RICE UNIVERSITY

Transitivity and Intonation:
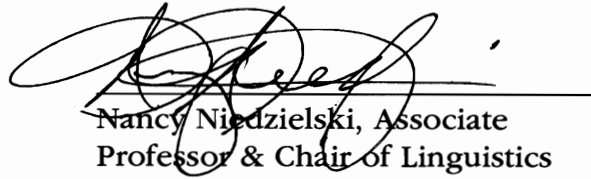A Preliminary Account of Transitive Lowering

By

David Andrew Katten

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Katherine Crosswhite, Assistant
Professor of Linguistics

Nancy Niedzielski, Associate
Professor & Chair of Linguistics

Michael Byrne, Associate Professor of
Psychology

Houston, Texas
May 2011

ABSTRACT

Transitivity and Intonation:

A Preliminary Account of Transitive Lowering

By

David Katten

Are subjects produced differently based on the transitivity of the clause they are embedded in? Based on data from a narrative reading experiment, it is shown that transitive subjects are produced with a lower f0 than intransitive subjects and that this difference is statistically significant ($p<0.05$). It is suggested that the purpose for such a difference originates from a propensity for English speakers to accent new referents, which are common in the object position. By lowering the f0 of the subject, speakers increase the efficacy of an accent on a new object later in the clause. Finally, the read narrative procedure is evaluated for its strict control of stimuli, while also reproducing known intonational phenomena.

ACKNOWLEDGMENTS

advisor Katherine Crosswhite. Without their encouragement in times of doubt (and there were many, many times of doubt), I would not have finished. Their patience and guidance was more appreciated than they will ever know. Thanks Katherine, you're the best.

TABLE OF CONTENTS

Chapter 1

1. Introduction

This dissertation explores the interplay between transitivity and speaker production of pitch, with findings that support the idea that linguistic patterns may best be explained by their function, rather than abstract formal parameters (Chomsky & Halle 1968). In the case presented here, a functional link between pitch realization and transitivity is proposed, defended, and experimentally validated.

The research question addressed in this dissertation is quite brief: is there a relationship between intonation and transitivity? More specifically, do speakers produce different intonation contours for transitive and intransitive

sentences? English is a language that employs suprasegmental structure to index information structure by placing pitch accents on referents new to a discourse, and avoiding pitch accents on referents that have been previously mentioned, or are inferable from context (that is, given). English also displays reliable distributions of given and new referents with respect to core syntactic roles. This provides a space for speakers to make their introductions of new information more prominent through manipulation of pitch contours. As a brief example, this dissertation explores if there is difference in the production of "Mary" in the following pair:

1) With her friends in attendance, **Mary** walked down the aisle.

2) After threatening to do so, **Mary** smacked the bar patron.

At first blush, it is difficult to imagine why there would be any **reliable** differentiation between such grammatical subjects with respect to intonation and pitch accenting. However, I will show that the subject "Mary" in sentence 2 (the transitive version) is consistently produced with a lower fundamental frequency than the "Mary" of sentence 1 (the intransitive version), and further that this qualifies as a unique case of grammaticalization, and could be analyzed as a type of tonogenesis.

The distinction of transitive vs. intransitive clauses is not overtly marked in English, but it is marked in some ergative languages. In an ergative language,

such as Basque or Dyirbal, subjects of transitive and intransitive clauses bear differing morphological or syntactic characteristics. This dissertation contributes to the understanding of ergative patterning by expanding the field of focus beyond morphosyntactic structures to include phonological categorization. Because I motivate the observed grammaticalized response in terms of usage, and efficacy, this dissertation takes a functional stance toward language production and structure; the observed shape of English is in part a product of communicative constraints that constitute the function of language.

To answer the research question, I will present the results of a language production experiment designed to elicit comparable transitive and intransitive sentences. I have framed the background, experiment, and results in terms of 4 propositions, further fleshed out in Chapter 2. These propositions are not hypotheses; they are postulates based on existing research, upon which a testable hypothesis is constructed. The propositions are:

1. As speakers talk, their baseline pitch decreases

2. Given two equal pitch accents (i.e. of the same absolute Hz), listeners perceive the accent that is later as more prominent

3. In English, new referents are accented

4. In English, new referents occur toward the end of the clause

---

Given the importance of information distribution and information structure, this experiment also sheds light on the production of transitive clauses and the encoding of information structure. The experiment shows that speakers do in fact articulate subjects of transitive and intransitive clauses differently, and in a way that is consistent with the existing literature—that transitive subjects have lower pitch than intransitive subjects.

1.1. A justification for cross-disciplinary exploration

Linguistics, like many social sciences, is to some extent compartmentalized. That is, it defines certain boundaries that delimit what the appropriate targets of study should be. Phonetics deals with sounds, phonology sound categories, syntax the structure of clauses. Over time, these boundaries

tend to break down as scholars find fruitful links across them: syntax and semantics, phonetics and phonology, discourse and grammar.

While these mergers tend to point out the artificiality of division, they are not arbitrary. Phonetics and phonology have a high degree of commonality, and trying to divorce syntax from semantics has proved basically impossible; virtually every syntactic theory is forced to recognize the value of the lexicon at some level, even if it is only at that of "interface." What is more rare is a joining of highly disparate linguistic fields—the not-as-obvious linkages that correlate structures that, at first blush, are totally unrelated.

This dissertation originates with the hope that it too may span disciplines of linguistics to reveal an unexpected truth. Phonetics and syntax are at opposite ends of the concreteness spectrum—phonetics being real and measurable signals and articulations, syntax internalized and abstract mental representation. That is not to say that there are no mentalist aspects to speech sounds, or no physical and quantifiable products of syntax, but that they are prototypically differentiated; phonetics is studied quantitatively, syntax qualitatively.

Linguists, such as Bollinger and Pierrehumbert to name a few, generally acknowledge that intonation, syntax, and information structure are mutually

influential in English. The rising tone of questions is an example where the phonetics of prosody plays a major role in constructing a syntactic parse. As early as Bolinger (1958), linguists have noticed the proclivity of English speakers to accent "new" information and deaccent "old" information, a pattern also observed in Dutch (Noteboom & Terken 1982). While the phenomenon, known in the literature as deaccenting, is well-attested vis-à-vis givenness, there is much to learn about how accentuation acts on referring expressions that occur in different syntactic contexts. Although researchers construct different models of information structure (see next chapter), I will employ the term givenness to denote how accessible a referent is in a discourse, where accessibility is measured by whether the referent has been previously mentioned or not.

In this chapter, I will discuss the assumptions behind this primary question, as well as the possible implications of both affirmative and negative answers. This is a chapter dedicated to the "big picture"—the whys but not yet the hows.

2. Transitivity and Grammar

From one point of view, transitivity is one of if not *the* central question of syntax. It is at the heart of two central areas of syntactic study: argumenthood

and the clause. Yet to say that this is what transitivity is "about" would be misleading; it is a far vaster object of study, tied to myriad aspects of syntax. Transitivity intersects a wide range of linguistic features or processes, such as animacy, relativization, agency, and affectedness. Transitivity in the context of this dissertation refers to the assignment of thematic roles to participants, as mediated by grammar. That is, transitivity provides a systematic linking for actors with particular relations to the events/situation described in a clause to the grammatical roles (such as subject or object) available in the language.

For the purposes of this dissertation, I adopt a highly simplistic operationalization of transitivity that is based solely on the number of arguments in the clause. Thus, within the context of my experiment, "intransitive" means the presence of one argument (a subject), and "transitive" the presence of two overt arguments (subject and object). As established by Hopper and Thompson (1980), this binary, strictly structural view misses key characteristics of transitivity, both crosslinguistically and within English. Aspects such as animacy, affectedness, and intention of the clausal participants are critical to understanding of clausal relations. Due to such considerations, it would be inappropriate to use such a simplistic, binary distinction as the formal definition of transitivity. However, the experiment reported on in this

dissertation carefully controls those factors, using sentences that are matched for animacy, intention, etc., and differing only in the number of arguments. In this respect, use of a binary distinction is justified.

A key aspect of English grammar for this dissertation is that of word order. English has fairly predictable word order in the unmarked case, SVO, and is right-branching. This is relevant because subjects are canonically and predictably 1) earlier in the clause than objects and 2) followed by a verb, which is in turn, optionally followed by an object. When word order in a clause is aligned with an intonation contour, this means that subjects are generally toward the beginning of the intonational phrase, and objects are toward the end. This alignment is dependent on clausal complexity, as well as other characteristics of spontaneous language such as false starts, but it is a fair initial observation.

For discussing word order and transitivity, this dissertation adopts the tripartite labeling distinction of S, O, and A (Dixon 1979). Ostensibly, these letters are abbreviations for Subject, Object, and Agent (or Actor). However, in practice, they do not denote grammatical roles or relations, but are merely symbols for participants in transitive vs. intransitive clauses; A for the nominal subject of a transitive clause, S for a nominal subject of an intransitive clause, and O for the nominal object of a transitive clause.

A discussion of the tripartite distinction also offers an opportunity to discuss ergativity. English exhibits nominative-accusative patterning, in which subjects of transitive clauses have similar morphological and syntactic properties with subjects of intransitive clauses.

3. He led her through the marsh as the alligators slept.

4. He spoke eloquently, albeit slowly.

5. She spoke eloquently, albeit slowly.

One way in which English expresses a nominative-accusative system is through pronominal form. In sentences 3 and 4, the subject of a transitive clause (3) is of the same form as the subject of an intransitive clause (4)—they are both realized as "he." However, the object of the transitive clause (her, in 3) is different than the subject of the intransitive clause (she, in 5).

This contrasts with other languages, such as Bardi (Bowern et al, 2008), which uses an ergative morphological marker (-nim) to convey which participant is the subject of a transitive clause in this free word order language.

Bardi expresses ergative patterning because the subject of an intransitive clause has the same morphological marking (namely, Ø, or no additional morphology) as the *object* of a transitive clause. It is the subject of the transitive clause that has different marking (through the suffix –nim). In contrast with English's nominative-accusative patterning, Bardi displays an ergative-absolutive system.

6.   Olooma-nim   barnamb   inoongoorribigal
        old woman-nim   stingray   3SG.chased.3SG
        "The old woman chased after the stingray."

7.   Oolooma   Barnamb-nim   inoongoorribigal
        old woman   stingray   3SG.chased.3SG
        "The stingray chased after the old woman."

8.   Aamba   inggamagal
        man   3SG.laughed
        "The man laughed."

The ergative marker *–nim* is applied only to the subjects of transitive clauses (6, 7), but is absent on the subject of the intransitive clause (8). Such is the ergative-absolutive alignment.

Using the tripartite distinction of A, S, and O, this can be visualized.

Nominative-Accusative
Like English

Ergative-Absolutive
Like Bardi
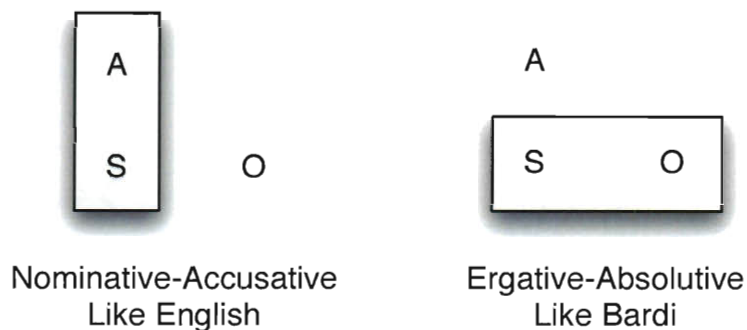
Figure 1. Comparison of nominative and ergative patterning.

Regardless of one's preference for grammatical theory, there is no question that transitivity is a critical dimension along which speakers shape their talk. The number and character of clausal participants is relevant for all sentences and grammatical theories. But simply by labeling the difference between a subject in a transitive vs. intransitive context, it is evident that not all subjects are exactly the same. The power of the tripartite distinction is that it can supply motivation for a number of hypotheses about how suprasegmental phonetics might interact with grammar, which will be taken up more fully in Chapter 2. A simple Subject/Object distinction does not readily do so.

The tripartite distinction has proven fruitful in describing a number of grammatical phenomena, among them Preferred Argument Structure, or PAS

(Du Bois 1987). PAS is a set of observations about naturally occurring discourse that can be distilled into 4 soft constraints that speakers observe in creating their talk. The constraints express speaker preferences with respect to argument realization—the constraints reveal preferred sites for new information, and for lexical arguments.

- Avoid lexical A

- Avoid more than one lexical argument

- Avoid new A

- Avoid more than one new argument (all from Du Bois 1987)

These constraints are based on pear story narratives of Sacupultec Mayan and speculated by the author to be universals. They are organized around two dimensions: quantity and quality. For quantity, the constraints posit that speakers avoid overloading the clause with "costly" constituents, where the introduction of new referents to a discourse, or producing a full lexical noun are considered more effortful than tracking an existing referent, such as pronominal or zero forms. The other, quality, restricts the type of referent that can be located in the A position—speakers avoid placing discourse-new or lexical nouns in the subject role of transitive clauses.

To reformulate these constraints in terms of S and O: speakers prefer S and O as loci for lexical arguments and new information; S and O share an availability for certain syntactic and information structure features.

The key insight for PAS study, however, is that naturally occurring discourse exhibits a patterning similar to ergativity—A's are different than S and O. Du Bois argues that this discourse pattern can ossify into a grammar; that over time, patterns in language may be overtly marked through syntax or morphology. This is a strongly functional perspective, and one that I will adopt for this dissertation. From a functional view, the structure of language is the result of competing communicative, cognitive, and social constraints. In the case of PAS, Du Bois suggests that the additional processing power required for lexical arguments or for new referents is the source for the soft constraints, which conspire to align A, S, and O in an ergative pattern that isolates A as a non-locus for new information or lexical noun phrases.

3. Deaccenting and unnaccentuation

The term "pitch accent" in this dissertation is used to describe the phenomenon of localized f0 excursions, i.e. pitch peaks.[1] Phonetic literature discusses a process by which a constituent that is expected to receive a pitch accent fails to obtain one, resulting in a lower fundamental frequency (f0) and perhaps a flat or linear pitch track rather than a peak. However, to be explicit and complete, this type of account must explain what would make such an accent "expected," and often does so by appealing to the information status of the referent. But the process of "deaccenting," a term favored in the literature, implies a cognitive process[2] where a so-called underlying accent is removed. I can find no evidence of such a process, only robust observations that referents in these conditions are unaccented. I will call this phenomenon unaccentuation; the term is non-committal regarding why certain tones appear or fail to appear in certain conditions. The referent is simply unaccented.

The interplay between focus and prosody can be approached from one of two directions. The more formal way originates with Chomsky and Halle (1968)

---

[1] Pitch may also be associated with duration or intensity. However, I will only concentrate on fundamental frequency as the prime phonetic indicator of pitch accent.
[2] The process must be cognitive (as opposed to articulatory) because the expectedness is a function of information status, which is strictly the domain of cognition.

who put forth a generative process by which sentential components receive stress through a Nuclear Stress Rule (NSR). The NSR posits an accent for constituents with certain discourse properties—i.e., where information is new or not expected from the context—but does so within the confines of generative grammar. NSR-based approaches take "newness" to be a feature of the syntactic parse rather than the discourse context that prompts speakers to introduce information. This view of accent is unabashedly formalist; the placement of accent is reliable and predictable, and violations of NSR-related rules are, in a sense, ungrammatical.

In practice, accent is not confined to nominal constituents. Allerton and Cruttenden (1979) review a debate from *Language* (Bresnan 1971, Bresnan 1972, Bolinger 1972, Berman and Szamosi 1972, Lakoff 1972) that examined stress from the NSR perspective. In noting the examples provided over this 4-5 year period, Allerton and Cruttenden propose a model that prefers predicative stress, where verbs are the most common site of accenting. They list three exceptions to this that deal exclusively with definite (and therefore likely given) subjects: when there is no verb to accept the accent, when the verb is one of (dis)appearance, or when the verb denotes a misfortune. While the NSR was the dominant prism through which researchers viewed intonation, linguists

situated stress on purely syntactic and semantic grounds, and overlooked the role of information structure and discourse status of referents.

An alternative is to examine the "exceptions" and outliers (such as verbs of disappearance) as indicative of a shortcoming in the theory. Perhaps the explanations are flawed—or at least fail to adequately predict sentential stress—because they set aside the purpose of stress within a linguistic system. If a theory permits speakers to produce pitch accents for a communicative function, such as identifying new referents, or making them more carefully pronounced, then such a theory would hold greater explanatory power for particular examples than one that posits a strict grammar of sentential stress.

