PROCESSING IMPLIED MEANING THROUGH CONTRASTIVE PROSODY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

DECEMBER 2010

By

Heeyeon Yoon Dennison

Dissertation Committee:

Amy J. Schafer, Chairperson
Victoria B. Anderson
Benjamin K. Bergen
William O’Grady
Richard Schmidt
© 2010

by

Heeyeon Yoon Dennison
To my family on both sides of the Pacific
ACKNOWLEDGEMENTS

This dissertation would not exist without the tremendous support and encouragement from organizations, colleagues, friends, and family from both sides of the Pacific.

Foremost, I am grateful to the faculty and staff of the Department of Linguistics at the University of Hawai‘i, who have supplied me with many wonderful opportunities to learn and teach. As much as it has been a privilege to work and live in the beauty of Mānoa valley, it has been a blessing to be a part of the UH ‘ohana. Thanks to the East-West Center for the invaluable support of the Degree Fellowship and enabling me to live and study as part of a multinational body of student representatives. Likewise, my research could never have come this far without the support of the NSF’s Doctoral Dissertation Improvement Grant (BCS-0921696). Additionally, many thanks to the following organizations for their contributions: the Center for Korean Studies, the Arts and Sciences Advisory Council, the Graduate Division, the Graduate Student Organization at the University of Hawai‘i, the Cognitive Science Society, and the Department of Cognitive Science at the University of California at San Diego.

Among the members of my committee, my deepest appreciation goes to my advisor and committee chair, Amy J. Schafer. Through your patient and perceptive mentorship, you provided more than just guidance. The knowledge and skills you have instilled in me have given me the confidence to pursue any academic challenge. Your integrity and your passion for empirical discovery and unbiased academic rigor inspire me to pay it forward.

Without Victoria B. Anderson’s infectious enthusiasm for intonation phonology, I would be at a loss to find a native speaker of English who could endure my countless questions, let alone the hours of contour recordings. Thank you for your positive encouragement, unwavering optimism, and uncanny ability to inject fun into everything (e.g., the ASA conference).

Thank you to Benjamin K. Bergen for consistently providing challenging yet invaluable feedback served with a healthy dose of wit. Thank you also for your part in ridding the world of thesis gnomes. Personally, I think they go best with kimchi.
When I first met William O’Grady, he quoted Emerson: “Every mind has a choice between truth and repose.” It’s then only fitting that he’s always been an influential role model; despite juggling an incredible amount of work each day as a department chair, a teacher, and a researcher, among his other roles, he always provided me with insightful and timely feedback. Thank you for always being there.

Also, my sincere thanks to Dr. Richard Schmidt for his contributions on my dissertation defense. I can’t help but think of you whenever hearing airport announcements—“Honolulu airport IS a smoke-free airport.”

Kudos to all the department and college staff members working behind the scenes who provided me with excellent assistance. Jennifer Kanda and Nora Lum’s efforts kept all logistics and processes running smoothly. Thanks to John Kawahara’s help for grant-related business and Laurie Durand for her editing services. I owe significant thanks to our lab technician, Kurt Brunner, who worked tirelessly to help develop a customized eye-tracking data analysis program. Without your patience and expertise, I would still be manually and painfully digging through fixation data.

While all the academic support kept me moving, I would be lost without the emotional support from family and friends that kept me grounded. Words alone cannot adequately express how grateful I am for the unconditional love, trust, and encouragement my parents Young-Moon Yoon and Sung-Ja Lee have given throughout my academic career. It could not have been easy to cheer me on from so far away, but you did so with patience and grace. 엄마 아빠, 정말 감사합니다. An enormous thank you to Eun-Kyung Yoon and Jae-U Kwon for being my stalwarts throughout this journey. Our family is blessed to have you both. And to my guardian angels Tae-Wook and Tae-hwan, I look forward to making up for the many precious moments I missed. To Ken and Phyllis, and the whole Dennison family, I am always in debt to you. Your many cards and letters helped keep my spirits high.

My heart-felt thanks also go to my lifelong buddies Eunhye Kim, Hyojin Kim, Min-Jung Kim, and Tae-Eun Kim. I am lucky to have evergreen friends like you. Sincere gratitude to Kwan-Hee Lee, Jong-Sook Lee, and Dr. Kyung-Sung Kim for their warm and persisting encouragement. My dear friends, Dr. Dena Beeghly, Dr. Martha Drobnak,
Dr. Hyoejin Yoon, Dr. Garry Molholt, and Walter and Kay Hipple, thank you all for believing in me and inspiring me to pursue a Ph.D. Mahalo to the kindest and most generous officemates, Kent Sakoda, Sumi Chang, and Yeonhee Yoon. Thanks to Sooyoung Hwang and Sorin Huh for being such warm and thoughtful friends, and to Paul Seokhoon Choi for the much needed and endless laughter. Much appreciation also goes to Hale Mānoans Sompon Naklang, Sawako Kinjo, Isidro Villaflor, Aileen, Francis, Livea, Yue, and Nawon Kim (to name a few) for keeping me sane over many chats and meals together. Thanks to both Chiara Logli and Yumi Matsumoto for sharing your wisdom, strength, and free spirits.

Of course, life would not be the same in Hawai‘i without the K group to help me escape from it all. A deep thanks to Jin-Sook Kim, Soyoung Kim, Jin Sun Choe, Onsoon Lee, Sang-Gu Kang, Chae Eun Kim, Cheng Eun Shim, Yoonsang Song, Kum Jeong Joo, and Sunju Kim. Likewise, xiexie to Nian Liu and Hao Zhang who shared so many uplifting stories (including my favorite one about the van). Adora is very lucky to have parents like you. Domo arigoto Manami Sato for showing me fun ways to study and relax. Thanks to Hyekyung Hwang for the many conversations and warm-hearted advice. Vielen Dank, Bodo Winter, for lending your thoughts in some of my toughest times and for sharing your endless love for cafes. Thank you to so many wonderful people who have influenced me along the way: Apay Tang, Kaori Ueki, Toshiaki Furukawa, Yumiko Enyo, Yukie Hara, Karen Huang, Hiroko Sato, Antonio Cheung, Elaine Lau, Hye-Young Kwak, Jung Hee Kim, Suann Robinson (as well as Archer and Shu), Hunter Hatfield, Megan Joelle Kirtley, James Grama, Jawee Perla, Carl Polley, and Mari Miyao.

Above all, my deepest gratitude is reserved for Matthew Dennison, who patiently held my hand through all the ups and downs along this journey. Although English is not my native language, your love and support is never lost in translation.
Understanding implicature—something meant, implied, or suggested distinct from what is said—is paramount for successful human communication. Yet, it is unclear how our cognitive abilities fill in gaps of unspecified information. This study presents three distinct sets of experiments investigating how people understand implied contrasts conveyed through prosody, which includes the accentual pattern of the sentence, known as intonation. Despite long-standing claims that intonation can convey contrastive meaning, empirical evidence for how it does so—especially at the tune level—remains scarce. An example is *The cookie jar WAS full…*, where a strong contrastive accent on *WAS* is followed by a rising tone at the end. This tune suggests that the jar is no longer full. However, to what extent do people actually perceive such meaning, and if so why? Moreover, what are the cognitive processes of interpreting such cues and meanings that arise beyond the word level meaning?

The Sentence Continuation studies show that adult native English listeners often infer both contrast and implicature from this example above (i.e., the jar is now empty) and that these meanings reflect an additive effect of sub-tonal elements in the tune. The Picture Naming studies reveal that the implicature is created approximately two seconds after the sentence offset. However, this latency does not reflect a simple two-step sequential processing where implicature is delayed due to the initial computation of the literal assertion. The results instead indicate that listeners initially access alternative meanings such as a full jar and an empty jar, and then select one dominant implication. Importantly, the degree of competition between multiple meanings depends on the nature and accessibility of the alternatives. The Visual World Eye-Tracking studies show that when alternatives are subtly represented only as background information, listeners initially attend to the asserted meaning before building an implicature. However, when sentences contain contradictory predicates that easily activate binary alternatives, an implied meaning is constructed from these competing alternatives. Together, the findings impact our understanding of how the mind maps information from tune, text, experiential knowledge, and discourse context to generate implicatures.
# TABLE OF CONTENTS

Acknowledgements  
Abstract  
List of Tables  
List of Figures

Chapter 1. INTRODUCTION .................................................................1  
1.1 Motivation ..............................................................................1  
1.2 Autosegmental framework for English prosody .........................8

Chapter 2. EXPERIMENT 1: SENTENCE CONTINUATION .....................12  
2.1 Background ...........................................................................12  
2.2 Experiment 1A .......................................................................16  
2.2.1 Materials ...........................................................................22  
2.2.2 Participants and procedures .................................................27  
2.2.3 Data coding ........................................................................27  
2.2.4 Results ................................................................................28  
2.2.4.1 Naturalness judgment .....................................................29  
2.2.4.2 Continuation meaning ....................................................30  
2.2.4.3 Correlation between naturalness and continuation .............33  
2.2.5 Discussion ..........................................................................34  
2.3 Experiment 1B ........................................................................35  
2.3.1 Design and hypotheses ........................................................37  
2.3.2 Materials ...........................................................................38  
2.3.3 Participants and procedures .................................................41  
2.3.4 Coding ...............................................................................41  
2.3.5 Results and discussion .........................................................43  
2.3.5.1 Naturalness judgment .....................................................44  
2.3.5.2 Continuation perspective ...............................................46  
2.3.5.3 Continuation meaning ....................................................49  
2.4 General discussion of Experiment 1 ..........................................58

Chapter 3. EXPERIMENT 2: PICTURE NAMING .................................65  
3.1 Background ...........................................................................66  
3.2 Experiment 2 ..........................................................................80  
3.2.1 Design ...............................................................................80  
3.2.2 Predictions .........................................................................85
3.2.2.1 Negative neutral sentences……………………………………..85
3.2.2.2 Contrastive sentences……………………………………..86
3.2.2.3 Affirmative neutral sentences……………………………..88
3.2.3 Norming study…………………………………………………….89
3.2.4 Materials…………………………………………………………90
3.2.5 Participants and procedures……………………………………91
3.2.6 Results………………………………………………………………92
  3.2.6.1 Results from the affirmative sentence model…………..94
  3.2.6.2 Results from the contrastive sentence model……………96
  3.2.6.3 Results from the negative sentence model………………97
  3.2.6.4 Results from two image models…………………………99
  3.2.6.5 Results from the full model…………………………….100

3.3 Discussion……………………………………………………………………103

Chapter 4. EXPERIMENT 3: VISUAL SEARCH……………………………112
  4.1 Background…………………………………………………………….113
  4.2 Experiment 3…………………………………………………………117
    4.2.1 Visual materials………………………………………………….118
    4.2.2 Participants and procedures……………………………………119
    4.2.3 Sentence materials and experimental conditions………………121
    4.2.4 Predictions…………………………………………………………126
      4.2.4.1 The affirmative set……………………………………….126
      4.2.4.2 The negative set…………………………………………129
      4.2.4.3 Comparison between affirmative and negative sets……130
    4.2.5 Results………………………………………………………………131
      4.2.5.1 Mouse click results……………………………………….132
      4.2.5.2 Eye movement results…………………………………….135
        4.2.5.2.1 The onset of the signal-driven eye gazes………135
        4.2.5.2.2 Results from the affirmative set…………………139
        4.2.5.2.3 Results from the negative set………………….160
  4.3 Discussion……………………………………………………………………165

Chapter 5. CONCLUSION……………………………………………………………170
  5.1 Summary of tasks and findings………………………………………170
  5.2 Implications of the findings…………………………………………174
  5.3 Remaining questions………………………………………………….178

Appendix A: Critical sentences used in Experiments 1A, 1B, and 2……………182
Appendix B: Critical pictures used in Experiment 2……………………………183
Appendix C: Critical items and three cohorts used in Experiment 3……………186

References…………………………………………………………………………187

ix
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Three sentence conditions in Experiment 1A (Regular capital letters mark the L+H* accent whereas small capital letters indicate the H* accent)</td>
<td>17</td>
</tr>
<tr>
<td>2.2. Mean duration (in milliseconds) of each phrase in each test condition in Experiment 1A; the underlined numbers represent values from the accented items</td>
<td>25</td>
</tr>
<tr>
<td>2.3. Mean F0 minimum and maximum values (in Hz) of each phrase in each test condition in Experiment 1A</td>
<td>26</td>
</tr>
<tr>
<td>2.4. Four experimental conditions as a product of two Pitch Accents and two End Contours in Experiment 1B</td>
<td>37</td>
</tr>
<tr>
<td>2.5. Mean F0 minimum and maximum values (in Hz) of each phrase in each test condition in Experiment 1B; Δ stands for pitch excursion (F0 max – F0 min)</td>
<td>40</td>
</tr>
<tr>
<td>2.6. Planned contrasts testing the significance of the paired differences with respect to the rate at which state-contrast implicature was perceived in Experiment 1B (** significance at α = .05, • at α = .10)</td>
<td>53</td>
</tr>
<tr>
<td>3.1. Six experimental conditions crossing 3 Sentence types with 2 Image types</td>
<td>82</td>
</tr>
<tr>
<td>3.2. Participants’ mean naming latencies/standard errors/standard deviations (in milliseconds) as a function of 2 Image types and 6 ISI conditions, collected after the Affirmative Sentences</td>
<td>94</td>
</tr>
<tr>
<td>3.3. The Affirmative model’s fixed effects; the intercept shows the estimated group mean for Image 1 (i.e., mentioned-state) at the 0 ms ISI</td>
<td>95</td>
</tr>
<tr>
<td>3.4. Participants’ mean naming latencies/standard errors/standard deviations (in ms) as a function of 2 Image Types and 6 ISI conditions, collected after the Contrastive Sentences</td>
<td>96</td>
</tr>
<tr>
<td>3.5. The contrastive model’s fixed effects; the intercept value shows the estimated mean naming times for Image 1 (i.e., mentioned-state) at the 0 ms ISI</td>
<td>97</td>
</tr>
<tr>
<td>3.6. Participants’ mean naming latencies/standard errors/standard deviations (in ms) as a function of 2 Image Types and 6 ISI conditions, collected after the Negative Sentences</td>
<td>97</td>
</tr>
<tr>
<td>3.7. The Negative model’s fixed effects; the intercept value shows the estimated mean naming times for Image 1 (i.e., mentioned-state) at the 0 ms ISI</td>
<td>98</td>
</tr>
<tr>
<td>3.8. The Image 1 model’s fixed effects; the intercept value shows the estimated mean naming times for Affirmative sentences at the 0 ms ISI</td>
<td>100</td>
</tr>
</tbody>
</table>
3.9. The Image 2 model’s fixed effects; the intercept value shows the estimated mean naming times for Affirmative sentences at the 0 ms ISI.

3.10. The full model’s fixed effects; the intercept value shows the estimated mean naming times for the mentioned-state Image (Image 1) for affirmative sentences at 0 ms ISI.

4.1. Two sets of conditions comprised of three types of tunes and two types of truth for the object location in the display.

4.2. Mean duration (in ms) for test items; the bold font marks values from the accented syllables.

4.3. F0 minimum and maximum (in Hz) for test items; “a” and “b” mark F0 maximum on the accented/non-accented syllable and at the rising boundary, respectively.

4.4. Fixed effects from the four condition model in the affirmative set; the intercept estimate shows the predicted mean first-fixation latency on the TARGET in the Contrastive True condition (measurement starting from the target word onset).

4.5. Fixed effects from the two slow condition model in the affirmative set; the intercept estimate shows the predicted mean first-fixation latency on the TARGET in the Contrastive True condition (measurement starting from the target word onset).

4.6. Fixed effects from the fast condition model in the affirmative set; the intercept estimate shows the mean first-fixation latency predicted for the Emphatic True condition (measured from the target word onset).

4.7. Scale comparison.

4.8. The affirmative set model; the Contrastive True condition as the base level; Analysis window = 501–3000 ms (from the target object noun onset).

4.9. The affirmative set model; the Contrastive True condition as the base level; Analysis window = 1050–1350 ms (from the target object noun onset).

4.10. The slow set model; the Contrastive True condition as the base level; Analysis window = 1050–3000 ms (from the target object noun onset).

4.11. Fixed effects from a weighted empirical logit model analyzing the likelihood of fixating the TARGET in the negative set, as a function of experimental condition and time between 400 ms and 1200 ms from the target word onset.

4.12. First fixation latency analysis for the negative set.
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Sample pitch tracks for the test sentences in (a) Contrastive, (b) Affirmative Neutral, and (c) Negative Neutral conditions, from top to bottom</td>
<td>24</td>
</tr>
<tr>
<td>2.2. Mean naturalness score for each sentence type tested in Experiment 1A</td>
<td>29</td>
</tr>
<tr>
<td>2.3. Percentage of each meaning type generated from each test condition in Experiment 1A</td>
<td>31</td>
</tr>
<tr>
<td>2.4. Percentage of the state-contrast implicature generated in each test condition by each participant in Experiment 1A</td>
<td>32</td>
</tr>
<tr>
<td>2.5. Mean duration of each region in each test condition in Experiment 1B; the x-axis represents time in milliseconds</td>
<td>39</td>
</tr>
<tr>
<td>2.6. Mean naturalness scores on a five-point Likert scale as a product of two Pitch Accents and two End Contours in Experiment 1B; the number 5 on the y-axis indicates the highest naturalness score</td>
<td>45</td>
</tr>
<tr>
<td>2.7. An interaction between Experimental List and Pitch Accent from the participant analysis regarding sentence naturalness judgments in Experiment 1B; the asterisk indicates a marginally significant pair (p = .052)</td>
<td>46</td>
</tr>
<tr>
<td>2.8. Mean proportions of trials for which continuations were generated from the speaker, listener, or either perspective in Experiment 1B</td>
<td>46</td>
</tr>
<tr>
<td>2.9. Individual variation on Continuation Perspective; the x-axis marks all 40 participants; the y-axis marks the proportion of either speaker or listener perspective taken by each participant</td>
<td>47</td>
</tr>
<tr>
<td>2.10. Effects of two Pitch Accents and two Edge Tones on the mean percentages of the speaker perspective taken by participants in Experiment 1B</td>
<td>48</td>
</tr>
<tr>
<td>2.11. Mean percentage of three meaning types coded in each test condition in Experiment 1B; the x-axis marks test conditions, differing only in the combinations of two Pitch Accents and two End Contours</td>
<td>49</td>
</tr>
<tr>
<td>2.12. Mean percentage of continuations encoding contrast over all meanings (i.e., neutral, other, contrast) as a function of two Pitch Accents and two Edge Tones</td>
<td>50</td>
</tr>
<tr>
<td>2.13. Percentage of trials for which contrast was perceived as a function of Experimental List and End Contour in Experiment 1B; asterisks mark the significant pairs</td>
<td>51</td>
</tr>
<tr>
<td>2.14. Mean percentages of all meaning types identified in each condition of Experiment 1B</td>
<td>51</td>
</tr>
</tbody>
</table>
2.15. Mean percentage of responses indicating participants’ perception of state-contrast implicature as a function of Pitch Accents and End Contours in Experiment 1B…52

2.16. Mean percentage of responses indicating participants’ perception of the state-contrast implicature as a function the four distinct tunes in Experiment 1B………52

2.17. Percentage of trials for which individual participants perceived the state-contrast implicature in each test condition in Experiment 1B……………………………55

2.18. Percentage of trials for which each item generated the state-contrast implicature in the L+H* L-H% condition in Experiments 1A and 1B; the item marked by a black dot received the lowest naturalness mean ratings……………………………………57

3.1. Reproduction of Kaup et al.’s (2005) results on picture-response latencies (in ms) collected after negative sentence processing; the pictures in the bars are added to show the counterfactual-match versus the factual-match conditions…………………73

3.2. Mean naming times in milliseconds (y-axis) collected after the Affirmative sentences as a function of 2 Image types at each of the 6 ISIs (x-axis)…………94

