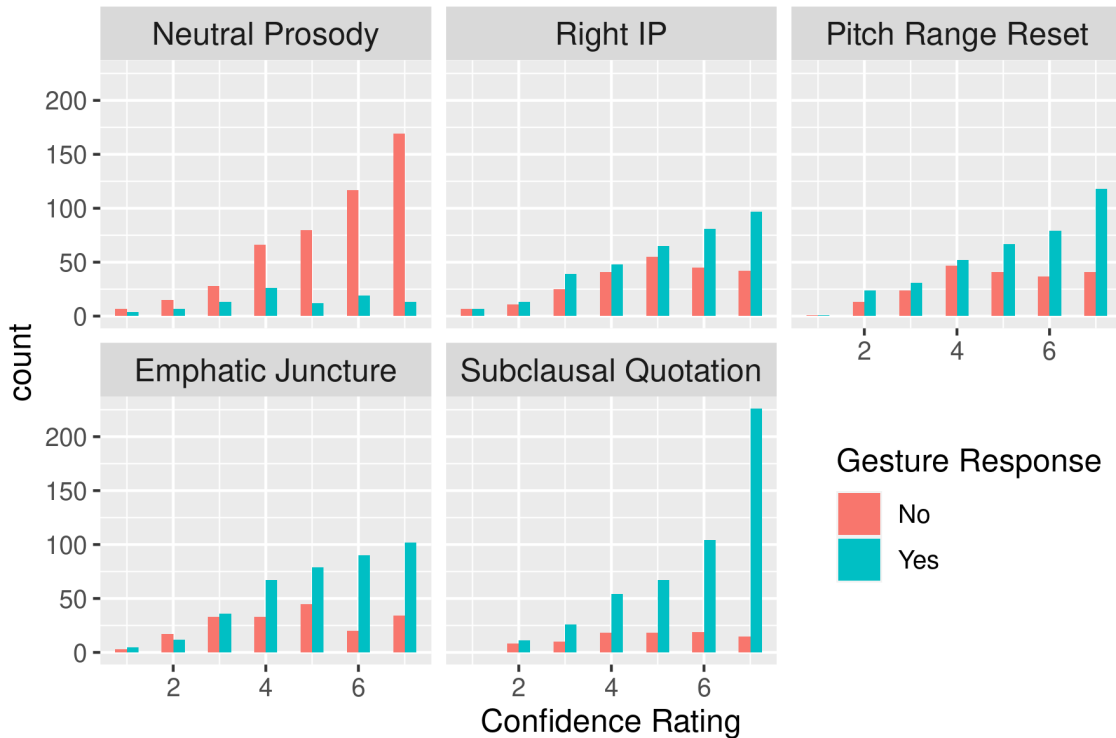
Figure 4.4: Total counts of confidence ratings by condition and gesture response (*no* in red or *yes* in blue).

condition is shown in Figure 4.5.

This high level view of the results once again confirms the separation of the control conditions into a relative floor and ceiling. The three experimental conditions still fall in the middle and are more quotation-like than not. Weighting the gesture responses with the confidence ratings positions the Emphatic Juncture as the most quotation-like of the three experimental conditions. This could be either because the Emphatic Juncture condition produced more confident "yes" responses or less confident "no" responses. On the other end, the Right IP condition has a lower mean weighted response than the other two experimental conditions. The Pitch Range Reset condition falls in between. Statistical modeling was used to investigate if these observations amounted to any significant differences between conditions.
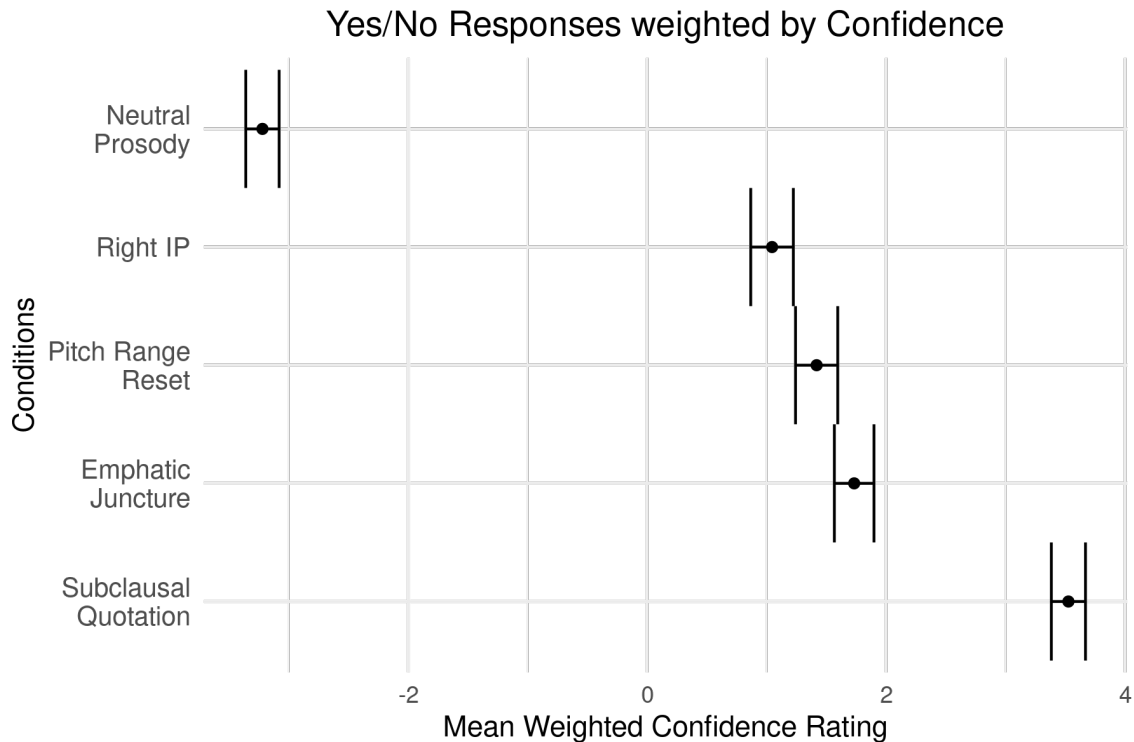
Figure 4.5: Mean weighted responses by condition

4.3.5 Statistical modeling

Bayesian methods were used to analyze the results of the experiment. The benefit of using Bayesian statistical methods instead of Frequentist methods is that the credibility of the model estimates is explicitly calculated, giving more information about the presence, size, and direction of an effect. In addition, explicit calculations of the probability of an effect avoid issues that arise from the arbitrary dividing line between "significant/non-significant" effects. Bayesian methods also provide more easily interpretable results, particularly when it comes to the Credible Intervals (CrI) that surround model estimates. These intervals contrast with Confidence Intervals (CI), which are often erroneously interpreted in how CrIs are meant to be interpreted (i.e., there is some probability that the parameter's true value is within the interval). We used R and the brms packages to conduct the analyses (R Core Team, 2019; **?**).

A generalized linear mixed-effects (glme) model was run to analyze the effect of condition on the gesture responses. The 5 conditions of the experiment were dummy coded

for contrasts. In addition to the fixed effect of condition, the model included random intercepts for participants and items. The R package brms was used to run the the glme model using Bayesian interference. The output of this model is shown in Figure 4.6.
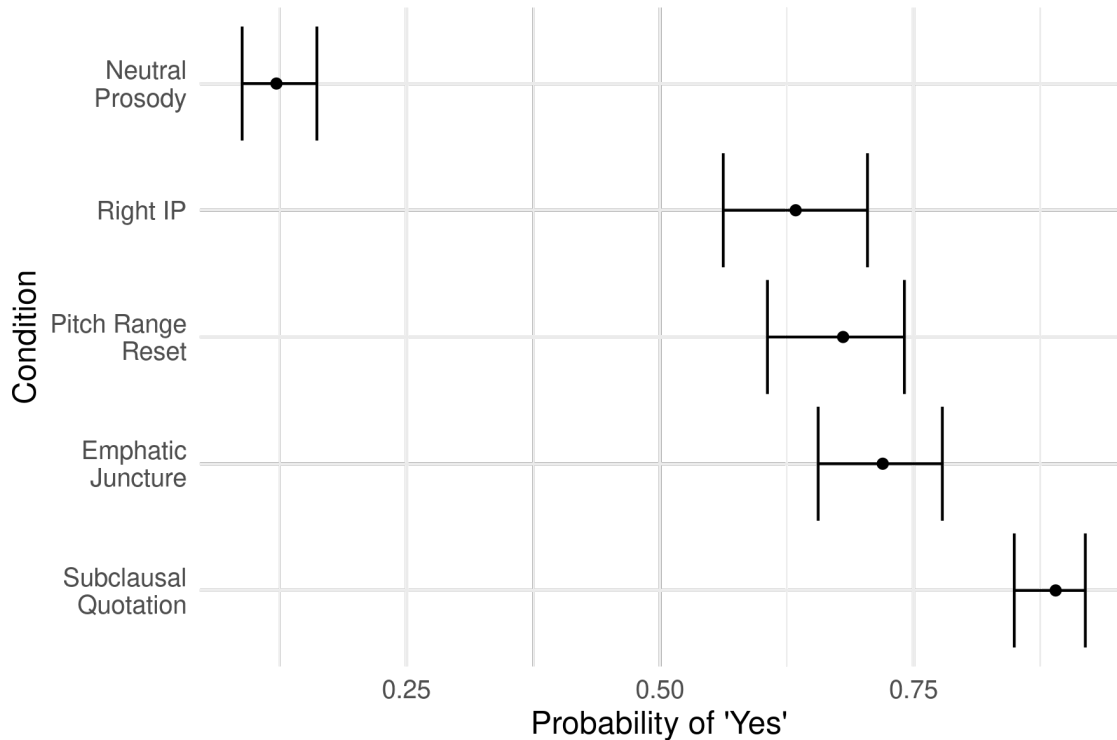
Figure 4.6: Model predicting the probability of "Yes" responses using Bayesian interference

This model predicts the probability of a given stimulus resulting in a "yes" response for each condition. The range of values which we can be 95% confident contain the true probability for each condition is represented by the 95% Credible Intervals.