3.1. The transcription of prosody

A central issue for intonation research is the written representation of prosody and prosodic features. Frequently, and especially in the pragmatics literature, pitch accent is simply transcribed by words in all caps or small caps (e.g., THAT'S what I said vs. That's what I SAID). Trager and Smith (1951) proposed a far more complex system that has four levels of intrasyllabic stress or strength (which correspond roughly to loudness), four intrasyllabic static levels of pitch, four types of variation for the static pitch level, and three types of "dynamic tone" that bridge non-adjacent syllables. The model makes several assumptions, including: (1) stress is directly correlated to loudness and exists at

discrete levels, and (2) pitch patterns are independent of stress and describable through four tone levels plus three dynamic tones (Terken & Hermes 2000:91-92). Moreover, this system is incredibly complex while not yielding much in the way of insight or discovery. How do these levels correlate with meaning, and how does an independent researcher replicate results when the assignment of the values is so difficult?

A later convention for representing English prosody is the Tone and Break Indices system (ToBI) developed over the past 30 years as an extension of autosegmental metrical theory (Pierrehumbert 1980, Pierrehumbert & Hirschberg 1990, inter alia). In discussing my work and the literature on unaccentuation throughout this dissertation, I will employ the ToBI system for pitch accents. ToBI diverges from previous accounts of intonation in that stress and accent are only perceptible in the context of a melodic tune rather than absolute levels of stress or pitch. Thus, there is no appeal made to several levels of tone, only to targets at the extrema of the pitch range: high and low. Tone-marked constituents influence pitch trajectory in a local domain, leaving the speaker wide latitude in producing utterances with the same ToBI structure.

The ToBI system provides a framework for the transcription and analysis of prosodic features on a language-by-language basis, the most well-developed description being for English. ToBI divides the prosodic transcription process into three major tiers: orthographic, tones, and breaks. Tones are associated with three analytic levels: words, intermediate phrases (also known as phonological phrases), and intonational phrases. Boundaries between each phrasing level are determined by the degree of disjuncture between words, with the least amount of disjuncture (e.g., clitics) being labeled 0-2, phonological phrases marked with a 3, and intonational phrases a 4. The relationship between these levels is hierarchical; intonational phrases are the maximal unit of analysis and are composed of a discrete number of intermediate phrases, which in turn consist of a discrete number of words. Every intonational boundary is also an intermediate boundary, which is important in understanding the notation system described below.

The tone tier in English is a sequence of High, Low, and composite (**H+L, L+H**) tones. These are not defined by invariant acoustic features, but are auditory impressions of f0 in relation to the rest of the intermediate or intonational phrase. There is a tone that spans the entire intonational phrase (H% or L%), marked at its right edge, and a tone for each constituent intermediate phrase (**H-** or **L-**), also marked at its own right edge. A pitch

accented word can receive one of six possible accents, with the relevant syllable marked with a "*" e.g., **H*** or **L+H***.

More recent studies of prosody (Pierrehumber 1980, Watson 2002) tend to concern themselves with aspects of intonation other than contrastive stress (Bolinger 1958) of constituents, instead focusing on the usage of prosodic phrasing as the mechanism for disambiguation. A notable exception is Schafer et al. (2000), who examined how the accent of a relative pronoun can influence how listeners parse the sentence. They found that in alternations of accent such as "I asked the girl who/WHO is cold," listeners interpreted the accented version as being a request for new information from the girl, whereas unaccented pronouns triggered a relative clause parse. The results refer back to the notion of givenness, and provide evidence that accented words are new. In this case, the question word interpretation is a request for new information, and thus is associated with the accented form.

3.2. Perception of accent

The perception of pitch accent is an important component in the investigation of transitivity and intonation based on the theory presented here. In essence, I assume a theory of phonology where the phonetic continuum of accenting may be used to index a grammatical relation. Without this

assumption, it would be impossible to map unaccentuation to a transitive or intransitive. However, this model should be consonant with a broader notion of speech perception, where hearers make use of related cognitive resources to be considered good listeners (Hawkins 2003). This section is a brief introduction to perception of prominence,[5] which is discussed more fully in the next chapter.

Pitch accent is associated with a localized peak in f0 with respect to the surrounding syllables or words (Pierrehumbert 1980). The qualifier "localized" is necessary because pitch accents may be associated with High or Low tones, thereby making accent more than a simple f0 spike or plateau. There are other phonetic correlates that are perceived as prominent, such as increased duration (Turk & Sawusch 1997) and amplitude (Terken & Hermes 2000), but the clear indicator of pitch accent is, unsurprisingly, pitch.

The perception of prominence is directly related to the abstract notion of baseline or reference; prominent syllables are those that have f0 maxima that are significantly higher than this baseline. Pitch perception research has sought to identify how the baseline is represented for listeners' models of pitch in speech. Gussenhoven et al (1997) discussed the role of f0 minima in gauging prominence, noting that:

"There is convincing evidence that f0 minima contain information that is relevant to making judgments about the prominence associated with f0 maxima, but it remains unclear to what extent variations in the frequency of f0 minima affect prominence judgments, and also whether all f0 minima contribute equally." (Gussenhoven et al 1997:3010)

f0 minima will play a critical role for this dissertation, as they are indicators of the baseline. Pitch baseline partly determines the perception of prominence, and measuring deviations from it will be the basis for much of the analysis presented in Chapter 3.

Baseline declination, which is a component of f0 downtrend over the course of an utterance, is a complicating factor in determining prominence. Indeed, it appears that listeners anticipate declination and build it into their model of prominence, such that a peak with a fixed frequency is perceived as being more prominent if it occurs later in the utterance than earlier (Gussenhoven et al 1997). The overall decline in f0 and how it pertains to prominence perception will be addressed in depth in Chapter 2.

Based on the binary tone distinction and on/off nature of pitch accent in the ToBI system, it might be assumed that intonation is a categorical phenomenon and that listeners perceive speech as either [+accented] or [-accented]. Researchers know intuitively that intonation is highly variable, but it

is an open question whether pitch accent is categorically perceived (accented/unaccented). Ladd and Morton (1997) addressed this question by performing a number of classical categorical perception (CP) tests, and found the characteristic S-shaped curve associated with CP when using a continuum of pitch peaks on a particular word. However, they were unable to find any sense of confusion at the putative category boundary. They interpreted this lack of confusion as the listeners' ability to perceive fine distinctions in pitch as they were using discrimination tasks. They therefore concluded that while there is a category of emphatic vs. normal pitch, listeners are sensitive to small changes in f0, which makes pitch "continuously perceived."

The issue of whether or not prominence is a +/- feature has little impact on the present study; gradient interpretation of prominence is orthogonal to the alignment of pitch contour and clausal structure. Yet the perception of prominence is important for various theories of prosodic phonology, e.g., autosegmental metrical theory. In asserting that prominence is binary, I intend only to state that being prominent confers a special status on a referent, not to argue for degrees of that status, or lack thereof.

As for boundary tones (as opposed to pitch accent), Remijsen and van Heuven (2003) performed categorical perception tests with Dutch listeners. They used resynthesized phrases that ended in an intonation that ranged from

clear statement (**L-L%**) to clear question (**L-H%**). Their data show that overall, there is a strong S-shaped curve, characteristic of CP. However, unlike other CP studies, there was a great deal of between-subject variability as to where the category boundary is. Although curious, this fact does not alter the authors' conclusion that perception of boundary tones is categorical, at least for Dutch speakers distinguishing declarative from question sentences.

Remijsen and van Heuven criticized the Ladd and Morton study for experimental design flaws. Specifically, they criticized that the steps on their accenting continuum were too close together as to induce the expected CP characteristics. They referred to Ladd and Morton's footnote that explained how perfect discrimination was possible when they doubled the distance between stimuli. Remijsen and van Heuven constructed their stimuli as having a distance between tokens larger than Ladd and Morton's reported (and analyzed) data but smaller than Ladd and Morton's footnoted study and found clear CP properties, which underscores the importance of experimental design. The difference in conclusions between the two experiments seems to be due to stimuli design, and that categorical perception experiments must include appropriately spaced stimuli, lest different results obtain (as in the case between Ladd and Morton's analyzed vs. footnoted studies). It is also worth

pointing out that Remijsen and van Heuven were studying a different intonational contrast than Ladd and Morton, but given that both studies were targeted at CP of a salient intonational feature, they are at least broadly comparable.

3.3. Accenting and Information structure

Pitch accenting is not the only acoustic or phonetic process associated with information status, or whether a referent has been mentioned before. Numerous studies have investigated the production and processing of referring expressions with respect to their duration (where the process is called "shortening") and intelligibility (sometimes called "attenuation," though that term is more comprehensive). The overarching theme of this research is that given information, or referring expressions that have already been mentioned recently in a discourse, is spoken in a different manner from its first appearance in the discourse. The next section will discuss the role of information structure in processes such as these.

The term "information structure" refers to the distribution and form of referring expressions as they surface over a discourse. I will use the term primarily as shorthand for the given-new dichotomy of a referent to a discourse, with an implicit recognition that such a dichotomy is continuous, not categorical. Thus, a mention of a referent entails assigning it a value on this

scale, where it is discourse-new, discourse-given, or somewhere in between. Notably, speakers and hearers may assign different values to a referent, and may model their beliefs of the other's determination of given or new for a particular referent.

One of the oldest directions for intonation researchers has been the interplay between intonation and information structure (Bolinger 1958). To that end, each study that refers to the discourse status of an entity must operationalize the notion of information structure, as I have above. While some studies prefer the coarse given/new dichotomy (Noteboom & Kruyt 1987, Terken & Noteboom 1987, Horne 1990, Hawkins & Warren 1991, inter alia), others have used the more fine-grained classification of Prince (1981) to look for an alignment between gradation of information structure and gradient responses in a task (Brown 1983, see below). Therefore, from the outset researchers must make decisions that shape the data and analysis, potentially rendering similar studies incomparable.

This potential incomparability arises when one tries to draw general conclusions about intonation and information structure from studies that differ in theoretical prerequisites. A study like Noteboom and Kruyt (1987), which was contextualized as contributing to text-to-speech applications, takes a

feature-based approach to information structure where referents are given or

new[3] because its goal is to map accent to focus, which they also define as

being plus or minus. Noteboom and Kruyt found that in general listeners

judged spoken sentences with accenting that matched information structure as

more acceptable than those where accent patterns did not match expectations.

This contrasts with a study like Baumann and Grice (1996), who unpacked

givenness as an experimental variable and sought gradient responses for

unaccentuation to mirror the gradient nature of givenness. Based on these two

studies, one could only hypothesize that a semi-active referent with an

intermediate accent would be judged as intermediately acceptable; the

plus/minus experiment had different aims and employed a different materials

design than the more continuum-based experiment, which makes them difficult

to compare.

Further complicating matters, Bard and Aylett (1999) and Terken and

Hirschberg (1994) have shown that the syntactic relation (subject, object, etc.)

of a referring expression affects how prosody and information structure

---

[3] The authors also mention the possibility of "implicit" givenness, where a referent is mentioned in the prior discourse, but referred to in the target sentence with a different, but synonymous, noun phrase.

interact. A more comprehensive survey of this research appears in the next section.

Different researchers using different criteria have also caused some confusion about what constitutes givenness in formulating their backdrop for unaccentuation. Givenness (Schwartzchild 1999, Bard & Aylett 1999), precedence (Barr & Keysar 2002), accessibility (Mithun 1996, Chafe 1996), relevance (House 2006) and predictability (Aylett & Turk 2004) have all been used to describe the discoursal force behind unaccentuation. Clearly related, each term and its suppositions seems to vary from researcher to researcher, making it different to characterize what unifies research in unaccentuation. The consensus seems to be that prior discourse or physical context can prime or pre-activate some referent, making it easier to identify or cognitively access. I will employ the given/new terminology and interpret givenness as a cline; referents may be more given or more new than others, and the degree of givenness, for current purposes, will be contingent upon their explicit mention in the prior discourse.

Delving into the myriad ways in which expressions may conjure other referents, Prince 1981 developed a typology of givenness where the manner of referent introduction plays a role. Brown (1983) tested the Prince hierarchy

experimentally by comparing the degree of deaccenting exhibited by a referring expression (RE) with its location in the taxonomy. She found a robust relationship in which the more given or activated the referent is, the more deaccented the anaphoric RE is. The hierarchical approach is also worthwhile because it accounts for the grammatical and discourse environment that leads to activation. With respect to Brown, it stands to reason that how a referent is introduced affects its activation state.

In English, REs that tend toward the given end of the cline are commonly generic, pronominal, or definite. Umbach (2001) investigated the realization of definite NPs, and argued that while definites may be used for given or new referents, there is a preference for listeners to understand unaccented definites as given information. Conversely, listeners surmised that accented definites introduce a new referent. Definite noun phrases function as an important ingredient for accessibility and reference theories, as psycholinguistic experiments often use definites as stimuli, and Umbach's research shows that accent can be a powerful influence on the interpretation of an RE as given or new.

Other researchers have focused on the realization of pitch accents on full NP anaphors based on the activation of the concept referred to or the relation of the referring expression to its precedent, e.g., part-whole, hyponymy,

synonymy, etc. (Baumann & Grice 2006). The impetus for this type of research was the complex notion of accessibility, where one entity evokes another. For instance, "waiter" makes "menu" very accessible. They examined how semantic relations could be activated and treated as given (i.e., unaccentuated) even if they were not strictly referred to in the past discourse. Accessibility in this sense is a gray area in intonation research, as so few NPs in discourse are completely new (and also so few are realized as full NPs) (Everett 2009). This dovetails nicely with Brown's results in finding that lexical semantics plays a role in argument realization, as mediated through information structure.

Another question regarding unaccentuation is the crosslinguistic validity of the process. Because givenness and information structure is a universally salient feature of discourse (Lambrecht 1994), and seems to be a major factor in unaccentuation, there is the possibility that all languages employ a similar prosodic strategy in marking given or new constituents. However, evidence suggests that unaccentuation is language-specific, and that it does not obtain for many languages in a typologically diverse sample (Cruttenden 2006).

Avesani and Vayra (2005) studied Italian as an initial inquiry to accenting in Romance languages. They found that NPs that were repeated in a discourse (and therefore given) were just as accented as the original mention. That is,

Italian speakers did not show unaccentuation in cases where English speakers are known to do so. This implies that the cognitive processes that underpin unaccentuation are language-specific, and not general universals.

Cruttenden (2006) provides the strongest evidence yet that unaccentuation is a language-by-language phenomenon, though when present, one that is sensitive to the givenness of a constituent. He examined 14 languages from a diverse set of families and constructed 10 pragmatic and syntactic scenarios where unaccentuation would be likely based on givenness. His results indicate almost mandatory unaccentuation for English and German, and almost mandatory accenting of repeated constituents in Romance languages, especially Spanish and Italian. Surveying across scenarios (which are designed to modulate the givenness of a referring expression), it is clear that accenting is not a binary typologic distinction—there is a cline of unaccentuation with no discernible relation between languages. Examining the extrema of his accenting/unaccenting cline, a tentative hypothesis might be that languages that makes grammatical or lexical use of stress (Spanish) or pitch (Swedish) are reluctant to adopt unaccentuation as an information packaging strategy while languages like English and German are more free to exploit unaccentuation, as they do not employ the suprasegmenal tier for lexical or grammatical relations.

However, the sample sizes are so small and the languages sampled so few that any such explanation would be highly speculative at best.

3.4. The Grammar/Phonetics Interface

As hinted at in the previous sections, there are various ways in which speech production may suggest a syntactic structure. In this section, I will further survey how syntax and information structure interact with phonetics. It should not be surprising that a speaker's production of clauses belies her parse or discourse model[4] of the status of referents, nor that listeners are mindful of prosody and the suprasegmental tier in performing parse construction or referent identification.

For an analyst to consider them significant, the phonetic processes that interact with syntax and discourse must either amplify or attenuate some acoustic measure(s). That is, if I am to claim there is a relation, then in the presence of some discoursal or syntactic feature (like a new referent, or a relative clause), there should be a measurable difference in an acoustic dimension. This is an obvious point, but it bears mentioning since the acoustic

---

[4] A speaker's discourse model is the mental representation of which referents are active in a scene, and their relative salience.

measures that have been studied in this context are few: f0 (which was discussed earlier), amplitude/loudness, duration, and breaks or disjunction. Additionally, some researchers (Bard chief among them) have extended potential indicators to a perceptual domain, but his methodology regarding intelligibility has not been widely adopted.

Beginning with duration, "shortening" has been a focus for Fowler (Fowler & Housum 1987, Fowler 1988) as a means of understanding how speakers and hearers of English modify repeated references. The 1987 paper showed that duration is strongly tied to whether a content word in a spoken monologue has been said before or not. They showed that previously mentioned words were shorter and less intelligible than new mentions. Perceptually, listeners were able to use this attenuation as a cue, preferring to treat attenuated tokens as anaphoric. Fowler et al (1997) examined how attenuation functioned across episodic boundaries in a narrative. Their results, which show only weak effects, indicate that durational shortening is blocked across boundaries. That is, once a new episode begins, speakers produce the token with a longer duration.

The 1988 paper adds an additional twist: the shortening process does not occur when speakers are reading lists of words, but it does occur in monologic narratives. This result could be the divorce of word and referent; a word in a

list has no contextual binding that places it as a reference. But it also raises the possibility that shortening is a hearer-oriented activity, engaged specifically to assist a conversational partner.