3.3. Mean naming times in milliseconds (y-axis) collected after the Contrastive sentences as a function of 2 Image types at each of the 6 ISIs (x-axis)………………96

3.4. Mean naming times in milliseconds (y-axis) collected after the Negative sentences as a function of 2 Image types at each of the 6 ISIs (x-axis)………………98

3.5. Mean naming times in milliseconds (y-axis) as a function of 3 Sentence Types and 3 Image Types at each of the 6 ISIs (x-axis)……………………………………..102

3.6. Mean difference in milliseconds (i.e., naming times for Image 1 minus naming times for Image 2) for each Sentence Type at each ISI………………………………103

4.1. An example of room display…………………………………………………….119

4.2. Mean mouse click times averaged across participants (in milliseconds, measured from the sentence offset); the four bars on the left represent data from the affirmative set; the four on the right represent data from the negative set……..132

4.3. Mean fixations on the target objects collapsed over four affirmative conditions and aggregated into 50 ms bins that were time-locked at the onset of the target object noun; the value 0 on a log odds scale (y-axis) corresponds to 0.5 on a proportion scale (Johnson, 2008)……………………………………………………….136

4.4. Mean fixations on the target objects collapsed over four negative conditions…..137

4.5. Proportion of fixations to the target objects at a millisecond grain in the four affirmative conditions; time on the x-axis is aligned to the target word onset……140

4.6. Proportion of fixations to the target object averaged across participants, aggregated into 50 ms bins in four affirmative conditions………………………………………140
4.7. Proportions of fixations to the TARGET, SAME ROOM COHORT, and OTHER ROOM COHORT in the Neutral True condition.............................143

4.8. Proportions of fixations to the TARGET, SAME ROOM COHORT, and OTHER ROOM COHORT in the Emphatic True condition.............................143

4.9. Proportions of fixations to TARGET and OTHER ROOM COHORT in the Emphatic True and Neutral True conditions.................................144

4.10. Fixation proportions to the TARGET, SAME ROOM COHORT, and OTHER ROOM COHORT in the Emphatic False condition..........................145

4.11. Comparison between Emphatic True and False conditions.........................146

4.12. Proportions of fixations to TARGET, SAME ROOM COHORT, and OTHER ROOM COHORT in the Contrastive True condition.............................147

4.13. Proportion of fixations to TARGET and OTHER ROOM COHORT in the Contrastive True and Emphatic False conditions...............................148

4.14. Time course of the fixations to the TARGET objects in the four affirmative conditions; log odds scale..................................................154

4.15. Proportion of fixations to the TARGET object averaged across participants, aggregated into 50 ms bins in four negative conditions.........................160

4.16. Proportion of fixations to the TARGET and OTHER ROOM COHORT in the Contrastive True and Emphatic False conditions in the negative set.........163

4.17. Proportions of fixations to TARGET and OTHER ROOM COHORT in the Emphatic True and Neutral True conditions in the negative set...............164
1.1 Motivation

Speakers often imply more meaning than what they directly describe in the sentence. In many cultures, jokes, sarcasm, denials, or counter-assertions are delivered indirectly rather than directly. As these are common ways in which people conduct daily conversations, it seems natural that people catch messages “between the lines” without much trouble. Consequently, the majority of previous studies on implied meaning have sought to understand the phenomenon from the perspective of interpersonal psychology, where speakers and listeners calculate implied meaning logically and rationally, under some general principles about communication and conversational interactions that may or may not be shaped by cultural practices (e.g., Brown & Levinson, 1987; Grice, 1975; Wierzbicka, 2003).

However, one issue with implied meaning that is largely underexplored involves the cognitive processes that are undertaken in the mind of the listener. That is to say, the current field of linguistics and cognitive science in general has asked only minimal questions about how the mind processes a message that is not fully described in the words themselves. This lack of interest seems surprising in light of the paradox that the field faces: instances of implied meanings are ubiquitous, while the cognitive processing load seems greater for implied meanings than described meanings. Once we place ourselves in a second language environment, we immediately realize how challenging it is to understand what the speaker implies. Even possessing an impressive amount of the second language vocabulary doesn’t really help to solve this problem, because implied meanings are conveyed in numerous ways. For example, one common method that is used by English speakers (and by speakers of many other languages, such as Korean) is to change the tone of voice, in the sense that they create variations on the melodic or accentual structure of the sentence, namely intonation. This suggests that intonation structure is one information source that is often needed to understand implied meanings.

The current dissertation addresses how native English listeners comprehend a speaker’s message that is merely implied through intonation, a nonlexical cue in the
language. Its particular focus is to investigate a specific type of implied message that expresses a notion of contrast, in order to delineate the ways in which intonation elements impact meaning activation and integration processes. An example case is The cookie jar WAS full..., where a strong, contrastive accent on WAS is followed by a slightly rising contour at the end of the clause. While one of the implications drawn from this tune indicates that “the cookie jar is no longer full and it may be empty now”, the cognitive processes of interpreting such cues and meanings are underexplored. Investigating the relationship between intonation and implicature bears significance in several ways. First, intonation is one component of speech prosody, which in turn is integral to spoken communication. That is, no part of natural speech can be delivered without some type of prosody, which encompasses intrinsic acoustic features in the sound signal (e.g., pitch, tempo, loudness, rhythm, voice quality) as well as the abstract structure that organizes these features into speech strings (e.g., intonation, prosodic phrasing; see Cutler, Dahan, and van Donselaar, 1997, and Speer and Blodgett, 2006, for a detailed review). Moreover, an accumulating body of research shows that, during language processing, listeners rapidly make use of prosodic information to build their hypotheses about incoming sentence structures as well as meanings. For example, listeners subconsciously keep track of the accentual prominence pattern in the sentence to predict the information status of the upcoming words (e.g., as topic or focus; see Sections 2.1 and 4.1 for more details). This suggests that understanding the relationship between prosody and speech is a significant issue not only to linguists but also to a broad range of other scientists, for example, those who are developing automatic speech recognition and generation systems. The current dissertation contributes to the understanding of this issue by showing how listeners utilize lexical information in the segmental channel together with prosodic information in the suprasegmental channel to construct a message that is not stated anywhere in the sentence.

Second, there is a lack of empirical understanding on precisely how an intonation contour implies more meaning than what is expressed by the segmental channel, despite the long-standing interest in this topic from various subfields in linguistics, including semantics (e.g., Büring, 2005; Jackendoff, 1972; Lee, 2000, 2006, 2007), intonational
phonology and pragmatics (e.g., Gussenhoven, 1983; Liberman & Sag, 1974; Pierrehumbert & Hirschberg, 1990; Ward & Hirschberg, 1985, 1986), and psycholinguistics (e.g., Ito & Speer, 2008; Sedivy, Tanenhaus, Eberhard, Spivey-Knowlton, & Carlson, 1995; Watson, Tanenhaus, & Gunlogson, 2008; Weber, Braun, & Croker, 2006). Intuition-based theoretical research has discussed that an intonation tune that resembles a rise-fall-rise pitch movement conveys contrast or contradiction, which is detachable from the lexical contents expressed in the words (e.g., Cruttenden, 1997; Gussenhoven, 1983; Jackendoff, 1972; Lee, 2007; Liberman & Sag, 1974). Subsequently, experimental studies (e.g., Bartels & Kingston, 1994; Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006) have found that one element in intonation that is responsible for evoking contrast is the pitch accent type that creates a sharp rising movement within the accented syllable (i.e., the rising peak accent; for a model of English intonation and its notations, see Section 1.2). However, these studies have tested the role of pitch accent in only one type of tonal context, where the post-accented materials create a low-falling contour shape that indicates a completion of the asserted message. This has limited the scope of our understanding on the dynamics behind the mapping between intonation form and meaning. The current dissertation shows that the role of the rising peak accent that evokes contrast can be further differentiated when it is followed by a slightly rising tone at the end, rather than a simple falling tone. Juxtaposing different pitch accents with different boundary tones illuminates the discussion on tonal processing mechanisms, e.g., is a compositional mechanism involved, where a tune’s meaning emerges from the meanings of sub-elements in the tune, or a holistic mechanism, where a tune in its entirety conveys a distinctive meaning? (see Chapters 2 and 4).

Third, the value of investigating intonation for implied meaning is that it will enhance our understanding on how pragmatic knowledge operates during listeners’ comprehension of speakers’ implicit messages. There is a growing body of literature in psycholinguistics that addresses this issue, although the majority of the work has concentrated on how contextual information assists in resolving lexical or syntactic ambiguities created when sentences contain expressions known as logical operators or
quantifiers, whose function is “to quantify over people, objects, events, locations, and
time to provide information about which entities and what number or proportion of them
contribute to the meaning of a sentence” (Paterson, Liversedge, Rowland, & Filik, 2003,
p. 264). For example, several recent studies experimentally demonstrated that a quantifier
like some generates a secondary meaning as in some but not all, not because a
generalized rule-based system automatically creates such inference, but because the
pragmatic system takes into account contextual information to calculate the additional
inference when needed (e.g., Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006;
Noveck, 2001; Noveck & Posada, 2003). This suggests that an inferred meaning requires
processing time and effort to incorporate contextual information that is given in a
particular conversation context.

That inference making is an effortful process has been claimed similarly by
Chevallier, Noveck, Nazir, Bott, Lanzetti, and Sperber (2008), who showed that
participants perceive an enriched meaning more often in an environment where making
more cognitive effort is possible. Thus, a sentence connector like or—which normally
produces an inclusive meaning as in a or b or both—induces a secondary exclusive
meaning as in a or b but not both more often when participants are given more time to
process, or when additional attention is locally drawn to the target lexical item via a
marker of visual focus (i.e., capital letters) or vocal focus (i.e., pitch accent). These
studies together indicate that perceiving implied meaning requires more processing
resources (in terms of both time and effort) and it is possible only when context provides
additional information that justifies the need for it. Indeed, these studies discussed
implied meaning in terms of “conversational implicature,” which refers to something
meant, implied, or suggested, as distinct from what is said in a given conversational
situation (Grice, 1975). Because conversational implicature arises from literal meaning in
consideration of contextual information, researchers have assumed that it is a type of
inference that requires time to calculate (Clifton & Frazier, 1989).

However, there are other types of implied meanings that make use of contextual
information in a strikingly efficient manner. For example, Sedivy and colleagues
demonstrated that people utilize presupposed contextual information immediately when
they encounter contrastive focus markers (such as the word *only* or contrastive pitch accent), in order to establish an interpretive link between what is in focus and what is in contrast (Sedivy, 1997; Sedivy, 2002; Sedivy, 2003; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Sedivy et al., 1995). Such pragmatic linking from a focus cue in the language to the presupposed contrast sets in the context is established on-line, such that listeners use that information to constrain their referential domain that unfolds over time (Ito & Speer, 2008; Sedivy et al., 1995; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006).

Reports on processing conversational implicature, versus processing presupposition associated with contrastive focus, together suggest that a clear answer has yet to be found in terms of the precise mechanism that can uniformly explain how pragmatic knowledge is used to compute an implied meaning. The current research attempts to contribute to the discussion of this issue by investigating a case introduced earlier (e.g., *The cookie jar WAS full...*), where different elements of an intonation contour such as contrastive pitch accent and the rising end contour evoke both contrast and implicature, thus called a “state-contrast implicature” (see Section 2.2). A question that is evaluated is whether different types of processing mechanisms are needed to account for different types of implied meanings, or whether a general processing mechanism can consistently explain how different implied meanings are computed.

In psycholinguistics and cognitive psychology, one view for a general processing mechanism that has been gaining considerable experimental support is that human language processing is accomplished through dynamic and continuous neural activities that utilize multiple information sources in parallel (e.g., Friederici & Weissenborn, 2007; Marslen-Wilson & Tyler, 1990; McClelland & Elman, 1986; Schlesewsky & Bornkessel, 2003; van Berkum, Brown, & Hagoort, 1999). That is, the language processor (i.e., the human brain) does not wait until a complete linguistic expression has fully arrived (Spivey, 2007). Rather, it works with partial and incomplete information as the language input unfolds over time, in order to continuously generate and test hypotheses on the incoming structure and meaning. For word recognition, for instance, several studies have shown that each unfolding segment and phoneme in the language input (e.g., k, æ, kæ)
activates and deactivates word candidates in listeners’ minds (e.g., k → kiss, cola, candy, kæ → candy, camera, camel, etc.), until enough evidence is accumulated to determine the matching target (e.g., Marslen-Wilson, 1975; Marslen-Wilson & Tyler, 1990; McClelland & Elman, 1986). Along with these subconscious processes, a range of factors that are linguistic and nonlinguistic—such as frequency of word or co-occurring syntactic frames (MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton & Sedivy, 1995), contextual plausibility (Warren, McConnell, & Rayner, 2008), recency of word form or concept (Griffin, 2002), and pragmatic knowledge on causal inference (Rohde & Ettlinger, 2010)—exert simultaneous and yet weighted influences on which word candidates should be more strongly activated than others at a given moment. Furthermore, accentual information on a temporarily ambiguous syllable that lasts for only a few hundred milliseconds also influences people’s predictions about the choice of information category that the upcoming word exhibits (Dahan, Tanenhaus, & Chambers, 2002; Watson, Tanenhaus, & Gunlogson, 2008). Related to the predictive aspect of processing, previous research has shown that when people hear a sentence fragment like The boy will eat..., they make predictions about the incoming word based upon previous knowledge of the verb and information in the visual display, thus looking at an image of cake much more frequently than an image of other objects like a train, indicating an anticipation that something edible will be mentioned (e.g., Altmann & Kamide, 1999).

These numerous research findings indicate that language processing proceeds in a manner that maximizes the probability of successful understanding of the message at the earliest possible opportunity (e.g., McClelland & Elman, 1986; Spivey, 2007). In other words, the language processor not only makes use of locally presented linguistic cues but also actively takes in various sorts of information from the global environment (either language or physical context) to build the most plausible analysis from the language input as early as possible. Along with those processes, previous experience with each of those cues creates differential effects on the analysis at each moment. This suggests that language processing operates in a way that satisfies multiple constraints (i.e., statistical basis of cues; MacDonald & Seidenberg, 2006).
An important aspect underlying constraint-based models is the substantial emphasis on how information is learned, represented, and used, because (a) language, (b) those concepts expressed by language, and (c) the factors affecting language processing are all learned through our physical and mental interactions with the world. Thus, a meaning of a word is not a fixed thing, but rather reflects our dynamic knowledge representation of the concept associated with the word, and the world. This idea is well captured in experimental findings by Barclay, Bransford, Franks, McCarrell, and Nitsch (1974) who showed that the concept that people activate upon processing a word like piano changes depending on the type of context in which the word appears. That is, when people comprehended piano in a sentence like The man lifted the piano, people were faster and more accurate in recalling the target word when the recall cues referred to the heaviness of the piano rather than the musical nature of a piano. However, when the sentence was The man tuned the piano, the results showed the opposite pattern. It is conceivable that this kind of differential conceptual representation would not be possible if our experience with pianos is limited to only one type, which further highlights the importance of our experience-based knowledge and its representation.

Despite this emphasis on how information is learned from experience, however, most of the constraint-based models in psycholinguistics to date have not explicitly discussed a fundamental question on how this experiential nature of meaning is specified in mental representation of meaning. In other words, there has been lack of explanations on how any aspect of perceived experience can get translated into the representation of meaning. Related to this issue, there is a growing body of behavioral and brain-imaging studies showing that perceptual traces of experience appear during language comprehension. Dahan and Tanenhaus (2005), for example, offered such a study, where they recorded eye movements of participants who used a computer mouse to move the mentioned object to another location on the computer screen. Besides the referent object (e.g., snake), the display included two unrelated objects, along with a visual competitor (e.g., rope). The shape of the competitor was different from the target object’s shape at the surface level (i.e., as shown in the display), but crucially, the visual representation of the concept shared some similarities (i.e., both snakes and ropes are often seen in a coiled
shape). The results showed that, despite the zero overlap in the names of the visual competitor and the target and in their surface visual forms, participants initially looked at not just the target but also the competitor significantly more often than they looked at the unrelated distractors, which suggests that participants’ conceptual knowledge formulated during experiences with those objects left perceptual traces that can be activated during language processing about those objects. This finding suggests a possibility that perceptual traces of our experiences with objects and events may act as one type of constraint with which language processing operates. On this view, I explore (especially in Chapter 3) a hypothesis that perceptual simulation of concepts expressed in language can aid inferences that people create from those simulated concepts.

Before presenting the current study, however, I first summarize a widely cited prosody annotation system called Tones and Break Indices (ToBI) for one practical reason and one theoretical reason. That is, using formalized notions will aid effective description of nonlexical information, which otherwise requires extensive explanations. More importantly, the current study provides empirical tests for some of the claims that were put forward by the original theory that initiated the ToBI system: the autosegmental model of prosodic phonology developed by Pierrehumbert and her colleagues (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980).

1.2 Autosegmental framework for English prosody

One of the essential claims behind the autosegmental model of prosody (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980) is the notion that tonal structure constitutes its own information string or channel that is separate from the segmental channel. A growing body of experimental evidence supports this conception by showing that prosodic information is processed separately from or in parallel with segmental information. In addition, the autosegmental prosodic model envisions two structures for prosody, each of which comes with distinctive characteristics, and yet their relationship is tightly linked. On the one hand, tonal structure deals with the distributed pattern of prominence signals over an utterance, where two types of tonal elements, $H$ (high tone) and $L$ (low tone), are linearly structured to represent the relative salience of words and
phrases within the utterance. On the other hand, phrasing structure deals with prosodic chunking, where words and phrases are grouped into levels of phrases that are hierarchically organized (e.g., intermediate phrase vs. intonation phrase). These two structures work in tandem, because the prominence pattern in the tonal structure often signals how prosodic phrases should be structured. The ToBI system was developed to describe these abstract structures by allowing them to be linked to the segmental string through alignment and association between the text and various tonal events (e.g., Beckman & Ayer, 1997; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Veilleux, Shattuck-Hufnagel, & Brugos, 2006). The current discussion focuses on the tonal structure while holding the phrasal structure constant, since all materials used in the current study represent the tonal pattern within one intonation phrase that consists of one intermediate phrase. Also, it focuses on the tonal structure of Mainstream American English (MAE)-ToBI, since American English is the target language for the current study.

In both Pierrehumbert’s model of English prosodic phonology and the subsequent MAE-ToBI system, the tonal structure consists of two types of prosodic events, pitch accents and edge tones, both of which make use of the two tonal primitives, H and L, to serve different purposes. On the one hand, pitch accents indicate which of the words receive prominence in the utterance. In this case, the H and L tones are associated directly with the prominent syllables, and this association is indicated via an asterisk symbol *, as in H* or L* (i.e., high peak and low valley accents). When the pitch-accented syllable carries a tonal movement within the accented syllable itself, thereby creating a dynamic tonal shape, the prominence-carrying tone is marked with * and the other tone that leads or follows the pitch movement is joined via the plus symbol, +. Hence, dynamic pitch accents such as L+H* (rising peak accent) and L*+H (scooped accent) can be represented, where the difference lies in whether the strongest prominence is heard with the H tone (as in L+H*) or the L tone (as in L*+H). Also, H+!H* exists, where the exclamation symbol ! indicates that the pitch of the associated tone is reduced (or downstepped) when compared to that of the preceding tone.
On the other hand, the H or L tones can also serve as edge tones to indicate prosodic grouping at the edge of a prosodic domain. The current MAE-ToBI posits two hierarchical prosodic domains; hence, two types of edge tones are postulated. When the H and L tones mark the edge of a smaller prosodic domain (i.e., intermediate phrase), they are called the *phrasal tones* (i.e., H- or L-). When they mark the larger prosodic domain (i.e., intonation phrase), they are called the *boundary tones* (i.e., H% or L%). Because a larger prosodic domain embeds a smaller prosodic domain by definition, the end of an intonation phrase is always marked via a sequence of stacked edge tones, creating different end contours (e.g., L-L%, L-H%, H-L%, H-H%).