The model output was run through the R package emmeans to calculate the estimated marginal means of conditions and conduct pairwise comparisons of all conditions. The results of these pairwise comparisons are shown in Figure 4.7.

Because the pairwise comparisons were made using Bayesian methods, we can be confident that any contrast estimate that does not overlap with zero is indicating that there is an effect of condition in that comparison. For each pairwise comparison, the condition listed first is the condition that is less likely to induce a "yes" gesture response. As a
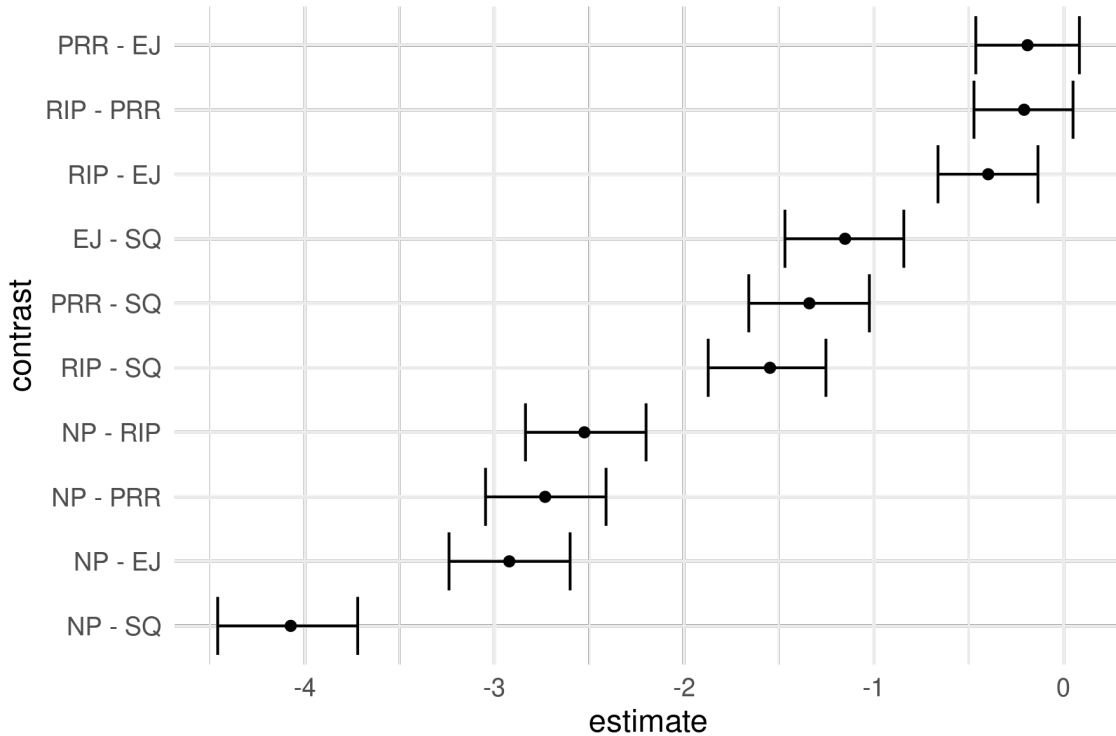
Figure 4.7: Pairwise comparisons of each condition (y-axis) conducted using estimated marginal means of conditions. A larger absolute value on the x-axis corresponds to a larger effect size.

reminder, the results of the experiment showed that "yes" gesture responses occurred least in the Neutral Prosody condition, then the Right IP condition, then the Pitch Range Reset condition, then the Emphatic Juncture condition, and were most common in the Subclausal Quotation Prosody condition.

As expected, the strongest effect is between the control conditions (NQ-SQ). Each of the experimental conditions also produced a strong effect when contrasted with the Neutral Prosody control condition. There is also an effect when contrasting the Subclausal Quotation control condition with each of the experimental conditions, though it is weaker than the effect of the experimental conditions contrasted with the Neutral Prosody control. This also makes sense given that the experimental conditions were all more quotation-like than no-quotation-like.

The contrasts between the experimental conditions showed much smaller effect sizes.

66

The model shows that the Emphatic Juncture condition produces slightly more "yes" gesture responses than the Right IP condition. This is the only contrast among the experimental conditions that does not straddle zero. Although the other two contrasts do include zero in their ranges, zero is towards the edge of the range, indicating that there is a small to negligible difference between those conditions. The effects are in the direction of the Pitch Range Reset condition producing more "yes" gesture responses than the Right IP condition and likewise for the Emphatic Juncture condition over the Pitch Range Reset condition.

4.4 Discussion

4.4.1 Evaluation of the methodology

Participants identified both the floor and ceiling control conditions as expected, indicating that the methodology was effective for distinguishing clear instances of subclausal quotation. Also, unsurprisingly, the three experimental conditions fell in between the floor and ceiling controls.

Based on the participants' feedback from the post-experiment debriefing, some of the participants struggled to understand what was expected of them with the task. Others caught on more easily and did not report any trouble. On the one hand, the struggles reported by some participants suggest that the methodology could have been clearer. In particular, more detailed or explicit instructions or training about various gesture types of interest (including air quotes, pointing, and a few other filler gestures) likely would have reduced confusion about the task.

On the other hand, these differences among participants about the obtuseness of the task are not unexpected in hindsight. Previous research has shown that there can be individual differences in sensitivity to prosodic features based on an individual's Autism Quotient (AQ) score (Jun and Bishop, 2015; Bishop, 2012). It seems possible that the methodology of this experiment unintentionally functioned as an indirect AQ diagnostic.

4.4.2 Interpretation of the results

The contrasts between the experimental conditions shown in Figure 4.7 were the most interesting effects to investigate. They most directly get at the research question undergirding the experiment: which intonational feature most readily conveys the presence of subclausal quotation?

4.4.2.1 Intonational meaning below the tune

All three of the experimental conditions signaled subclausal quotation to a weaker extent than the ceiling control condition, which used all three of the intonational features. It is interesting that for all three of these conditions, omitting the other two intonational features only weakened the perception of gesture rather than completely destroying it. I interpret this as evidence that the tune (i.e. the tonal contour that results from combining the three key intonational features) associated with subclausal quotation is not all or nothing. Although the results suggest that it becomes harder to reliably recover the full-tune meaning (i.e. the presence of subclausal quotation), there is still meaning being conveyed below the tune level in the presence of any one of the key intonational features. This contrasts with work by Goodhue et al. (2016) that claims that there are no meaningful sub-parts of intonation below the tune level. In other words, the meaning of tunes is non-compositional. The results from this experiment show, however, that sub-parts of the tune are still conveying the intended meaning, albeit more weakly.

As an analogy, consider the role of phonemes and morphemes in communicating meaning. The smallest unit of referential meaning is the morpheme. The individual phonemes that make up a morpheme do not carry referential meaning on their own. As such, a simplex morpheme is non-compositional at the phonemic level. Goodhue et al. (2016) are claiming that full tunes are analogous to morphemes, whereas intonational features that make up a tune are analogous to phonemes. The results from this experiment show that this analogy does not hold. The referential meaning of the full tune **is** recoverable from its sub-parts, namely its intonational features. Perhaps even more surprisingly, some intona-

tional features (e.g. the Emphatic Juncture) convey this meaning more reliably in isolation than others.

There is an important caveat that ought to be noted here. The intonational features associated with each of the experimental conditions were more complex than the primitive tones of the MAE_ToBI inventory. Thus, although there is evidence for intonational meaning below the tune level, the results of this experiment are agnostic to what the smallest meaningful component of intonation is. In other words, these results are not endorsing the recoverability of meaning at the level of individual notational elements (e.g. basic tones H and L, or abstract prosodic categories like pitch accent, phrase accent or boundary tone) a la Pierrehumbert and Hirschberg (1990).

4.4.2.2 The Emphatic Juncture signals the start of subclausal quotation

Of the three experimental conditions, the Emphatic Juncture condition was marginally closer to the Subclausal Quotation condition compared to the other two. I interpret this to mean that the Emphatic Juncture is slightly more obvious as a signal to the presence of quotation. I believe that the Emphatic Juncture being a slightly better subclausal quotation signal makes intuitive sense as well. The Emphatic Juncture consists of an unexpected combination of acoustic features. On the one hand, the boundary tone sequence is a pitch plateau, which indicates an impending continuation and lack of completion (Pierrehumbert and Hirschberg, 1990). On the other hand, the presence of an obligatory pause forces a certain level of discontinuity within the utterance. These features are naturally in tension and thus form an intriguing pair; but I propose that their combination is remarkably sensible for marking the start of subclausal quotation. The pause not only gets the listener's attention (Diachek and Brown-Schmidt, 2022), but also clearly denotes acoustically where the subclausal quotation begins, on par with using the lexical marker *quote*. Meanwhile, the plateau portion of the Emphatic Juncture signals that the quoted material ought to still be integrated into the rest of the utterance, corresponding to the semantic use of the quoted content in addition to its mention. (I will discuss this concept in greater detail in the fol-

lowing chapter.) The plateau may also be serving to signal an interruption of the current IP due to the start of an embedded IP, as discussed in Chapter 3, Section 4. Further research could aid in confirming these two functions of the Emphatic Juncture plateau and pause.