Other researchers, most centrally Bard, have focused on intelligibility as a dimension to be manipulated by speakers seeking to layer their speech with information structure cues. Intelligibility is more difficult to measure, as it requires a first experiment to generate tokens, and a second perception experiment to measure how accurately a hearer can identify the word when excised from context, which is inherently subjective. However, there is value in this double measuring—it measures actual hearer perception, as opposed to acoustic metrics. The benefit is that the results implicitly address the question of whether or not listeners actually pay attention to how intelligible a token is. However, it should be noted that the results indicate a difference in signal quality in the two conditions: degraded "intelligibility" measured in a lab setting does not necessarily imply that listeners in a real-time speech event experienced difficulty interpreting the speaker.

Like shortening and unaccentuation, intelligibiltity attenuation expects "less" of something in the repeated, activated, or given context. This might initially be hypothesized to be due to effort minimization—the speaker just

doesn't want to achieve the phonetic targets more than once. However, the reductions can also be seen as a productive resource for signaling. The suprasegmental tier can be used for morphemic (as in Mandarin) or syntactic (as in a number of African languages, or English in denoting yes/no questions) purposes. Aylett (2000) identifies an inverse relationship between redundancy (his term for repetition) and length/care of articulation. In both Aylett (2000) and Aylett & Turk (2004), he promotes a smooth signal redundancy hypothesis, where the inverse relationship between redundancy and intelligibility is the result of a processing tradeoff; token recognition is more difficult when the signal is degraded, and referent identification is more difficult when the referent is new to the discourse. By balancing these competing processes, speakers "smooth" their talk through redundancy and phonetic reduction. Like unaccentuation, though, it is unclear whether this would be a cognitive universal, or a byproduct of the non-accent on given information in the certain languages that have been found to demonstrate it.

The role of prosody in disambiguating syntactic structures has also emerged as a productive area of intonation research. Researchers have focused on how prosodic breaks may trigger syntactic and semantic wrap-up, causing listeners to align constituent boundaries with breaks (Schafer 1997, German et al 2006). These studies show that speakers and hearers orient to the dual

structures of intonation and syntax in parsing, and therefore that prosodic organization may influence the listener's representation of syntactic phrasing. Various models exist for predicting where in a clause speakers will produce a break or how prosodic phrasing will inform online processing of ambiguous clauses (see Watson & Gibson 2004).

Parallel to the disambiguation research are studies that look for prosodic effects in anticipatory referent resolution strategies (Snedeker & Trueswell 2003). These studies differ from ambiguity resolution experiments in that they investigate how listeners use prosody to identify referents in syntactic environments that do not lead down a garden path. The goal of this research is to learn how processors employ accent and break information online to resolve a referent, rather than a parse. Experiments test how quickly listeners are able to perform a task—either identifying the referent or a side task like lexical decision—in environments that either conform to a known prosodic patterning (e.g., new referents are accented) or deviate from it (cf. Terken & Noteboom 1987).

Other researchers have investigated syntactic ambiguity (choosing among a number of valid parses) as an area that may be influenced by prosody. Syntactic ambiguity is of two major types: garden-path and attachment. The

difference between the two is whether or not the ambiguity is temporary and resolves once all the elements of the sentence are known. Garden-path ambiguities, as in 1) and 2), are the result of incrementally processing each element of the sentence, which is necessary due to temporal linearity. Attachment ambiguities arise from multiple felicitous attachment sites, as in 3).

9)   Whenever the maid checks the room is dirty.

10)   Whenever the maid checks the room it's dirty.

11)   The spy saw the birdwatcher with binoculars.

Comparing 9 and 10, the noun phrase *the room* is temporarily ambiguous; it may function as the direct object (DO) of the verb "checks," as in 10, or it may be the subject of a sentence complement (SC), as in 9. In written English, punctuation has a significant effect on how people prefer to parse the sentence (Hirotani et al 2006). But in spoken language, the presence or absence of a prosodic break after *"checks"* is the most salient cue. This contrasts with sentences like 11 where there is no grammatical cue for the correct interpretation (i.e. the binoculars could belong to either the spy or the birdwatcher).

Cross-modal naming tasks (Boland & Blodgett 2001) have been useful in demonstrating the incorporation of prosodic breaks in online syntactic

processing. In these experiments, subjects heard a stimulus up to "room" (using 9 and 10 as examples), which either had a prosodic break after "checks" or not. Subjects then saw a visual target (either "it's" or "is") that they were supposed to treat as the next word in the sentence. Subjects then read the word aloud and supplied a self-generated completion for the sentence. The authors predicted that having a break after "check" would induce speakers to wrap-up their parse to that point and consider "the room" as the anaphoric subject of an upcoming clause. If there was no break, the authors anticipated that "the room" would be parsed as an object of the verb "checks." Reaction times to the visual target would indicate whether the target was aligned with the listener's parse or forced a syntactic reanalysis. This hypothesis was borne out, and later work (Blodgett 2004) showed that the lexical bias of the verb did not diminish this effect.

There is evidence that referent tracking involves two linguistic dimensions (referring expression form and prosody) that correspond to two psychological dimensions (identifiability and activation). Lambrecht (1994) posited that referring expression (pronoun, lexical NP, zero) is correlated with identifiability, or the ease of picking out a specific referent among those in the discourse. Simultaneously, pitch accenting conveys the activation of the

referent in memory. While activation and identifiability are related concepts, they are distinct in that they can be modulated independent of each other. For instance, a person known to two interlocutors is certainly high on the identifiability scale, yet if introduced into the discourse, their activation of that person will decay with time. As an example, both Fred and Carol know Barbara as a mutual friend. Fred and Carol start a conversation about gardening, with no reference to Barbara. At this point, if Fred were to say, "You know, Barbara likes gardening," Barbara would be high on the identifiability scale (because she's a common friend) but low on the activation scale (because she hadn't been mentioned in the prior discourse).

Prosody has been shown to be a useful resource for listeners in referent resolution experiments. Listeners readily use speaker disfluency to anticipate an upcoming new argument (Arnold et al 2003, 2004). German listeners also used contrastive pitch accent (L+H*) to anticipate a switch from the topic referent to a new referent (Weber et al 2006). Together, these results show that speech rhythm can be a perceptual cue for the direction of a discourse and topic.

4. Summary

This chapter has been a brief introduction to prosody and the syntax-phonetics interface. It began by laying out the research question for this

dissertation: is there a relationship between intonation and transitivity? That potential relationship, to be hypothesized in the next chapter, required a number of working definitions and a survey of research into how intonation intersects with other components of the linguistic system.

The single criterion for transitivity, as employed here, is the number of overt arguments, where argumenthood is defined as a noun phrase immediately following a verb and not governed by a preposition. As was acknowledged, the actual linguistics of transitivity is far more nuanced and far more complex, but this definition will permit a hypothesis that is more directly testable than a fully developed factor of transitivity that accounts for animacy, affectedness, etc.

I briefly touched on a set of soft constraints that rely on a tripartite distinction between core arguments: A, S, and O. These constraints were grouped under the heading of Preferred Argument Structure, which appear to be valid across a variety of languages and speech situations. PAS was useful for its ratification that transitive and intransitive subjects in English are syntactically fused, yet in discourse may have different distributions (although other researchers suggested this difference in distribution was due to animacy of

actors and not an underlying constraining on the character of transitive and intransitive clauses).

I then examined a commonly studied property of English: some nominals receive a pitch accent and others do not, a feature that is correlated with the discourse status of the referent. For various reasons, I eschewed the more common "deaccented" for the more neutral "unaccented" for speech that did not have a pitch accent on it.

Lastly, I conducted a brief survey of experimental work that correlated phonetic features with the production and comprehension of syntax and referent resolution. It is fair to say that speech is designed with phrasing in mind, and that prosodic phrasing offers a domain for mirroring syntactic phrases, whether they are set off by breaks or emphasis. I also pointed out that there are a number of dimensions along which production may be modified, and these need not be simple accent or duration.

In the next chapter, I will develop testable hypotheses regarding the nature of the relationship between transitivity and intonation. I will continue to examine the state of intonation and syntax research, but in more structured manner. These hypotheses are built directly from the 4 propositions stated at the outset of the chapter, and Chapter 2 will discuss the prior work that justifies them. In the third chapter, I will report an experiment that directly

tests that hypothesis. In the final chapter, there will be a discussion of what the

experiment can and cannot explain, as well as directions for future research.

Chapter 2

As mentioned in Chapter 1, the researched presented here rests on the following propositions:

---

1. As speakers talk, their baseline pitch decreases

2. Given two acoustically equal pitch accents (i.e. of the same absolute Hz), listeners perceive the accent that is later in the clause as more emphatic

3. In English, new referents are accented

4. In English, new referents are later in the clause

These four propositions suggest the following hypothesis:

- Subjects of transitive clauses will have lower pitch accents than subjects of intransitive clauses.

In this section, I will discuss the body of evidence supporting each proposition. I assert that propositions 1-4 and the research that supports them are sufficient to motivate the hypothesis. The experiment detailed in chapter 3 is designed to specifically test this hypothesis using a language production procedure.

Proposition 1

1. As speakers talk, their baseline pitch decreases

This is a noncontroversial principle in prosody and intonation, not only because of its cross-linguistic attestation but because of its simple explanation: speakers have to exhale to speak, which decreases air volume in the lungs, which in turn decreases the rate of expiration and f0 (Pierrehumbert 1979). However, fluid dynamics cannot entirely account for the normally observed rate of declination over time (Maeda 1976, see below), so I will examine other factors that may lead to pitch downtrend.

Essentially, this proposition notes that as speakers move from the beginning toward the end of an utterance, their fundamental frequency (f0) will trend downwards. There will be peaks and valleys over the course of a pitch contour, which may themselves carry meaning. The baseline, in this context, refers to the lower bound of f0 through those peaks and valleys. Complementarily, topline is the upper bound. The baseline may be thought of as an imaginary line that connects the valleys of fundamental frequency, and topline as the connection of peaks and plateaus. Topline and baseline form a theoretical, downward-trending envelope for f0.
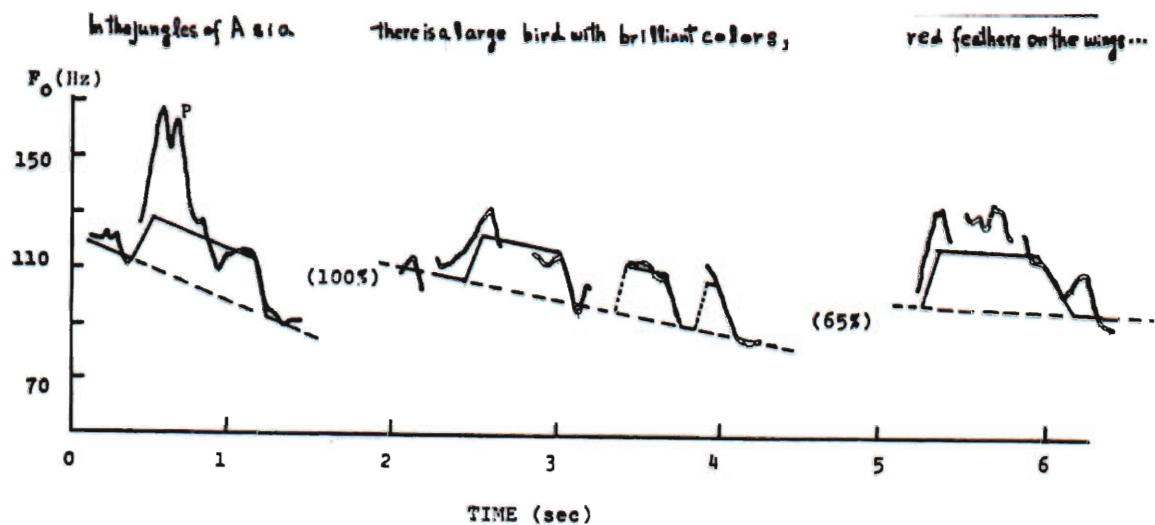
Figure 2. Hand-drawn models of baseline and peaks in f0, from Maeda (1976).

Appeals to physiology and the pressure differential between each side of the glottis are common in attempting to explain f0 declination. However, the findings of Maeda (1976) regarding declination and muscle activity with respect to the hyoid and larynx are worth discussing. Maeda found declination in his experiment to be greater than predicted by decreases in subglottal pressure alone. Thus, he formed the hypothesis that another physiological factor is in play, called "tracheal pull," in addition to a decrease in subglottal pressure.

Based on a video X-ray of a single subject, Maeda finds evidence to isolate the physiological factors that govern declination. Although f0 was affected by larynx height, laryngeal configuration and trajectory seemed to be far more robust in explaining amplitude contours (Maeda 1976, p.219), and while laryngeal height was in general correlated with higher f0, it did not track with various peaks. In short, the larynx generally ended up lower as a sentence progressed, but not enough to account for the variation of declination. The factor that did correlate was that of ventricle length, or "the gradual shortening of the vocal fold length" (Maeda 1976, p. 233). However, the musculoskeletal cause of such shortening (as it continues generally throughout an utterance) was not fully determined.

It is worth reiterating that Maeda was only looking at English speakers, and only one of them. However, even though the exact reasons for declination are still murky, the phenomenon clearly does exist and does seem to be physiologically controlled, with various physiological mechanisms and physical principles responsible, primarily located at the glottis.

Given that pitch contours exhibit peaks, plateaus, and valleys in the context of an overall decline, a model could be built for predicting realized f0 values as the result of global and local effects. Global effects in this model would include speaker's vocal fold mass—a factor that would contribute to all measurements in a phrase. Local effects would be more short term, and only affect smaller regions in a contour, such as pitch accenting. Strik and Boves (1995) quantitatively modeled the relationship between subglottal pressure and f0 by analyzing the values and trajectories within this context of global/local effects. They specifically dealt with two counter-arguments offered to the subglottal pressure theory of declination: the first that subglottal pressure alone is not sufficient in predicting f0 downtrend, and the second that downtrend is linguistically relevant (Breckenridge 1977) and therefore in some sense controlled. Regarding the first objection, there is sufficient evidence (Strik & Boves, p. 213-217) to show that respiratory muscles actively regulate fundamental frequency, i.e. subglottal pressure alone does not govern f0. That

is, muscle activity does not have global effects; once the local effects of muscle activity are controlled for, the declination component of f0 is almost entirely accounted for by the changes in subglottal pressure. In considering the second possibility, that declination is linguistically relevant and predictable, Strik and Boves point out that declination is generally seen as "automatic," even if it is "part of the linguistic system." Mere status as "linguistic" does not entail that subglottal pressure is irrelevant for declination.

One complication of acoustically determining baseline declination is that of low boundary tones and pitch accent, denoted as **L-**, **L+**, **L%** and **L*** in the ToBI notation (discussed in the previous chapter). The low tones contrast with their high counterparts (e.g., **H***) perceptually, but are difficult to algorithmically separate; it would be virtually impossible to identify an f0 level that was always considered low, and contrast it with a level that was always high.

These tones seem to bear some linguistic meaning when situated in a discourse context, which means that any experimental stimuli may be designed to exclude them. By designing stimuli that suggest **H*** to the speaker, and measuring pitch peaks and variation via f0, it is more likely that any observed difference in fundamental frequency values is a result of a pitch accent on a

baseline, and not interaction with an unlikely or rare boundary tone plus low tone combination. By also controlling the syntactic context (see above), the local effect of boundary tones becomes predictable and thus an independent variable for later post hoc analysis. In the experiment presented in Chapter 3, the measured tokens tend toward the beginnings and middles of utterances and are not affected by **L%** or **L+** or **L-**. However some initial mentions of story participants are phrase final, and a post hoc analysis will be required to understand their pitch measurements.

It should also be noted that Herman et al. (1996) find evidence for "final lowering" in English declarative sentences, where **L%** tones have steeper drops in subglottal pressure than in f0. They tie this to Pierrehumbert and Hirschberg's (1990) positing of the "end of an interpretive unit" that "cues finality." Final lowering is outside the scope of this dissertation, as I am more concerned with early-contour and mid-contour peaks.

Related to the issues surrounding measurements coming from similar contexts is the concern that utterance length is correlated with steepness of declination (Yuan & Liberman 2010). With sufficiently long text, the declination of f0 over the course of an utterance will be slight, not steep. However, it should be expected that there still would be some degree of baseline declination, even if it is less exaggerated.

Proposition 1 is offered as a valid proposition for all languages; as speakers talk, his fundamental frequency trends downwards. There is evidence that the cause of this declination is a function of physiology and pneumatics, but the precise reason is orthogonal to the goal of this study. Proposition 1 is discussed in terms of the baseline, the imaginary bottom of f0 tracking. In theory, it could be phrased merely in terms of f0—"As speakers talk, pitch decreases"—but that would be overly vague as the actual f0 track does exhibit excursions and peaks. Clearly, f0 is not monotonically decreasing, but the baseline is. Furthermore, the baseline plays a role in perception of pitch accent, a critical component of Proposition 2, and it will be best if all propositions share common terminology and objects of study, to the extent possible.