This theory posits that intonation—variation in the pitch or perceived fundamental frequency (F0)—is one important cue to prominence and phrasing in the spoken language. More important, the theory predicts that the proposed tonal units (pitch accents and edge tones) bear psychological reality, in the sense that each unit is perceived categorically and signals independent functions. In consequence, the ToBI system requires a phonological transcription of the tonal structure to validate the proposed tonal inventories (Veilleux et al., 2006). In a way, this idea resembles efforts in segmental phonology where functionality (i.e., distinguishing word meanings) is considered as the core criterion for categorical or phonemic distinctions (Gussenhoven, 2004; Ladd, 1996). However, while the categorical perception of individual phonemic segments is relatively well established in segmental phonology (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Harris, Hoffman, & Griffith, 1957; Newport, 1982), the extent to which prosodic units—in particular, pitch accents like H* vs. L+H*—are perceived categorically is a matter of ongoing debate (e.g., Arvaniti & Garding, 2007; Bartels & Kingston, 1994; Dainora, 2001, 2002; Ladd & Schepman, 2003; Watson, Tanenhaus, & Gunlogson, 2008). Moreover, there is a paucity of experimental research testing the exact mapping between tonal categories and their functions (Veilleux et al. 2006), even though it has been two decades since an insightful theoretical claim was made regarding the mapping between prosodic forms and meanings (Pierrehumbert & Hirschberg, 1990).
The current dissertation aims to provide new discoveries in this regard. It uses three distinct perceptual experiments to test some of the fundamental claims made by the autosegmental theory of English intonation, namely the psychological reality of pitch accents and edge tones, as well as claims regarding the underlying mechanism that constructs intonational meaning. Chapters 2, 3, and 4 present these tests. Because each experiment investigates the topic from a completely different angle, each chapter begins with its own introduction to give the necessary theoretical background.
CHAPTER 2

EXPERIMENT 1: SENTENCE CONTINUATION

2.1 Background

Understanding language involves building a mental representation of the discourse (e.g., Barsalou, 1999; Fletcher, 1994; Garrod & Sanford, 1994; Gernsbacher, 1990; Glenberg, Meyer, & Lindem, 1987; Graesser, Singer, & Trabasso, 1994; Johnson-Laird, 1983; Zwaan & Radvansky, 1998). While the precise representational format taken by the mental discourse model differs among various theoretical frameworks (see Chapter 3 for an example model), language processing researchers generally agree that any well-constructed discourse representation will include information about the discourse participants, their relations, and the states or actions those participants take (Ferretti, McRae, & Hatherell, 2001; McRae, Hare, Ferretti, & Elman, 2001). In addition, more abstract information, such as time and perspective or viewpoint of the event, becomes part of those mental representations (e.g., Morrow & Clark, 1988; Zwaan, 1996). It is also known that such representations are not only built when understanding an entire passage or dialogue. Even when we read or hear a sentence presented in isolation, we use general knowledge to assist our understanding of the sentence meaning (e.g., Glenberg, 1997; Lakoff, 1987; Zwaan, Stanfield, & Yaxley, 2002). Thus, no “isolated” meaning representation exists, as our cognitive ability enables us to situate the meaning of even an “out-of-the-blue” sentence in plausible contexts (e.g., Fauconnier, 1997).

Recent studies in spoken language comprehension have established that, among many other cues, prosody is one important cue that guides the processes of constructing discourse models. Key findings are that accentuation itself, as well as the distribution of pitch accents over an utterance, modulates how the individual concepts are accessed and integrated into the discourse structure. For example, accentuation increases the clarity of acoustic cues (e.g., via enhanced pitch, duration, amplitude, and spectral cues), which in turn facilitates cognitive processes involved in segmental processing or lexical access. Therefore, when compared to unaccented information, accented information is processed more quickly (as evidenced for instance by a word-initial phoneme that is detected more quickly when it is accented than when it is not; e.g., Cutler & Foss, 1977; Shields,
McHugh, & Martin, 1974), and more deeply (that is, accenting a logical connector like *or* leads to more frequent activation of the so-called weak meaning, as in *A or B but not both*; Chevallier et al., 2008). Furthermore, listeners use accentual structure to identify the type or status of information in the discourse. Thus, pitch accents are often used to establish anaphoric relations among the concepts mentioned in the discourse. Dahan et al. (2002), for instance, showed that listeners consider the absence of an accent as a signal for previously mentioned or given information, whereas the presence of an accent is taken as signaling new information, or previously mentioned but nonfocal information (see Section 4.1 for a detailed review of this study). This tendency to link accentual structure to information structure in turn influences various behavioral measures. For example, when asked to judge prosodic appropriateness of sentences, listeners give high scores when new information is accented and old information is not (Birch & Clifton, 1995). Also, comprehension time is reduced when new information is accented while old information is not (Bock & Mazzella, 1983; Nooteboom & Kruyt, 1987; Terken & Nooteboom, 1987). Finally, recent studies have further verified that discourse representation is influenced by even the specific shape of accent that is realized on a syllable. Thus, while a plain high peak accent \( H^* \) is often perceived as signaling new information, a rising peak accent \( L+H^* \) mainly highlights contrastive information in the discourse (Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006; see Section 4.1 for a detailed review).

The previous findings summarized thus far indicate that constructing meaning representation from spoken language is tightly related to the extent to which the prosodic form is processed. Moreover, what is informative to listeners is not just the individual instances of accents (or absence of accents), but the accentual pattern as a whole, i.e., the intonation tune or contour. However, does this mean that listeners achieve a successful understanding of meaning from an intonation contour in its entirety (i.e., a holistic view)? Or, can each prosodic element such as individual pitch accents or boundary tones convey independent meanings (i.e., a compositional view)?
The goal of the current dissertation is to investigate these questions in greater detail. Several (nonexperimental) studies have suggested that certain tunes convey meanings in their entireties (e.g., Bolinger, 1958; Cruttenden, 1997; Ladd, 1980; O’Connor & Arnold, 1961). For example, Jackendoff (1972) noted that, when the intonation contour L+H* L-H% appears on some background information, the tune evokes a sense of contrast. Thus, in example (2.1) below (Jackendoff’s intonation contour as reiterated in Pierrehumbert & Hirschberg, 1990), the meaning conveyed by the utterance in (2.1b) is that “as for the beans, Fred ate them, but as for the other food, other people may have eaten it.”

(2.1) a. What about the beans? Who ate them?

\[ H^* L - L+H^* L-H% \]

b. Fred ate the beans.

Since the L+H* L-H% tune raises attention to the previously mentioned topic and suggests contrast between possible subtopics, Lee (2000, 2006, 2007) claimed that this tune in English is a linguistic device that marks a particular information category called “contrastive topic,” which divides a total sum topic into subtopic categories. Lee further claimed that, when this tune is applied to a predicate, this expresses a meaning of “contrastive predicate topic.” Referring to the tune’s function, Lee call this tune the “contrastive tune.”

(2.2) Examples of contrastive sentences (partially adapted from Lee, 2006, p. 390)

a. How did your sister do on the exam?

\[ L+H^* L-H% \]

b. She PASSED. (…but she didn’t ace it)

\[ L+H^* L-H% \]

b’. She didn’t ACE it. (…but she passed)
(2.2) shows examples of contrastive tunes, which are expressed on the predicates that form a scale of, in this case, something like: fail < pass < ace. Lee explained that the L+H* L-H% tune on the predicate evokes alternatives of the accented predicate, whose activation strength is determined by either semantics (i.e., lexical meaning) or pragmatics (i.e., contextual meaning). These processes then generate scalar propositions. When the tune is realized on an affirmative sentence like (2.2b), the implied proposition involves contradiction as in …but not q, where q indicates an alternative item that is positioned higher on the scale than the uttered predicate. When the target tune is realized on a negative proposition like (2.2b'), the implication is …but q, where q indicates a lower or weaker predicate. These implied propositions appear in the parentheses in (2.2).

One immediate implication from this claim is that subtle changes in the shape of the tune may eliminate or dramatically reduce the contrastive (predicate) topic meaning, since this meaning is the function of the whole contrastive tune, L+H* L-H%. However, there are two reasons to question this possibility. First of all, it is unknown how the contrastive tune is actually interpreted by people; there is no behavioral evidence on the degree to which contrastive topic is perceived from this tune. Second, there are other (theoretical) studies suggesting that contrastive meaning evoked by this tune may actually result from the combination of meanings indicated by individual prosodic units. Pierrehumbert and Hirschberg (1990), for example, attributed the contrastive function of the L+H* L-H% contour mainly to the L+H* accent, advocating a compositional hypothesis for tune meaning. Their hypothesis stated that each prosodic unit (i.e., pitch accent, phrasal tone, and boundary tone) is associated with one or more meanings, which combine to contribute to the overall tune’s meaning. Thus, the meaning of the L+H* L-H% tune was considered as the composite of meanings from the rising pitch accent (L+H*) and the sequence of a low phrase tone followed by a high boundary tone that creates a rising end contour (L-H%). On this view, the authors claimed that the L+H* is the “contrastive pitch accent,” which functions to (a) evoke a salient scale for the discourse-proper alternatives and (b) convey the speaker’s commitment to add the accented item into the interlocutors’ mutual belief space. The L-H% was called the “continuation rise,” since it signals a close link between intentions underlying the current
and subsequent utterances (regardless of the actual realization/presence of a subsequent utterance).

While Pierrehumbert and Hirschberg (1990) did not explicitly discuss what kinds of contributions this L-H% contour would make for the perception of the contrastive (topic) meaning, their theory implies that the scope of prosodically driven interpretation does not need to be an entire clause or sentence. Rather, it can be reduced to smaller phonological domains such as words and phrases that contain individual prosodic categories such as pitch accents and edge tones. Thus, this view provides a potential way to achieve an on-line mapping between prosodic form and meaning during sentence comprehension (see Section 4.1 for more background on this point).

Based on this insight, the two experiments presented in this chapter aimed to gather detailed experimental evidence on three related and yet largely unexplored questions. First, is it true that an intonation pattern such as L+H* L-H% conveys an abstract meaning that contradicts what was expressed in the lexical contents, and if so, how general is this meaning (e.g., do native English speakers agree on this)? Second, what is the exact form or nature of the meaning conveyed by this L+H* L-H% tune? That is, what kinds of contrastive meaning are evoked by this tune, and, among these contrastive meanings, are any of them dominant? If so, why? Third, does an intonational meaning reflect the meaning of the tune in its entirety, or does it emerge from meanings of individual prosodic units? How consistent is the mapping between intonation form and meaning? Is there any systematic pattern between the form of the prosodic element in the sentence and the form of contrast that is perceived?

These questions will be addressed by Experiments 1 and 2 presented in Sections 2.2 and 2.3, respectively. Then, the last section of this chapter (Section 2.4) discusses their implications.

2.2 Experiment 1A

In everyday conversation, we often hear sentences such as *The jar WAS full, The car WAS new*, or *The door WAS open*, where people emphasize the auxiliary verb in the sentence for various reasons. One reason might be to highlight the state of affairs
mentioned in the sentence by asserting, for example, that the car was indeed new, as if somebody is questioning how new it was at the previous time that is under discussion. However, another reason might be to imply that there is a change in state, e.g., the car was new at some previous point in time, but it is no longer that way (henceforth, the *state-contrast implicature*). How do listeners understand which of these meanings is intended by the speaker?

One useful cue to this problem will be the tune of the sentence. In this study, test sentences like these were prepared with the so-called contrastive tune L+H* L-H%, or an affirmative or negative form with a neutral tune (see Table 2.1), in order to address how the tune, text-to-tune mapping, and lexical choices work together to create implied meaning. The participant’s job was to listen to sentences spoken in one of these forms and create a follow-up sentence that would naturally continue the discourse (see Section 2.2.2 for more details). I hypothesized that if participants successfully created a discourse model to represent meaning that was perceived from the auditory sentence, then it should be fairly easy for them to expand or adjust the mental model and express it in a natural way. Thus, continuation sentences created by participants can serve as a window to evaluate the type of meaning constructed in their mental models, based upon the form of the sentence that they heard.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Contrastive tune</td>
<td>L+H* L-H% The pencil WAS sharp...</td>
</tr>
<tr>
<td>C2 Affirmative Neutral tune</td>
<td>H* H* L-L% The PENCIL was SHARP.</td>
</tr>
<tr>
<td>C3 Negative Neutral tune</td>
<td>H* H* H* L-L% The PENCIL was NOT SHARP.</td>
</tr>
</tbody>
</table>

The test sentences used in this experiment are extremely simple and yet possess interesting properties, which have largely been left unexplored. First, the auxiliary *was* simultaneously delivers two kinds of abstract information. On the one hand, it encodes the state of affairs taken by the subject entity. On the other hand, it indicates time of the
event. These “dual” aspects in the meaning of was seem to interact in an intriguing way when it is spoken with a contrastive pitch accent (L+H*) as in C1. This accent is known to evoke alternative or contrast sets with respect to the accented concept. On this view, an L+H* on was will evoke contrast on state (e.g., wasn’t), as well as time (e.g., is, will, etc.) because the word itself delivers both of these concepts. Furthermore, the alternative set of the L+H*-accented was might include negated versions of the auxiliary in different time frames as in wasn’t, isn’t, will not, etc. This is because was expresses information on state and time simultaneously; thus, the contrast set evoked by the L+H* accent might reflect an immediate incorporation of these dual meanings (see Chapter 4 for a direct test for this hypothesis).

The second notable property is that the proposed sentences include predicate adjectives that denote binary attributes. According to Gross, Fischer, and Miller (1989), among others, binary attributes are thought of as two ends of a scale or a pole, and depending on the type of attribute (e.g., contradictory as in alive/dead, contrary as in hot/cold, or reverse type as in tied/untied), the attribute can be gradable or nongradable. Also, it has been claimed that accessing one end of the scale of a set of binary attributes allows easy access to the opposite end (Deese, 1964, 1965; Frazier, Clifton, & Stolterfoht, 2008; Gross et al., 1989; Kaup, Lüdtke, & Zwaan, 2006; Kennedy & McNally, 2005). This suggests that a sentence with an adjective that denotes something that might be one of a pair of binary attributes may always evoke some sense of contrast, regardless of the form of intonation.

Lastly, the sentence in C1 contains the continuation rise L-H% (Pierrehumbert & Hirschberg, 1990). As mentioned above, this rising end contour has been claimed to bring a hierarchical structure into the discourse, such that it signals the dependence of the current interpretation on what was said before or what is coming next (i.e., a marker of “backward or forward reference”). If this signal is used in the absence of any surrounding context, this might evoke an inference about what could be coming next. In this sense, the function of L-H% can be compared to the role of lexical negation at the discourse level, which expresses a discrepancy in discourse expectation surrounding the target utterance.
and encourages an inference about a possible factual situation (e.g., Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999; Kaup et al., 2006; Lüdtke & Kaup, 2006).

Given these properties in the test sentences, the first research question asked whether the type of meaning evoked by each sentence will vary as the function of the sentence form, either by intonation contour (i.e., contrastive vs. neutral) or the positive/negative valency (i.e., affirmative or negative). Intuition-based research suggests that the contrastive tune L+H* L-H% in C1 should solely or dominantly evoke contrastive meaning because that is the function of the tune (Jackendoff, 1972; Lee, 2000, 2006, 2007; Liberman & Sag, 1974). Considering that the accent L+H* will evoke contrast in terms of both state and time (either separately or simultaneously), I hypothesized that the resulting contrastive meaning will be of a type that expresses a state-contrast (e.g., sharp vs. dull). Moreover, the tune includes the continuation rise L-H% that signals continuity in the discourse. On the demand of continuing the discourse in the absence of any given context (i.e., target sentences are presented in isolation), this cue will encourage participants to look for or infer a sequence of events that can be talked about (i.e., the participants will create an implicature). In this process, concepts that have been already activated in the mind will serve as useful resources, which include the mentioned concepts (e.g., was sharp) as well as the alternative concepts evoked by L+H* (e.g., wasn’t sharp, isn’t sharp, is sharp, will not be sharp, etc.). Of those alternatives, I predict that the one that fits with world knowledge will be the most frequently selected, because familiar concepts are processed more easily than unfamiliar, novel, or unknown concepts (Miller, 2003). Considering all of these factors, I predict that sentences delivered with the L+H* L-H% tune will most plausibly generate continuation sentences expressing a state-contrast implicature (e.g., The pencil was sharp → “but it is now dull”).

Negation in C3 (e.g., The pencil was not sharp) may also evoke state-contrast easily, since the function of negating one concept is to indicate the discrepancy between the presupposed expectation and a factual state. Moreover, the contrastive predicates used in the sentences might easily evoke an opposing attribute, thus allowing an easy inference for a possible factual state against the negated counterfactual state (e.g., “The pencil was in fact dull”). In the task of generating natural continuations, however, talking about a
factual state suggested from the negative sentence does not seem to be the most natural way to continue the discourse, because it seems somewhat redundant. Rather, participants might prefer to talk about some reasons for why it was not the case (e.g., “because I used it all day”), or some consequences of the negated state (e.g., “so I had to use the sharpener to sharpen it again”). This predicts that the proportion of continuations that indicate state-contrast meaning might be lower after people hear negative sentences than contrastive sentences.

As for the affirmative neutral sentences (C2), it has been shown that the neutral tune (e.g., H* L-L%) is often used to introduce new topics into the discourse (e.g., Dahan et al., 2002). Hence, after hearing these sentences, listeners may focus more on the asserted state of the entity, rather than any implied contrastive state. If this is the case, the affirmative neutral prosody condition should not result in continuations expressing any contrast. However, all sentences describe events in the past, and hearing about an event in the past, especially in isolation, might create some level of contrastive sense, perhaps because we live in present time and situate our perspective according to our current reference time. In this case, even affirmative neutral sentences might generate some continuations indicating contrast, but the probability of contrastive continuations may be much lower in this condition than in the contrastive prosody condition.

The second research question addressed in Experiment 1A is related to the exact nature or form of the meaning that indicates contrast, since contrast is a broad notion that requires further clarification. Consider the examples in (2.4), which lists some possible continuation sentences in response to the lexical string given in (2.3).

(2.3) The pencil was sharp
(2.4) a. ... but I was using it all day and now it’s dull.
    b. ... but the eraser was useless.
    c. ... but I still couldn’t poke a hole in the paper cup.
    d. ... and it was useful for filling out bubbles on the form.
The continuation in (2.4a) expresses the state-contrast implicature, where the asserted state no longer holds true. In (2.4b), however, the entity’s asserted state is agreed to (i.e., the sharpness of the pencil is accepted). What is in contrast with the (satisfactory) state of the pencil, though, is the undesirable state of another entity, the eraser; henceforth, I call this type of contrastive meaning subject contrast. The next continuation example in (2.4c) also expresses a meaning that accepts the sharp state of the pencil. However, the meaning also suggests that the sharpness of the pencil was still insufficient for completing a subsequent task. Because the meaning in this continuation presupposes a sequence of expectations with respect to the stated event (e.g., if a pencil is sharp, it should suffice to make a hole), I call this type event contrast. Finally, the last continuation in (2.4d) expresses that the pencil’s (satisfactory) state led to a desirable consequence. Although this continuation could result from emphatically accepting the asserted state of the pencil, and hence hinting at a sense of contrast (e.g., wasn’t sharp vs. WAS sharp), the continuation itself doesn’t strongly suggest contrast. Therefore, I call this type of meaning neutral.