Although the Emphatic Juncture condition was only slightly more quotation-like than the other two experimental conditions, the experimental results combined with the putative role of the Emphatic Juncture make it the best candidate of the three intonational features to include in the semantic analysis of subclausal quotation. I will endeavor to do so in the next chapter.

4.4.2.3 The Right IP is the weakest signal of subclausal quotation

The final observation based on the experimental results is the relative weakness of the Right IP in signaling subclausal quotation compared to the other two conditions. There are several plausible explanations for this. For instance, speakers can insert additional IPs for a variety of reasons, including to signal syntactic structure (Selkirk, 2002; Fodor, 1998; Jun, 2003) or to break up long utterances into same-sized sisters for processing purposes (Fodor, 1998). Because there are several reasons a speaker might choose to insert an IP boundary, it is more difficult to recover the presence of subclausal quotation based solely on the presence of an IP juncture.

Rather than signaling the presence of subclausal quotation, however, I contend that the Right IP has a different function in the intonational marking of quotation. Specifically, it functions to mark the end of the quoted material once an instance of subclausal quotation has already been established through either a lexical marker or an emphatic juncture. I will discuss the in more depth in the following chapter.

4.5 Summary

In this chapter, I presented an experiment designed to determine to what extent each of the three key intonational features convey the presence of subclausal quotation. Results

showed that each of the three intonational features is capable of conveying the presence of subclausal quotation, but to a weaker extent than when the full tune (all three features) is present. This is evidence that intonational meaning is recoverable from elements below the full tune.

Of the three intonational features, the Emphatic Juncture was the strongest signal to the presence of subclausal quotation. It is thus the most attractive of the three features to attempt to incorporate into a semantic analysis of subclausal quotation.

CHAPTER 5

Semantics

5.1 Setting the stage

5.1.1 Quotation as metalinguistic mention

The phenomenon of quotation has long been of interest to both semanticists and philosophers of language (see Cappelen et al. (2020) for an overview). Some of the earliest work (Tarski, 1956; Quine, 1980) focuses on the metalinguistic aspect of quotation, noting the distinction between use and mention. A quoted expression functions to mention that particular linguistic expression itself rather than its denotation (i.e. its use). Specifically, quotations are unstructured proper names of the quoted expressions (hence the name of this theory, Proper Name Theory). An example of subclausal quotation being used purely for metalinguistic mention can be seen in 22:

(22)   a. "UCLA" is an acronym.

         b. "Linguistics" has eleven letters.

An important feature of this sort of metalinguistic quotation is its opacity. A substitution of a synonymous or co-referential expression does not preserve the truth value of the sentence, as shown in 23:

(23)   a. "UCLA" is an acronym. (*True*)

         b. "University of California, Los Angeles" is an acronym. (*False*)

Proper Name Theory cleanly accounts for this lack of co-referential substitution. It also has several weaknesses, however. First of all, there seems to be no limit in speakers'

capacity to generate new quotations, which suggests that there are some principles or structure undergirding their generation (Davidson, 1979). Under Proper Name Theory, however, we would expect hearing a quotation to be equivalent to learning a new proper name. This does not seem to match with the ease with which speakers can understand and even generate new quotations.

Davidson (1979) proposes a refinement to this sort of approach, what he calls the "demonstrative theory" of quotation. The key principle of his theory is that quotes are definite descriptions containing demonstratives. Under his analysis, the quotation marks themselves do the heavy lifting and can be interpreted as "the expression a token of which is here" (Davidson, 1979). Demonstrative Theory accounts for speakers' seemingly infinite capacity to produce and understand quotation, since there is a direct link between the demonstration (metalinguistic) and the typical use of an expression. One weakness of Davidson's theory is its total reliance on orthographic quotation marks for the interpretation of quoted expressions (Reimer, 1996), when examples like those in 22 can just as easily be understood when delivered orally or manually (i.e. signed) rather than orthographically.

5.1.2 Subclausal quotation: More than just mention

These earlier accounts all seek to analyze quotation writ large. They serve as a helpful starting place, but quotation has a multitude of uses that warrant more nuanced treatments. In particular, the type of subclausal quotation being investigated in this dissertation clearly contributes more than simply to mention the linguistic expression within the quotation. There are several sources of evidence that show subclausal quotation can function beyond pure mention. First of all, subclausal quotations can saturate whatever arguments the quoted material does (Cappelen and Lepore, 1997). In other words, the content within the quotation is grammatically incorporated into the rest of the sentence as it would be if the quotation marks were absent. Several examples from the SQ corpus demonstrating this are shown below:

(24) Romeo Mattison was "training" Goron Pezar's wife. *SQ as V, from Barry*

(25) We're throwing a "fundraiser" on your birthday. *SQ as N, from Schitt's Creek*

(26) It's almost offensive when we're criticized on how quote-unquote "terrible" healthcare is in this country. *SQ as Adj, from C-SPAN*

(27) Fanny and Freddie were found to "have been cooking the books." *SQ as VP, from C-SPAN*

The above examples show that subclausal quotations are not automatically collapsed into a single semantic type simply by virtue of being quotations. Additionally, evidence from *wh*-movement and NPI licensing shows that subclausal quotations are not syntactically opaque metalinguistic objects (Maier, 2014). Consider the following fabricated examples:

(28) What did Adam want to "give to Rebecca?" *wh-movement out of SQ*

(29) Pierre never "earned anything" for his good behavior. *NPI licensed from outside SQ*

Semantically, the lack of syntactic opacity correlates with the fact that subclausal quotation employs both the use and mention of the quoted expression (Davidson, 1979).

The complexity of subclausal quotation is not new to semanticists and philosophers (see De Brabanter (2010) for an overview). In fact, it has been difficult for scholars in this area to even agree on a term for the phenomenon (or perhaps phenomena) at issue. It has been referred to as hybrid quotation (Recanati, 2001), mixed quotation (originating from mixed direct and indirect quotation (Partee, 1973)), incorporated quotation (Clark and Gerrig, 1990), double-duty quotation (García-Carpintero, 2003), impure quotation (Gómez-Torrente, 2003) or subclausal quotation (Potts, 2007). There has also been a distinction in the literature regarding scare quotes, distinguishing them from other types of subclausal quotation. For the purposes of this dissertation, I am referring to the phenomenon as subclausal quotation. This is an attempt to be theory-neutral, since the phenomenon of interest is quite literally a quotation that occurs on the subclausal level.

74

5.1.3   Semantic analyses of subclausal quotation

Several semanticists have put forth analyses for just the sort of subclausal quotation described above. I will briefly explore two such analyses, one from Cappelen and Lepore (Cappelen and Lepore, 1997), and another from Potts (2007). As we will see, both of these analyses rely on semantic operators that are either silent or represented only in the orthography. As such, my intended contribution is not to innovate an analysis from scratch or even propose a new semantic function for quotation. Rather, I will show that existing theories of subclausal quotation can be enriched by including intonation.

5.1.3.1   Subclausal quotation as both direct and indirect discourse (Cappelen and Lepore, 1997)

Cappelen and Lepore (1997) are the one of the first to directly discuss the phenomenon of what they call mixed quotation (and what I will henceforth continue to call subclausal quotation), as shown in 30.

(30)   Alice said that life "is difficult to understand."[1]

They demonstrate how theories of other types of quotation (pure, direct, or indirect) are insufficient to account for subclausal quotation and put forth a unified account of all four types of quotation. They analyze subclausal quotation as having two meaning contributions joined by conjunction (combining Davidson's remarks on direct discourse (Davidson, 1979) and indirect discourse (Davidson, 1968)). Davidson defines the sametokening operation as a (silent) demonstrative responsible for direct quotation (i.e. a verbatim reproduction of an utterance that another speaker has said). Similarly, samesaying is the indirect quotation version, where the speaker uses a demonstrative to report the meaning of something someone else has said without necessarily using precisely the same words.

---

[1]It is interesting to note that the form of this quotation resembles only a small number of the quotations from the subclausal quotation corpus discussed in Chapter 2. Preceding the quotation within the same sentence, there is a source for the quotation (i.e. Alice) and a verb of saying. There are a variety of types of subclausal quotation that do not fit these criteria but should still be accounted for in a comprehensive theory of subclausal quotation.

Cappelen and Lepore combine these two functions for subclausal quotation. In particular, cases of subclausal quotation can be used to attribute both the sametokening and the samesaying relation between the same two utterances. For the sentence in 30, this works out to the following denotation:

$$\exists u(\textbf{says}(a, u) \& \textbf{samesays}(u, \text{that}_1) \& \textbf{sametokens}(u, \text{this}_2))[\text{Life}[\text{is difficult to understand}]_2]_1$$

The bracketed material in this denotation is marked with subscripts to indicate which of the demonstratives it corresponds to. The indirect quotation (*that*, marked with subscript 1) marks the full proposition that Alice said (i.e. samesaying). Only a portion of what Alice said is being directly quoted (i.e. sametokening). The direct quotation (*this*, marked with subscript 2) indicates the portion of the utterance that the speaker is reproducing from the source verbatim.