Proposition 2

2. Given two acoustically equal pitch accents (i.e. of the same absolute Hz), listeners perceive the accent that is later as more prominent.

Proposition 2 relies heavily on the first proposition regarding baseline declination. In short, Proposition 2 is a restatement of existing research proposing that listeners incorporate declination into judgments about prominence, where prominence is understood as a phonological rather than semantic or discoursal phenomenon.

As was discussed in the previous chapter, Pierrehumbert (1979) frames declination in terms of fundamental frequency, utterance length, and amplitude. Her study is one of the first to attempt to model the perception of prominence in the context of baseline declination. By isolating the acoustic factors influencing the perception of prominence, Pierrehumbert concluded a) that "speakers normalize for declination in judging the relative height of peaks in the intonation contour," b) that "speakers expect more declination in wide pitch range utterances than in narrow pitch range utterances," and c) that "the expected slope of declination is less for a longer utterance than for a shorter one." While (b) and (c) dovetail with Yuan and Liberman (2010), who demonstrated the relationship between utterance length and baseline declination with naturally occurring data, (a) has been an impetus for examining the phonological mapping between acoustic properties like f0 and the presence of prominence.

Pierrehumbert (1979) asked respondents to judge the relative pitch of two pitch accents in a sequence of syllables. The figure below shows 5 stimuli from her experiment.
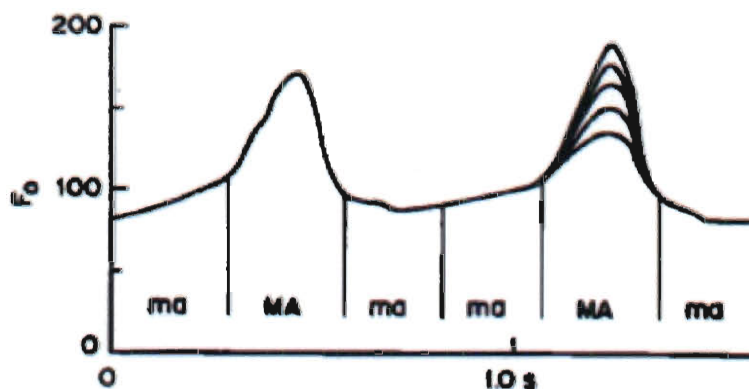
Figure 3. Stimuli from Pierrehumber (1979)

In the above figure, there are 5 pitch contours: 1 where the second peak is higher than the first, one where the peaks are equal, and 3 where the second pitch peak is below the first.

Pierrehumbert found that listeners judged the second peak as having equal pitch when the second syllable was, on average, ~ 9 Hz lower than the first peak when there was a wide pitch range[5].

Pitch range, i.e. the range of values that a pitch contour takes over the course of an utterance, is a complement to prominence. A narrow pitch range

---

[5] Pierrehumbert also found curious results in her narrow pitch range condition, but concluded that they were an artifact of experimental design and the statistical test employed. She nevertheless states "that speakers normalize for declination in judging the relative height of peaks in the intonation contour."

affords little room for f0 peaks and valleys, and thus little acoustic space for generating contrasts. Conversely, a large pitch range suggests that the speaker has employed a variety of f0 values for the purposes of communication, be those purposes affective or otherwise.

Compared with "declination," the term pitch range is more neutral in a temporal sense as it does not convey a sense of downtrend over time. Additionally, pitch range is understood as fundamental frequency range; it does not account for duration or loudness, two other key dimensions in the perception of pitch accent. It is for these reasons that I have phrased Proposition 2 as I have. The notion of progression and temporal sequencing of tokens is critical to understanding why English subjects are realized differently in transitive vs. intransitive contexts.

Pitch range does affect perception, though not in the same manner as peaking. Knight & Nolan (2006) suggested that speakers were sensitive to rapid changes in f0, and not as sensitive to levels when distinguishing peaks vs. plateaus. They asserted that speakers employing a compressed or expanded pitch range did seem to affect peak alignment and plateau length of emphatic syllables.

However, the perception of prominence is multi-dimensional, and cannot be reduced to a purely signal-based phenomenon. Prominence is a linguistic

judgment that imparts meaning and has the opportunity to interact with other linguistic domains. Focus, information structure, and syntax and semantics have all been studied as loci for prominence (Foraker & Mcelree 2007). The act of emphasizing particular words, or constituents, is yet another case of the productivity of foreground/background asymmetry.

Terken (1991) further built on the incorporation of baseline declination in making judgments, but also tested whether listeners performed differently based on if they were instructed to compare the "pitch" of two peaks, or the "prominence" of the two peaks. Regardless of instruction, listeners judged second peaks that were lower than the first as being equal to the first. But in the "pitch" instruction, the difference between peaks judged to be equal was smaller than in the "prominence" instruction. For instance, when the first peak had a value of 125 Hz, with the "pitch" instruction listeners on average judged a second peak of 123 Hz to be have equal pitch. But when judging prominence, listeners judged a peak of 115 Hz to be equally prominent.

The critical distinction for this proposition is between pitch and prominence. Although both are perceptual, pitch is more closely associated with f0 in the literature, while prominence seems to encompass something more context dependent, and is more closely associated with larger linguistic

structures, such as focus. To put it another way, pitch judgments are made based on acoustic signals alone, while prominence judgments may incorporate information from other domains such as syntax or pragmatics. How listeners tend to prosodic constructs such as pitch or prominence matters for cross-study comparison. While some studies seek judgments on equality of pitch (notably Pierrehumbert 1979), others ask subjects to determine prominence (Gussenhoven & Rietvald 1988, Terken 1991, Gussenhoven et al 1987). Further, it is known that listeners do judge pitch and prominence differently, with prominence judgments being more susceptible to baseline declination slope than are pitch judgments (Terken 1991).

The basic experimental setup for a perception of prominence question uses resynthesized stimuli and a procedure that has subjects judge when two syllables are equally prominent (or one greater than the other). This methodology provides a foundation for motivating Proposition 2; I am primarily concerned with how declination models interact with the perception of accents toward the middle and end of a phonological phrase, not the general characteristics of accenting. Essentially, Proposition 2 claims that the perception of accent is influenced by position in the clause.

The studies presented in this section are also conducted with this comparison methodology, where one accent is compared to another in the

same phrase, a configuration which necessarily requires them to share an order relation. The order relation is critical for Proposition 4 below, regarding where new referents are likely to appear in English. For multiple accents in a phrase to be applicable for my research, I appeal to the broader concept of "tune" in prosody, where intonation is seen as the interaction of High and Low tones that are applied to relevant constituents. That is, I will assume that tones are applied to constituents in sequence, which is typically subject-verb-object. This assumption links English grammatical structure to the tonal tier, and permits application of the comparison methodology ("are these two tones equal?") to production of sentences.

This proposition concerns itself with speech perception, and has been framed as the perception of prominence. Like any phonological feature, its interpretation as category or continuum is of interest. Do listeners conceive of prominence as a binary choice of on/off, or do they ascribe a range of values to a referent's prominence based on the continuous nature of the signal?

As mentioned earlier, Ladd and Morton (1997) used a forced-choice design to hypothesize a "categorically interpreted but continuously perceived" status for emphatic speech, which differs from standard phonological perception, where contrasts are typically categorically perceived. Remijsen and van Heuven

(2003) reevaluated and criticized the stimuli of Ladd and Morton on the basis that the steps between stimuli were too small. When there was a greater difference between f0 steps on the unaccented/accented continuum, the results took on the familiar S-shaped response curve. It would thus seem that accent is, at a minimum, categorically understood.

While there has been interest in whether prominence is perceived categorically or continuously, there is a gap in the literature as to whether the same kind of prosodic prominence is *produced* categorically. That is, it seems to be taken for granted that a prominent syllable or referent bears a feature of [+prominent]. Every instance of prominence is assumed to be accented or non-accented. Yet Baumann and Grice found that speakers exploit the inherent continuity of factors pertaining to prominence (f0, loudness, etc.) to signal variable levels of an intermediate level of referent identifiability. Grammatical prominence (e.g., clefting) does have a plus/minus impact; a constituent is either clefted or is not. But variation of prosodic prominence enables a speaker to encode more or less prominence, and consequently a range of values for givenness.

Proposition 2 is critical for linking word order and prosody through a temporal dimension. Its value pivots on the fact that hearers are sensitive to time when making prominence judgments. And while not explicitly stated in

the proposition, it is evident that baseline declination as a function of time is sensitive to the starting point. A pitch contour that starts lower than another will generally be shallower than the one that begins higher. And consequently a pitch peak of X Hz in the lower-starting contour will be interpreted as more accented than the same X Hz peak in the higher-starting contour.

Proposition 3

3. In English, new referents are accented

Clearly, information structure and givenness have a central role in this dissertation, and some operationalization of these terms is in order. In keeping with the spirit of simplicity, I will hew to the given/new terminology and simply recognize that intermediate states exist, but are beyond the scope of this inquiry.

Even a dichotomous distinction of given/new is of sufficient complexity to investigate the nexus of information status, prosody, and transitivity (although, it should be noted, transitivity is prime for my purposes). The motivation for incorporating a factor of information status derives from Proposition 3, which in turn is required for the dissertation's hypothesis. It would be incomplete to not address how, or if, information status interacts jointly with prosody and transitivity.

The study of intonation and newness has also intersected the study of focus (vis-à-vis topic) in the literature. Selkirk (2002) models a distinction between presentational focus and contrastive FOCUS. This is crucial as contrastive focus might be the more tended to form of prominence (Ghomeshi et al, 2004). Selkirk's experiment uses the right-node-raising construction to elicit contrastive FOCUS (e.g., "The Santa Lucia fir is confined to even though it didn't originate in the North American continent.") She found that the second of such contrastive pairs often received a pitch accent on the nuclear vowel (typically an **L+H***) and had a strong prosodic break following it, while presentational focus (e.g., "The catalpa tree originated on the North American continent") was associated with **H*** and lower break indices (i.e. less disjunction between words). She proposes a phonological constraint to compete within a constraint-based framework, ultimately concluding that the **L+H*** might be specific to an IP boundary (which would align with the break observations.). However, with such a specific construction as right-node-raising, the finding of the paper does not generalize to other observed contrastive focus phenomena.

On the other hand, the scalar nature of information structure does seem to map to a scale of tones associated with accent, i.e. High, Low, and Composite (**H+L*** or **L+H***). Baumann and Grice (2006) reported that participants placed a

H+L* accent on information that is accessible through another mention, such as with synonyms or hyponyms. Lexical relations modify the character of accenting on overt NPs; while given information lacks an accent, and new information receives an H*, words whose referents are not explicitly mentioned or are re-encoded are prosodically marked somewhere in between, with the H+L* tone. H+L* tones exhibit a moderate f0 peak and more gradual fall-off compared with H*, which produces the perception that, comparatively, they are "less" than H*.

Listeners also use prominence in constructing a parse. Schafer et al. (2000) demonstrated that prominence may cause hearers to interpret accented information as being a request for new information. Consider this pair, reprised from Chapter 1:

- I asked the girl who is cold
- I asked the girl WHO is cold

The prominence in the second example prompted listeners to interpret the wh-phrase as a sentential complement to "ask," whereas the absence of prominence caused listeners to interpret a relative clause.

The results refer back to the notion of givenness, and provide evidence that accented words are new, or in the case of questions, refer to new information not yet situated in the discourse. In this case, the question-word interpretation is a request for new information, and thus is associated with the accented form. The result is notable because it demonstrates that hearers tend to intonation and prominence, and that such paramters are not "noise" or epiphenomenal. Instead, prominence plays an integral role in the construction of a syntactic parse.

As a discourse proceeds, speakers adhere to turn-taking with their interlocutors. They produce information, which may be a completely new contribution, or more often tethered to prior discourse. Frequently, this asymmetry of information status is labeled as the difference between given and new information, but has also seen terms such as "active," "primed," "precedence," "relevance," "theme/rheme," or "accessible," among others. The role of semantic frames has also illuminated the typology of information structure; Prince (1981) constructs a hierarchy of givenness, which Brown (1983) successfully demonstrated could be used to elicit gradient response times for a task. Brown demonstrates external validity to partitioning the givenness spectrum for the laboratory, suggesting listeners may produce fine

gradations of responses when presented with stimuli that are constructed with givenness as a variable.

In English, prosody is a key resource for signaling new information; such information is accented, lengthened, and more clearly articulated. As discussed earlier, this is commonly thought of as a form of prominence, where the new information (and specifically the nuclear vowel of the head noun) receives special acoustic characteristics, such as increased duration and higher f0 (Fowler & Housum 1987, Bard & Aylett 1999, Pierrehumbert 1979, *inter alia*).

The presence or absence of prominence on a referent has further influences on listener interpretation (Noteboom & Kruyt 1987) and verification (Terken & Noteboom 1987, Bock & Mazzella 1983). Hearers can use prominence to determine a speaker's signaled information status, and even topic changes (Shaffer 1984). Prominence seems to be so salient that not only will speakers reliably withhold it from referents known to the hearer, but hearers are able to reliably anticipate referents based on the pitch contour (Snedeker & Trueswell 2003). It has even been suggested that tone and tune are so reliable as to be phonemic and index givenness (Steedman 2000).

Since subjecthood is such a critical component of the present study, it is worth examining the realization of accent on subjects when they are given vs.

new. Horne 1990 is such a study, which unfortunately does not offer much in the way of insight. Although he claimed "the height of the f0 peak ... did not vary significantly in 'given' vs. 'new' position[s]," he presented no statistical test to validate that conclusion.

Proposition 3 is restricted to English for theoretical as well as practical reasons. Practically, I only have access to English speakers, and thus can only test in English. But it is worth recalling that unaccentuation does not seem to be a universal feature of language; only a limited number of languages in the literature exhibit it (cf. Cruttenden 2006). This is so despite its seemingly universal capabilities to mark new information, where new information is present in all languages. Instead, it appears that only a small number of related languages exhibit unaccentuation: English (the work of Bollinger and numerous others), Dutch (the work of Noteboom), and German (the work of Baumann). Indeed, these languages have been a rich locus for unaccentuation research, and almost all that is known or postulated about this phenomenon derives from evidence in these three languages. For these reasons, no conclusion may be drawn about unaccentuation generally beyond English.

The purpose of Proposition 3 is to establish that it is reasonable to expect English speakers to accent new referents upon introducing them in a discourse. Going forward, it will be assumed that speakers pronounce new information

with emphasis, and a primary carrier of that emphasis is the pitch contour. The crudeness of the given/new model adopted here notwithstanding, information structure presents an important lens through which to view fine phonetic detail and explains a significant component of variation in how speakers choose to accent words.

Proposition 4

4. In English, new referents tend to occur later in the clause.

An earlier draft of this proposition read "New referents in English are typically realized as direct objects." However, this is not really the case. New referents may also be invoked through semantic frames, or as obliques or adjuncts. They are not tied to a grammatical locus, unlike the relationship between English subjects and given information. In PAS, this is expressed as a constraint to "Avoid new A." And while an English speaker could adhere to PAS by situating new information as the subject of an intransitive clause, this appears to be a dispreferred strategy. Regardless of the grammatical role of new information, it is postulated here that it will **in general** not be at the head of the clause or IU.

Proposition 4 crucially ties new information to a clausal position. Ideally, it would be supported with ample evidence from corpora, but I am unable to

obtain readily available data coded for information status. As for conducting

such a study for the purpose of supporting Proposition 4, the assembly and

coding of such a corpus is far beyond the scope of this study.

However, there is related evidence for English as it pertains to PAS.

Kumagai (2006) presented data from English "pear stories" that suggests that A

and S are in fact loci of given information:

	New n	New %	Acc. n	Acc. %	Given n	Given %	Total n
A	4	.9	0	0	440	99.1	444
S	46	8.6	0	0	492	91.4	538
O	114	22.1	12	2.3	390	75.6	516
Oblique	61	20.9	15	5.1	216	74.0	292
Other	6	50	0	0	6	50	12
Total	231	12.8	27	1.5	1544	85.7	1802

Table 1. Information Distribution by Grammatical Role (Kumagai 2006)

The above table shows that among A and S, referents are overwhelmingly

given. Furthermore, the distribution of new referents tilts markedly toward non-

subject positions—only 50 of 231 (22%) new referents were A or S. Tellingly,

the object argument is the most preferred locus of new information. The

distribution of new information seems to suggest discourse accusativity in

English, where A and S have similar properties, while O is different. In this

case, A and S are sites for (almost) exclusively given referents, while O is

available for introducing new referents. Kumagai also reaffirms the findings of Kärkkäinen (1996), who observed a similar ~20% of new referents realized as A or S in a corpus of English.