These continuation examples illustrate that a meaning that expresses contrast can span a wide range, and the focus of contrast changes depending on the discourse representation that was built from the previous discourse. The critical question is which of those contrastive meanings is likely induced when the preceding sentence, like the one in (2.3), carries the contrastive tune L+H* L-H%. Finding an empirical answer to this question bears theoretical significance in several ways. First, all existing claims about the contrastive nature of this tune are built upon intuitive examples, and currently there is no direct evidence on the claim that this tune conveys contrast or contradiction. Second, the cognitive processes for using this prosodic cue to perceive a contrastive meaning are yet to be understood. But a more fundamental issue is that the theoretical discussions on contrast do not necessarily spell out the details of the contrastive notion in a way that makes sense for its real use and processing. Recent psycholinguistic research shares the idea of “evoking alternative sets” as the notion of contrast. However, there are hardly any discussions on more interesting questions such as what kinds of alternatives are activated for a given contrast, and which alternatives are selected, how, and why (see Sedivy, 2007,
for a similar point). The value of finding answers to these questions is not limited to the matter of language comprehension, since the notion of contrast is tightly related to much broader issues on human perception and cognition. For example, contrast and/or polarity or opposition has been claimed to be one of the basic conceptual relations that assist humans to perceive and learn about the world (e.g., Israel, 2004).

Therefore, I predict that investigating continuation sentences provided by the listeners will be useful because the continuations should potentially indicate: (a) contrast that was perceived, if any; (b) the type of contrast and justifications about the possible causes for that contrast; (c) explicit or implicit replacements of the original state with one of some alternative states, in the case where the state-contrast implicature was perceived. The aspects of meanings expressed in the continuations can then be evaluated by native English–speaking coders to seek generalizations.

Besides these theoretically motivated questions, Experiment 1A also addressed a practical question, namely the degree of naturalness of the L+H* L-H% tune when compared to the H* L-L% tune. This experiment also served as a pretest for the materials used in Experiment 2 (described in Chapter 3), which investigated the speed of implicature processing when triggered by each sentence condition presented in Table 2.1.

### 2.2.1 Materials

Each experiment list contained 33 critical sentences, 11 of which represented one of the three conditions shown in Table 2.1 above. All target sentences included a definite subject noun phrase, the past auxiliary was, and a predicate adjective, in order to describe a state of affairs of an entity in the past. Moreover, the predicate adjectives belonged to one of the three subtypes of contrastive adjectives—contradictory (e.g., sharp vs. dull), contrary (e.g., full vs. empty), and reversal (e.g., tied vs. untied)—all of which exert opposition relations (Gross et al., 1989; Israel, 2004). It was crucial that the subject entities be carefully chosen so that the sentential meaning as a whole could easily induce a binary scale that would limit the degree of gradable attributes. Some example materials

---

1 Gross et al. (1989) suggest that the contradictory type can subsume the reversal type, which mostly includes deverbal adjectives such as tied versus untied.

2 My sincere thanks to Dr. Barbara Kaup, who kindly shared her original materials prepared in German. For
include: *The car was new, The cage was locked, The necklace was latched, The candle was lit*, and *The mailbox was full* (see Appendix A for the complete set of materials).

I posited that this binary scale manipulation would be beneficial for the current study for three reasons. First, this manipulation might reduce the cognitive demand involved in accessing alternative meanings, and ease the processes of drawing an implicature about the actual state of affairs. It has been proposed that in the mental lexicon, predicate adjectives are organized around antonym pairs, each member of which is then clustered by its synonyms (Gross et al., 1989; see Chapter 3 for more on this point). Moreover, a previous study by Kaup et al. (2006) showed promising evidence that reading sentences like *The umbrella was not open* enables people to access the implied factual state of the umbrella (i.e., a closed umbrella), even in the absence of additional contextual information.² Thus, the current experimental materials may successfully induce state-contrast implicatures. The question of why binary scale might allow easy access to alternative meanings is further discussed in the General Discussion section (Section 2.4). The second reason for this binary scale manipulation was to help coders to recognize as clearly as possible meanings expressed in the continuations as selections from a plausible alternative set, since meaning assessment that is perspicuous and objective is crucial for drawing any generalizations. Lastly, this manipulation satisfied the design of the second set of the experiments, described in Chapter 3.

All sentences were recorded at a sampling rate of 11025 Hz by a female phonetician who is proficient in the MAE-ToBI system. Acoustic measurements of the duration and pitch excursion (Tables 2.2 and 2.3 below, respectively), along with the ToBI annotations by two independent transcribers, verified that the sentences were produced as intended. In particular, the hallmark features of an L+H* pitch accent (i.e., a sharp rise from the valley to the high peak during the accented syllable, and an extended duration as a result) were well captured, despite the fact that it was the first accented word in the C1 utterances. The top pitch track in Figure 2.1 presents an example of the L+H* accent, where the pitch rise happened within the accented syllable.

² My sincere thanks to Dr. Barbara Kaup, who kindly shared her original materials prepared in German. For the current study, I adapted some of those materials and added new items (these materials appear in Appendix A).
Three one-way ANOVA tests found that the mean duration of each phrase differed significantly across the conditions: subject ($F_{2}(2,64) = 10.39, p < .01$), auxiliary ($F_{2}(2,64) = 348.59, p < .01$), and predicate ($F_{2}(2,64) = 3.47, p < .05$). Overall, phrases containing pitch accents showed longer durations than the counterpart phrases in the unaccented cases. For example, the L+H* accent on *was* in the Contrastive condition (C1) contributed to the longer duration in this condition than in the unaccented auxiliary
phrases of the other two conditions (C1−C2 = .145, 3 SE = .007, p < .01; C1−C3 = .156, SE = .008, p < .01). Also, the H* accents on the predicate phrases in the Affirmative Neutral (C2) and Negative Neutral (C3) conditions resulted in longer phrase durations when compared to the nonaccented predicate in the Contrastive condition (C2−C1 = .019, SE = .008, p < .05; C3−C1 = .017, SE = .008, p < .05). However, the results on the subject phrase duration were somewhat different, i.e., the accented subject in C3 was significantly shorter than both the unaccented subject in C1 and the accented subject in C2 (C3−C1 = −.026, SE = .009; p < .01 C2−C1 = −.038, p < .01, SE = .007, p < .01). This finding is not surprising though, given that negative sentences were longer overall, due to the additional word not. The phonetics literature reports that speakers naturally reduce syllable durations when there is more material to say in a given time frame (Ladefoged, 2001). In the current materials’ recording sessions, the speaker maintained the timing and rhythm of her utterances to keep them as similar as possible across the conditions, and this attempt likely resulted in the reduction of the subject phrase duration in the negative sentences. The predicate phrase in C3 didn’t undergo the duration reduction because this is the phrase containing a nuclear pitch accent (i.e., the last pitch accent within an intermediate phrase, which is usually the most prominent in that phrase; e.g., Veilleux et al., 2006).

Table 2.2. Mean duration (in milliseconds) of each phrase in each test condition in Experiment 1A; the underlined numbers represent values from the accented items.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subject</th>
<th>Auxiliary</th>
<th>(not)</th>
<th>Predicate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 (Contrastive)</td>
<td>421</td>
<td>304</td>
<td>0</td>
<td>443</td>
<td>1168</td>
</tr>
<tr>
<td>C2 (Affirmative Neutral)</td>
<td>433</td>
<td>159</td>
<td>0</td>
<td>462</td>
<td>1054</td>
</tr>
<tr>
<td>C3 (Negative Neutral)</td>
<td>396</td>
<td>148</td>
<td>273</td>
<td>461</td>
<td>1278</td>
</tr>
</tbody>
</table>

Like the duration measure, pitch values also showed intercondition differences at each phrase. Table 2.3 presents the mean maximum and minimum pitch values. The pitch excursion value (in Hz) at each phrase (i.e., the difference between the maximum and minimum values) was used as the dependent measure for three one-way ANOVA tests: subject ($F_2(2,64) = 8.96, p < .01$), auxiliary ($F_2(2,64) = 150.67, p < .01$), and predicate

---

3 The difference scores here (and hereafter) indicate mean duration difference in seconds in the respective conditions.
Despite the fact that both the Affirmative and Negative Neutral conditions contained the H*-accent subjects, the mean excursion of the Affirmative-condition subjects resulted in greater value than those of both the unaccented subjects in the Contrastive condition (C2−C1 = 14.27,\(^4\) \(SE = 3.7, p < .01\)) and the accented subjects in the Negative condition (C2−C3 = 8.1, \(SE = 3.11, p < .05\)). This finding confirms that the subject phrases in the Negative condition underwent overall acoustic reduction. As for the Auxiliary phrases, the F0 excursion was greater in the Contrastive condition than in the other two conditions (C1−C2 = 63.33, \(SE = 4.33, p < .01\); C1−C3 = 62.86, \(SE = 3.77, p < .01\)) due to the L+H* accent being present only in the Contrastive condition. Finally, the predicate phrases in the Contrastive condition showed smaller mean excursions than those in the other two conditions (C1−C2 = −11.06, \(SE = 5.18, p < .05\); C1−C3 = 9.34, \(SE = 4.91, p = .07\)), because the Contrastive condition did not contain the H* accents but the other two conditions did. Overall, these results on acoustic measurement are in line with the phonological descriptions of the tunes from the ToBI analysis.

Table 2.3. Mean F0 minimum and maximum values (in Hz) of each phrase in each test condition in Experiment 1A.

<table>
<thead>
<tr>
<th></th>
<th>Subject</th>
<th>Auxiliary</th>
<th>(not)</th>
<th>Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>accent</td>
<td>F0 (Hz)</td>
<td>accent</td>
<td>F0 (Hz)</td>
</tr>
<tr>
<td></td>
<td>min</td>
<td>max</td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>C1</td>
<td>—</td>
<td>138</td>
<td>201</td>
<td>L+H*</td>
</tr>
<tr>
<td>C2</td>
<td>H*</td>
<td>153</td>
<td>230</td>
<td>—</td>
</tr>
<tr>
<td>C3</td>
<td>H*</td>
<td>154</td>
<td>223</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) F0 maximum value reflecting the final rise
\(^b\) F0 maximum value on the accented syllable

In addition to the test sentences described above, 20 filler sentences were prepared with various other types of non-neutral prosody (e.g., L* H-H%, where L* was placed on the subject phrase). This was done to diversify the types of intonation used in the experiment since two-thirds of the test stimuli included neutral intonation. Also, 12 of

\(^4\) The mean difference here (and hereafter) indicates pitch excursion difference in Hz in the respective comparison conditions.
the 20 fillers intentionally contained minor mistakes (e.g., hesitation, mispronunciation) to allow a test for sentence naturalness (as a pretest for Experiment 2).

2.2.2 Participants and procedures

Eighteen native English speakers at the University of Hawai‘i at Mānoa were recruited either for course credit or $5 compensation. After signing the consent form, each participant sat down in front of a computer screen, wore headphones, and received written instructions as well as four practice trials. Once the main session started, participants listened to one sentence at a time, presented in random order. After listening, they first rated the naturalness of the sentence on a Likert scale ranging from 1 to 5. A natural sentence was defined as one that an adult speaker would naturally say in normal, everyday conversations. Once participants typed the naturalness score, the sentence was replayed and participants could change the score if they wanted to. Otherwise, they typed a continuation for the sentence. They were told to do this as quickly as possible, providing whatever continuation naturally came into mind first and not worrying about producing a “fancy” continuation. Overall, the experiment was controlled by E-prime (version 1.2, Psychology Software Tools, Inc.), which collected the naturalness ratings, times that people took to make those ratings (i.e., reaction times to the first rating instances), continuation sentences, and reaction times that people took to finish those continuations. On average, the entire experiment session took approximately 20 minutes.

2.2.3 Data coding

For an objective assessment of the meanings expressed in the continuations, three other native English speakers independently coded the continuations. Coders were provided with a coding scheme that included test sentences as well as the continuations produced by the participants. How the test sentences were pronounced was also explained.

The coding scheme required that the continuation meanings should be classified into four distinct categories: (a) state-contrast, (b) other contrast, (b) neutral, and (d) other. For example, given a target sentence like *The pencil was sharp*, the coder would mark the *state-contrast* category when the meaning expressed in the continuation
contradicted the state of affairs asserted in the target sentence and expressed an alternative opposite state (e.g., “...but now it’s dull”). However, if the continuation expressed contrast with respect to other parts of the sentence such as the subject (e.g., “…but the eraser was useless”) or the entire proposition (“...but I still couldn’t poke a hole in the paper cup”), the coders would mark it as other contrast. Since this experiment was the first test for the contrastive meaning driven by the L+H*L-H% contour, the focus was to find out if the state-contrast is the most dominant contrast type, against all other types of contrast.

Besides the two contrastive types, a continuation was coded as neutral when it expressed a meaning that accepted the state of affairs asserted in the test sentence and continued it without any obvious contrast (e.g., “…and it was useful for filling out bubbles on the form”). The last category, other, included any continuations that coders thought of as ambiguous or nonsequiturs (e.g., *His beard was long → “I don’t have a beard”).

Additionally, coders were instructed to establish clear, logical delimitations for each meaning category and to be as consistent as possible. Coders were also asked to be conservative in determining contrastive meaning, and were encouraged to use the other category to mark any unclear cases. Once the coders were finished with their job, all coding charts were merged and majority decisions were determined for each trial (i.e., coding agreement from two or more coders). The high intercoder reliability, which reached 99.9%, indicated the usefulness of this coding scheme.

### 2.2.4 Results

All data analyses in Experiment 1 (A and B) were done through the following procedures. First, given that all analyses were planned for the repeated-measures ANOVA tests, raw data were first assessed with respect to two basic assumptions required by the test: normality and sphericity in the data distribution across experimental conditions and lists, using the Kolmogorov–Smirnov (K-S) test for normality and Mauchly’s test for sphericity (e.g., Baayen, 2008; Johnson, 2008; Meyers & Well,
2003). These results were not reported unless significant. When significant, data were transformed using standard methods (e.g., arcsine transformation for skewed proportion data) or degrees of freedom were corrected using the Greenhouse–Geisser correction (Field, 2009).

2.2.4.1 Naturalness judgment

The dependent measure for the first analysis was the naturalness judgment scores that participants assigned to the auditory target sentences. As explained above, participants were given an opportunity to modify their initial judgment, although they rarely changed their original decisions (i.e., changes were made in only 2% of the trials). The final judgment scores were used for analysis. Figure 2.2 below shows the mean naturalness scores collected in each condition from all items averaged across participants, since individual participant’s and item’s means fell within the range of ± 3 standard deviations (SD) from the respective group’s grand mean (i.e., mean = 4.32; SD = 0.48 for participants, SD = 0.34 for items), except one item whose grand mean was smaller than 3 SD from all items’ mean (i.e., The zipper was done: mean = 3.28). While this item could be considered as an outlier, and hence be removed from the analysis, I included it with the following two considerations. First, this item fell outside the −3 SD range in the Affirmative Neutral condition only. Second, the naturalness rating itself is a holistic measure that can be influenced by (a) the sentence’s tune, (b) lexical contents, or (c) a combination of both. Therefore, including this item for further analyses on other measures (e.g., types of continuations following this sentence) should be informative in determining the exact source of the perceived naturalness. The y-axis on Figure 2.2 represents the five points used in the Likert scale, with the number 5 indicating the highest naturalness.

An alternative analysis method is to use multinomial logistic regression models (e.g., Yaeger, 2008), which I will pursue in a follow-up report.
Figure 2.2. Mean naturalness score for each sentence type tested in Experiment 1A.

As seen in the figure, each sentence type was perceived as displaying a similar degree of naturalness, except for the filler items that contained mistakes. One-way ANOVA tests found a main effect of the sentence type ($F_1(4, 68) = 14.54, p < .01$; $F_2(4, 114) = 14.52, p < .01$), and pairwise comparisons indicated that the “mistake” fillers generated significantly lower mean scores when compared to the scores from each of the other sentence types all at $\alpha = .01$. Also, pairwise comparisons from the by-item analysis (but not by-participant analysis) found significantly lower mean values in the Contrastive condition when compared to the Affirmative Neutral condition (mean difference: $C1 - C2 = -.5, SE = .12, p < .01$). This finding perhaps reflects that the L+H* L-H% tune occurs less frequently than other neutral tunes like H* L-L% in real world corpora (e.g., analysis on the tunes in radio commentators’ speech; Dainora, 2001, 2002). However, it could be due to the difference in the end contour type. Whereas the neutral sentences used the falling tone indicating sentence finality, the contrastive sentence included the continuation rise L-H%, which suggests discourse dependence between the current utterance and preceding or subsequent one (Pierrehumbert & Hirschberg, 1990). Given that sentences were presented in isolation in the experiment, it is conceivable that the L-H% contour that generates a sense of incompleteness contributed to the lower naturalness judgment.

2.2.4.2 Continuation meaning

The next analysis focused on the types of meanings that coders identified from the continuation sentences, which were produced by participants after hearing target
sentences in three different forms. Figure 2.3 shows the percentage at which continuations indicated each of the four meaning types per test condition.

As seen in Figure 2.3, sentences that differed only in form (either by prosody or negation) generated continuations encoding different meaning types. The grid-patterned portions of the bars show that the Contrastive condition, which delivered target sentences in the L+H* L-H% tune, induced continuations expressing the state-contrast implicature significantly more often (about 63% of the trials) than both the Affirmative and Negative Neutral counterparts (7% and 3% of the trials, respectively). This difference was borne out in one-way repeated-measures ANOVA tests (with test condition as a within factor for both by-participant and by-item analyses, and experiment list as a between factor for the by-participant analysis), which considered the dependent measure as the proportion of continuations expressing the state-contrast implicature in a given condition: $F_1(1.18, 17.72) = 65.51, p < .01; F_2(1.63, 52.14) = 165.77, p < .01$. Planned contrasts showed that the mean proportion of the state-contrast implicature differed between the Contrastive condition (C1) and each of the other two conditions, all at $p < .01$: C1 vs. C2 ($F_1(1,15) = 69.2, F_2(1,32) = 178.28$); C1 vs. C3 ($F_1(1,15) = 69.6, F_2(1,32) = 223.83$). The Affirmative and Negative Neutral sentences showed marginal difference from each other in the rate at which they evoked the target meaning ($F_1(1,15) = 3.571, p = .08; F_2(1,32) = 3.09, p = .09$).

---

6 Degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\eta = .59$ for the $F_1$ analysis; $\eta = .82$ for the $F_2$ analysis), since Mauchly’s test indicated that the assumption of sphericity had been violated: $\chi^2(2) = 16.53, p < .01$ for the $F_1$ analysis; $\chi^2(2) = 8, p < .02$ for the $F_2$ analysis.
Despite this dramatic difference among sentence types, one surprising finding was that the Contrastive tune (L+H* L-H%) produced the state-contrast implicature in notably less than 100% of the trials. As seen in Figure 2.3, this tune readily generated neutral meaning for about 24% of the trials, whereas all other types of contrastive meaning (besides the state-contrast implicature) were found in only 3% of the trials. This finding challenges a previous claim that the L+H* L-H% tune is a linguistic device for encoding a contrastive (predicate) topic in English at all times (e.g., Lee, 2000, 2006, 2007). Various meanings found from the Contrastive condition indicate that conveying contrast is not the only function of the L+H* L-H% tune. This finding supports the argument that one type of intonational tune can serve multiple functions, as previously suggested by the one-to-many mapping relationship between a tune’s identity and its meanings (e.g., Gussenhoven, 2004; Ladd, 1996; Pierrehumbert & Hirschberg, 1990).

Moreover, further analyses on individual participants showed variation in performance. Figure 2.4 presents the proportion of the state-contrast implicature generated in each test condition by each of the 18 participants, where the participants on the x-axis were sorted by the percentage of instances in which they provided continuations encoding the implicature in the Contrastive condition.