Through this combination of Davidson's two accounts of direct and indirect discourse, Cappelan and Lepore claim to achieve their stated objective of analyzing subclausal quotation in a way that overlaps with the three other forms of quotation (i.e. pure linguistic mention, direct, and indirect quotation). It should be noted, however, that the examples of mixed quotation they examine and analyze all occur as substrings of utterances that contain indirect discourse. In other words, there is always a specified source and a verb of saying preceding the quotation within the same sentence. In 30, *Alice* is the source and *said* is the verb of saying. It is unclear to me how their analysis would handle a quotation which does not have a source and a verb of saying higher in the sentence.

Additionally, their proposal runs into a similar issue to Davidson's (1979) in that the meaning contribution of quotation is contributed by the quotation marks. Moreover, when there are two sets of brackets corresponding to each of the demonstratives and the contents of those brackets partially overlap, such as in the denotation of 30 above, how can a speaker indicate what content belongs in which set of brackets in a way that is recoverable by the listener? While they acknowledge that quotation happens outside of written speech, they do not expound on how their theory would be extended to domains beyond orthography.

Presumably, they would think that there is either a silent correlate in spoken quotation or that it is designated by some non-lexical means. As I will show, I believe the latter option is the case, with intonation marking spoken quotation.

5.1.3.2 Multidimensional analysis of subclausal quotation (Potts, 2007)

Potts (2007) puts forth a proposal that seeks to unify the analysis of full clause and subclausal quotation. He accounts for the metalinguistic (i.e. mention) aspect of quotation by introducing a new semantic type *u* for linguistic entities. He begins by doing so for full clause quotation, such as the example in 31.

(31) Lisa said, "Homer is bald."

Potts proposes that the interpretation of quotation hinges on the covert use of the verb **utter**, which has a new type $\langle u, \langle e, t \rangle \rangle$. Covert **utter** pairs individuals with natural language expressions, indicating that a particular speaker uttered a particular linguistic string. As such, it functions similarly to the same-tokening operator of Cappelen and Lepore (1997) in that it only occurs in the presence of a direct quotation. Potts emphasizes that the **utter** operator is not analogous to a speech act operator of the performative verb hypothesis (Ross, 1970; Krifka, 1999; Geurts and Maier, 2005) because the utterance relation is agnostic to the communicative intent of the speaker. Rather, **utter** is simply a relation between individuals and linguistic objects.

Potts handles the simultaneous use and mention of quotation by analyzing it as having two dimensions of meaning. The two dimensions are kept independent throughout the denotation by using angled brackets. The first element in the ordered pair is the usual semantic content, and the second element is the metalinguistic content. This is similar to but contrasts with earlier theories, such as Cappelen and Lepore (1997), where the metalinguistic portion of the meaning is a conjunct in the denotation. In this framework, the interpretation of the sentence in 31 is shown below:

⟦Lisa said, "Homer is bald."⟧ = ⟨⟦**utter**⟧("Homer is bald")(l), ⟦**say**⟧(⟦**bald(Homer)**⟧)(l)⟩

Introducing the type u also allows for the complex interpretations of subclausal quotation involving both use and mention. In order to simultaneously represent both the use and mention meanings of subclausal quotation, he introduces a function **quote-shift** of the type $\langle u, \langle e, \sigma \times \tau \rangle \rangle$. Crucially, the first element of this tuple, which corresponds to the regular semantic denotation (i.e. its use), is relativized to the utterance worlds of the entity argument. This allows the speaker employing the quotation to utilize another person's expression-to-meaning mappings that do not obtain for the speaker. The interpretation of Potts's quotation function **quote-shift** is shown below:

⟦**quote-shift**⟧$(P)(d)$ = ⟨the X such that **say**(⟦X⟧ = ⟦$SEM(P)$⟧)(d)), ⟦**utter**⟧$(P)(d)$⟩

for any $P \in D_u$ and $d \in D_e$

In other words, using **quote-shift** allows the speaker to not only mention a linguistic expression from another speaker, but also to use it to contribute to their meaning even if they would not typically use that expression in the way the other speaker did.

At this point, there is an important way in which Potts's theory of subclausal quotation appears to be impoverished. Potts himself acknowledges this in his concluding remarks:

"It seems clear also that a complete theory of quotation will reference specific intonation contours as the auditory equivalent of quotation marks."
(Potts (2007), p. 24)

. Although this is a high-level prediction, it speaks directly to the contribution I am seeking to make here. In the next section I will lay out how the intonational correlates of subclausal quotation described earlier in this dissertation integrate into Potts's analysis.

5.2 Incorporating intonation into a semantic analysis

Both of the analyses of subclausal quotation discussed above would be enriched through the inclusion of intonation. Both accounts do not articulate how the semantic operators they introduce are realized outside of written language, when orthographic quotation marks are unavailable. In particular, the machinery of Potts (2007) that needs to be marked is the **quote-shift** operator, which includes covert **utter** in its denotation. For Cappelen and Lepore (1997), the implicit demonstratives corresponding to **same-saying** and **same-tokening** need to be realized.

While I believe that it is possible to incorporate intonation into either of these analyses, accomplishing this for Cappelen and Lepore constitutes a more substantial challenge. There are two independent operators that could be marked (i.e. **same-saying** or **same-tokening**). The content that is being same-said and the content that is being same-tokened could also be marked. Despite the similarity between Cappelan and Lepore's two operators, **same-tokening** only happens in cases of direct quotation. It seems stipulative to mark only one of them with intonation while leaving the other entirely covert. A proper attempt to expand the analysis of Cappelen and Lepore to include intonation should also consider how intonation is used to mark indirect quotation (i.e. **same-saying**), which is beyond the scope of this dissertation. As such, I will focus on demonstrating the incorporation of intonation into Potts's analysis.

5.2.1 Exploring the marking of subclausal quotation

5.2.1.1 Why does quotation need to be explicitly marked?

In one sense, subclausal quotation is an optional process. Speakers can make an utterance with or without including a substring as being quoted. One could imagine the counterparts to the sentences in 24-27 that lack quotation. The non-quotation versions are still fully acceptable, grammatical sentences, though lacking the meaning contribution of the quotation. Because of this optionality, it is paramount that quotation be somehow

externalized in order for the speaker to make the presence of quotation clear to the listener.

5.2.1.2   What sort of marking do we expect for quotation?

Based on Potts's theory of subclausal quotation, we can predict what we expect ought to be marked in order to clearly communicate subclausal quotation to the listener. Three indicators come to mind: (1) the speaker must signal the presence of subclausal quotation (i.e. that the utterance contains a subclausal quotation), (2) the start of the subclausal quotation, and (3) the end of the subclausal quotation. The first indicator corresponds to the inclusion of **quote-shift** in the interpretation of the utterance. The second and third indicators mark the quoted string, which is essential for evaluating **quote-shift**. In the following sections I will consider how the observed intonation corresponds with these predictions.

5.2.1.3   Available strategies for marking quotation

As the subclausal quotation corpus discussed in Chapter 2 showed, English speakers have several strategies available for marking the presence of quotation. Speakers can mark the start of quotation using a lexical marker like *quote* or *quote-unquote* or by making an air quotes gesture with their fingers. Speakers also mark subclausal quotation intonationally.

As discussed in Chapter 2, Section 4, there are three key intonational features that speakers use to mark subclausal quotation. First, speakers mark the start of a quotation using an emphatic juncture, which is a large juncture that includes a pitch plateau and an obligatory pause (Sturman, 2019). The second feature is a pitch range reset on the quoted material. This results in the High targets within the quotation being higher than they would be otherwise, making them more prominent acoustically (Ayers, 1996). Lastly, speakers mark the end of a quotation using an IP break. While these three features are all serving to mark intonation, they have different roles in doing so, which correspond to the marking indicators predicted by the semantic theory in the previous section.

The quantitative analysis of the corpus revealed that overlap among cues occurs fre-

quently but is not required. Table 5.1 (repeated from Chapter 2) shows the occurrence of each of the three intonational features. The results are separated into rows based on the absence or presence of a lexical marker, as well as the overall rates.

	EJ	Pitch Range Reset	Right Edge IP
No lexical marker (N=20)	90%	85%	90%
Lexical marker (N=70)	73%	79%	89%
Overall (N=90)	76%	80%	89%

Table 5.1: Occurrence rates for each of the key intonational features found in the Subclausal Quotation Corpus, separated by row based on whether the token included a lexical marker (*quote* or *quote-unquote*) as well as the overall rates.

What can we glean from these quantitative results as it relates to the marking of quotation? I propose that the occurrence rates of each of these features, and in particular how they vary in the presence of lexical markers, gives insight into the role each intonational feature is playing in marking subclausal quotation.

5.2.1.4 The Emphatic Juncture is the intonational externalization of quote-shift

Let us begin by looking at the Emphatic Juncture. The Emphatic Juncture is the only feature that had a significant difference in usage based on whether the speaker used a lexical marker such as *quote* or *quote-unquote* at the start of a quotation, with less frequent usage when a lexical marker was present. I interpret this to mean that the Emphatic Juncture is the intonational manifestation of **quote-shift**. The speaker employs the Emphatic Juncture to signal two things: (1) that they are using a subclausal quotation construction (i.e. the presence of **quote-shift** in the semantic representation of the utterance) and (2) that the quoted string is beginning following the juncture.

The results of the experiment described in Chapter 4 provide additional evidence for the Emphatic Juncture's role in externalizing quotation. Of the three intonational features, the Emphatic Juncture condition produced the highest rate of quotation-like responses.

81

Though it was a less reliable signal than when all three features were present, participants were still able to detect the presence of quotation fairly reliably when they heard the Emphatic Juncture. I take this as further support from another source that the Emphatic Juncture functions to externalize **quote-shift**.