If new information is predisposed to appear in the O or Oblique position, as in Kumagai's data, can it therefore be assumed that these constituents appear at the end of the clauses, or at least post-verbally? At first glance, the answer would appear to be no; the word order could deviate from standard SVO in any number of ways. However, Kumagai did claim that for coding purposes "only the post-verbal referents in the clause-core position classified into the O category" (Kumagai 2006, p. 680), which suggests that English speakers prefer a post-verbal object. Further, Roland et al (2007) performed a corpus analysis of object position (a category which also included object clefts and relative clauses) to show that objects are post-verbal ~85% of the time (and ~95% in spoken speech). Together these studies provide strong evidence that new information is situated later in clauses, and not randomly distributed around the verb.

As for the lateness component of Proposition 4 (i.e. that new objects will be later in the clause), I base this on the fact that clauses exhibit structure along two dimensions. First, there is the syntactic structure; the hierarchical

mental model that allows parsers to organize the relationships between words. But there is also a temporal dimension. Speakers must present tokens linearly in speech or writing. The grammar of English conspires to create a number of patterns for nominal expressions along these dimensions, but the one that is the most relevant pattern for the present study is that SVO word order places objects (and frequently adjuncts) later in the clause, while subjects are typically toward the beginning. Other researchers have posed cognitive or deep-structure motivations for this order, which mostly concern the privileged status of known information and hypothesize that starting a clause with what is known makes it easier, in some sense, to process the new information later. It is simply an easily made observation that, in general, English speakers place new information later in a clause.

While clausal structure generally aligns with intonational structure (Watson & Gibson 2005, Watson, et al 2006), it need not. Whereas I am comfortable with the assertion that new information be later in the clause, I cannot with any reliability posit where new information occurs within an intonation unit (IU). There is a limit to what may be posited about how information structure, syntax, and prosody intersect, and this limit is exceeded in statements like "new information is right-aligned within a prosodic phrase." The more general "new referents occur later in the clause" removes prosody

from the equation, but still allows a tenable relationship between novelty and lateness, lateness which applies to both clausal structure and intonation.

Hypothesis

My primary hypothesis can be phrased thusly:

> **Subjects of transitive clauses will have consistently lower pitch accents than subjects of intransitive clauses.**

As laid out in this chapter, this hypothesis flows from the following propositions:

---

1. As speakers talk, their baseline pitch decreases

2. Given 2 equal pitch accents, listeners perceive the one later in the clause as more prominent

3. In English, new referents are accented

4. In English, new referents occur later in the clause

New referents come later in transitive clauses, and receive a pitch accent (**H***). By lowering the start of the intonation contour, speakers decrease the

acoustic range for deploying an accent, thereby maximizing the efficacy of the accent in carrying a meaning of "new information."

For transitive subjects, it would behoove a speaker to begin their prosodic phrase at a low level. Doing so would permit the speaker to hit a low(er) f0 target on a new, accented syllable later in the clause, but still have it be perceived as prominent and therefore interpreted as referring to a new referent. That is, **low A's make it easier for later O's to be interpreted as accented.**

Conversely, intransitive clauses do not share the same syntactic or information packaging constraints. Should the subject of an intransitive clause bear new information (which, again, is more rare), it would occur toward the beginning of the clause prior to the listener building a perceptual model that includes declination. Thus, I expect the acoustic correlates of pitch for transitive subjects (A's) to have mean pitch lower than intransitive subjects (S's).

Having laid out the framework for the hypothesis and stated exactly what I expect to find, the experiment in Chapter 3 is designed to elicit talk with S and A, situated in discourse. These tokens may be measured via a variety of acoustic properties and compared across different conditions and within subjects. That data may then determine if there is any systematic intonational difference between A's and S's.

Chapter 3

This chapter presents an experiment designed to investigate whether subjects of transitive and intransitive clauses are produced differently by English speakers, specifically via the manipulation of intonation and accentuation. In this chapter, I test for the modulation of pitch accent on subjects, examining whether the subjects of transitive clauses have reduced f0 excursions when compared with those of intransitive clauses, as predicted by my hypothesis set forth in the previous chapter.

If this hypothesis is correct, I would expect subjects of transitive clauses to possess lower f0 values than subjects of intransitive clauses. For instance, in the

following examples, the word "Louie" should have lower f0 values when produced in narratives like 1), compared with narratives likes 2).

1) Of all of his friends, Louie was the most likely to take on a dare. There basically wasn't anything he wouldn't do for a few bucks. With very little prodding, <u>Louie</u> **drank** until he passed out. Of course, it was just a matter of time before he did another dare and ended up in the hospital.

2) Of all of his friends, Louie was the most likely to take on a dare. There basically wasn't anything he wouldn't do for a few bucks. With very little prodding, <u>Louie</u> **drank** a concoction made by his jokester friend. Of course, it was just a matter of time before he did another dare and ended up in the hospital.
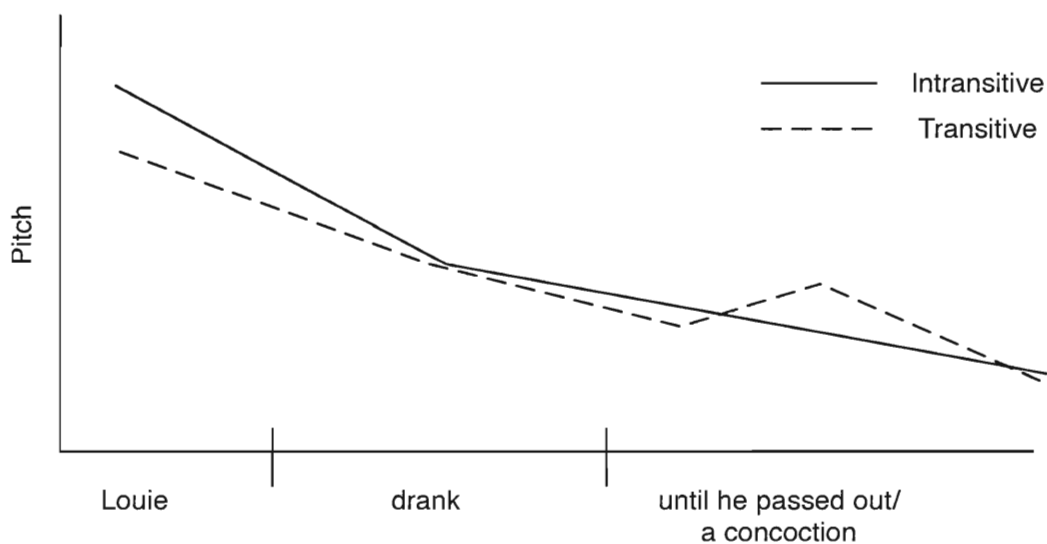

Figure 4. Hypothetical schematic pitch contours for transitive and intransitive clauses

As illustrated in Figure 4, starting the transitive sentence earlier decreases the distance between the initial peak (P1, on "Louie") and the later peak on the object (P2, "concoction"). The smaller the value of P1-P2, the more accented P2 will be perceived (see Chapter 2).

For studying accenting, a common procedure has been to have participants read sentences and dialogues (Gussenhoven et al, 1997). Reading experiments are easily replicated and straightforwardly address the research question of this dissertation. Passages allow the investigator to easily alternate verbs for transitive vs. intransitive constructions, which is critical for validity— otherwise any effect noted could be epiphenomenal and resulting from differences in matrix verbs. Thus, it is required that any procedure generate data points that use the same verb in transitive and intransitive contexts. There are only a limited number of verbs that meet this requirement with much felicity when situated in a given context. Controlling the conditions for producing transitive and intransitive pairs is essential to ensuring repeatability and validity of the results.

It should also be pointed out that this procedure illuminates intonation research from a unique perspective. The use of read narratives of the type presented here is a relatively novel experimental technique (Yaeger-Dror 1996).

In cases where information structure is applied to prosody, researchers have elicited tokens by having participants read dialogues (Horne 1990), answer questions, or produce spontaneous speech (Oliviera 2000). Narratives provide a rich context for information structure yet can be designed so that speakers produce tokens in precisely the conditions of interest (here, transitive vs. intransitive). Given the state of intonation research, constructed reading examples serve a useful purpose, as the reading of isolated sentences and dialogues has been fruitful for the field. The efficacy of the read narrative procedure will be evaluated in Chapter 4.

1. Participants

For this experiment, 21 individuals (18 female) participated in a production exercise. Participants ranged in age from 18-30. For 21 speakers, each producing 34 stories, there should have been 714 possible target sentences. However, two subjects skipped a story, so there were 712 stories available for analysis. Participants were paid[6] $10 for their time and given a debriefing document.

---

[6] Participants were offered course credit or $10, and all elected the money.

2. Materials

In creating stimuli, special attention was paid to the verbs in the target sentence of each story. Lexical verb bias (the propensity for a verb to occur in a particular construction given other felicitous alternatives) has been shown to be a factor in comprehension exercises involving parsing. Rather than conduct a post hoc investigation into whether lexical bias influenced the results of this production experiment, I opted instead to control for that variable and design stimuli to balance it. Based on the work of Gahl et al (2004), I selected a total of 34 verbs, including 12 transitive-biased, 12 intransitive-biased, and 10 equi-biased.

For each of the 34 verbs selected, I constructed a scenario that would permit felicitous uses of the verb in a transitive and intransitive context. I created two versions with transitive target clauses: one with a new object and one with a given object. A participant would see exactly one of the versions of each scenario. One example of a scenario is below, for the verb "lean," an intransitive biased verb.

Intransitive Target	Transitive Target (new object)	Transitive Target (given object)
The construction site was quiet as Lyle pulled into the parking lot. He had arrived early, so with the extra time he got out and looked around the building that was going up. On his walk, **Lyle leaned against a door** to get a good look at the insulation. But soon enough, some other guys showed up and they had to get to work.	The construction site was quiet as Lyle pulled into the parking lot. He had arrived early, so with the extra time he got out and looked around the building that was going up. On his walk, **Lyle leaned a shovel** against the wall to keep things neat. But soon enough, some other guys showed up and they had to get to work.	The construction site was quiet as Lyle pulled into the parking lot. He had arrived early, so he got out and unloaded his shovel and tools. On his walk, **Lyle leaned the shovel** against the wall to keep things neat. But soon enough, some other guys showed up and they had to get to work.

In this design, a participant's response engenders two variables: the transitivity of the clause, and the information status of the object. My focus is on the transitivity; I am concerned with how the grammatical structure of the clause affects pitch excursions on the subject. However, I want to ensure that any results obtained can be more confidently tied to grammar, and not information structure. To that end, I have controlled for the information status of the object as an experimental variable

To avoid a final lowering effect (Liberman & Pierrehumbert 1984) in f0 measurement, none of the measured tokens[7] were placed adjacent to a phrase-final boundary. Additional context was added to the end of target stimuli, as seen below.

- Lyle leaned a shovel **against the wall...**

By adding "against the wall" to the clause, the prosodic phrase boundary is pushed right, away from the token "shovel," which was measured for accenting (see below). Thus, any measurements made on "shovel" are not complicated by the phrase boundary tone or final lowering/lengthening. Instead, I expect words like "shovel" to exhibit some degree of downtrend and, if a new referent, to also exhibit an H* pitch accent.

In the transitive case, the variable of transitivity is confounded with a variable of newness of object. That is, a single target stimulus may engender one or two variables: an intransitive clause may only be analyzed for

---

[7] I measured object tokens to test for validity of the read narrative procedure. See chapter 4.

comparison against transitive clauses, while a transitive stimulus provides data that may be used in analyses of both transitivity and establishing if object newness is a significant factor in the production of subjects.

Grammatical subjects were fully sonorous; given that pitch tracking algorithms can be influenced by obstruent consonants and especially sibilants, when constructing the stories for a scenario I always chose a proper name consisting of only [+sonorant] segments. This maximized the length of the token, as well as the accuracy of the pitch tracking algorithm when analyzing segments.

3. Procedure

Participants were taken to a soundproof lab and fitted with a head-mounted microphone that was attached to a solid-state recorder. For recording level calibration purposes, they were asked to read the instructions aloud, and then proceed with three practice stimuli. The practice stimuli were in the same format of four sentences plus a comprehension question (described below) used with the target stories. At the end of the practice session, participants were asked if they had any questions, and if there were none, were left to complete the experiment in a sound booth while a lab technician monitored their progress.

A session was composed of 34 stories and 34 comprehension questions, or 34 trials. A trial had the following structure of:

1. The story was displayed on the screen

2. The participant read the story silently to himself.

3. The participant read the story aloud into the microphone.

4. The participant pressed a key to move to the comprehension question

5. The story disappeared and the comprehension question was displayed.

6. The participant read the comprehension question silently.

7. The participant spoke the comprehension answer aloud.

8. The participant pressed another key to move to the next trial.

Trials were presented randomly, and there were no distractors in the list. Participants took anywhere from 25-35 minutes to complete the task.

When a participant began a trial, they were presented with a story where the target sentence was one of three variations (see description of scenarios, above):

1) Intransitive clause [Intr]

2) Transitive clause, object is discourse-new [Tr-N]

3) Transitive clause, object is mentioned in prior discourse [Tr-G]

Participants were only presented with one version of each story; approximately 1/3 of the stories read by a participant had an [Intr] target sentence, 1/3 had [Tr-N] and 1/3 [Tr-G]. A participant was randomly assigned to one of three lists of stimuli. Each verb/story was produced by seven different participants for each variation (Intr, Tr-N, Tr-G).

After collecting the recordings, the audio file of each participant was segmented into tokens for the following regions on a per-story basis:

- First mention of target subject (in the first sentence of the story) [s1]

- Second mention of target subject (in third sentence) [s2]

- Vowel of target verb (segmented by syllable) [v]

- Vowel of target object, if present (segmented by syllable) [o]

- Answer to comprehension question [ans]

Each of these tokens was coded, along with information about story/verb and variant of the story (Intr, Tr-N, or Tr-G). For verbs and objects, the syllable

carrying primary lexical stress was also coded. The tokens were labeled using the TextGrid features of Praat. Praat extracted all pitch and intensity contours from the recording session, which were then used as data to perform the analyses described below. For the first and second mentions of the target subject, the entire word was selected for analysis, as grammatical subjects were fully sonorous by design. However, for verb and object vowels, the extracted token also included any post-vocalic liquids, such as r-coloring, that made determining the terminus of the vowel difficult.

Using the audio analysis software Praat, quantitative statistics were gathered for each of the extracted tokens. Pitch was extracted via an autocorrelation method for all tokens, including those that exhibited creaky voice. The pitch detection algorithm reported a Hertz value every 1ms. Similarly, intensity was measured every 5ms in decibels. For each token, the following was recorded:

- Duration

- Mean f0

- Standard deviation of f0

- Mean intensity

- Standard deviation of intensity

- f0 range

- f0 delta (change from beginning to end of token)

- Intensity range

- Intensity delta

Due to the methods of autocorrelation and intensity detection, there were some instances when a token did not have a pitch or intensity value at its beginning or end. In these cases, the first data point in the token was recorded as the beginning (and similarly for the end). Since the initial and terminal points were to be analyzed for pitch/intensity trajectory over the course of the token, this had little effect on the outcome.

These measures, which I will refer to as "intuitive" as they are the most obvious indicators of pitch and pitch trajectory, are known to be influenced by other exogenous variables, such as location in the narrative (Oliviera 2000, Oliviera 2003) or length of utterance (Fowler 1988). Determining if grammatical subjects exhibit different pitch contours requires a more sophisticated analysis that incorporates context. That is, accent can only be judged with respect to the overall trendline against which a pitch peak stands in relief.

To detect this type of accent, I created a linear regression model of the local pitch contour around a token and measured the mean residual generated

by the target word itself. I will refer to this as the linear regression data, or LR data.
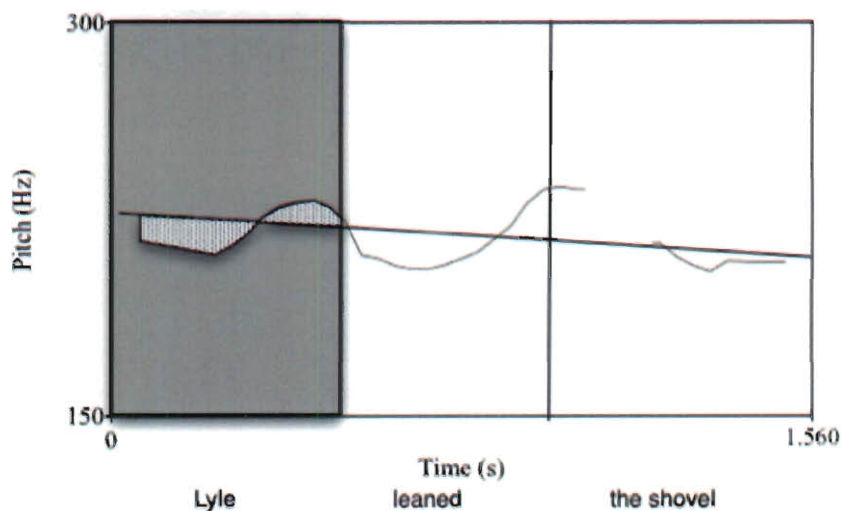

Figure 5. f0 track of a target sentence (Tr-G). The shaded rectangle represents the grammatical subject (A), while the hatched region is the residual of the linear regression.