Figure 2.4. Percentage of the state-contrast implicature generated in each test condition by each participant in Experiment 1A.
The figure shows that participants rarely produced continuations indicating the state-contrast implicature in the Affirmative and Negative Neutral conditions. In the Contrastive condition, 13 out of 18 participants (i.e., 72% of all participants) produced continuations expressing the target implicature on more than half of the trials (range: 60% [1 person]–100% [3 people]), for an average of 80% of the continuations. However, the remaining 5 people (marked in the grey box) produced the target meaning for only 17% of the trials (range: 0% [1 person]–45% [1 person]). Analysis of continuation response times (that were measured from the onset to the offset of typing) revealed that those 5 people spent substantially less time creating continuations in general. Whereas the 13 people spent an average of 17.4 seconds, the 5 people spent an average of only 10.4 seconds generating continuations across all test conditions. This suggests that these 5 people were perhaps less attentive to the task, or perhaps they produced continuations before they had had time to fully process the implicature. These findings together raise further questions about the extent to which implicature perception requires processing time and effort. In Chapter 3, I present a set of experiments addressing the time course of building the state-contrast implicature, using the same sentence materials with a different task.

2.2.4.3 Correlation between naturalness and continuation

A few other findings include a negative correlation between the naturalness ratings and the continuation production times. Across all participants, the higher naturalness scores the participants had given to the target sentences, the quicker they were to create follow-up sentences (across all three conditions: $r = -.596, p < .01$; C1 only: $r = -.427, p = .08$). This suggests that building a discourse representation is easier when the sentence that was just heard sounded natural. This result is complemented by another finding that people who took longer to make naturalness judgments on the target sentences also took longer to generate continuations of those sentences ($r = .567, p < .05$).
Finally, the target sentences’ naturalness per se didn’t exert much influence on the rate at which participants produced continuations for the state-contrast implicature. The only significant result was the one found from the item analysis using the 13 people’s data in the Contrastive condition, where items that received higher naturalness ratings produced more state-contrast implicature ($r = .46, p < .01$). More analysis on the individual items is discussed at the end of the Results section in Experiment 1B.

2.2.5 Discussion

In sum, Experiment 1A established empirical evidence that the L+H* L-H% tune frequently elicits contrastive continuations, particularly those expressing a state-contrast implicature. While the contrastive tune also generated other types of meanings, it evoked state-contrast implicature substantially more often than did both its affirmative and negative neutral counterparts.

One remaining question is whether the frequent perception of the target implicature in the L+H* L-H% condition is due to the contrastive pitch accent L+H*, the rising end contour, L-H%, or both of these together. While the results from Experiment 1A can only support the claim that the targeted meaning is generated by the whole tune, the existing literature presents a claim that the functional unit of intonation—the so-called “intonational morpheme”—can be as small as a single pitch accent or a boundary tone (Pierrehumbert & Hirschberg, 1990). Accordingly, recent studies have provided experimental evidence that the main function of the L+H* accent is to evoke contrast (by activating alternative sets), whereas the H* accent can convey both contrastive and noncontrastive meaning (e.g., Dahan et al., 2002; Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006). As for the edge tones, however, there is much less empirical evidence for their functions in generating contrastive meaning. Experiment 1B, presented in the following section, explored the exact function of individual prosodic elements by factorially manipulating two types of pitch accents and two types of end contours.
2.3 Experiment 1B

The main purpose of Experiment 1B was to investigate the extent to which the perception of the state-contrast implicature is systematically related to the identity of each prosodic element comprising the L+H* L-H% contour. As mentioned above, Experiment 1A’s finding that the L+H* L-H% tune often generates the state-contrast implicature could be due to the presence of the contrastive pitch accent (L+H*), the rising end contour (L-H%), or the combination of both. Paired comparisons between the L+H* L-H% tune and any of the following examples in (2.5)–(2.7) generate different intuitions about the degree to which the state-contrast implicature is perceived.

First, consider the sentence in (2.5) below, where the high boundary tone H% is replaced with a low one L%, resulting in a falling end contour L-L%.

(2.5) L+H* L-L%

The pencil WAS sharp.

Intuition suggests that this sentence emphatically affirms the state of affairs mentioned in the sentence by evoking a contrast between was and wasn’t. However, one might also infer that the current state of the pencil may be different from the previous state (was vs. isn’t: state-contrast implicature). The extent to which this tune (L+H* L-L%) generates state-contrast implicature when compared to L+H* L-H% will be informative in determining how much of the target meaning is generated by the rising boundary tone (H%) when it is preceded by a low phrase tone (L-).

Second, the perception rate for the state-contrast implicature might decrease if the contrastive pitch accent L+H* is replaced with the more common high pitch accent H*, known as the “presentational” accent, as in (2.6) below.

(2.6) H* L-H%

The pencil was sharp…
On one extreme hand, we might expect a complete elimination of the contrastive meaning when H* is used. These two accents have been claimed to make different contributions to interpretations (Pierrehumbert & Hirschberg, 1990), and accordingly, the current MAE-ToBI system lists them as two phonologically distinct pitch accents; that is, they are discussed as separate intonational morphemes or tonemes (e.g., Beckman & Ayers, 1997; Beckman et al., 2005; Veilleux et al., 2006).

However, recent research in sentence processing suggests that the high peak accent (H*) can serve multiple functions. For example, it is used to mark a new item as well as a discourse-given but nonfocal item (Dahan et al., 2002). It can also indicate contrast. Watson, Tanenhaus, and Gunlogson (2008) demonstrated that people derive contrastive interpretations not only from L+H* but also from H*, although the degree to which they do so is stronger with L+H*. From this finding, Watson, Tanenhaus, and Gunlogson (2008) claimed that the interpretive domains of these two pitch accents overlap, which further questions the category membership of these accents. These recent reports therefore indicate that sentences like (2.6) might still evoke state-contrast implicature, but to a lesser degree than sentences with L+H*. By exploring whether or not H* can induce state-contrast implicature, we can glean important information about this accent’s phonological status and functionality.

The last comparison made to an example like (2.7) below suggests that the function of the L+H* L-H% tune found in Experiment 1A could reflect an effect of that particular tune in its entirety. That is, the tune’s effect could be more than just the sum of the pitch accent’s effect and the end contour’s effect, i.e., a super-additive effect. In fact, if both of these are replaced with their counterparts (i.e., H* and L-L%, respectively), the resulting tune itself is one of the neutral tunes, which is normally used for plain statements (although the accent on the auxiliary may evoke a sense of narrow focus). Therefore, sentences like (2.7) should not create the state-contrast implicature, or should greatly reduce it.

(2.7) H* L-L%

The pencil was sharp.
By considering the meanings of these three additional tunes, Experiment 1B addressed three specific research questions: (a) Is the contrastive accent (L+H*) itself sufficient to evoke the state-contrast implicature when the continuation rise (L-H%) is absent (i.e., replaced with L-L%)? (b) Will the perception of state-contrast implicature be eliminated or reduced when L+H* is replaced with an H* accent in the tonal context of L-H%? and (c) When both L+H* and L-H% are absent, will the rate at which state-contrast is implied decrease dramatically?

2.3.1 Design and hypotheses

This experiment adopted the lexical items as well as the sentence continuation paradigm from Experiment 1A. Table 2.4 below presents four test conditions fully crossing two Pitch Accents (L+H* and H*) with two End Contours (L-H% and L-L%). Both Pitch Accents have a high prominence, meaning that the prominence on the syllable is indicated by a tone that is high compared to the speaker’s local pitch range. The two End Contours both contain a low phrase tone (L-) but they differ in terms of the boundary tone type (i.e., either H% or L%). Here, I use the term End Contour, instead of boundary tone, to indicate that the factor that is manipulated in this experiment is the sequence of a phrase tone and a boundary tone as a unit.7

Table 2.4. Four experimental conditions as a product of two Pitch Accents and two End Contours in Experiment 1B.

<table>
<thead>
<tr>
<th>Pitch Accents</th>
<th>End Contours</th>
</tr>
</thead>
<tbody>
<tr>
<td>L+H*</td>
<td>L-H%</td>
</tr>
<tr>
<td></td>
<td>L+H* L-H%</td>
</tr>
<tr>
<td></td>
<td>C1: The pencil was sharp…</td>
</tr>
<tr>
<td></td>
<td>L-L%</td>
</tr>
<tr>
<td></td>
<td>L+H* L-L%</td>
</tr>
<tr>
<td></td>
<td>C2: The pencil was sharp.</td>
</tr>
<tr>
<td>H*</td>
<td>L-H%</td>
</tr>
<tr>
<td></td>
<td>H* L-H%</td>
</tr>
<tr>
<td></td>
<td>C3: The pencil was sharp…</td>
</tr>
<tr>
<td></td>
<td>L-L%</td>
</tr>
<tr>
<td></td>
<td>H* L-L%</td>
</tr>
<tr>
<td></td>
<td>C4: The pencil was sharp.</td>
</tr>
</tbody>
</table>

If L+H* evokes alternatives by default, whereas H* mainly marks new items or less salient items in the discourse (while occasionally evoking contrast), then the state-contrast implicature will be more frequently generated when the target sentences contain L+H* than H*. In other words, I predict a main effect of the Pitch Accent type.

7 Pierrehumbert and Hirschberg (1990) claim that the functions of L-H% and L-L% are essentially those of the boundary tones. The current study does not make this assumption.
If the rising end contour (L-H%) signals a forward reference (i.e., an indication that there are more things to be said about the uttered proposition) whereas the low falling tone (L-L%) indicates an independent assertion (Pierrehumbert & Hirschberg, 1990), then the state-contrast implicature will be more frequently created with L-H% than L-L%. That is, there will be a main effect of an End Contour.

Moreover, it has been claimed that tunes themselves have important semantic and pragmatic functions (e.g., Cruttenden, 1997; Dianora, 2001, 2002; Gussenhoven, 1983; Jackendoff, 1972; Lee, 2000, 2006, 2007; Liberman & Sag, 1974; Ward & Hirschberg, 1985, 1986). Therefore, I predict a super-additive effect when the contrastive pitch accent L+H* and the rising end contour L-H% are combined. This tune should generate continuations indicating state-contrast implicature significantly more frequently than the summed amounts that are generated from the pitch accent or end contour alone.

2.3.2 Materials

All sentences were recorded at a 22050 Hz sampling rate by the same female phonetician who recorded the materials in Experiment 1A. The initial F0 of the utterances was carefully controlled to phonetically represent the mid part of the speaker’s pitch range. This was done to provide participants with a perceptual basis for the speaker’s local pitch range, which in turn should assist the identification of the first Pitch Accent type in the sentence. All test items were recorded in all four test conditions (i.e., with four different tunes), in order to obtain the most natural token for the Pitch Accent (on the auxiliary) and the End Contour (on the predicate phrase) in each sentential context. Then, those tokens that were determined to be the best representations of each prosodic category were cross-spliced across the conditions for that item. This was done to provide the same acoustic signal for the same prosodic category across all conditions to the greatest extent possible while still maintaining naturalness. For example, if the best token of L+H* was found in C1, then this token was used to replace the L+H* in C2 for that item, so that L+H* in both conditions represented the same acoustic signal.

Figure 2.5 below shows the mean duration of each region—the subject, the auxiliary, and the predicate—averaged across the items per test condition.
Two-way repeated-measures ANOVA tests found that each region duration across the conditions differed as a function of the Pitch Accents and End Contours. For all phrases, there was a main effect of Pitch Accent (Subject: $F(1,31) = 16.46, p < .01$; Auxiliary: $F(1,31) = 151.4, p < .01$; Predicate: $F(1,31) = 6.5, p < .05$), meaning that the phrases were significantly longer in the conditions containing L+H* rather than H*. The longer auxiliary duration with the L+H* accent is expected, given that speakers normally need more time to realize the rising pitch movement (from low to high) in the accented syllable (Veilleux et al., 2006). The longer duration of the subject and predicate regions when they surround an L+H* accented phrase (rather than an H* accented phrase) indicates the speaker’s strategy to use speech tempo and rhythm to control her production of the L+H* and H* accents. One challenge for recording the current test materials was to instantiate the phonetic differences required by L+H* versus H* (i.e., the presence or absence of the preceding low tone target) on the one-syllable-length word was. Moreover, the acoustic characteristics of the segment /w/ itself (i.e., the lowered first formants associated with approximants) can contribute to the perception of an initial low tone. Thus, the speaker controlled her tempo and rhythm in order for the sentences with the H*-accented was to reach the high tone target quickly but gradually from the word onset, whereas the sentences with the L+H* accents have enough time to realize the low leading tone. These aspects of production contributed to the overall longer durations for the L+H* accented sentences. Besides the effect of Pitch Accent, there was also a main effect of the End Contour for the Predicate phrase only ($F_2(1,31) = 91.52, p < .01$), which was due to the final lengthening associated with the L-H% contour.
As for the pitch values, Pitch Excursion (i.e., the difference between the maximum and minimum F0 values) was used as the dependent measure for two repeated-measures ANOVA tests (Table 2.5). The results on the subject phrase found a main effect of Pitch Accent ($F_2(1,31) = 10.27, p < .01$), where the subject phrases exhibited greater excursion when they preceded the L+H* accent than when they preceded the H* accent. This is due to the lower F0 value on the last syllable of the subject phrase when it preceded an L+H* accent rather than an H* accent. For the auxiliary phrase, there was a main effect of Pitch Accent ($F_2(1,31) = 38.39, p < .01$), as well as a marginal interaction of Pitch Accent and End Contour ($F_2(1,31) = 3.33, p = .08$). That is, the pitch excursion was greater overall when the auxiliary received an L+H* accent rather than an H*. In addition, this difference was bigger when the L+H*-accented was preceded the L-H% tone rather than L-L%, whereas the pattern was the opposite for the H*-accented was. Lastly, the results on the Predicate phrase found a marginal effect of Pitch Accent ($F_2(1,31) = 3.68, p = .06$) and a full effect of End Contour ($F_2(1,31) = 54.79, p < .01$), where the L-H% contour resulted in greater pitch excursion than L-L%.

Table 2.5. Mean F0 minimum and maximum values (in Hz) of each phrase in each test condition in Experiment 1B; Δ stands for pitch excursion (F0 max – F0 min).

<table>
<thead>
<tr>
<th>Subject F0 (Hz)</th>
<th>Auxiliary F0 (Hz)</th>
<th>Predicate Edge F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max</td>
<td>min</td>
<td>Δ</td>
</tr>
<tr>
<td>—</td>
<td>229</td>
<td>132</td>
</tr>
<tr>
<td>—</td>
<td>229</td>
<td>133</td>
</tr>
<tr>
<td>—</td>
<td>224</td>
<td>141</td>
</tr>
<tr>
<td>—</td>
<td>225</td>
<td>141</td>
</tr>
</tbody>
</table>

One experimental list consisted of 32 test items (8 tokens x 4 conditions) and 64 filler items. The majority of fillers (62.5%) contained various types of neutral tunes to balance out the marked prosody in the test conditions, though several fillers did share some properties similar to those in the test conditions. In terms of the End Contour, 44 filler sentences ended with L-L%, while 12 fillers ended with H-H% and 8 fillers with L-H%. For more variability, 20 negative-sentence fillers were also included (i.e., 21% of all materials). In this experiment, all fillers contained natural signals (i.e., no fillers included any intended mistakes).
2.3.3 Participants and procedures

Forty native English speakers at the University of Hawai‘i at Mānoa participated in this experiment and received $10 compensation. All participants were carefully screened to ensure that they were native English speakers who learned only English before age 5. The experimental procedure was identical to that of Experiment 1A, except that this experiment included more materials, and hence took longer to finish. One experimental session took 40 to 50 minutes on average.

For each session, participants listened to sentences one at a time (presented in random order) and gave a naturalness score for the sentence they heard. They then heard the sentence a second time, and were given the option of changing the naturalness score if they wanted to. After this, participants typed a sentence that they felt would naturally continue the discourse as quickly as possible. To provide flexibility in meaning perception, participants were instructed that they could continue the discourse as if they were the speaker of the previous sentence (i.e., speaker perspective), or as if they were listeners (i.e., listener perspective). Participants were also instructed to indicate if they didn’t understand the target sentence by typing something like “I don’t understand.” As in Experiment 1A, the experimental software E-prime (version 1.2, Psychology Software Tools, Inc.) collected data on naturalness ratings, reaction times for naturalness ratings, continuations, and reaction times for continuations.

2.3.4 Coding

For consistency, the same coders who evaluated the data from Experiment 1A were recruited again as coders for the present study. These three native English-speaking coders were provided with a new coding scheme that included target sentences as well as continuation sentences typed by the participants. Unlike in Experiment 1A, however, the coding chart showed target sentences without any information about how they were spoken (i.e., prosody). As a result, coders had to rely solely on the lexical items that were used in the target sentences to evaluate the meanings expressed in the corresponding continuations. This procedure challenged coders and completely ruled out the possibility that a coder might bring in any previous expectations or biases concerning the types of
meanings that they were evaluating; hence, this ensured the most conservative evaluation of the type of meaning expressed in the continuations.

The coding scheme specified two major coding categories, as well as a category indicating an error trial. A trial was identified as an error when: (a) the continuation was absent or incomplete, (b) participants explicitly indicated a difficulty in understanding the target sentences (e.g., “I don’t understand the sentence”), and (c) the continuation was grammatically well formed, but semantically irrelevant to the target sentence due to a misunderstanding of the lexical contents (e.g., *The door was ajar* → “The jar was filled with peanut butter”).

All non-error continuations were then qualified for the two major coding categories: Continuation Perspective and Continuation Meaning. First, Continuation Perspective concerned whether the meaning expressed in the continuations represented the speaker perspective (i.e., as if the continuation producer was the speaker of the auditory sentence), the listener perspective, or whether the perspective was ambiguous (i.e., either).

The second category—Continuation Meaning—include three subcategories: neutral, other, and contrast. The neutral category was used when a continuation was a natural follow-up of the test sentence without any indication of obvious contrast (e.g., *The pencil was sharp* → “and I was ready for the test”). The other category was used when the continuations were grammatically and semantically well formed but the coders were unsure of the type of meaning expressed in the continuations, or when the continuation itself expressed uncertainty (e.g., *The car was new* → “What happened to it?”).

Contrast, the last category for Continuation Meaning, included four subtypes to further clarify the form of the contrastive meaning evoked by different tunes. The first subtype represented the focus of the current study, state-contrast implicature—when the continuation expressed a change from the asserted state to an implied alternative state (e.g., *The candle was lit* → “But then the wind blew it out”). The second subtype was emphatic acceptance—when a continuation emphatically accepted the state of affairs asserted in the test sentence (e.g., *The tree was leafy* → “It was so leafy that I couldn’t
see the sun”). The third subtype expressed subject contrast—when a contrast was made with respect to the subject entity (e.g., The pencil was sharp → “but the eraser was junk”). Lastly, event contrast was used when the continuation indicated that the experiment participant formed a sequence of expectations with respect to the stated event but that one of the expectations was not fulfilled (e.g., The belt was buckled → “but his pants still fell down”).

Besides these major coding categories, coders were asked to mark if the continuation expressed any change in the dimension of time (e.g., change from past to present tense), orthogonal to other properties such as Continuation Perspective and Continuation Meaning. Intercoder reliability showed 96% agreement.

2.3.5 Results and discussion

Before the main analyses, grand mean continuation typing times (averaging across all sentence conditions) were examined to determine any outlying participants. Because each participant’s mean response times fell within 3 standard deviations of all participants’ mean, all data were included in the analyses described below. The coding summary indicated that coders identified 1.1% of the continuations in all trials (i.e., 14 out of 1,280 trials) as errors due to the reasons explained in the coding section above. Due to their irrelevance to the discussion of prosody and meaning, they were removed from the subsequent analyses. Among the 1,266 non-error trials, 51 trials failed to generate coder agreement regarding the types of meanings expressed in the continuations. An additional 13 trials were identified as not having generated coder agreement on Continuation Perspective. These trials were therefore excluded from the analyses as well, in order to allow data comparisons across different dependent measures (e.g., naturalness vs. perspective vs. meaning). These procedures resulted in the loss of a total of 6% of the data, and the remaining 1,202 trials were included in the main analyses for target sentence naturalness, as well as continuation perspective and meaning.