Another strategy speakers have to externalize **quote-shift** is using a lexical marker. When speakers opt to use a lexical marker, that word can fulfill both of the functions related to signaling the start of a quotation. This renders the use of the Emphatic Juncture optional, which results in the lower (but still common) use of the Emphatic Juncture in the Lexical Marker segment of the corpus shown in Table 5.1.

One puzzle that arises from the quantitative analysis of the corpus is the overlap between marking strategies. Speakers can mark the presence of subclausal quotation with the Emphatic Juncture, a lexical marker, or with an air quotes gesture aligned with the quotation. In fact, the corpus included an example where all three of these marking strategies are attested, shown in Figure 5.1.[2]
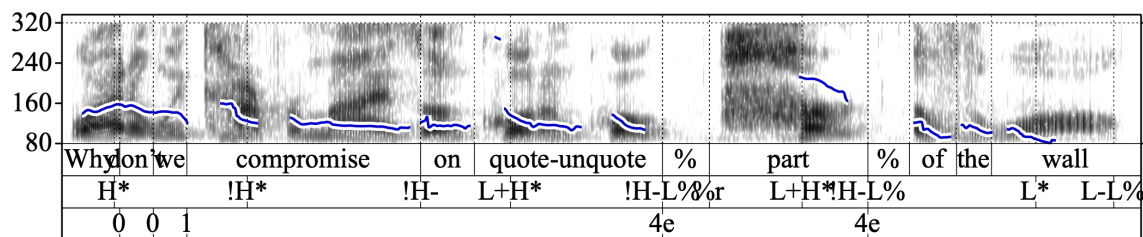
Figure 5.1: A token from the subclausal quotation corpus where the speaker marks the quotation with all three modalities (lexical marker, all three intonational cues, and gesture) Transcript: *Why don't we compromise on quote-unquote % part % of the wall* (data from C-SPAN)

The puzzle here is that each of the three marking strategies (Emphatic Juncture, lexical marking, and the air quotes gesture) is independently capable of signaling the presence of subclausal quotation and the start of the quoted string. Why, then, do speakers employ

---

[2]This token was taken from a video recording of a C-SPAN broadcast, allowing us to observe the air quotes gesture through the video.

more than one of these strategies to mark the same quotation? And why is the presence of all three together not perceived as redundant?

While I do not have a full-fledged theory for how and why these three strategies occur simultaneously without redundancy, I take this as evidence that the different modalities represented (e.g. words, intonation, gesturing) are not making identical, entirely equivalent contributions. Further work investigating the nuances of the meaning contributions of each strategy would be an interesting direction for future research.

5.2.1.5 The pitch range reset boosts the prominence of the quoted material

The second intonational feature that marks the presence of quotation is the pitch range reset. The pitch range reset occurs after the presence of subclausal quotation has been established and the quoted string has begun, but before the end of the quotation. I hypothesize that the role of the pitch range reset is to mark the quoted material as noteworthy by increasing its acoustic prominence.

5.2.1.6 The right IP boundary marks the end of the quotation

Finally, we come to the role of the right IP boundary. Once a subclausal quotation has been established and the quoted string begins, the next signal that needs to be marked in the signal is where the quoted string ends. While it may seem trivial, this is an important aspect of computing the meaning of a sentence that includes quotation. Semantic analyses of subclausal quotation often take for granted what exactly is contained within a given quotation because it is clearly marked in the orthography with the close of the quotation marks. But how is the end of a quotation signaled in oral speech, when orthographic quotation marks are unavailable?

This is the function of the right IP boundary. An IP boundary is a large juncture, and speakers naturally tend to align large intonational junctures with large syntactic or information structural boundaries (Selkirk, 1996), making an IP boundary an ideal candidate for marking the end of a subclausal quotation. The quantitative results from the corpus

shown in 5.1 indicate that the right IP boundary serves to delimit the end of the quoted string regardless of whether the start of the quotation includes a lexical marker (*quote* or *quote-unquote*).

5.2.2 *Quote* versus *unquote*

One interesting observation from the corpus is the asymmetry in the use of the lexical markers *quote* and *unquote*. While speakers often use *quote* to mark the beginning of a quotation, using *unquote* when the quotation ends is much more rare. Interestingly, speakers do say *unquote* but typically do so compounded with *quote* (to form *quote-unquote*) at the beginning of the quotation. Regardless of whether they used an Emphatic Juncture or a lexical marker at the start of the quotation, though, speakers use an IP boundary to mark the end of the quotation.

If *unquote* is available as an option to mark the end of quotation, why is it not used more often? I posit that there is an informativity difference between *quote* and *unquote*. When *quote* is used, it performs two of the functions proposed in Section 5.2.1.2 above: (1) signaling the presence of a subclausal quotation and (2) indicating the start of the quoted string. In the case of *unquote*, however, it is only serving to mark the end of the quoted string. Thus, *unquote* has an impoverished meaning relative to its counterpart *quote*. This informativity asymmetry results in *unquote* being marked in comparison to the intonational alternative, inserting an IP boundary to signal the end of the quotation.

5.3 Communicative uses of subclausal quotation

Recall that the literature makes a three-way distinction about the types of quotation that happen at the subclausal level: (pure) metalinguistic, mixed (a combination of direct and indirect discourse), and scare quotes. In the subclausal quotation corpus, however, utterances do not always map neatly into one of these three categories. Instead of trying to shoehorn the data into the literature, I will start from generalizations about the com-

municative uses of subclausal quotation based on the corpus and then extrapolate up to theoretical insights.

5.3.1 Descriptive categories for subclausal quotation

Descriptively, I observed three broad categories of subclausal quotation content. The first category is used to flag neologisms or uncommon terms in an effort to appeal to authority (e.g. a term an expert knows but non-experts have probably never heard). The second category of content can be characterized as using quotation to explicitly report what someone said. This kind of quotation happens frequently in news reporting, courtrooms, and other places where people are concerned with replicating the exact words someone used as accurately as possible. The third usage of quotation can be broadly construed as a means for the speaker to create epistemic distancing. The content of this type of quotation is often controversial, something that when used might be characterized by some as provocative or offensive. Below are examples of each type of quotation from the subclausal quotation corpus:

(32) The ones who were really into quote-unquote "juuling" could often do a pod in a day. *Neologism/Appeal to authority, from C-SPAN*

(33) President Trump directed us to quote "talk with Rudy." *Speech report, from NPR*

(34) More and more people are quote-unquote "home-grown Americans" who don't come from other countries. *Epistemic distancing, from C-SPAN*

As exemplified in 33, the Speech Report use typically involves a specified source earlier in the discourse and a verb of saying preceding the quotation. This stipulation seems reasonable, particularly if the purpose of this usage is to directly report what someone has said.

5.3.2 Ambiguity among communicative uses

In many cases, however, it is difficult to cleanly categorize an utterance into one of these communicative use bins. Consider for example, this extended quotation that includes more context around one of the corpus tokens ("good old neighbor") from an NPR broadcast:

> Putin in fact recently expressed some relief that now it's Ukraine which is being accused of interfering in the US elections and not Russia. He's also said that Ukraine shouldn't rely on faraway friends and instead turn to its "good old neighbor" Russia, but sarcasm aside, you know, the Kremlin is enjoying this focus on Ukraine in the US. (*NPR WLRH broadcast from 11.26.19*)

In this passage, the subclausal quotation "good old neighbor" occurs following a specified source (Putin) and a verb of saying. Both of these would support the Speech Report use of subclausal quotation. But immediately following the quotation, the speaker indicates that he has just used sarcasm, which aligns with the Epistemic Distancing use. This suggests that subclausal quotations are potentially ambiguous among the available communicative uses.

Unfortunately, speakers are not usually so transparent about their stylistic or communicative intentions. Moreover, it should be noted that there is no observable intonational differences that could provide empirical grounds for associating a given token with a particular communicative use of subclausal quotation. In the absence of any external differentiators on the quotation itself, it seems prudent to appeal to context (and therefore pragmatics) to account for the ways speakers use quotation.

5.3.3 Pragmatic content of subclausal quotation

As always, the guiding force of Gricean pragmatics is the Cooperative Principle (Grice, 1978). In the case of subclausal quotation, we assume that the speaker is following the Cooperative Principle, and that means that using the quotation is relevant to the commu-

nicative purposes of the conversation. The speaker tries to be as relevant as possible in all circumstances (Wilson and Sperber, 1986). In particular, it is relevant that x said u, where x is the source of the quotation and u is the content of the quotation. The listener then is tasked with figuring out why it is relevant that x said u. There are a few reasons that this could be, which point us to different communicative uses of quotation.

I propose that the relevance implicatures generated by two of the three types of subclausal quotation center around content (un)endorsement. Subclausal quotations that are appealing to authority result in a positive endorsement, whereas epistemic distancing quotations result in content unendorsement. Speech report quotations do not generate an implicature related to content (un)endorsement because the speaker's use of the quotation is relevant for another reason in the contest. In this case, it is relevant that x said u because in the context it is relevant to know exactly what x said.

In the case of appealing to authority, the use of quotation is a means for the speaker to justify their use of the term or phrase (i.e. positive endorsement of the content). In other words, it is relevant that x said u because x is an expert and trustworthy regarding the topic of u.

In the case of epistemic distancing, the implicature generated has the opposite effect. Interestingly, there are many cases of epistemic distancing where there is no particular individual who the quotation can easily be attributed to. Rather, the interlocuter must infer the sort of person who would say the quotation based on what they know of the speaker and the context. These inferences often produce caricature individuals or stereotypes (e.g. Trump supporter, liberal, leftist, etc.). This caricature individual serves as a foil for the speaker, who uses the quoted content specifically to unendorse that content and distance themselves from it in a stylistic manner that might be characterized as sarcastic or derisive. The orthographic example of subclausal quotation in Figures 5.2 illustrates the sort of caricature foil I am describing.