Figure 5 above has 4 components: an f0 contour, a linear regression fit line, a shaded region of time denoting the target word, and a hatched region illustrating the residual of the linear regression over the course of the target word. In the example, the majority of the area of the hatched region is below the regression fit line, which is realized as a negative mean residual.

A linear regression model is advantageous because it estimates a speaker's natural f0 based on physiology or prosodic phrase boundary resets (the intercept) and the downtrend for the particular prosodic phrase being spoken (the slope). Mean residual for a token is a measurement of how much that token deviates from an "expected" intonation contour. That is, the fit line provides a baseline against which a listener may project a speaker's intention to highlight or downplay a constituent.

For this analysis, three linear regression models were based on three samples of f0: the target subject plus 1s of speech after, the 1s of speech after alone, and the minima of f0 from the word plus the 1s of speech following. One second was chosen because it would capture most if not all of the target intonation phrase and thus be an appropriate model of the local f0 contour. The following graphs are for illustrative purposes only, and do not represent actual linear regression fits.
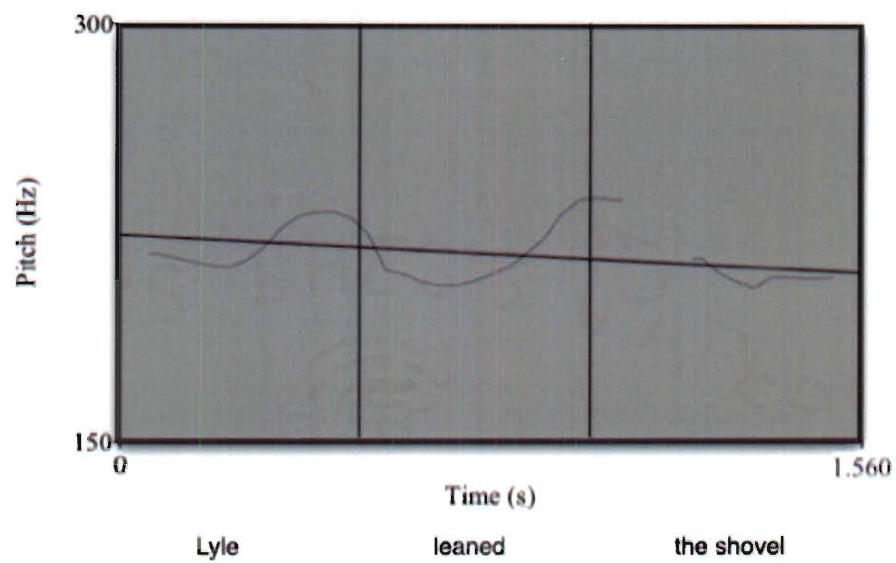
Figure 6. Sample regression fit based on the target word plus 1s after

84
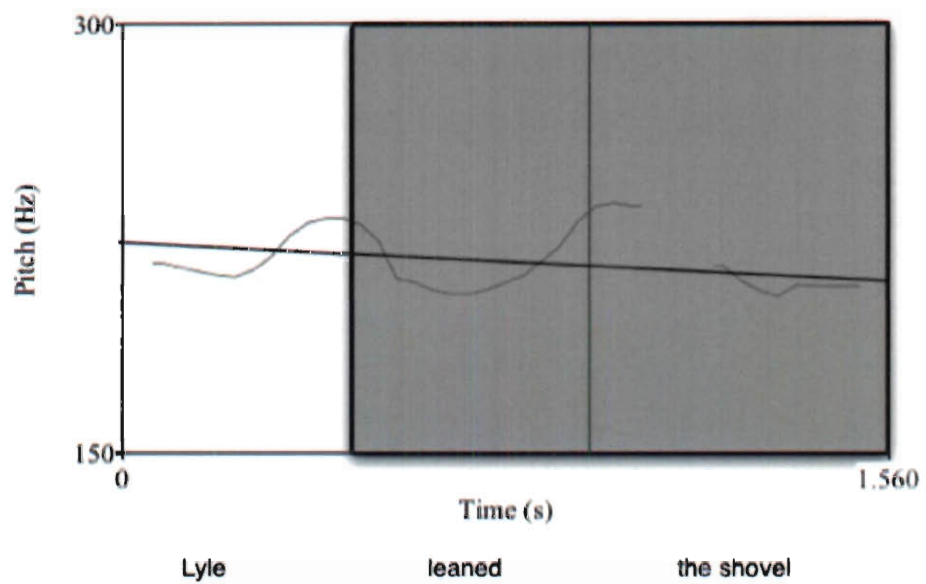
Figure 7. Sample regression fit based on 1s following the target word

Figure 8. Algorithmically determined f0 minima and sample regression fit based on minima in the target word + 1s region

4. Results

After collecting and coding all tokens from participants, I discarded 29 tokens (4.1% of 710) because the participant incorrectly answered the comprehension question for the story. Accordingly, all repeated measures ANOVAs presented below are based on aggregated data.

4.1 Effect of Transitivity on Intonation of Grammatical Subjects

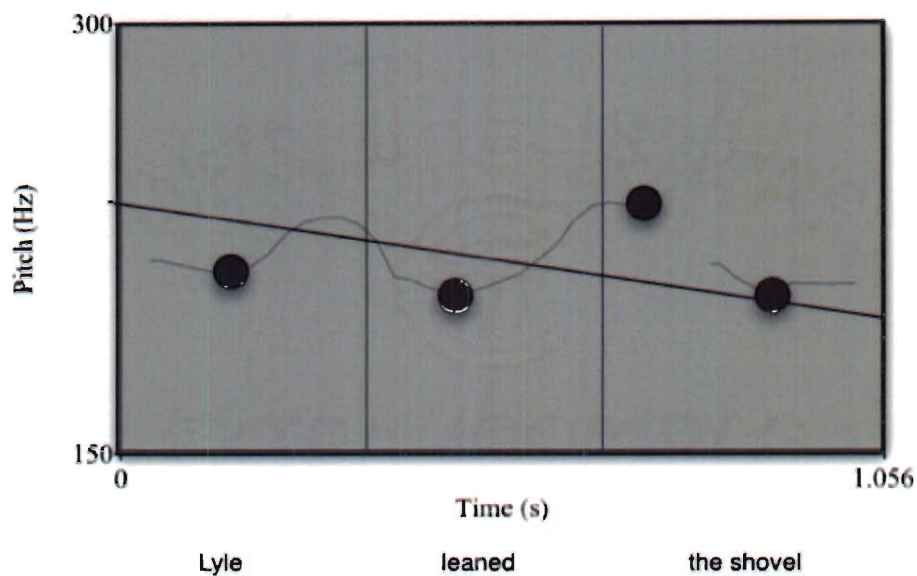As expected, the intuitive measures did not reveal a significant effect for transitivity. Using the intuitive measures of grammatical subjects as the response (i.e. A or S), a one-way repeated measures ANOVA was performed using transitivity as the main factor, using aggregated data.

Measure	$F(1,20)=$	$p<$
Mean of f0	2.057	0.167
Mean of Intensity	2.805	0.110
Standard Deviation of f0	0.560	0.463
Standard Deviation of Intensity	1.504	0.234
f0 Range	0.424	0.522
f0 Delta	1.524	0.231
Intensity Range	1.587	0.222
Intensity Delta	0.054	0.818
Duration	0.218	0.646

Table 2. Results for intuitive measures, testing main effect of transitivity.

As is clear from Table 2, transitivity was not a significant effect for any of the intuitive measures.
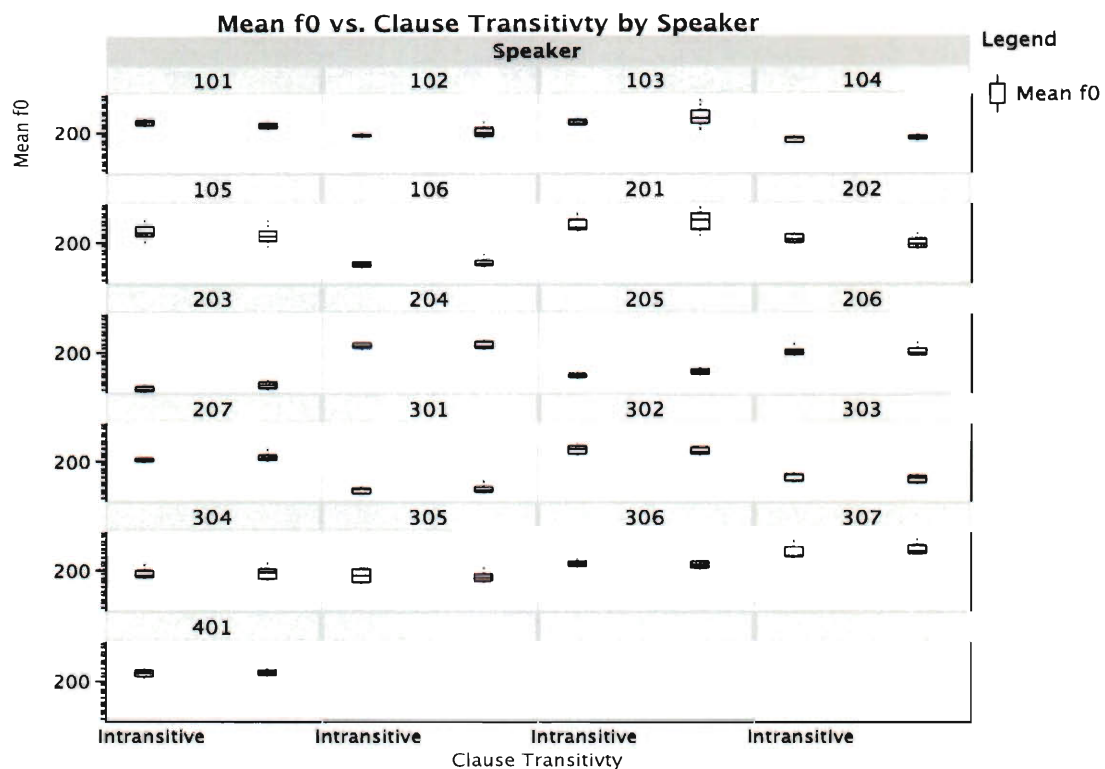
Figure 9. Mean f0 vs. clause transitivity, by speaker. For each speaker (e.g., 101), box plot for intransitive is on the left, transitive on the right.

When examining the linear regression (LR) data, however, transitive tokens have lower f0 contours than their intransitive counterparts. For completeness, LR fits were computed for multiple subsets of the pitch data: the target subject + 1s after, the minima of the subject + 1s data, and only the 1s after the target subject (see Figures 6-8). Using mean residual of the

grammatical subject (A or S) as the response, a one-way ANOVA with repeated measures and using aggregated data was performed with transitivity as a main effect.

Regression Basis	Intransitive Mean (Hz)	Transitive Mean (Hz)	$F(1,20)$	$p<$
Subject + 1s after	-0.430	-2.230	12.163	0.002
Only 1s after subject	-5.149	-10.512	10.359	0.004
Minima of subject + 1s after	8.049	5.343	7.315	0.014

Table 3. LR Data results by LR model using both one-way and two-way mixed effects models. Values are for clause transitivity effect.
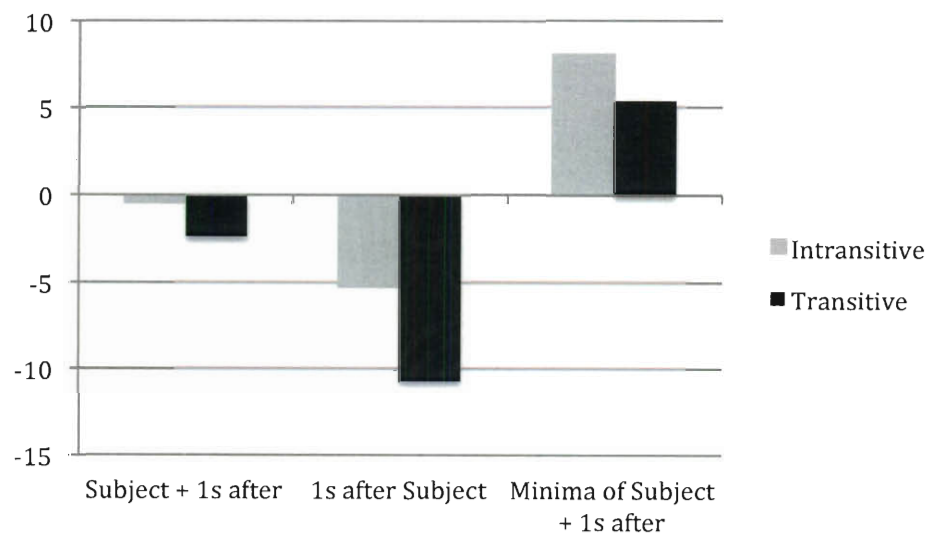

Figure 10. Comparison of transitive and intransitive mean residuals by LR fit basis.

Figure 10 displays the key insight of this study—across all measures, transitive subjects are lower than their intransitive counterparts. The critical comparison is between the left and right bars in each group—regardless of the basis for the linear regression, transitive subjects are lower.

4.2 Effect of Givenness on Intonation of Grammatical Objects

To compare with prior studies, the accent of objects was compared in the given and new condition. For this analysis, only object vowels were segmented for analysis since syllabus nuclei are carriers of accent, which contrasts with the subject analysis where the entire word was extracted and measured as a token. Unlike with grammatical subjects, where I could be very deliberate in choosing the phonetic characteristics of the segments, I was more constrained in word choice with verbs and objects. For verbs, I limited myself to only those verbs that I could ascertain a lexical bias for transitivity, and for objects, I was constrained by the semantics of the verb and the context of the scenario.

As with the subject data, the intuitive measures were not significant. Table 4 reports a two-way mixed effect ANOVA, with object status (new vs. given) as the fixed effect, and speaker as the random effect, as well as a repeated measures model.

Measure	F(1,20)=	p<
Mean of f0	0.683	0.419
Mean of Intensity	0.856	0.366
Standard Deviation of f0	0.610	0.444
Standard Deviation of Intensity	0.514	0.482
f0 Range	0.437	0.516
f0 Delta	0.013	0.911
Intensity Range	0.448	0.511
Intensity Delta	0.423	0.523
Duration	0.039	.0846

Table 4. Intuitive measures for all object vowels.

Employing the linear regression procedure to measure mean residual for words was also performed with objects testing for object information status as the main effect. However, the fit data was based on the previous 1s before the word so as to capture the portion of the contour that was aligned with the core grammatical words (the subject and verb).

There was not a significant effect for object status (given vs. new) when mean residual was calculated for all object vowels, regardless of basis (see below). However, multisyllabic objects could be introducing undesirable variation in pitch. Consider a monosyllabic object like "tree" vs. a multisyllabic

"parasailer." The multiple vowels in "parasailer" will be influenced by the order in which they appear. To eliminate this bias, I performed an analysis of only the lexically stressed vowels. The results are displayed in Table 5.