This section begins by presenting naturalness judgment results. Then, results on continuation perspective and meaning are presented. For the meaning part, I first focus on the analysis of contrast as a general category, followed by the analysis of the specific type
of contrast, the state-contrast implicature. Results on participants’ individual variability are also reported for the state-contrast implicature, along with correlation analyses comparing the results from various dependent measures. Finally, item analysis is provided to see differences among individual items in both Experiments 1A and 1B.

2.3.5.1 Naturalness judgment

The dependent measure was the naturalness rating scores (from 1 to 5) collected on a Likert scale. As in Experiment 1A, final judgment scores were used (as judgment change occurred only for 0.5% of all trials). The grand mean values from each participant’s and item’s data (across all conditions) fell within the range of ±3 standard deviations (SDs) from all participants’ and items’ means (mean = 3.6; SD for participants = 0.76, SD for items = 0.34), except for the one item zipper. I included all data in the analyses for the reasons explained in the discussion of Experiment 1A.

Figure 2.6 presents mean naturalness scores (across participants) collected in each of the four test conditions as a product of two Pitch Accents and two End Contours. The grand mean value averaging across all test conditions (i.e., 3.6) was lower than that in Experiment 1A (i.e., 4.3), perhaps due to the fact that all test materials in the current experiment placed pitch accents on the auxiliary, inducing a sense of narrow focus. In contrast, two thirds of the test materials in Experiment 1A placed pitch accents on the subject and predicate phrases, which is a neutral way to introduce new items into the discourse. The implication of pitch accent location on building discourse representations is further discussed in the General Discussion section of Experiment 1 (section 2.4).

A comparison of mean values in each condition indicates that participants gave slightly higher naturalness scores when the target sentences ended with a falling tone (L-L%) rather than a rising tone (L-H%). The raw data passed initial tests for data normality and distribution, and were subsequently submitted to two-way repeated-measures ANOVA tests crossing two factors (Pitch Accent and End Contour) for within-participants and within-items analyses, and Experimental List for a between-participants analysis only.
The results showed that the small differences observed in the figure were statistically reliable. There was a main effect of End Contour ($F_1(1,36) = 5.1, p < .05$, $F_2(1,31) = 4.6, p < .05$); sentences were judged more natural when they ended with a low tone than a rising tone (mean difference: (L-L%)−(L-H%) = .16, SE = .07, $p < .05$). This is likely because each test sentence was presented in isolation. As mentioned above, the low falling contour L-L% normally signals sentence finality for isolated sentences (except when it is used for wh-questions), whereas the rising contour L-H% signals continuation (i.e., the continuation rise: Pierrehumbert & Hirchberg, 1990).

As for the Pitch Accent, there was no main effect from participant or item analysis. However, a marginal interaction between Pitch Accent and Experimental List was found from the participant analysis: $F_1(3,36) = 2.8, p = .053$. Posthoc paired $t$-tests indicated that only those participants in List 3 gave higher naturalness scores for the sentences with H* than for those with L+H* ($t_1(9) = 2.24, p = .052$). As seen in Figure 2.7, participants in the other lists perceived the two accents as equally natural (with numerically higher scores for L+H* than H*). This indicates that the difference between these two accents is indeed marginal with respect to the perceived naturalness, as both of them are frequently used in English (Dainora, 2001, 2002).
Figure 2.7. An interaction between Experimental List and Pitch Accent from the participant analysis regarding sentence naturalness judgments in Experiment 1B; the asterisk indicates a marginally significant pair ($p = .052$).

2.3.5.2 Continuation perspective

The first dependent measure was the proportion of trials (across all conditions) for which coders indicated that continuations were generated from the speaker perspective. As seen in Figure 2.8, participants strongly preferred to provide continuations from the speaker perspective (one-way ANOVA: $F_1(2,72) = 100.98, p < .01; F_2(2,62) = 869.01, p < .01$).

However, there was also individual variation such that some participants showed a heavy preference for the speaker perspective, whereas other participants showed less preference for it. Only 10% of the participants generated continuations mainly from the listener perspective. The $x$-axis in Figure 2.9 represents all 40 participants in Experiment 1B. The $y$-axis shows the proportion of trials for which continuation sentences indicated either speaker or listener perspective.
The fact that most continuations were generated from the speaker perspective suggests a possibility that coders could have been biased to use that category, unless there was clear evidence to the contrary. However, it could also suggest that during language comprehension, comprehenders often mentally situate themselves as the protagonists of the narrative plot, except when there is a specified perspective that they have to take, e.g., indicated by proper nouns or third-person singular pronouns (e.g., Brunyé, Ditman, Mahoney, Augustyn, & Taylor, 2009; Bruzzo, Borghi, & Ghirlanda, 2008). At the same time, the fact that about 20% of the participants showed a more balanced preference suggests that comprehenders are flexible and can adopt different perspectives (e.g., Borghi, Glenberg, & Kaschak, 2004; Zwaan, 2004).

For an evaluation of how prosody influenced continuation perspective, the second dependent measure used the proportion of trials for which continuations were created from the speaker perspective in each test condition. Figure 2.10 shows mean proportion converted into percentage. Two two-way repeated-measures ANOVA tests crossing Pitch Accent and End Contour as the within-participant and within-item factors showed that participants more often took a speaker perspective when the target sentences were delivered with the final rising tone L-H% rather than the falling tone L-L%: a main effect of End Contour \( (F_1(1,39) = 9.41, p < .01; F_2(1,31) = 21.55, p < .01) \). Also, there was a significant interaction effect from the participant analysis, such that participants’
preference for creating continuations from the speaker perspective was stronger when the sentences with the L-H% contour contained the L+H* accent rather than the H* accent ($F_1(1,39) = 4.38, p < .05; F_2(1,31) = 1.33, p = .257$). Lastly, there was a marginal preference for the speaker perspective whenever the target sentences included an L+H* rather than an H*: a marginal main effect of Pitch Accent: $F_1(1,39) = 2.91, p = .096; F_2(1,31) = 4.02, p = .054$).

Figure 2.10. Effects of two Pitch Accents and two Edge Tones on the mean percentages of the speaker perspective taken by participants in Experiment 1B.

These results lead to the following insights. First, the main effect of the end contour type provides experimental evidence that L-H% and L-L% fulfill distinct functions. On the one hand, participants more often took the speaker perspective with the L-H% contour perhaps because this contour signals continuity in discourse. On the other hand, the listener perspective was taken more with the L-L% contour because it signals sentence finality (Pierrehumbert & Hirschberg, 1990). Speakers perhaps use L-H% as a marker for keeping the conversation floor, i.e., to signal to the listeners that there are more things to be said. While the continuity signal can also invite the listeners to take the floor for their contributions, it is perhaps more frequent that the listeners wait for the speaker to finish her turn. Second, the interaction effect suggests that the perspective type encoded in the continuations could be related to the type of meaning expressed in those continuations. For example, continuations generated after participants heard the L+H* accent together with the L-H% contour, which resulted in more speaker perspective, might express more contrastive meanings, in particular the state-contrast implicature. This meaning represents cognitive processes happening in one’s mind (e.g., activating contrast sets, creating an implicature); thus, taking a speaker perspective will be helpful
in creating a discourse model where the described state can undergo some adjustment. Comparison between continuation perspective and meaning is provided after the main analysis for meaning in the next section.

2.3.5.3 Continuation meaning

Figure 2.11 shows, per test condition (x-axis), the percentage of trials (y-axis) for which coders identified continuation meanings in one of three general categories (i.e., contrast, neutral, and other). The dotted portions of the bars indicate the percentage of continuations expressing contrast, whereas the white portions indicate continuations coded for the neutral meaning. The grey portions indicate the other cases.

![Figure 2.11. Mean percentage of three meaning types coded in each test condition in Experiment 1B; the x-axis marks test conditions, differing only in the combinations of two Pitch Accents and two End Contours.](image)

One unexpected finding was that the L+H* L-H% tune generated substantially fewer contrastive meanings (i.e., 50% of all trials), compared to the results of Experiment 1A, which found about 66% contrastive meaning from the same condition (when the rates of the state-contrast implicature and other contrastive meanings were summed). Possible reasons for this finding are discussed in the General Discussion section below. What is more important, however, is that the percentage of contrastive meanings varied as a function of the test conditions, which differed by Pitch Accents and/or End Contours but not by lexical items.
The proportion of continuations encoding *contrast* was used as the dependent measure for two-way repeated-measures ANOVA tests (that crossed Pitch Accent and End Contour as two within-subjects and within-items factors, and Experimental List as a between-subjects factor). The results indicated a strong main effect of Pitch Accent ($F_1(1,36) = 7.27, p < .01; F_2(1,31) = 9.11, p < .01$), as well as End Contour ($F_1(1,36) = 19.12, p < .01; F_2(1,31) = 12.76, p < .01$). Pairwise comparisons indicated that participants generated more continuations expressing *contrast* after listening to the target sentences spoken with L+H* than after listening to those with H*, and they also generated more such continuations after sentences with L-H% than after those with L-L%. Figure 2.12 below shows mean proportions of contrastive meaning converted into percentages in each test condition.

![Figure 2.12](image)

Figure 2.12. Mean percentage of continuations encoding *contrast* over all meanings (i.e., *neutral*, *other*, *contrast*) as a function of two Pitch Accents and two Edge Tones.

While the main effect of Pitch Accent was robust, there was a marginal interaction between End Contour and Experimental List in the participants analysis ($F_1(3, 36) = 2.8, p = .054$). As indicated in Figure 2.13, posthoc paired $t$-tests found that the difference between L-H% and L-L% in generating contrastive meaning was larger and significant in List 1 ($t_1(9) = 2.31, p < .05$) and List 2 ($t_1(9) = 3.69, p < .01$), while the difference in List 3 and List 4 was smaller and didn’t reach statistical significance, although the difference pattern across all test conditions was the same.
Figure 2.13. Percentage of trials for which *contrast* was perceived as a function of Experimental List and End Contour in Experiment 1B; asterisks mark the significant pairs.

While this finding could potentially indicate a weaker effect of End Contour for evoking contrastive meaning, we must see, before drawing any conclusions, to what extent Pitch Accents and End Contour influenced the rate at which participants perceived a particular type of contrastive meaning, namely the state-contrast implicature, to which I now turn my focus.

The bottom four portions of the bars in Figure 2.14 below represent the percentage of each subtype in the *contrast* category. The other portions of the bars show the percentage of *neutral* and *other* meanings. As indicated by the grid-patterned portions at the bottom, the most dominant type of *contrast* across all conditions was the **state-contrast implicature**. The black portions of the bars indicate that the L+H* L-H% tune, unlike the other tunes, didn’t generate any continuations of the **emphatic acceptance** type.

Figure 2.14. Mean percentages of all meaning types identified in each condition of Experiment 1B.
For statistical analysis, the dependent measure was the proportion of continuations indicating the state-contrast implicature (over all meaning types). Two-way repeated-measures ANOVA tests (crossing Pitch Accent and End Contour as within-subjects and within-items factors and Experimental List as a between-subjects factor) found two robust main effects without any interaction effect: Pitch Accent \( (F_1(1,36) = 5.06, p < .05; F_2(1,31) = 5.76, p < .05) \) and End Contour \( (F_1(1,36) = 16.03, p < .01; F_2(1,31) = 16.08, p < .01) \). This shows that the rate at which participants perceived the target implicature from test sentences was affected orthogonally by Pitch Accent and End Contour types as seen in Figure 2.15 below.

![Figure 2.15. Mean percentage of responses indicating participants’ perception of state-contrast implicature as a function of Pitch Accents and End Contours in Experiment 1B.](image)

While these findings establish equally important roles of Pitch Accents and End Contours in evoking the state-contrast implicature, another important question that Experiment 1B addresses is whether a tune in its entirety also contributes to the perception of that meaning. Hence, I rearranged the data in order to test if there is any effect that goes beyond the additive effect of two individual prosodic elements, Pitch Accents and End Contours (Figure 2.16).

![Figure 2.16. Mean percentage of responses indicating participants’ perception of the state-contrast implicature as a function the four distinct tunes in Experiment 1B.](image)
As seen in the figure, each tune produced a significantly different amount of the state-contrast implicature (a main effect of Tune: \( F_1(3,108) = 6.90, p < .01; F_2(3,93) = 5.91, p < .01 \)). Table 2.6 below presents results on the planned contrasts among all possible tune pairs. I labeled each pair using T (Tune), A (Accent), and E (End contour), in order to capture the source of the differences that each of the paired comparisons indicates.

Table 2.6. Planned contrasts testing the significance of the paired differences with respect to the rate at which state-contrast implicature was perceived in Experiment 1B (** significance at \( \alpha = .01 \), ** at \( \alpha = .05 \), · at \( \alpha = .10 \)).

<table>
<thead>
<tr>
<th>Label</th>
<th>Sig.</th>
<th>Pairs in planned contrast</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T^1 )</td>
<td>***</td>
<td>( L+H^* L-H% ) vs. ( H^* L-L% )</td>
<td>( F_1(1,36) = 27.81, p&lt;.01; F_2(1,31) = 19.42, p&lt;.01 )</td>
</tr>
<tr>
<td>( T^2 )</td>
<td>n.s.</td>
<td>( L+H^* L-L% ) vs. ( H^* L-H% )</td>
<td>( F_1(1,36) = 0.81, p=.37; F_2(1,31) = 1.13, p=.30 )</td>
</tr>
<tr>
<td>( A^1 )</td>
<td>·</td>
<td>( L+H^* L-H% ) vs. ( H^* L-H% )</td>
<td>( F_1(1,36) = 3.52, p=.07; F_2(1,31) = 2.01, p=.12 )</td>
</tr>
<tr>
<td>( A^2 )</td>
<td>n.s.</td>
<td>( L+H^* L-L% ) vs. ( H^* L-L% )</td>
<td>( F_1(1,36) = 2.36, p=.13; F_2(1,31) = 2.36, p=.14 )</td>
</tr>
<tr>
<td>( E^1 )</td>
<td>**</td>
<td>( L+H^* L-H% ) vs. ( L+H^* L-L% )</td>
<td>( F_1(1,36) = 7.68, p&lt;.01; F_2(1,31) = 5.52, p&lt;.05 )</td>
</tr>
<tr>
<td>( E^2 )</td>
<td>**</td>
<td>( H^* L-H% ) vs. ( H^* L-L% )</td>
<td>( F_1(1,36) = 9.63, p&lt;.01; F_2(1,31) = 7.54, p&lt;.01 )</td>
</tr>
</tbody>
</table>

The results presented in Table 2.6 indicate two notable findings. First, the influence of the rising end contour L-H% in generating the state-contrast implicature is stronger and more reliable than the effect of the contrastive pitch accent L+H*. The pairs that differed by the End Contour type (i.e., \( E^1, E^2 \)) all produced significant results, whereas the pairs differing by the Accent type (i.e., \( A^1, A^2 \)) resulted in weaker and somewhat relative effects. That is, when the Accent type was controlled, more state-contrast implicature was produced whenever a tune contained L-H% rather than L-L%.

However, the effect of L+H* was stronger than that of H* only when the accents were embedded in the tonal context of L-H% (i.e., in pair \( A^1 \)), and this effect reached a marginal significance in the participants analysis only. This is in line with a previous finding that both of these pitch accents in the high tone category—L+H* and H*—can evoke contrastive meaning (Watson, Tanenhaus, & Gunlogson, 2008). However, the functions of the two End Contours are distinct, such that the L-H% evokes the state-contrast implicature frequently, whereas the L-L% does so at a much lower rate, even in the presence of the contrastive pitch accent L+H*. 

53
The second notable finding in Table 2.6 is that the effect of a tune on meaning perception is strongest when the individual prosodic elements constituting the tune each carries functional relevance to the target meaning. Moreover, the contribution that is made by each prosodic element is additive and not super-additive in nature. In pair T¹, for example, the contrastive tune L+H* L-H% contained the stronger contrastive element of each prosodic category—Pitch Accent and End Contour—and hence, the tune’s effect was significantly different from that of the neutral tune H* L-L%, which contained the weakest contrastive elements of the prosodic morpheme counterparts. In contrast, the two tunes in pair T² (i.e., L+H* L-L% versus H* L-H%) each included only one of the stronger contrastive elements from either prosodic category (i.e., L+H* or L-H%). Therefore, each of these tunes generated less state-contrast implicature when compared to the contrastive tune L+H* L-H%. Moreover, the paired comparison between the two tunes in pair T² showed that they were not statistically different from each other. In fact, the numerical differences between the percentages of state-contrast implicature from three tunes—L+H* L-H%, H* L-H%, and L+H* L-L%—are strikingly similar (see Figure 2.16 above). These patterns together indicate that the size of the L+H* L-H% tune’s effect is additive to those of L+H* and L-H%, which in turn provides strong support for the compositionality hypothesis for tune meaning (e.g., Pierrehumbert & Hirschberg, 1990; see the General Discussion section).

Despite this significant influence of individual prosodic elements, the rate at which continuations encoded state-contrast implicature in Experiment 1B was also subject to individual variation. In Figure 2.17 below, the y-axis lists all 40 participants’ data, which were sorted according to the percentage of trials for which each participant generated continuations expressing the target meaning in the L+H* L-H% condition. The four columns on the x-axis represent data collected in each of the conditions.
Overall, people who generated continuations encoding more state-contrast implicature in the L+H* L-H% condition also generated more continuations expressing that meaning across the conditions \((r = 0.92, p < .01)\). In the L+H* L-H% condition itself, 18 out of 40 participants (above the red-dotted line) produced continuations expressing state-contrast implicature on more than half of the trials, for an average of 73% (henceforth, High Producers). On the other hand, the other 22 people (below the red-dotted line) produced the target meaning for an average of 21% of the trials in the L+H* L-H% condition (henceforth, Low Producers). Analysis on the reaction times to generate continuations showed that the mean typing times from these two groups—13.7 seconds for High Producers and 15.7 seconds for Low Producers—didn’t differ as much as what.
was found in Experiment 1A (i.e., 7 seconds difference between two equivalent groups). However, when only the 9 people at the bottom tier of the Low Producers—who generated continuations expressing the implicature in less than 13% of the trials—were considered, they spent much less time in creating continuations overall (11.7 seconds, versus 16.5 seconds for the rest of the 31 participants).

In addition, a positive correlation was found between continuation production times and the rate at which overall contrastive meanings were perceived. That is, the longer it took for people to generate continuations, the more meanings those continuations expressed in the contrast category in general (Low Producers: \( r = .53, p < .05 \); High Producers: \( r = .41, p = .09 \)) and the state-contrast implicature subcategory in particular (Low Producers, \( r = .55, p < .01 \); High Producers: \( r = .39, p = .11 \)), with the stronger results coming from the Low Producer group. In contrast, more quickly generated continuations expressed more neutral meanings (Low Producers: \( r = -.44, p < .05 \)) as well as more other meanings (High Producers: \( r = -.46, p = .056 \)).

These findings together indicate that perceiving contrast in general and the state-contrast implicature in particular may indeed require more processing time and effort than what is needed for neutral or other meanings. Interestingly, continuation generation times were also correlated with the types of perspective encoded in the continuations, when either the grand mean data or the data from the L+H* L-H% condition (C1) were considered. For example, people who took longer to create continuations took the Speaker perspective more often (grand mean: \( r = .30, p = .065 \); C1 only: \( r = .32, p < .05 \)), and more Speaker perspective was correlated in turn with fewer other meanings (grand mean: \( r = -.39, p < .05 \), C1: \( r = -.32, p < .05 \)). In contrast, people who quickly produced continuations took the Listener perspective more often (grand mean: \( r = -.31, p = .053 \); C1: \( r = -.33, p < .05 \)), and more Listener perspective was correlated in turn with more other meanings (grand mean: \( r = .33, p < .05 \); C1: \( r = .28, p < .08 \)). These findings provide insights on the relationship between continuation perspectives and the types of meanings expressed with those perspectives.