In this example, the speaker (PragerU, a conservative think-tank and propaganda generator) references a generic leftist as the source for the quoted content to set up a foil for

Figure 5.2: Image of subclausal quotation from PragerU demonstrating epistemic distancing from the caricature of a leftist

their perspective (presumably that they don't think racism is inherently or always bad). The implicature generated, then, is that it is relevant that x said (or would say) u because the speaker wants to distance themselves from x and therefore unendorse u.

If the appeal to authority is positive endorsement and epistemic distancing is unendorsement, the speech report use case forms the middle ground, where the speaker is neither endorsing nor unendorsing the content. Rather, the speaker has some other motivation for employing subclausal quotation. For example, in the case of newscasters or legal witnesses, it might be very important to clearly communicate that they are precisely reporting an utterance made by someone else. This could be to avoid perjury or a defamation lawsuit. In this case, reporting that x said u might be directly relevant to the Question Under Discussion (Roberts, 1996). As such, the interlocuter is more likely to remain agnostic about whether the speaker endorses or unendorses u because the relevance of the quotation does not need to be inferred from the context as it does in the other two cases. Instead, the implicature generated in this case is that it is relevant that x said u because it is relevant to know exactly what x said.

5.4   What else can we learn from the Emphatic Juncture?

Interestingly, investigating how intonation fits as part of a complete theory of subclausal quotation also illuminates other phenomena which might be semantically con-

nected to quotation. In particular, if the Emphatic Juncture externalizes a metalinguistic type-shifter in the context of quotation (i.e. **quote-shift**), the presence of the Emphatic Juncture in other phenomena might be signaling the same kind of metalinguistic shift.

One such phenomenon worth exploring is the transparent free relative. Transparent free relatives (Wilder, 1998; Grosu, 2003) are a special type of free relative that embeds a syntactic constituent (usually a DP) within a free relative. As can be seen in 35, they consist of a non-sortal use of *what*, an optional source, an embedding verb that selects an equative structure or small clause, and the pivot (bolded in the following examples). The pivot is the content embedded within the TFR that is being attributed to the TFR source.

(35)   . Subject source TFR: The pitcher mastered what he calls **the slurve**.

No-source TFR: The pitcher mastered what is called **the slurve**.

Crucially, speakers often insert an emphatic juncture directly preceding the noun within the pivot (Sturman, 2019). An example of a TFR that includes an emphatic juncture can be seen in Figure 5.3.
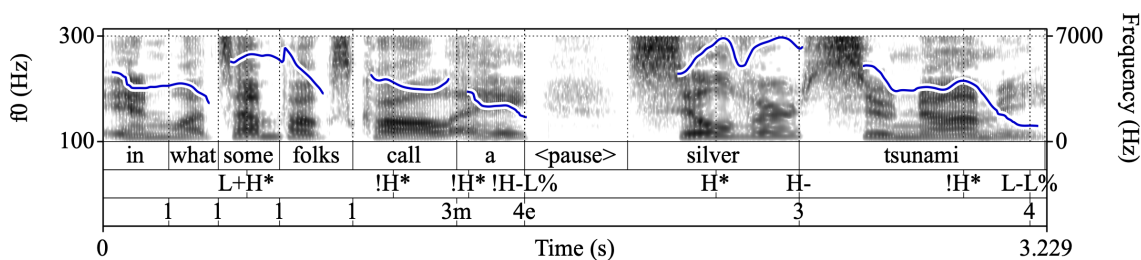
Figure 5.3: Pitch track (including an emphatic juncture marked with a box) of *in what some folks call a % silver tsunami.* This example shows the placement of an emphatic junctures in a transparent free relative (Data from National Public Radio)

Harris (2014) shows experimentally that the pivot of a subject source transparent free relative is very likely to be interpreted as relative to the source's beliefs (e.g. the pitcher in 35), whereas a no-source transparent free relative does not receive a shifted interpretation. This evidence, along with the presence of the Emphatic Juncture, suggests that an operator

like **quote-shift** is likely part of the meaning of transparent free relatives. In fact, perhaps the operator Potts calls **quote-shift** ought to be renamed to something less tied to quotation specifically since it is functioning more broadly as a meta-linguistic type-shifter (often) externalized by the Emphatic Juncture. I propose renaming the operator ***meta-shift***.

5.5 Conclusion

5.5.1 Summary

In this chapter, I briefly summarized the semantic literature surrounding subclausal quotation, including analyses by Cappelen and Lepore (1997) and Potts (2007). I pointed out how these theories are impoverished when considering natural language beyond orthography, and what we would expect to be marked related to subclausal quotation in the speech signal. I then showed how the intonational features found in the subclausal quotation corpus (Chapter 2, Section 4) correspond to the indicators of subclausal quotation we expected to be marked. Specifically, the Emphatic Juncture marks the presence of subclausal quotation and the start of the quoted content; the right IP boundary marks the end of the quoted content, and the pitch range reset on the quoted content increases the prominence of the quoted content to set it apart from the rest of the utterance.

I then outlined the descriptive categories of how subclausal quotation is used and discussed the pragmatic content associated with the various uses. In particular, the listener must determine why it is relevant that x said u, where x is the source of the quotation and u is the content of the quotation. Based on contextual factors, the relevance implicature may be that the speaker is endorsing or unendorsing the content, or there may be some contextual pressure to precisely report an utterance made by someone else.

Finally, I explored another construction that involves the use of the Emphatic Juncture, Transparent Free Relatives. The presence of the Emphatic Juncture in this phenomenon might be signaling the same kind of metalinguistic shift we see in subclausal quotation.

5.5.2 Future research

There is much to be done in the realm of future research. I foresee a cross-linguistic investigation of subclausal quotation to be of interest, particularly as it relates to the role of intonation in marking quotation. Psycholinguistic experimentation could also be helpful to further investigate the potential ambiguity among the various communicative uses of subclausal quotation as well as the presence of various implicatures.

Lastly, there is the fact that an operator like **quote-shift** has a clear means of intonational externalization when it was previously analyzed as silent. Semanticists and syntacticians would do well to look for intonational externalizations of other "silent" operators, particularly in phenomena that appear optional, like subclausal quotation.

CHAPTER 6

Conclusion

6.1   Big picture

The two primary goals of this dissertation were to understand how subclausal quotation is marked intonationally and to show how semantic analyses of subclausal quotation could be enriched through the incorporation of intonation. The work presented here used descriptive and experimental methods to conduct a thorough analysis of subclausal quotation intonation. I then showed how the key intonational features of subclausal quotation can supplement an existing semantic analysis of subclausal quotation (Potts, 2007). Along the way, the natural speech data collected for the subclausal quotation corpus surfaced interesting insights related to both the theory of intonation and the semantics of subclausal quotation.

In the subsequent sections, I will recapitulate the key findings in each of these areas. Following this, I will discuss how I see these findings contributing to the field more broadly. Finally, I will conclude by suggesting areas for future research.

6.2   The intonational marking of subclausal quotation

There were two sources of empirical evidence for identifying and analyzing the intonation associated with subclausal quotation. The first stage was the assembly and analysis of a subclausal quotation corpus. Following the analysis of the corpus, I conducted an experiment to determine which intonational feature(s) most readily conveyed the presence of subclausal quotation.

6.2.1 Corpus findings

The corpus consisted of 90 total tokens, sourced from radio, podcasts, television and YouTube (via the YouGlish transcript search engine). The intonation of the corpus tokens was transcribed in the MAE_ToBI transcription system. Analyzing these transcriptions surfaced three key intonational features of subclausal quotation: an initial emphatic juncture at the start of the quotation, a pitch range reset on the quoted material, and an IP boundary at the end of the quotation.

The presence of these features was remarkably consistent throughout the corpus. When no lexical marker (*quote* or *quote-unquote*) was present, each of the three features occurred in at least 85% of the corpus tokens. When the speaker used a lexical marker, there was no significant change in the occurrence of the pitch range reset or the right IP boundary. There was, however, a decrease in the usage of the emphatic juncture, from 90% down to 73%. I interpret this decrease to be related to the process of externalizing the metalinguistic type-shifting operator (what Potts calls **quote-shift**). When a lexical marker is used, it can serve the function of externalizing this operator, rendering the Emphatic Juncture optional.

6.2.2 Intonation perception experiment

Following the corpus analysis, I designed and ran an experiment isolating each of the three key intonational features. The goal was to learn more about how each feature contributes to communicating the presence of subclausal quotation.

6.2.2.1 Gesture methodology

The experiment employed a novel methodology which was an adaptation of a classic phonetic categorization task. Rather than listening to a single sound and deciding which phoneme category it belongs to (as in the vowel categorization task), participants heard a full sentence and were asked whether they thought the speaker used a gesture. After this question, participants were asked how confident they were in their gesture answer on

93

a 1-7 Likert scale. The "Gesture" and "No Gesture" categories were introduced to the participants via a brief training at the start of the experiment.

The experiment consisted of five conditions, including two control conditions. The control conditions used either all three of the intonational features associated with Subclausal Quotation Prosody or none of them (i.e. Neutral Prosody). The experimental conditions each isolated one of the intonational features and excluded the other two: the Emphatic Juncture condition, the Pitch Range Reset condition, and the Right IP boundary condition.