Regression Basis	Overall mean with new object (Tr-N)	Overall mean with given object (Tr-G_	F(1,20)	p<
Target vowel + 1s previous	-7.594	-10.889	4.804	0.040
1s before vowel	-14.913	-22.342	6.125	0.022
Minima of target vowel + 1s previous	-4.041	-9.579	6.086	0.023

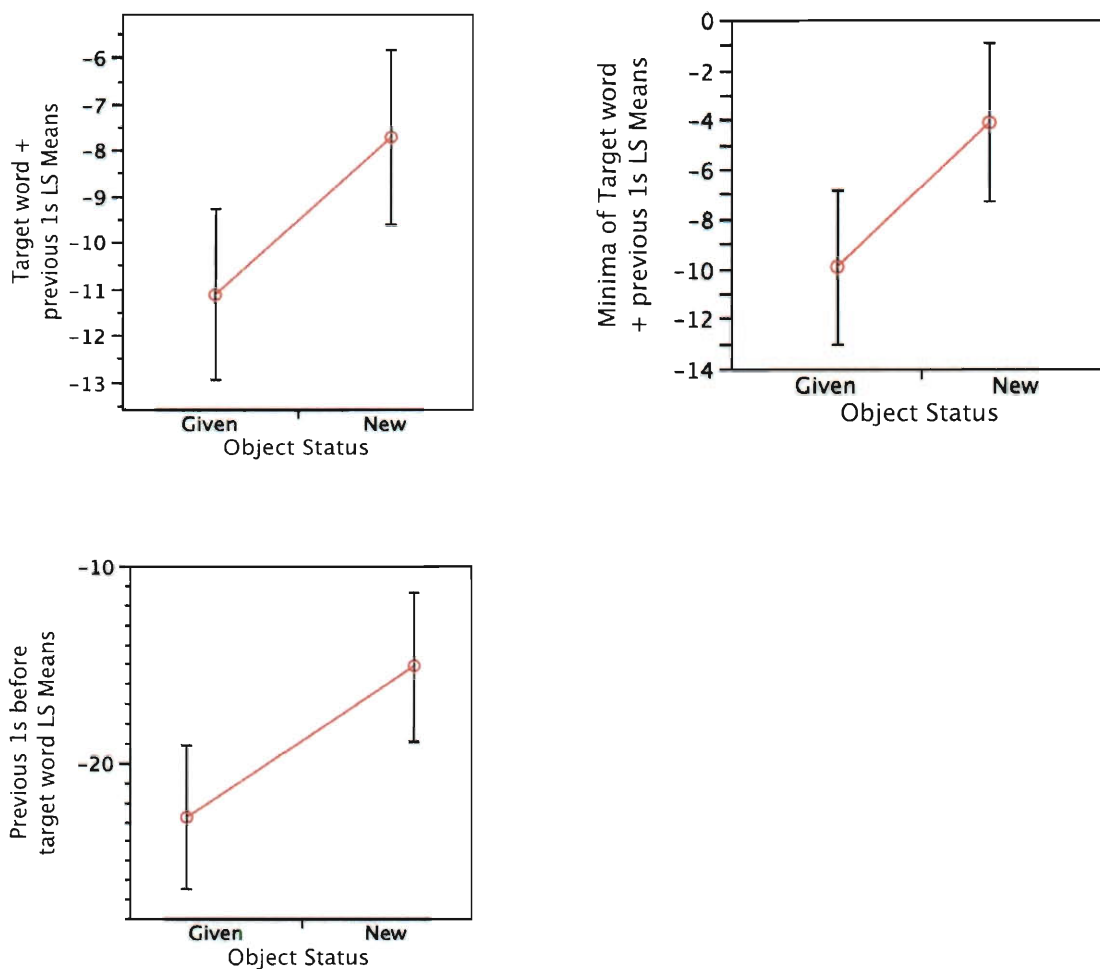**Table 5. Object accent by status using LR data**

Figure 11. Means and standard error for lexically stressed vowels using LR data.

5. Discussion

In sum, there is evidence to support 2 conclusions:

The first is that speakers systematically employ lower pitch excursions for

subjects in transitive clauses than they do for intransitive ones. This is evidence

by Table 3, which shows a statistically significant main effect of transitivity. When coupled with Figure 10, the interpretation of the significance is clearly that intransitive subjects had lower (in absolute terms) mean residual than transitive subjects.

The second conclusion is that the narrative reading experiment produces similar results regarding the accenting of new information, as compared with other studies, with given objects lower. Although the measurement of f0 was restricted to single vowels, it is clear from Table 5 and Figure 11 that new object vowels did indeed have larger pitch excursions than given object vowels.

Critically, the first conclusion is not easily accounted for by a formal account of transitivity or of intonation. The alignment of the communicative function of low transitive subjects with the results suggest a functional explanation for transitive lowering. This preference for a functional appeal will be further explored in the following chapter.

Chapter 4

In the previous chapter, I presented the results of a narrative reading experiment that demonstrated a statistically significant difference in mean residual of f0 models between subjects in transitive and intransitive clauses. The data demonstrated a clear preference for subjects of transitive clauses (A's) to have lower pitch accents than subjects of intransitive clauses (S's). I will refer to this effect as "transitive lowering." In this chapter, I will interpret these results and explore how they relate to the hypothesis and research question presented earlier.

At the beginning of this dissertation, I established the guiding research questions as "Is there a relationship between intonation and transitivity?" It is fair to say that the experimental findings support an emphatic "yes" to this

question. The data support the hypothesis that speakers employ different pitch accents for grammatical subjects based on the transitivity of the clause. This difference is predicted by known principles that are both language-dependent (such as English siting new information in the object position, and being an SVO language) and language-invariant (e.g., baseline declination).

1. Analysis as a Case of Grammaticalization

An interesting point that emerges from this research is that transitive lowering appears to be grammaticalized. That is, I have framed the results as finding alignment between grammatical roles (A and S) and intonation. The results from Chapter 3 show such an alignment, but in this section I will explore those results from other perspectives to show conclusively that transitive lowering has grammaticalized and now exists regardless of any external licensing factors that brought about f0 lowering for transitive subjects.

The hypothesis I laid out in Chapter 2 suggested that an inquiry into the accenting of transitive and intransitive subjects might reveal systematic differences. The hypothesis also was built on propositions that highlighted the importance of information structure in prosody with respect to the accenting of new information. Given the salience of new object accenting, perhaps the

relevant difference in subject accenting isn't transitive vs. intransitive, but clauses introducing a new referent vs. not.

By design, the reading experiment in Chapter 3 provides enough data to address this question directly. The following table presents only the transitive tokens, Tr-G and Tr-N, using LR data.

	Overall Mean with new object (Tr-N)	Overall mean with given object (Tr-G)	$F(1,20)=$	$p<$
Subject + 1s after	-1.943	-2.647	1.377	0.254
Only 1s after subject	-9.034	-11.920	1.275	0.272
Minima of subject + 1s after	5.439	5.256	0.016	0.901

Table 6. Results for subject accenting vs. information status of object referent.

The results demonstrate that the newness of the object is not the relevant variable for transitive lowering—there is no appreciable difference between Tr-G and Tr-N tokens. Instead, the only variable that seems to induce a significant effect is whether the clause is transitive or intransitive. With this finding, I can more strongly conclude that speakers exhibit different accenting patterns for

grammatical subjects, regardless of the information status **or** presence/absence of an object.

I therefore contend that this is properly understood as a kind of grammaticalization (Hopper & Traugott 2003). One hypothesis regarding transitive lowering is that speakers lower the pitch of subjects in an effort to mark (via accent) for a particular purpose—to increase the efficacy of pitch accents on later, new objects. However, the data show that efficacy alone cannot account for the results, as transitive subjects were not distinguished based on the information status of the object. If information structure was the guiding force in subject pitch realization, then I would have expected significant results in Table 6.[8]

Yet transitive lowering occurs **regardless** of whether or not there is a later new object. Given the results of Chapter 3 and the table above, I must conclude that the trigger for lowering is transitivity, and not whether or not there is actually a new referent to introduce. This can be seen as a case of grammaticalization.

---

[8] Obviously, the results in Table 6 are subject to Type II error, but given the data I have chosen not to reject the null hypothesis.

In grammaticalization, a particular form, either phonetic or syntactic, is used repeatedly in various contexts, and then is reanalyzed as having a different meaning (or function) based on the characteristics of the contexts the form had been used in. A classic example is "going to" in English, which originally was strictly used for motion ("going to town") and was then reanalyzed to imply purpose ("going to eat in town"), and now conveys future time ("it's going to rain").[9] The usage, while retaining a relationship to the original, licensing form, is clearly independent and provides a more generalized function/meaning than the initial context. To map that onto transitive lowering, the initial form (low pitch accent) is introduced by a function (efficacy) and comes to be associated with a concomitant form (transitive subjecthood). Speakers reanalyze the lowering to be associated with part of the grammar (i.e., the subject) and not the information packaging of the clause.

To go further, transitive lowering may also be seen as part of an ergative phenomenon because it applies a grammaticalized feature to A (and not S), which happens to be the opposite of a feature English applies to O (namely, accenting of new information). This is indirect evidence that irrespective of

_____

[9] For this example, a phonetic reanalysis also occurred, c.f., *gonna*.

English's adherence to PAS, the constraints seem to conspire to mark English core arguments in an ergative pattern—S and O are available for pitch accenting, while A is not. PAS constraints predict that A will not be new, while S and O might be new. New referents are regularly accented, and so A is not subject to the accenting of new information in English[10].

Figure 12 illuminates this distinction:

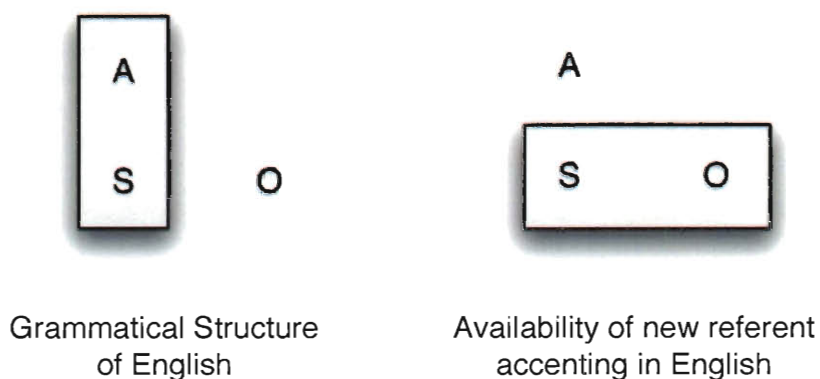| Grammatical Structure of English | Availability of new referent accenting in English |

Figure 12. Split ergativity in English.

It is this ergative pattern in English that I believe is a reflex of the grammaticalization of transitive lowering. Within the context of pitch accenting, A's are subject to different information structure constraints than S's and O's. In

---

[10] However, A's could receive a pitch accent from another process, such as emphatic speech.

terms of the experiment in Chapter 3, this is realized as a lower mean pitch for

A vis-à-vis O.

Although the results of the experiment from Chapter 3 are predicted by

the hypothesis and four propositions of Chapter 2, it is not possible to say

conclusively that maximizing the efficacy of new objects' pitch accents is the

reason for low f0 on A's. It is clear that the results align with the hypothesis,

and speakers produce A's with a lower pitch than their S counterparts.

Proposition 3, which dealt with the perception of sequential pitch accents,

is a useful frame for analyzing the grammaticalization. Starting a clause with a

lower pitch has the effect of making later pitch accents more perceptually

prominent. That is, lowered A's come about for a perceptual *function*—to assist

in the identification of a new object referent. Yet, the low-pitch phenomenon

occurs regardless of whether there actually *is* a new object referent. This is the

essence of the grammaticalization analysis. The phonetic form of transitive

lowering is introduced in a particular situation for a particular purpose, namely

transitive clauses that introduce new referents through the object role. But over

time, that form comes to be associated with the A role itself, regardless of the

object information status. This kind of reanalysis is a case of "bleaching," where

the original, licensing context of a form becomes bleached, and the form (here,

transitive lowering) becomes a part of the larger syntactic construction. Yet this

case differs from the more standard semantic bleaching of Hopper and Traugott in that the domain of "loss," i.e. the component of the construction that is lost over time, is information structure rather than semantics. While pragmatics and usage mediate the grammaticalization in both the standard semantic analysis of grammaticalization, and the information structure analysis here, the lack of historical evidence for intonation patterns makes the case for diachronic change significantly more difficult. However, with the advent of recording technology and the growth in spoken corpora over the past decades, I anticipate that future researchers will find sufficient evidence for intonational change, both in general and for transitive lowering.

In the experiment, speakers tended to newness in generating pitch accents for new objects. Thus, it is reasonable to postulate that speakers were aware of an object's information status when it was produced. However, the only test for determining if the speaker's discourse model *at the time they produced the subject* is if speakers produced subjects differently in Tr-G and Tr-N conditions. Since speakers produced them without reliable differences, then the evidence supports the claim that speakers lower the subject f0 if the clause is transitive; the information status of the object has no impact on the production of the subject.

Formalist models of language would likely judge this to be epiphenomenal, given that the most plausible explanation for the distribution of f0 accenting is closely tied to a communicative function. But this is insufficient for two reasons. First, transitive lowering could be most simply explained by positing a syntactic feature on the subject that reflects the transitivity of the clause, which in turn causes speakers to lower the pitch. This model is advantageous for the formalist because it neatly maps observed behavior to an existing framework of features and phrases. However, it suffers from a critical flaw: if there is feature spreading from IP to subject NP, why is transitive lowering the *only* reflex of it? That is, given the primacy of transitivity in formal accounts, and the sundry methods of testing for parameters and phrasing, why is the only evidence for such a feature a) phonetic and b) hitherto unnoticed by researchers? A formal account of this nature simply cannot explain why a feature that intersects such a key component of grammar (i.e. transitivity) has such a faint, and strictly phonetic, reflex which interacts with no other syntactic processes.

Second, to even acknowledge transitive lowering, one must appeal to usage and experimental data. While psycholinguistics is not opposed to formal accounts and indeed may bolster them, experimental data is most welcome when it builds off existing formal models. Unfortunately, formal models have

failed to predict transitive lowering as the phenomenon does not appear to interact with other posited syntactic principles or parameters. Yet transitive lowering is clearly tied to syntax because transitivity is the only relevant factor in its appearance. Thus, it is incoherent to call transitive lowering epiphenomenal because it has failed to be predicted or detected with formal means, yet is robustly exhibited based on a key dimension of formalist accounts (namely transitivity).

Transitive lowering demands a functional account that recognizes the importance of communication as a purpose of and shaping force for language. The communicative function of pitch accenting in carrying information about a referent's information status in English clearly links transitive lowering to interaction and perception.

2. New Experimental Techniques

To obtain the statistically significant findings reported here, a transform was performed from the intuitive measures of f0 to a more enriched response that contextualized the token—the LR data. The LR data was justified in the previous chapter as being superior for providing a backdrop of declination. It is worth exploring the implications of such a transform, and elaborating on its methods.

At its heart, this experiment is about accent, which cannot be defined without reference to the surrounding context. The intuitive measures simply do not account for that context. Instead, what is needed is a metric of deviation, not from a mean, but from expected values. The LR data supplies this, and allows easy adjustment of the definition of context by calculating the LR parameters from different subsets of the surrounding f0 contour. It is notable that the results held regardless of context, which suggests a robustness to the procedure.

A possible alternative to the LR method of contexualization is to measure the deviation from a mean f0 for some surrounding context. In practice, this is similar to the LR method, with the fit line having a slope of 0. But examination of target phrases shows f0 always trending downward, for all the reasons mentioned regarding Proposition 1 (i.e. as speakers talk, their baseline pitch decreases). The LR data captures this information to paint a more accurate picture of what the background pitch contour is. It is against this backdrop that accent is more accurately measured.

The discrepancy between results for the raw data (the "intuitive" measures) and the LR data is actually unsurprising. Speaker variability contributes to f0 values for physiological reasons such as vocal fold mass and anatomy, yet for long reading procedures, the calculation of means of f0

measured away from a prosodic phrase boundary are not a particularly good metric. For example, while there are pitch resets at prosodic phrase boundaries (Oliviera 2003), the size of those resets is not predictable. Measuring f0 further into the utterance then has to incorporate prosodic resets as well as the utterance length. Further, all of the propositions that this study is based on deal with f0, not intensity or duration. Pitch is a complicated perceptual phenomenon with facets that reach into these other domains. Given the research question, they should be investigated. However, intensity and duration do not share the same downdrift that f0 does, and so there is dramatically less experimental and theoretical evidence that suggests they will exhibit differences. The perception of prominence is critical in motivating the hypothesis, and the clearest indicator of prominence is f0 (Vainio & Järvikivi 2006).

As mentioned in the description of the procedure for LR data, linear regression models built on f0 data combine downtrend and speaker variability into a single response. The variation of production from speaker to speaker and utterance to utterance is similar to the variation of formant structure, except that f0 contours are considerably more ballistic over short (~200ms) time frames. That is, formants tend to hold a stable value throughout the token

(using a long-window analysis) while f0 will rapidly change within the same token. That stability makes statistical tests more straightforward and easier to model; rapidly changing responses in the context of a global f0 level and declination are too susceptible to their environment and variation to be a reliable response. The LR data eliminates this statistical noise and unaccountable variation, producing a more direct measure of accent.

The difference in means between transitive and intransitive subjects certainly seems small at 2-3 Hz. Given the small effect size, it could be argued that the transformation of the response via the linear regression model introduced the statistical significance. However, 2-3 Hz difference over a 200-300ms range (typical for most vowel nuclei) should be compared with the general rate of declination, which is dependent on utterance length (Maeda 1976, Pierrehumber 1979), but is generally estimated at ~20 Hz/s.

If speakers tend to baseline declination in making prominence judgments, then it follows that perception is sensitive to small changes in f0 (c.f. Gussenhoven, et al 1997). A 2-3 Hz difference[11] in this context is significant, since it pushes down a hearer's model of the baseline. That is, the LR data

---

[11] Although 3 Hz might seem a small difference, it is enough to be perceptually relevant at the typical frequency range of f0 (100-250 Hz) (Roederer 1973)

measures a mean response for a region, and the greater the magnitude of a

negative mean LR response, the more the overall model has been shifted.

According to the hypothesis, the advantage of lower pitched subjects is that

they help with a later accent on a new object, should there be one.

2.1 Benefits and limitations of read narrative procedure

The results for accenting of new information support narrative reading as

a valid experimental design for intonation research. While not spontaneous, the

most significant benefit of the procedure is the standardization of embedded

contexts, which control the discourse model. Considering the diversity of

possible effects on f0, it only makes sense to limit these factors as much as

possible by eliciting the most directly comparable tokens possible. When the

narrative procedure reproduced accenting for new referents, it was clear that

narratives were invoking processors' discourse model, which was a primary

goal in creating stimuli. Otherwise, isolated clauses would have been sufficient.