Some other findings that are similar to those found in Experiment 1A include that the longer it took for participants to judge the target sentences’ naturalness, the longer it
took for them to come up with the follow-up sentences. Additionally, coders identified a pattern wherein continuations frequently expressed a change in time (e.g., past to present) when the encoded meanings were of the state-contrast implicature type \((r = .88, p < .01)\), but not when the continuations expressed neutral meanings \((r = -.84, p < .01)\) or emphatic acceptance \((r = -.30, p = .057)\). This points to the unique nature of the state-contrast implicature; it expresses contrast in terms of state of affairs as well as time. Section 2.4 discusses why listeners might arrive at this meaning.

Finally, individual items also differed somewhat in the rate at which they generated continuations encoding the state-contrast implicature. In Figure 2.18, the y-axis shows the percentage of trials that produced the state-contrast implicature in the L+H* L-H% condition. The x-axis presents all items sorted according to the averaged mean values from the results of Experiment 1A (i.e., the darker grey bars) and 1B (i.e., the lighter grey bars). The vertical red dotted line in the middle divides the items into two sets. The set on the left side includes the items whose averaged mean values were greater than 50%, whereas the other set, on the right side, includes the items with averaged means of less than 50%. The item marked by a black dot (next to the sentence) is the one with the lowest grand mean for the naturalness rating in both experiments.

![Figure 2.18](image_url)

Figure 2.18. Percentage of trials for which each item generated the state-contrast implicature in the L+H* L-H% condition in Experiments 1A and 1B; the item marked by a black dot received the lowest naturalness mean ratings.
As seen in the figure, 19 out of the 32 items (i.e., 59% of all items) produced the state-contrast implicature for more than 50% of the trials, for an average of 64% of the continuations. The other 13 items (i.e., 41% of all items) produced the state-contrast implicature for an average of 41% of the trials. The fewest instances of the target meaning were found from the sentence *The door was ajar*, for which three participants indicated their difficulty in understanding this sentence due to the lexical item *ajar*. On the other hand, the item that received the lowest naturalness rating (i.e., *The zipper was *done*) produced the target meaning for an average of 48% of the trials, although nine participants also expressed misunderstanding or uncertainty about this sentence’s meaning due to the combination of lexical items. A correlation analysis found that, across all test conditions, target sentences that received higher naturalness scores induced more continuations with the state-contrast implicature at a marginal level ($r = .28, p = .086$). However, this correlation was not found when only the data in the target condition (L+H* L-H%) were considered.

In summary, Experiment 1B replicated Experiment 1A’s finding that participants construct meaning that expresses contrast in general, and the state-contrast implicature in particular, when they hear sentences spoken with the L+H* L-H% tune. More importantly, Experiment 1B further established that this is due to the separate functions of the tune’s sub-elements, L+H* and L-H%. The next section discusses implications from this finding.

### 2.4 General discussion of Experiment 1

This chapter presented two sentence continuation experiments examining the role of intonation in perceiving contrastive meaning. In both experiments, participants listened to the target sentences spoken with various prosodies one at a time and judged how natural those sentences sounded to them. Upon the replay of the target sentences, participants typed continuations that would naturally continue the discourse. While participants were encouraged to make quick and casual responses, there was no technical constraint restricting their response times. After the experiments were finished, three

---

8 The frequency of all words for the subject entities (but not the predicates) was carefully matched (see Experiment 2 in Chapter 3).
native English-speaking coders classified the meanings expressed in each of the continuations, using a coding scheme that specified multiple meaning categories. The coders were instructed to decide on only one meaning category for each continuation, and when in doubt, they were encouraged to use the category called *other*, to indicate the uncertainty. High intercoder reliability in both experiments verified the usefulness of the coding scheme.

The first experiment, 1A, tested three broadly distinct intonation contours, namely the contrastive contour (L+H* L-H%) together with the affirmative and negative forms of the neutral counterpart (H* (H*) H* L-L%). The results showed that when the target sentences were delivered with the contrastive tune, participants produced continuations expressing that there is a change in the state mentioned in the sentence (i.e., state-contrast implicature) for at least 63% of the trials. However, people rarely produced such continuations when the sentences were delivered with either the affirmative or negative neutral tunes. Assuming that the percentage of state-contrast continuations is a comparable estimate across conditions of how often an implicature was induced, these results provide empirical support for a previous claim that the L+H* L-H% tune in English is a linguistic device encoding contrast (Lee, 2000; Lee, 2006; Lee, 2007). And yet, the results also indicate that evoking contrast is not the only function of this tune, since it readily permitted other interpretations as well. Moreover, individual variability was found in the rate at which participants generated the state-contrast continuations from the L+H* L-H% condition.

Experiment 1B further examined the exact source of contrastive meaning in the L+H* L-H% tune. To this end, either each part of the prosodic unit was replaced—L+H* with H*, or L-H% with L-L%, or both were replaced, creating four factorial conditions. Moreover, the coding scheme included more detailed subcategories for the contrastive meaning, thus allowing coders to identify subtle differences in the type of contrast expressed in the continuations. This experiment resulted in two separate findings, each of which is discussed below.

First, among the continuations collected in the L+H* L-H% condition, the overall percentage of trials that coders identified for contrastive meaning was lower in this
experiment than in Experiment 1A. The first of three possible explanations for this is the fact that Experiment 1B was twice as long as Experiment 1A. It is known that experiment length can negatively influence the results, due to the decrease in participants’ attention (e.g., Baayen, 2008). If perceiving implicature from a prosodic cue is a product of processing effort, then we should observe a decrease in the rate at which participants perceived the implicature as a function of participation time. This is a testable hypothesis, and accordingly, I plan an additional analysis, using a mixed-effects model treating the trial number (i.e., the order of the trial in the experiment) as a random-effect term (Baayen, 2008).

The second possibility is that participants might have failed to consistently catch the identity of the L+H* accent. Perhaps some participants in some trials perceived the L+H* accent as H* or even L*+H. While the material analyses—both the acoustic measurements and the ToBI transcriptions—indicated successful production of the L+H* tokens, to what extent those tokens will lead to clear categorical perception for different individuals is a matter that requires further empirical research. In particular, the accent-bearing word in the current experiment (i.e., was) constituted only one syllable, and the word’s first segment [w] contributes to a perception of low tone (due to its segmental characteristics). Hence, identifying the accent type on this word might have required more careful listening, which could have been more difficult in a longer experiment setting. One follow-up study that can clarify this issue is to use the materials in Experiment 1B and run a categorical perception task, not only with naive English-speaking participants but also with trained English ToBI transcribers. Comparing results of these two participant groups will illuminate the discussion on the pitch accent’s category membership, which is an issue at the heart of an ongoing debate in intonational phonology regarding intonational morpheme inventories and their functions (Beckman et al., 2005; Gussenhoven, 2004; Ladd, 1996). Similarly, another follow-up study to further clarify the relationship between intonation form and meaning is to have participants rate whether or not a particular tune is suitable for a given meaning.

The third possible explanation for the reduced rate at which participants produced the state-contrast implicature in Experiment 1B is perhaps related to the coders’ blindness.
to the target sentence prosody. As mentioned earlier, coders in Experiment 1B were not informed about how test sentences were spoken, thus relying solely on the lexical information to evaluate the meanings expressed in the continuations. Therefore, unless the continuations explicitly stated contrast, coders might have classified them as having some other meanings such as neutral or other. In contrast, coders in Experiment 1A were given auditory examples that showed them how the target sentences were delivered to the participants. If coders themselves perceived contrastive meaning from the target sentence prosody, then that additional knowledge might have encouraged them to classify some of the ambiguous continuations as contrast, hence boosting the rate of contrastive meaning in Experiment 1A. In this regard, it will be useful to systematically manipulate coders’ awareness of the target sentence prosody in another experiment, in order to see how it influences their perception of contrastive meaning in the course of semantic classification.

Despite the reduced rate at which continuations indicated state-contrast implicature, the most important finding from Experiment 1B was that the identity of both Pitch Accent and End Contour orthogonally influenced the percentage of continuations expressing contrast in general, as well as state-contrast implicature in particular. Target sentences induced more contrastive continuations when they were delivered with L+H* rather than H*, and with L-H% rather than L-L%. Furthermore, the percentage of state-contrast implicature with the L+H* L-H% tune was an additive effect of the L+H* pitch accent and the L-H% continuation rise, and more strongly influenced by the continuation rise than the accent. The fact that there was no evidence for a super-additive effect that goes beyond the summed effects of the two individual prosodic elements suggests that the state-contrast implicature does not result from interpreting the L+H* L-H% contour holistically. Rather, the results suggest that intonational meaning emerges from the meaning of individual prosodic elements, which supports the autosegmental approach to intonation theory (e.g., Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980).

The results thus speak for an interpretive processing mechanism that takes each cue’s function probabilistically and composes meaning incrementally. However, the current dependent measure (i.e., the rate of producing continuations with the state-
contrast implicature) is rather insensitive in detecting on-line meaning activation; it represents some level of conscious and effortful production processes aiming at naturally continuing discourse. Therefore, it is conceivable that the state-contrast implicature was perceived but never realized in some trials, or that this meaning resulted from an effortful inference processing in some other trials. To clarify this issue and provide a complementary perspective, Chapter 3 will present the second set of experiments that tested meaning activation on-line. Also, Chapter 4 presents an eye-tracking experiment that evaluates incremental processing of the prosody-induced implicature.

On the recognition of the effortful process involved in the continuation paradigm, the last discussion point about Experiment 1B is the finding that people who took longer to generate continuations produced more continuations encoding state-contrast implicature, whereas people who spent less time contributed more continuations with neutral or other meanings. This perhaps indicates subtle differences in the discourse representations that people built for each meaning type. In the present task, the participants’ job was to listen to isolated sentences one at a time and provide a sentence that would naturally continue the discourse. This required participants to situate themselves into a plausible context where the meaning of the just-heard sentence would fit. Thus, the amount of background information needed for the particular (mental) context would have influenced how long it took for people to come up with a subsequent meaning.

Crucially, the type of meaning encoded in the continuations requires different amounts of contextual information. For instance, a neutral meaning accepts what was asserted in the sentence and naturally continues the conversation as in *The jar was full… → “so I was able to make the peanut butter sandwich.”* In this case, all that is needed by the participants is the ability to come up with a natural consequence of the asserted state. In contrast, producing a state-contrast implicature requires additional information. First, contrast between possible alternative states has to be perceived, based upon the accent location and its lexical content. Thus, listening to a sentence fragment like “The jar WASL+H*…” would activate alternatives for the contrastively accented word *was*, which would include contrast sets in terms of both state and time as in *wasn’t*, *is*, *isn’t*, *will be,*
among others. Then, an inference has to be made as well, in order to provide a
continuation of the perceived contrast. A unique aspect of the state-contrast implicature is
that the details of that inference are not linguistically specified. There are overt linguistic
markers signaling contrast (L+H*) and continuation (L-H%), but detailing the
inference—i.e., the implied factual situation—is solely up to the comprehender.
Constructing the state reversal implicature would thus assign more cognitive work on the
participant’s part, which may naturally increase processing time. One way to verify this
possibility will be to compare continuation response times in each of the identified
meaning categories (e.g., times to generate neutral continuations vs. state-contrast
implicature).

Importantly, the perception of state-contrast implicature as a cognitive process
highlights the following two points. On the one hand, individual variation is expected;
constructing state-contrast implicature is not possible if a participant fails to either
perceive contrast or build an inference from it. On the other hand, finding a general
pattern across the participants is also expected, given that cognitive processes we
undertake as humans possess similar properties. An intriguing finding from the current
experiment was that many of the continuations expressing the state-contrast implicature
included similar contents, even though no contextual information was provided in the
task. For the state-contrast of the *jar* sentence, for instance, many participants discussed
how it became empty after their brothers ate all the contents, the most frequently
mentioned contents being cookies, candies, or peanut butter. For the sentence *The candle
WAS lit*..., many people continued that the wind then blew it out. This suggests that
people can not only work out implied meanings, but that they do so in similar ways. How
is it that participants build similar discourse representations and talk about similar
meanings in the absence of any overt contextual information?

More elaborate discussion on this point will be provided in the last chapter, but
one possibility is that the sentences included adjectives that denoted binary attributes,
which represent important conceptual relations that we all experience in the world (e.g.,
Barsalou, 1999; Israel, 2004; Lakoff & Turner, 1989). That is to say, people were able to
construct similar implied meanings perhaps because the counterfactual and factual

63
meanings (e.g., a full jar versus an empty jar) are tightly associated, and are thus stored in the same categorization based upon our own experiences. If we grew up in a culture that stores cookies in a jar, for example, we would naturally experience a full jar becoming empty as we (and others) eat up those cookies. Likewise, a candle-lighting experience must be accompanied by an unlit candle as well, since they are part of a connected event/situation. These kinds of frequently co-occurring events—as instances of episodes—would then become part of what we know about those objects, as we continue expanding our own experiences with the world. The domain-specific knowledge stored in our declarative memory can then be accessed and used as a basis for inferences and learning (Griffiths, Dickinson, & Clayton, 1999; Marshall, 1995; Tulving & Markowitsch, 1998).

Importantly, many of our perceptual experiences accompany linguistic exchanges as well. Thus, hearing a word like jar can activate experiences associated with the referent (e.g., Martin & Chao, 2001; Pulvermüller, 1999; Zwaan, 2004). Under this assumption, how does sentence prosody influence the cognitive processes involved in activating lexical items and those traces of the associated experience? In other words, how are the prosodic units and the text mapped together to activate not only meanings that are stated but also those that are implied? Furthermore, what form of representations would the activated meanings take? Would it be the case that only explicit meanings can take some traceable representations, or would explicit and implicit meanings share a similar representational format?

The next chapter presents Experiment 2, which investigated these questions. Using a picture-naming paradigm that is sensitive to on-line meaning activation, I explore a particular hypothesis that contrastive prosody evokes multiple meaning representations that are experiential in nature.
CHAPTER 3
EXPERIMENT 2: PICTURE NAMING

This chapter presents the second set of this dissertation’s experiments—end-of-sentence picture naming—that investigated cognitive processes involved in computing a prosodically motivated implicature. While the sentence-continuation experiments in Chapter 2 showed that sentence prosody can evoke an implicit contrast for the state of affairs mentioned in the sentence, how listeners arrive at such meaning is not understood. For example, do listeners first extract the literal meaning expressed in the sentence, and then use prosody to revise the initial interpretation for an implicature? Or do people immediately activate both mentioned and implied meanings but resolve the competition at a later point in time? In either case, how long does it take for listeners to arrive at a prosody-based implicature? Are there differences in computing inferences based on different linguistic cues? Moreover, what form of mental representation does an implied meaning versus an asserted meaning take?

The current chapter explores these questions based on the proposal that language comprehension engages mental simulation, a cognitive operation that enables language users to have vicarious experience of what is described or implied in the language (Barsalou, 1999). Section 3.1 begins by reviewing key experimental findings supporting mental simulation during language comprehension. Then, I provide an overview of previous research that has investigated how mental simulation might assist the processing of negated information. I also present two reasons that the literature on lexical negation bears significance for the current study. In doing so, I discuss research on an associative distributed network for lexical and conceptual organization to evaluate its consequence on meaning representation. Section 3.2 then details the current dissertation experiment, and Section 3.3 discusses the implications of the findings.
3.1 Background

The study of how exactly we extract meaning from utterances is still in its infancy. However, in the recent decade, scholars have accrued a vast amount of behavioral and brain imaging evidence showing that one mechanism utilized by language processors routinely and mostly subconsciously is mental simulation, which allows vicarious experience of what was described (or implied) in the language (Barsalou, 1999; Zwaan, 2004). For example, when a sentence describes actions, people engage their motor systems to simulate what it would be like to perform the described actions (e.g., Bergen & Wheeler, 2005; Bub, Masson, & Cree, 2008; Buccino et al., 2005; Glenberg & Kaschak, 2002; Kaschak & Borreggine, 2008; Tettamanti et al., 2005). Similarly, when a sentence describes a state of affairs, people engage their visual systems to simulate what the mentioned entities would look like in a fairly detailed manner—what shape they would take (Zwaan et al., 2002), how they would move (Kaschak et al., 2005), and in what color (Connell, 2007), orientation (Stanfield & Zwaan, 2001; Zwaan, Madden, Yaxley, & Aveyard, 2004), and even visual resolution (Yaxley & Zwaan, 2007).

An intriguing aspect of mental simulation that these studies have uncovered is that people’s knowledge of the grammar and the world work together to guide the simulation processes, which lead to subconscious access to subtle semantic features that are neither required nor specified by the lexical contents themselves. For example, Zwaan and his colleagues (2001, 2002) showed that object shape and orientation enter into mental simulations not because those features are mentioned in the sentence but because the words and phrases in the sentence are combined in such a way that it triggers our dynamic world knowledge about the object. Therefore, reading a sentence about an eagle in the sky will evoke an image of an eagle with widely stretched wings, whereas a sentence about an eagle in a nest will activate an image of a perched eagle. This highlights the fact that language processing involves activation and integration of meanings that are grounded in our perceptual experience and knowledge (e.g., Barsalou, 1999; Chambers, Tanenhaus, & Magnuson, 2004; Glenberg, 1997; Lakoff & Johnson, 1999; Spivey, 2007; Zwaan et al., 2002).
The underlying mechanism proposed and tested for mental simulation is that simulation recruits the stored patterns of neural activities in the brain that have been formed during our perceptual experience of the world (e.g., called “perceptual symbols” and “simulators” in Barsalou, 1999, “state spaces” in Spivey, 2007, and “functional webs” in Pulvermüller, 1999, 2002 and Zwaan, 2004). This includes the sensory-motor aspects of experience, as well as aspects of proprioception and introspective aspects from any perceived experience (Barsalou, 1999), e.g., recalling the attributes of a watch that was lost (“representational state”), remembering how happy we were when we passed the driver’s license test (“emotional state”), or retracking the criteria that helped us to pick the sweetest watermelon before and using them again for another success (“cognitive operation”). Because our attention is limited, aspects of these experiences that trigger neural activities are stored as schematic neural pathways, which can then be reactivated upon similar experiences that come either directly or indirectly through mental simulation. Thus, mental simulation involves a schematic and yet multimodal and dynamic knowledge structure that is strongly grounded in our experience.

Although the question of whether mental simulation is an epiphenomenon or whether it actually assists meaning comprehension still awaits direct tests and answers, several theoretical and experimental claims to date hint that mental simulation may fulfill an important function, especially in building various inferences. For example, Kaup and colleagues explored how mental simulation might assist in comprehending negative sentences, especially regarding the conceptual understanding about the deviation between what is expected and what is true. As noted in Kaup et al. (2006), among many others (e.g., Givón, 1978; Horn, 1984; Wason, 1965), a distinct aspect of a negative sentence (that is different from an affirmative sentence) is that it implicitly refers to two states of affairs. A sentence like The door is not open, for example, indicates one state where the door is open (i.e., the negated state or “counterfactual” state of affairs), and another state where this is not the case (i.e., the actual or “factual” state of affairs). Moreover, an additional inference can be drawn with respect to the factual state. For example, one likely state for a door that is not open is a closed state. On this view, previous researchers on sentential negation across different theoretical approaches have generally posited that
processing a negative sentence would require language comprehenders to first mentally represent the counterfactual state (e.g., with the door open) and then reject or revise that representation to arrive at the factual state (e.g., with the door closed). It was suggested that such a process might involve, for example, tagging an abstract negation operator on the negated proposition or discourse representation to indicate that this state should not be part of the described world representation (e.g., Kamp, 1981; Kintsch & van Dijk, 1978; McKoon & Ratcliff, 1992; van Dijk & Kintsch, 1983). Indeed, existing literature presents evidence that negated information loses its availability, especially when the described world representation suggests the absence of the negated concepts (as in *Elizabeth bakes some bread but no cookies*; MacDonald & Just, 1989). Moreover, a possible factual state might be inferred, depending on the amount of surrounding contextual information (e.g., De Villiers & Flusberg, 1975; Glenberg et al., 1999; Kaup et al., 2006; Lüdtke & Kaup, 2006; Wason, 1965).