6.2.2.2 Results

The rate of "yes" gesture responses for each of the three experimental conditions fell in between the rates for the control conditions (all three features present versus none of the three features). Of the three features, the Emphatic Juncture condition had the highest rate of "yes" responses, slightly higher than the other two. This indicates that the Emphatic Juncture is the best of the three features at signaling the presence of subclausal quotation.

The experimental conditions all elicited more "yes" than "no" gesture responses, resulting in them being closer to subclausal quotation control than to the neutral control. In other words, each intonational feature served as a weaker signal than the full tune that subclausal quotation was present, but a signal to its presence nonetheless. This constitutes evidence that there is meaning present in intonational features below the tune level.

6.2.3 How each feature marks subclausal quotation

Based on the results of both the corpus and the experiment, the Emphatic Juncture serves the roles of signaling the presence of quotation and marking the beginning of the quoted material. In the framework of Potts (2007), the Emphatic Juncture is the intonational externalization of **quote-shift**. While the Emphatic Juncture is the intonational externalization, **quote-shift** can also be externalized using a lexical marker (*quote* or *quote-unquote*) or a finger quotation gesture.

94

The right IP boundary marks the end of the quoted string. In order for the meaning of a subclausal quotation to be calculated, the string must be clearly defined. As a large juncture, the IP boundary is often employed to signal syntactic or information structure boundaries. As such, it is a natural candidate to externalize the end of a quotation.

The pitch range reset does not have an obvious direct correlate in the semantic analysis the way the other two features do. Instead, it functions to increase the acoustic prominence of the quoted material. This helps the quoted material to stand out in relation to the surrounding content.

6.3   Contributions to intonation theory

Analyzing the intonation of natural, spontaneous speech often surfaces interesting data points that expand our understanding of intonational theory. The speech that made up the subclausal quotation corpus was no exception, illuminating two meaningful theoretical insights in addition to the descriptive value: the Emphatic Juncture and embedded IPs.

6.3.1   Emphatic juncture

First reported in Sturman (2019) as part of work that set the stage for this dissertation, the Emphatic Juncture is a specialized type of Intonation Phrase boundary. As IP boundaries go, the Emphatic Juncture stands out as a theoretical oddball. The boundary tone is a plateau, but it also includes an obligatory pause. In some cases, such as when it marks subclausal quotation, it is followed by a pitch range reset. But in other cases, such as in emphatic speech, downstep is licensed across the juncture. This is a significant departure from the usual convention for the domain of downstep: a High target must be downstepped in reference to a previous High target within the same intermediate phrase.

In addition to its role in subclausal quotation, the Emphatic Juncture has been observed in transparent free relatives, which also have a metalinguistic flavor. Finding the Emphatic Juncture other places ought to be a signal to consider a metalinguistic analysis for whatever

phenomenon it is associated with. Although I did not include an analysis of the intonation of other sorts of quotation, I would expect to find the Emphatic Juncture marking them as well.

6.3.2   Embedded IPs

The second major finding related to the theory of intonation is the discovery of evidence for embedded IPs in English. In particular, a subclausal quotation forms an IP which is nested inside a larger IP. The existence of embedded IPs is supported by three sources of evidence. First, splicing out the quotation leaves a continuous-sounding pitch track in many cases. The second and third forms of evidence come from predictions made by the possibility of embedding IPs. If the nested IP is embedded after the nuclear pitch accent of the larger IP, the content after the embedded IP will appear to be a headless IP since it has been stranded from its nuclear pitch accent. This prediction is born out in the corpus data. There is a second prediction that an embedded IP would be inserted in a location that splits a bitonal pitch accent. This was attested for L+H* in the corpus, though not for the other two bitonal pitch accents in English.

6.4   Semantic contributions

In addition to the descriptive work and the intonational theory contributions, there were some meaningful insights related to semantic theory.

6.4.1   Intonation externalizes quotation

As detailed in Section 6.2.3 above, each of the three intonational features plays a specific role in the externalization of subclausal quotation. I detailed how these features correspond to a particular semantic analysis of subclausal quotation, using Potts (2007). Although this is the analysis I used to demonstrate integrating intonation into a semantic analysis, that is by no means because it is the only analysis I see as viable for including

intonation. Rather, Potts's analysis serves as a convenient example for demonstrative purposes. Future work could include extending this intonation integration process into other analyses of subclausal quotation. Above and beyond that, though, would be to examine other phenomena where the Emphatic Juncture appears and work out a similar mapping of semantic operators to intonational features.

6.4.2 Communicative uses of subclausal quotation

The subclausal quotation corpus demonstrated that speakers employ subclausal quotation for a number of different communicative functions. I proposed three different communicative uses for subclausal quotation: appealing to authority, speech reporting, and epistemic distancing.

Perhaps surprisingly, however, there was no intonational or other observable surface characteristic to differentiate between the various uses. It was also possible to assign more than one communicative use to a particular utterance. As such, the situation seemed ripe for a pragmatic analysis. Following the Maxim of Relevance, the speaker employs a subclausal quotation because it is relevant that x said u, where x is the source of the quotation and u is the quoted content.

The particular reason for the relevance differs for each of the three uses with respect to content endorsement. In the case of appealing to authority, it is relevant that x said u because x is an authority concerning u. Thus, the speaker uses quotation to justify their use of the phrase and positively endorse the content through an appeal to x's authority.

In the case of epistemic distancing, it is relevant that x said u because the speaker wants to contrast themselves with x in a satirical or derisive way. Since x is the source of u and they want to distance themselves from x, the speaker uses subclausal quotation to unendorse u in a stylistically interesting way.

The case of speech reporting does not involve direct endorsement or unendorsement on the part of the speaker. Instead, there is some other reason why it is relevant that x said u. For example, in the case of newscasters or legal witnesses, it might be very important to

97

clearly communicate that they are precisely reporting an utterance made by someone else. Thus, the implicature generated in this case is that it is relevant that x said u because it is relevant to know exactly what x said.

6.5   General discussion

As can be expected in any project that includes a descriptive component, the exploration of subclausal quotation intonation surfaced several unexpected insights related to intonational theory. In order to enable a richer analysis, both for the phenomenon of interest and more broadly, finding the phenomenon in natural speech is an important part of the process. From a qualitative perspective, speech produced in a lab for an experiment is simply not as vibrant and natural as speech found "in the wild," particularly when it comes to intonation. This is why I found it worthwhile to invest the time and effort needed to assemble a corpus of subclasual quotation rather than relying on lab-based productions.

6.6   Future research

Finally, I will conclude by pointing out some areas that I think would prove fruitful for future research. First and foremost, there are plenty of other semantic phenomena that the methodology I detailed above would be useful for investigating. Transparent free relatives seem like low-hanging fruit given that the Emphatic Juncture is often part of the externalization.

There is also a host of interesting follow-up work related to the contributions to intonational theory described here. Embedded IPs in particular open up an entire realm of research questions. How does the presence of an embedded IP affect speech planning or speech processing? How many levels of embedding are permissible, and relatedly, how marked are embedded IPs? Are there any examples of the other two bitonal pitch accents being split by IP embedding? There are also some convention-level questions related to labeling embedded IPs. Do new tone or break labels need to be added to the working

inventory of MAE_ToBI?

Speakers encode so much rich meaning through intonation, and linguists are only just beginning to unlock how intonation contributes at both the semantic and pragmatic levels. I hope that the work presented in this dissertation serves as a starting point for the strengthening of the intonation-semantics interface and the start of a broad research program to discover more of the ways that intonation delivers and enriches meaning.