The possibilities for isolated sentences producing the observed transitive

lowering effect are not explored here, but would be a useful extension of this

work. Especially interesting would be if there is a difference in definite vs.

indefinite object phrases in the absence of a discourse model. Initially, in

considering this hypothetical, I believed that there would be no difference. But

having produced results indicating a grammaticalization that imbues subjects with information about possible objects, I would not be surprised to see articles project that same type of information to the phrases that they govern.

One of the results I expected to occur alongside unaccentuation was the shortening of given information, based on the previous work of Fowler (1988). With my definition of "given" requiring repeated mention of a referent, I expected that grammatical subjects in the target sentence would exhibit shortening, as they were previously mentioned in the first sentence in the narrative.

Yet, this was not the case. There was no statistically significant effect for the length of first vs. second mentions for referents that served as grammatical subjects in the target sentence. As this undermines the validity of the read narrative design by failing to replicate an observed phenomenon, it is worth addressing.

The most straightforward possibility is that narratives fail to trigger shortening of referents in the subject position. That is, if there is an underlying process of shortening, it is not reliably triggered using passages for stimuli. This is a different result than found by Fowler (1988), who found shortening when participants read narratives, but not when they read lists in which some words were repeated.

3. Concluding Remarks

In this study, I have demonstrated with a high degree of confidence that English speakers produce grammatical subjects differently based on whether or not an object follows the verb. This investigation was based on 4 propositions, each buttressed by an existing corpus of research and observation, which then led to a testable hypothesis. Using relatively novel procedures for eliciting and interpreting tokens, I showed a clear preference for subjects of transitive verbs to have lower mean pitch relative to a baseline than subjects of intransitive verbs.

One hypothesis that was left unexplored relates to the perception of accent. It remains an open question whether hearers tend to early subject accent in projecting a syntactic parse. The use of fine-grained phonetic detail and prosody by hearers has been established in online parsing and referent resolution, but the results here suggest an opportunity for researchers to explore the combination further. For instance, do listeners expect an object (and thus a transitive parse) when they hear a particularly low subject? Furthermore, transitive lowering's perceptual effect is inexact. The magnitude of the impact on perception could be small, large, or non-existent, and would be a very fruitful line of further inquiry.

Regarding S's, it would be useful to know how accents on a new S (although rare in English) compared with accents on new O (and vice versa). It would strengthen the ergative patterning conclusion if the accent on S was somehow less than accents on O, and greatly strengthen it if new S's were not any more accented than given S's.

Another goal for this dissertation was to connect traditionally separate areas of linguistic inquiry. The grammaticalization of transitive lowering illustrates that. That the production of subjects would be systematically affected by the syntactic structure of the verb phrase following it is not immediately apparent for English and its highly grammaticized category of subject.

Ultimately, I had one overarching goal in this experiment, and that was to demonstrate that it is worth exploring the subtle ways that syntax may impact phonetics. Previously this had only been noticed in terms of prosodic phrasing, and the study of accent was confined solely to the domain of information structure. This study expands the boundaries the understanding of intonation, illustrating how speakers plan their talk in such a way that maximizes the efficacy of other phonetic processes.

Bibliography

Allerton, D. J, and A. Cruttenden. "Three reasons for accenting a definite subject." *Journal of Linguistics* 15.1 (1979): 49-53. Print.

Arnold, Jennifer, M. Fagnano, and Michael K Tanenhaus. "Disfluencies signal theee, um, new information." *Journal of Psycholinguistic Research* 32 (2003): 25-36. Print.

Arnold, Jennifer et al. "The old and, theee, uh, new: Disfluency and reference resolution." *Psychological Science* 9 (2004): 578-582. Print.

Avesani, Cinzia, and Mario Vayra. "Accenting, deaccenting, and information structure in Italian dialogue." *6th SIGdial Workshop on Discourse and Dialogue*. Ed. Laila Dybkjaer & Wolfgang Minker. 2005. 19-24. Print.

Aylett, Matthew, and Alice Turk. "The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech." *Language and Speech* 47.1 (2004): 31-56. Print.

Aylett, Matthew Peter. "Stochastic Suprasegmentals: Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech." 2000: n. pag. Print.

Bard, E. G, and M. P Aylett. "The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech." *Proceedings of ICPhS-99.* 1999. Print.

Barr, Dale J, and Boaz Keysar. "Anchoring comprehension in linguistic precedents." *Journal of Memory and Language* 46 (2002): 391-418. Print.

Baumann, Stefan, and Martine Grice. "The intonation of accessibility." *Journal of Pragmatics* 38 (2006): 1636-1657. Print.

Berman, Arlene, and Michael Szamosi. "Observations on sentential stress." *Language* 48.2 (1972): 304-325. Print.

Blodgett, Allison Ruth. "The interaction of prosodic phrasing, verb bias, and plausibility during spoken sentence comprehension." 2004: n. pag. Print.

Bock, J. Kathryn, and Joanne R Mazzella. "Intonational marking of given and new information: Some consequences for comprehension." *Memory and Cognition* 11.1 (1983): 64-76. Print.

Boland, Julie E, and Allison Blodgett. "Understanding the constraints on syntactic generation: lexical bias and discourse congruency effects on eye movements." *Journal of Memory and Language* 45 (2001): 391-411. Print.

Bolinger, Dwight L. "A theory of pitch accent in English." *Word* 14.2-3 (1958): 109-149. Print.

---. "Accent is predictable (if you're a mind-reader)." *Language* 48.3 (1972): 633-44. Print.

---. "Contrastive accent and contrastive stress." *Language* 37.1 (1961): 83-96. Print.

Breckenridge, J. *Declination as a phonological process.* Bell System Technical Memorandum, 1977. Print.

Bresnan, Joan W. "Sentence stress and syntactic transformations." *Language* 47.2 (1971): 257-281. Print.

---. "Stress and syntax: A reply." *Language* 48.2 (1972): 326-342. Print.

Brown, Gillian. "Prosodic Structure and the Given/New Distinction." *Prosody: Models and Measurements*. Ed. A. Cutler & D. R Ladd. Berlin: Springer-Verlag, 1983. 67-77. Print.

Chafe, Wallace. "Inferring Identifiability and Accessibility." *Reference and Referent Accessibility*. Ed. Thorstein Fretheim & Jeanette K Gundel. John Benjamins, 1996. 37-46. Print.

Chomsky, Noam, and Morris Halle. *The Sound Pattern of English*. New York: Harper & Row, 1968. Print.

Cruttenden, Alan. "The de-accenting of given information: A cognitive universal?." *Pragmatic Organization of Discourse in the Languages of Europe*. Ed. Giuliano Bernini & Marcia L Schwartz. Mouton de Gruyter, 2006. 311-355. Print.

Dahl, Östen. "Animacy and Egophoricity: Grammar, Ontology and phylogeny." *Lingua* 118.2 (2008): 141-150. Print.

Dixon, Robert. "Ergativity." *Language* 55 (1979): 59-138. Print.

DuBois, John. "The discourse basis of ergativity." *Language* 63 (1987): 805-855. Print.

Everett, Caleb. "A reconsideration of the motivations for preferred argument structure." *Studies in Language* 33.1 (2009): 1-24. Print.

Foraker, Stephani, and Brian McElree. "The role of prominence in pronoun resolution: Active versus passive representations." *Journal of Memory and Language* 56 (2007): 357-383. Print.

Fowler, Carol A. "Differential shortening of repeated content words produced in various commnicative contexts." *Language and Speech* 31.4 (1988): 307-319. Print.

Fowler, Carol A, and Jonathan Housum. "Talkers' signaling of "New" and "Old" words in speech and listeners' perception and use of the distinction." *Journal of Memory and Language* 26.489-504 (1987): n. pag. Print.

Fowler, Carol A, Elena T Levy, and Julie M Brown. "Reductions of spoken words in certain discourse contexts." *Journal of Memory and Language* 37 (1997): 24-40. Print.

Gahl, Susanne, Dan Jurafsky, and Douglas Roland. "Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons." *Behavior Research Methods, Instrumens, and Computers* 36.3 (2004): 432-443. Print.

German, James, Janet Pierrehumbert, and Stefan Kaufmann. "Evidence for phonological constraints on nuclear accent placement." *Language* 82.1 (2006): 151-168. Print.

Ghomeshi, Jita et al. "Contrastive Focus Reduplication in English (The Salad-Salad Paper)." *Natural Language and Linguistic Theory* 22.2 (2004): 307-357. Print.

Gussenhoven, Carlos. "Intonation and interpretation: phonetics and phonology." *Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, 2002. 47-57. Print.

Gussenhoven, Carlos, and A. C.M Reitveld. "Fundamental frequency declination in Dutch: testing three hypotheses." *Journal of Phonetics* 16 (1988): 355-369. Print.

Gussenhoven, Carlos et al. "The perceptual prominence of fundamental frequency peaks." *Journal of the Acoustical Society of America* 102.5 (1997): 3009-3022. Print.

Hawkins, Sarah. "Roles and representations of systematic fine phonetic detail in speech understanding." *Journal of Phonetics* 31 (2003): 373-405. Print.

Hawkins, Sarah, and Paul Warren. "Factors affecting the given-new distinction in speech." *Proceedings of the XIIth International Congress of Phonetic Sciences*. Aix en Provence, 1991. 66-69. Print.

Hirotani, Masako, Lyn Frazier, and Keith Rayner. "Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements." *Journal of Memory and Language* 54.425-443 (2006): n. pag. Print.

Hopper, Paul J, and Sandra A Thompson. "Transitivity in grammar and discourse." *Language* 56.2 (1980): 251-299. Print.

Horne, Merle. "Accentual patterning in 'new' vs 'given' subjects in English." *Lund University Working Papers* 36 (1990): 81-97. Print.

House, Jill. "Constructing a context with intonation." *Journal of Pragmatics* 38 (2006): 1542-1558. Print.

Kärkkäinen, Elise. "Preferred argument structure and subject role in American English conversational discourse." *Journal of Pragmatics* 25 (1996): 675-701. Print.

Knight, Rachel-Anne, and Francis Nolan. "The effect of pitch span on intonational plateaux." *Journal of the International Phonetic Association* 35.1 (2006): 21-38. Print.

Kumagai, Yoshiharu. "Information management in intransitive subjects: Some implications for the Preferred Argument Structure Theory." *Journal of Pragmatics* 38 (2006): 670-694. Print.

Ladd, D. Robert, and Rachel Morton. "The perception of intonational emphasis: continuous or categorical?." *Journal of Phonetics* 25 (1997): 313-342. Print.

Lakoff, George. "The global nature of the nuclear stress rule." *Language* 48.2 (1972): 285-303. Print.

Lambrecht, Knud. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents.* Cambridge: Cambridge University Press, 1994. Print.

Liberman, Mark, and Janet Pierrehumbert. "Intonational invariance under changes in pitch range and length." *Language Sound Structure.* Ed. Mark Aronoff & Richard T Oehrle. Cambridge, MA: MIT Press, 1984. 157-233. Print.

Maeda, Shinji. "A characterization of American English Intonation." 1976: n. pag. Print.

Mithun, Marianne. "Prosodic Cues to Accessibility." *Reference and Referent Accessibility.* Ed. Thorstein Fretheim & Jeanette K Gundel. Amsterdam: John Benjamins, 1996. 223-233. Print.

Noteboom, S. G, and J. G Kruyt. "Accents, focus distribution, and the perceived distribution of given and new information: An experiment." *Journal of the Acoustical Society of America* 82.5 (1987): 1512-1524. Print.

Noteboom, S. G, and J. M.B Terken. "What makes speakers omit pitch accents?." *Phonetica* 39 (1982): 317-336. Print.

Oliveira, Miguel. "Prosodic Features in Spontaneous Narratives." Nov. 2000: n. pag. Print.

Pierrehumbert, Janet. "The perception of fundamental frequency declination." *Journal of the Acoustical Society of America* 66.2 (1979): 363-369. Print.

---. "The Phonology and Phonetics of English Intonation." Sept. 1980: n. pag. Print.

---. "Tonal elements and their alignment." *Prosody: Theory and Experiment.* Ed. Merle Horne. Kluwer Academic Publishers, 2000. 1-36. Print.

Pierrehumbert, Janet, and Julia Hirschberg. "The Meaning of Intonational contours in the Interpretation of Discourse." *Intentions in Communication.* Ed.

P. Cohen, J. Morgan, & M. Pollack. Cambridge, MA: MIT Press, 1990. 271-311. Print.

Prince, Ellen. "Toward a taxonomy of given–new information." *Radical Pragmatics*. Ed. Peter Cole. New York: Academic Press, 1981. 223-256. Print.

Rebecca Herman, Mary Beckman, and Kiyoshi Honda. "Subglottal Pressure and Final Lowering in English." *Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings*. 1996. 145-148. Print.

Remijsen, Bert, and Vincent J van Heuven. "On the categorical nature of intonational contrasts: An experiment on boundary tones in Dutch." *The Phonological Spectrum*. Ed. Jeroen van de Weijer, Vincent J van Heuven, & Harry van der Hulst. John Benjamins, 2003. 225-246. Print.

Roederer, J.G. *Physics and psycophysics of music*. 1973. Print.

Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. "Frequency of Basic English Grammatical Structures: A Corpus Analysis." *Journal of Memory and Language* 57.3 (2007): 348-379. Print.

Schafer, Amy et al. "Focus and the interpretation of pitch accent: disambiguating embedded questions." *Language and Speech* 4.1 (2000): 75-105. Print.

Schafer, Amy J. "Prosodic Parsing: The Role of Prosody in Sentence Comprehension." Sept. 1997: n. pag. Print.

Schafer, Amy J et al. "Intonational disambiguation in sentence production and comprehension." *Journal of Psycholinguistic Research* 29.2 (2000): 169-182. Print.

Schwartzchild, Roger. "Givenness, AvoidF, and other constraints on the placement of accent." *Natural Language Semantics* 7 (1999): 141-177. Print.

Selkirk, Elisabeth. "Contrastive FOCUS vs. presentational focus: Prosodic evidence from right-node raising in English." *Speech Prosody 2002: Proceedings*

of the 1st International Conference on Speech Prosody. Aix-en-Provence, 2002. 643-646. Print.

Snedeker, Jesse, and John Trueswell. "Using prosody to avoid ambiguity: Effects of speaker awareness and referential context." *Journal of Memory and Language* 48 (2003): 103-130. Print.

Steedman, Mark. "Information Structure and the syntax-phonology interface." *Linguistic Inquiry* 31.4 (2000): 649-689. Print.

Strik, Helmer, and Loe Boves. "Downtrend in F0 and Psb." *Journal of Phonetics* 23 (1995): 203–220. Print.

Terken, J., and S. G Noteboom. "Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information." *Language and Cognitive Processes* 2.3/4 (1987): 145-163. Print.

Terken, Jacques. "Fundamental frequency and perceived prominence of accented syllables." *Journal of the Acoustical Society of America* 89.4 (1991): 1768-1776. Print.

Terken, Jacques, and Dik Hermes. "The perception of prosodic prominence." *Prosody: Theory and Experiment*. Ed. Merle Horne. Kluwer Academic Publishers, 2000. 89-127. Print.

Terken, Jacques, and Julia Hirschberg. "Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position." *Language and Speech* 37.2 (1994): 125-145. Print.

Trager, George L, and Jr. Smith. "An outline of English structure." *Studies in Linguistics: Occasional Papers* 3 (1957): 1-91. Print.

Turk, Alice, and James R Sawusch. "The domain of accentual lengthening in American English." *Journal of Phonetics* 25 (1997): 25-41. Print.

Umbach, Carla. "(De)accenting definite descripitons." *Theoretical Linguistics* 27 (2001): 251-280. Print.

Watson, Duane, Mara Breen, and Edward Gibson. "The role of syntactic obligatoriness in the production of intonational boundaries." *Journal of Experimental Psychology* 32.5 (2006): 1045-1056. Print.

Watson, Duane, and Edward Gibson. "Intonational phrasing and constituency in language production and comprehension." *Studia Linguistica* 59.2/3 (2005): 279-300. Print.

---. "The relationship between intonational phrasing and syntactic structure in language production." *Language and Cognitive Processes* 19.6 (2004): 713-755. Print.

Weber, Andre, Martine Grice, and Matthew W Crocer. "The role of prosody in the interpretation of structureal ambiguities: A study of anticipatory eye movements." *Cognition* 99 (2006): B63-B72. Print.

Yaeger-Dror, Malcah. "Register as a variable in prosodic analysis: The case of the English negative." *Speech Communication* 19 (1996): 39-60. Print.

Yuan, Jiahong, and Mark Liberman. "F0 declination in English and Mandarin broadcast news speech." *Proceedings of Interspeech 2010.* 2010. Print.