APPENDIX A

Appendix: Stimuli

Critical items

1. The actress wore a "light chartreuse" dress to the fundraiser gala.

2. The chessboard was made of "dark zebrawood" and walnut.

3. The new lipstick color resembled a "dusky mauve" in natural light.

4. Miranda bought "eggshell brown" paint at the hardware store.

5. The diver took photos of "aquamarine coral" during the expedition.

6. The couple danced a "salon tango" to end the performance.

7. The quilt featured an "antique calico" pattern on the front.

8. Akeem wore a "maize yellow" jacket to the track meet.

9. The accountant used an "oxblood red" folder for invoices.

10. The potter listed a "terracotta vessel" for sale.

11. The decorator gave the "plastic veneer" a coat of paint.

12. The conductor ended the concert with a "tremolo" string arrangement.

13. The pet shop bred "emerald catfish" in the back.

14. The ship caught a "sawtooth eel" by mistake.

15. Allen likes listening to "anti-folk" music when driving.

16. The baker made a "chocolate chiffon" cake for her birthday.

17. Leroy wants "Tonkotsu ramen" for dinner.

18. The Philosophy professor's "transcendental idealism" lecture was very dense.

19. The author is known for "magic realism" novels that capture the imagination.

20. Maria enjoys playing "rogue-like" video games.

21. Damien writes in "Spencerian calligraphy" with fountain pens.

22. The antique store sells "ironstone china" at a great price.

23. The artist debuted an "anti-modernist" mosaic at the gallery show.

24. Rebecca wore a "faux alexandrite" ring with pride.

25. Fiona rented a "bohemien chic" cabin for New Year's Eve.

26. George packed a "Belgian picnic" for his date with Antuan.

27. Daniel proudly displayed their "owl collectables" during the dinner party.

28. The server placed a "charcuterie tray" in the center of the table.

29. Amelia requested a "submarine cake" for dessert.

30. The actor slipped in and out of a "valley girl" accent during the play.

Filler items

1. That's Lauren's favorite restaurant.

2. William told his partner to meet him there.

3. Erin put that book in her backpack.

4. The mayor was there last week.

5. Owen took a class with that professor.

6. Anna heard the concert would be there.

7. That's the intersection where the accident happened.

8. The artist will display a painting there next month.

9. Naomi loves the smoothies at that cafe.

10. Walter wanted to eat there instead.

11. Darin loves to dance at that studio.

12. Keisha went to school there for a few years.

13. Edwin bought that coat on clearance.

14. The children hoped the kite would land there.

15. Maria found that rock on the beach.

16. The sailboat capsized there during the storm.

17. Nina watched that movie all the time when she was young.

18. Evan put his passport there for safekeeping.

19. The cook used that pot to make her favorite stew.

20. The hunter hid there hoping to spot a deer.

21. The librarian made that display for Black history month.

22. Hannah lived there when she was a girl.

23. Keith went to that club right when it opened.

24. Andy studied there during finals week.

25. Lily climbed that high on her last hike.

26. The admiral saw a boat this big sink at the start of her career.

27. The mover fit a box that wide through the doorway.

28. The dog jumped over a fence this short a week ago.

29. The player who scored the most points last season was that tall.

30. A bowl this wide cracked in the kiln last time.

31. Daniel ate a cake that big on his birthday.

32. Rhonda had a client this tall last month.

33. Alex found a lizard that small in Mexico.

34. Pierre fit through a doorway this narrow a few days ago.

35. Dorian saw a hedge that tall on their walk.

36. Carmen finished a hot dog that long last summer.

37. The neighbors' dog buried a bone this big in the backyard.

38. Elvira found a bone this small in her salmon fillet.

39. The fisherman caught a crab that big in the bay.

40. The submarine had a passageway that narrow before the renovations.

41. Warren found a Christmas tree that short in the local lot.

42. Yolanda wore a hat this wide to church last Sunday.

43. Becky dug up a clam that small on the beach.

44. Larry hung a sign that wide off the overpass.

45. The nurse saw a patient this short at the vaccine fair.

46. Lavena drank a coffee this big after their all-nighter.

47. Ivan rode a bike that short for the circus act.

48. Alma slept on a bed this narrow in her first apartment.

Bibliography

Ayers, G. M. (1996). *Nuclear accent types and prominence: Some psycholinguistic experiments*. Thesis, Ohio State University.

Beckman, M. and Ayers-Elam, G. (1997). Guidelines for ToBI labelling (version 3.0).

Beckman, M., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Jun, S., editor, *Prosodic Typology: The Phonology of Intonation and Phrasing, Volume 1*, pages 9–54. Oxford University Press.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in japanese and english. *Phonology Yearbook*, 3:225–310.

Bishop, J. (2012). Focus, prosody, and individual differences in "autistic" traits: Evidence from cross-modal semantic priming. *UCLA Working Papers in Phonetics*, 111(1):26.

Boersma, P. and Weenink, D. (2022). Praat: doing phonetics by computer [computer program] (version 6.2.14).

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, pages 633–644.

Bruce, G. (1977). Swedish accents in sentence perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, 12.

Calhoun, S. (2007). *Information structure and the prosodic structure of English: A probabilistic relationship*. PhD thesis, University of Edinburgh.

Cappelen, H. and Lepore, E. (1997). Varieties of quotation. *Mind*, 106(423):429–450.

Cappelen, H., Lepore, E., and McKeever, M. (2020). Quotation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2020 edition.

Clark, H. H. and Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, 66:764–805.

Crystal, D. (1969). Prosodic systems and intonation in english. cambridge studies in linguistics 1.

Davidson, D. (1968). On saying that. *Synthese*, pages 130–146.

Davidson, D. (1979). Quotation. *Theory and decision*, 11(1):27–40.

De Brabanter, P. (2010). The semantics and pragmatics of hybrid quotations. *Language and Linguistics Compass*, 4(2):107–120.

Diachek, E. and Brown-Schmidt, S. (2022). The effect of disfluency on memory for what was said. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Fodor, J. D. (1998). Learning to parse? *Journal of psycholinguistic research*, 27:285–319.

García-Carpintero, M. (2003). Double-duty quotation: the deferred ostension account. *Belgian Journal of Linguistics*, 17(1):89–108.

Geurts, B. and Maier, E. (2005). Quotation in context. *Belgian Journal of Linguistics*, 17:109–128.

Gómez-Torrente, M. (2003). Remarks on impure quotation. *Belgian Journal of Linguistics*, 17(1):129–151.

Goodhue, D., Harrison, L., Su, Y. C., and Wagner, M. (2016). Toward a bestiary of english intonational contours. *The Proceedings of the North East Linguistics Society (NELS)*, 46:311–320.

Grice, H. P. (1978). Further notes on logic and conversation. In *Pragmatics*, pages 113–127. Brill.

Grosu, A. (2003). 'transparent' free relatives as a special instance of 'standard' free relatives. *From NP to DP*, 1:139–178.

Harris, J. (2014). Signaling non-speaker commitment in transparent free relatives: A paired speaker-hearer judgment study. In *The Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue*, pages 71–79. Citeseer.

Holliday, N. (2021). Prosody and sociolinguistic variation in American Englishes. *Annual Review of Linguistics*, 7:55–68.

Holliday, N. R. (2016). *Intonational Variation, Linguistic Style and the Black/Biracial Experience*. PhD thesis.

Jun, S. (1998). The accentual phrase in the korean prosodic hierarchy. *Phonology*, 15(2):189–226.

Jun, S. (2003). Prosodic phrasing and attachment preferences. *Journal of psycholinguistic research*, 32:219–249.

Jun, S. and Bishop, J. (2015). Priming implicit prosody: prosodic boundaries and individual differences. *Language and speech*, 58(4):459–473.

Jun, S.-A. (2019). The goal of focus prosody is to violate rules of language-specific default prosody. Plenary speaker at the Intonation Workshop, University of Toronto.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, 3(3):129–140.

Krifka, M. (1999). At least some determiners aren't determiners. *The semantics/pragmatics interface from different points of view*, 1:257–291.

Kutik, E. J., Cooper, W. E., and Boyce, S. (1983). Declination of fundamental frequency in speakers' production of parenthetical and main clauses. *The Journal of the Acoustical Society of America*, 73(5):1731–1738.

Ladd, D. R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology*, 3:311–340.

Ladd, D. R. (1988). Declination "reset"and the hierarchical organization of utterances. *The Journal of the Acoustical society of America*, 84(2):530–544.

Ladd, D. R. (1990). Metrical representation of pitch register. *Papers in laboratory phonology I: Between the grammar and physics of speech*, pages 35–57.

Liberman, M. and Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range. In Aronoff, M. and Oerhle, R., editors, *Language sound structure*, pages 157–233. Cambridge, MA: MIT Press.

Liberman, M. Y. (1975). *The intonational system of English*. PhD thesis, Massachusetts Institute of Technology.

Lieberman, P. (1967). Intonation, perception, and language. *MIT Research Monograph*.

Maier, E. (2014). Mixed quotation: The grammar of apparently transparent opacity. *Semantics and Pragmatics*, 7:7–1.

Nespor, M. and Vogel, I. (1986). Prosodic phonology. *Dordrecht: Foris*.

O'Connor, J. D. and Arnold, G. (1961). *Intonation of colloquial English*. Longman.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in english. *The journal of the Acoustical Society of America*, 54(5):1235–1247.

Partee, B. (1973). The syntax and semantics of quotation.

Peperkamp, S. A. (1997). *Prosodic words*, volume 34. Holland Academic Graphics The Hague.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. PhD Thesis, MIT.

Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In Cohen, P., Morgan, and Pollack, editors, *Intentions in Communication*, pages 271–311. MIT Press.

Pike, K. L. (1945). The intonation of american english.

Potts, C. (2007). The dimensions of quotation. *Direct compositionality*, pages 405–431.

Quine, W. V. O. (1980). *From a logical point of view: 9 logico-philosophical essays*, volume 9. Harvard University Press.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Recanati, F. (2001). Open quotation. *Mind*, 110(439):637–687.

Reimer, M. (1996). Quotation marks: demonstratives or demonstrations? *Analysis*, 56(3):131–141.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In *Speech and language*, volume 10, pages 243–335. Elsevier.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics.

Ross, J. R. (1970). On declarative sentences. *Readings in English transformational grammar*, 222:272.

Roustan, B. and Dohen, M. (2010). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. In *Speech Prosody 2010-5th International Conference on Speech Prosody*, pages 100110–1.

Royer, A. J. and Jun, S.-A. (2019). Prominence marking in kazan tatar declaratives. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, pages 2995–2599.

Selkirk, E. (1984). *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, 3:371–405.

Selkirk, E. (1996). The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187:214.

Selkirk, E. (2002). Contrastive focus vs. presentational focus: Prosodic evidence from right node raising in english. In *Speech prosody 2002, international conference*.

Shattuck-Hufnagel, S. and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, 25(2):193–247.

Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., Pierrehumbert, J. B., and Hirschberg, J. (1992). Tobi: A standard for labeling english prosody. In *ICSLP*, volume 2, pages 867–870.

Smith, C. S. (2003). *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press.

Sturman, B. (2019). The emphatic juncture: A novel use of the IP boundary in English. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors, *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 2489–2493, Melbourne, Australia. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Tarski, A. (1956). The concept of truth in formalized languages. *Logic, semantics, metamathematics*, 2(152-278):7.

Wells, R. S. (1945). The pitch phonemes of english. *Language*, pages 27–39.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3):1707–1717.

Wilder, C. (1998). Transparent free relatives. *ZAS papers in linguistics*, 10:191–199.

Wilson, D. and Sperber, D. (1986). On defining relevance. *Philosophical grounds of rationality: Intentions, categories, ends*, pages 243–258.