Kaup and her colleagues (Kaup et al., 2006; Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007) have proposed that dynamic mental simulation will allow people to represent not only the counterfactual state but also the factual state in an experiential-based format, in order to assist the comparison between the two in a more realistic manner. That is to say, instead of positing an operator that is abstractly tagging the counterfactual information, the authors suggested that the negated or counterfactual state is simulated but not integrated with the representation of the described world, thereby signaling mental rejection of the simulated counterfactual state. For this, the authors proposed an auxiliary or secondary representational system that holds the simulation of the counterfactual state separately from simulations of affirmative information (however, the authors do not address how these separate representational systems might handle lexical negations like *unhappy* or nonmorphological cases like *deny* that can express a

---

9 The rejection and resulting availability of the negated information depend on various factors including discourse purpose and the scope of the message that is being negated. For example, Kaup (1997, 2001) showed that negated information maintains its availability when the described world suggests the presence of the negated concept as in *Elizabeth burns the letters but not the photographs*. Also, Giora, Balaban, Fein, & Alkabets (2005) argue that, in sentences like *Fred didn’t see a cyclist who was coming down the hill and hit him* (Tottie, 1994, p. 414), the negation marker should reduce availability of *see* but not of *cyclist*. Given the radius of the current dissertation and the breadth of the negation literature, I limit my discussion to the processing of simple sentential negation as in *The door is not open*. 

68
negative notion in an affirmative sentence such as *Jim is unhappy* or *Jim denies the story*). While the authors described a juxtaposition of two simulated states (i.e., the negated state and the described world’s state) before any inference takes place, they did not discuss its implication for processing. Instead, they suggested that a sequential, two-step process will lead to a shift of attention from the counterfactual state to a factual state, if there is sufficiently constraining information for making an inference.

Below I provide a detailed review of Kaup and her colleagues’ studies (Kaup, Lüdtke, & Zwaan, 2005; Kaup et al., 2006, Kaup et al., 2007) for two reasons. First, while many studies have focused on what properties in the language cause mental simulation and what kinds of semantic features appear during simulation, Kaup and colleagues’ studies raised important questions about the function of mental simulation, using a linguistic device that has been considered to be abstract and difficult in most theories of meaning representation (e.g., Carpenter & Just, 1975; Clark & Chase, 1972; Trabasso, Rollins, & Shaughnessy, 1971), especially when no supporting discourse context is provided (see the later part of Section 3.1 for more on this point). Second, the aspect of multiple meaning representations involved in negation is somewhat similar to the type of meaning investigated in the current dissertation, namely the state-contrast implicature. In fact, a state-contrast implicature evoked by prosody can be conceived of as a type of “pragmatic negation”; there is a state of affairs asserted by the lexical items, but the prosody suggests that the state no longer holds. Understanding how explicit negation is processed would benefit the understanding of how pragmatic negation operates. Indeed, this is the reason that the current dissertation’s experiments all compared prosody-induced state-contrast to sentential negation using the lexical item *not* (see Section 3.2.1 for the current experiment’s design).

**Processing negation**

To investigate whether negation is understood through two distinct stages of mental simulation, Kaup et al. (2006) presented native speakers of German with German negative (and affirmative) sentences that contained contradictory predicates such as *open* and *closed*, each of which expresses one of two binary states, thus easily allowing an
inference about a possible factual state; e.g., Der Schirm war nicht aufgeklappt ‘The umbrella was not open’ would easily induce the idea that it was actually closed.

Participants read each sentence (left adjusted on the screen) and pressed a button to indicate that they understood the sentence. Upon the button press, a picture appeared in the center of the computer screen, and the participants’ job was to name the depicted entity as quickly as possible; e.g., to say Schirm aloud. For critical trials, the picture always showed the entity mentioned in the sentence, but the depiction matched either the counterfactual state (e.g., an open umbrella) or an alternative factual state (e.g., a closed umbrella). The dependent measure was the naming response times, which were recorded from the moment that participants saw the picture (i.e., picture onset) to the moment that they began to name the entity (i.e., naming onset). Kaup et al. reasoned that if participants mentally simulated what was negated in the sentence during sentence reading, then their naming times to a subsequent picture would be smaller when the image depiction matches the negated or counterfactual state (e.g., an open umbrella) as opposed to the factual state (e.g., the closed umbrella). On the other hand, if processing a negative sentence activated mental simulation about the factual state, then the naming times would be smaller when the picture depiction matches the factual state than when it matches the counterfactual state.

In fact, this end-of-sentence picture naming paradigm is a well-known technique in psycholinguistics, which enables a subconscious measure of meaning activation and integration. While the participants’ job is to simply say what the depicted entity is, numerous studies have shown that the subtle aspects of meaning representations constructed during sentence comprehension facilitate naming times to the subsequent pictures when the depicted state is compatible with the mentally simulated state (this is called a “match-facilitation effect”; Zwaan et al., 2002). This is because mental simulation recruits neural records of perceived experience in the sensory-motor systems that are actually used in real perception. Thus, when compatible perception follows mental simulation, the patterns of neural activities that had been recruited during mental simulation will benefit the subsequent activation of similar neural circuits needed for actual perception. This match-facilitation effect cannot be a perceptual artifact, since half
Thus, participants do not have any clear motivation to constantly formulate a visual simulation of what was described in the sentence, as their chance of experiencing a match between imagery and the actual image is much lower than .5 (usually .25 or less, depending on the number of test conditions and fillers). Furthermore, evidence for similar match-facilitation from mental simulation to perceptual processing comes from other behavioral tasks such as a Stroop task (Connell & Lynott, 2007) and a part-verification task (Borghi et al., 2004), as well as from brain imaging studies (Pulvermüller, 1999).

Crucially, Kaup et al. (2006) manipulated how quickly the picture appeared after the sentence was finished, by employing two types of Inter-Stimulus Interval (henceforth, ISI) between the sentence offset and the picture onset: 750 milliseconds (ms) and 1500 ms ISIs. This was to test the sequential process posited for the simulation of two different states. The authors hypothesized that, soon after processing a negative sentence, participants will simulate the counterfactual state, but then after some time has passed, they will be able to infer the factual state. The different ISIs allowed different amounts of processing time to elapse between the sentence and the picture and thus enabled the researchers to investigate what meaning representations participants had formed at a certain point in time. With this design, Kaup et al. (2006) tested both affirmative and negative sentences to see how the representations and the amount of processing time needed for each representation differed between these two sentence types.

The authors found that, when participants received pictures 750 ms after the offset of an affirmative sentence (e.g., Der Schirm war zusammengeklappt ‘The umbrella was closed’), their naming times were significantly shorter when the depicted image matched the state that was described in the sentence (e.g., of a closed umbrella) than when it matched a contradictory state (e.g., of an open umbrella), indicating a match facilitation from sentence understanding to picture processing based upon the literal meaning. However, when it was a negative sentence that preceded a picture in this 750 ms ISI condition, participants’ naming times did not indicate any significant match facilitation toward either the counterfactual-state image or the factual-state image. Different results emerged from the 1500 ms ISI condition, however. After processing negative sentences,
participants became significantly faster at naming pictures that depicted the factual-state rather than the counterfactual state, suggesting that they had simulated the factual state from comprehending the negative sentences. As for the affirmative sentences, this 1500 ms ISI condition did not produce any significant effect, and as the authors explained, this is perhaps due to the fact that participants had already finished comprehending the simple sentence and turned their attention away from its meaning.

These findings show that mental simulation is dynamic and yet transient, in a way that can reflect a change in meaning representation. Furthermore, the authors argued that this change in mental simulation provides evidence for a two-stage sequential processing of negation, which includes a tipping point from the simulation of the counterfactual state to the simulation of the factual state. However, careful evaluation of their results and materials, in conjunction with information from other areas of the sentence processing literature, raises several questions that require further investigation.

First, the conclusion for the two-stage negation model drawn in Kaup et al.’s (2006) study was grounded on the combined findings that, at 750 ms ISI, negation did not produce any effect but at 1500 ms there was a match-facilitation effect toward the factual-state image. While this null effect at 750 ms does indicate no dominant meaning representation at that point, this finding cannot guarantee that there was actually a dominant representation for a counterfactual state at an earlier ISI. A cautious reading of the report suggests that evidence for a match-facilitation effect toward the counterfactual state comes from a few other studies of Kaup and colleagues. For example, Kaup et al. (2005) employed a picture identification task at 0 ms and 1500 ms ISIs to test meaning representation induced by negative sentences like *The X is not above the Y*. After the sentence, a picture containing two objects (e.g., a picture of X above Y) appeared in the center of the screen, and participants’ job was to decide as quickly as possible if both of the objects had been mentioned in the sentence. Figure 3.1 represents their results, showing that when the pictures were presented right after the sentence offset (i.e., 0 ms ISI), participants’ response times for correctly indicating that both objects were mentioned were significantly smaller when the picture depiction matched the counterfactual state (e.g., there was a faster reaction to a picture of an elephant above a
giraffe after processing *The elephant is not above the giraffe*) than when the picture depicted one possible factual state (e.g., a giraffe above an elephant). At 1500 ms ISI the opposite pattern was found (i.e., numerically shorter reaction times to the factual-state image than to the counterfactual-state image) although the paired comparison at this ISI didn’t reach significance.

![Figure 3.1](image)

Figure 3.1. Reproduction of Kaup et al.’s (2005) results on picture-response latencies (in ms) collected after negative sentence processing; the pictures in the bars are added to show the counterfactual-match versus the factual-match conditions.

Also, in another study, Kaup et al. (2007) tested both indefinite and definite negations (e.g., *There was no eagle in the sky* versus *The eagle was not in the sky*) at 250 ms ISI, and found that for both sentence types, response times in a picture-recognition task were significantly smaller when the picture depicted the target entity in a shape that matched the counterfactual state (e.g., an eagle with outstretched wings) than when it mismatched (e.g., an eagle with wings drawn in). Taking these results together, the two studies (Kaup et al., 2006; Kaup et al., 2007) concluded that all types of negative sentences evoke a representation of the counterfactual state at an earlier stage of meaning integration.

However, the differences in the materials and methods across the experiments mentioned above make it uncertain whether the same results could have been actually obtained if Kaup et al. (2006) had tested the negative sentences with contradictory predicates at an early ISI. In the absence of additional contextual information, perceiving a concrete factual state seems easier and highly probable when sentences contain contradictory predicates as in *The umbrella was not open* than when sentences merely
describe that a state of affairs does not hold, without sufficiently constraining the likely alternatives, e.g., *The elephant is not above the giraffe* (Kaup et al., 2005), *There was no eagle in the sky* or *The eagle was not in the sky* (Kaup et al., 2007). With this in mind, the current study tested sentences with contradictory predicates at six different ISIs (i.e., 0 ms, 500 ms, 1000 ms, 1500 ms, 2000 ms, and 2500 ms) between the sentence offset and the picture onset (see Section 3.2.1 for details), in order to provide a coherent investigation of whether and when the counterfactual and factual information is activated in the mind of listeners.

The second reason to evaluate the two-step model of negation processing is the evidence from other studies hinting that people might not represent both counterfactual and factual meanings at all times of negative sentence processing. That is, it has been claimed that, depending on the sentence properties and surrounding contextual information, people’s meaning representation can reflect either the counterfactual or factual state most dominantly (Brewer & Lichtenstein, 1975; Gannon & Ostrom, 1996; Lea & Mulligan, 2002). On this view, Mayo, Shul, and Burnstein (2004) used a probe judgment task to test two specific questions. First, are different types of negative statements encoded through different processes, thus resulting in different meaning representations? For example, when a negative sentence describes an attribute that has a clear antonym as does *guilty* in *Tom is not guilty* (i.e., a “bi-polar” attribute in their term), do people encode the intended message to be that Tom is innocent, whereas when an attribute (e.g., *romantic*) does not necessarily imply or embed an opposite relation (i.e., a “uni-polar” attribute), do people understand a negative statement like *Tom is not romantic* by representing some possible romantic behaviors initially and then negating them? Second, what are the implications for memory regarding different types of negation processing? Would processing negation of a bi-polar attribute as in *Tom is not guilty* assist people to store and remember the innocence of Tom (assuming this is the meaning represented upon comprehension)? However, people might still remember that the original description contained a negation even though their mental representation contained only the factual meaning. If so, upon the request for a recall of the original description, negative sentences with bi-polar attributes might increase a specific type of
memory error where people accidentally apply the negation marker to the factual meaning that they had in mind, thus recalling *Tom is not guilty* as *Tom is not innocent*. On the other hand, processing negation of a uni-polar attribute as in *Tom is not romantic* might lead people to remember Tom as a romantic person because processing such types of sentences will evoke the counterfactual meaning initially, for example, due to lack of an easily accessible antonymic schema.

To answer these questions, Mayo et al. (2004) in their Experiment 2 had participants first read a description like *Tom [is / is not] a [tidy / romantic] person* and then decide whether a subsequent probe commenting on the character’s behavior (e.g., for the tidy case, *Tom’s clothes are folded neatly in his closet* versus *Tom forgets where he left his car keys*) is congruent or incongruent with his personal traits, using one of three keys indicating “congruent,” “incongruent,” and “impossible to tell.” Half of the time, the adjective described bi-polar attributes, whereas for the other half, they were uni-polar attributes. The authors measured reaction times to process the descriptions as well as the probe statements, among other measures such as accuracy. Also, participants completed a memory task after they finished the congruency decision task plus a 5-minute distraction task. The most relevant finding for the current discussion is that after processing negative sentences describing a bi-polar attribute (e.g., *Tom is not a tidy person*), participants were faster (by 191 ms) to correctly indicate a congruency judgment when the subsequent probe described an attribute that is antonymic to the mentioned trait tidy (e.g., *Tom forgets where he left his car keys*) than they were when it described an attribute that is compatible with the mentioned trait (e.g., *Tom’s clothes are folded neatly in his closet*), which suggests that people successfully inferred a factual attribute of Tom (i.e., that he is messy). This pattern, however, did not emerge with negative sentences using the uni-polar descriptions (e.g., *Tom is not a romantic person*). In this case, participants were faster (by 35 ms) to judge the congruency when the probe described the attribute that matched the trait mentioned in the sentence (i.e., a romantic trait as in *Tom brings his wife flowers on every anniversary*) than when it mismatched, although the difference was smaller in this case.\(^{10}\) Importantly, these patterns of results not only came

\(^{10}\) The paired \(t\)-tests values were not provided in this study.
from participants who were actually instructed to consider negation as an “opposition relation,” but also from those who did not receive any specific instruction for interpretation. As for the memory task, only the bi-polar descriptions induced reversal errors, like remembering *not guilty* as *not innocent*, whereas uni-polar descriptions induced negation-detaching errors, like remembering *not romantic* as *romantic*. With these findings, Mayo et al. (2004) argued that the availability of the opposite schema is critical in determining how a negated message is processed; that is, people can easily infer the factual state when opposition relations can be easily accessed, whereas in other cases, understanding the counterfactual state will benefit the understanding of the following opposition. This claim is informative because it suggests that negative sentence processing exerts probabilistic characteristics that take multiple factors (e.g., the type of attribute that is described) into consideration.

Along this line, a body of literature concerning the semantic memory structure suggests a possibility for parallel meaning activations, especially regarding the antonymic meanings associated with the contradictory predicates (e.g., Frazier et al., 2008; Gross et al., 1989; Kennedy & McNally, 2005). This is the third reason to evaluate the two-stage model of negation. All test sentences in Kaup et al. (2006) included contradictory predicate adjectives, but the authors did not discuss the implication of these materials on on-line processing. However, Gross et al. (1989), among others, tested a claim that antonymy and synonymy are important parts of semantic relations that organize predicate adjectives (Deese, 1965; Gross et al., 1989; Miller, Fellbaum, Kegl, & Miller, 1988; Osgood, Suci, & Tannenbaum, 1957); that is, predicate adjectives are organized through bi-polar attributes in the mental lexicon, with synonymous adjectives clustering around the two direct antonym poles (e.g., *wet*, which is clustered with synonyms like *damp*, *moist*, *soggy*, *waterlogged*, is connected to *dry*, which is clustered with *baked*, *arid*, *parched*, *dehydrated*, *thirsty*, etc.). If this is the case, predicate adjectives that are part of direct antonym pairs (e.g., *wet/dry*) should be recognized faster than other adjectives whose antonymic relations are indirect (e.g., *wet/arid*). Also, an opposition relation should be recognized more quickly when an antonym is compared to another adjective that is a near or close synonym of its antonym rather than a far synonym (e.g., *wet/arid*).
vs. *wet/baked*). As Gross et al.’s experimental results supported these predictions (see the original study for the details of the method), an implication for the current discussion is that processing a contradictory predicate like *open* might activate its antonym *closed* right away due to the associated link between them. Unfortunately, Kaup et al.’s (2006) experiment design did not incorporate a way to test this possibility, since they probed meaning representation at best after 750 ms had passed from the sentence offset.

The question of precisely when the respective meaning (e.g., either negated meaning or opposite meaning) is accessed or encoded, however, is not clearly understood. One limitation of Mayo et al.’s (2004) study is that it did not specify precisely when the opposite schema is accessed to constrain meaning representation of the negated bi-polar attributes, even though the authors’ claim is for an immediate access to the opposite schema. One finding that was not discussed is the reading times collected for the initial descriptions such as *Tom is (not) a tidy/romantic person*, which were presented as whole sentences for self-paced reading by the participants. The results showed that reading times were always faster for the affirmative sentences than for the negative sentences regardless of the attribute type (i.e., bi-polar or uni-polar). While the authors simply explained that this suggests more complexity associated with negative sentence processing, the source of slower reading times must be explicated, since they could indicate (a) a mere fact that negative sentences were longer by one word, thus taking longer to integrate meaning, (b) an initial encoding of the negated attribute (during sentence reading), which was then suppressed upon the retrieval of the opposite schema (e.g., upon processing the subsequent probe), or (c) concurrent activation of both negated/counterfactual attribute and an antonymic attribute, followed by resolution toward the antonymic attribute before getting to the behavioral probe. Depending on which possibility holds, the authors’ interpretation of the probe response times must be reevaluated to incorporate the prior processing during negative sentence reading.

The experiment described in this chapter aims to provide further information regarding exactly when an opposite schema is accessed in the context of sentence processing, as well as when and how negation impacts resulting meaning, by using an auditory sentence presentation that avoids the problem of reading times. This is because
the current literature, which is mainly composed of studies with reading paradigms, presents mixed information regarding the question of whether initial encoding of negative sentences is more difficult than that of affirmative sentences. While some studies have answered yes to this question (e.g., Carpenter & Just, 1975; Clark & Chase, 1972; Mayo et al., 2004; Trabasso et al., 1971), a growing body of research shows that negation processing is no more difficult than processing affirmative sentences when discourse context supplies pragmatic information that increases felicity of the negated sentence (e.g., De Villiers & Flusberg, 1975; Glenberg et al., 1999; Lüdtke & Kaup, 2006). In particular, Glenberg et al. (1999) argued that previous claims for the elevated difficulty associated with negation processing are largely due to the use of a sentence verification task, which requires truth-value computation. MacDonald and Just (1989) provided evidence for this view, reporting that the difficulty associated with negation was borne out only when they measured truth-value computation and not when they measured the reading times on the negation region, as well as the sentence end region including the last three words in the sentence (in a self-paced moving-window paradigm). However, Kaup et al. (2006), who tested mental simulation of negative sentences from reading, did not report information on reading times; thus, it is difficult to know whether the factual-state match-facilitation effect found from the 1500 ms ISI condition represents correct timing information.

Regarding meaning activation and representation related to negated concepts, MacDonald and Just (1989) proposed that negative sentence processing includes discourse-level processes, where the resulting representation reflects contextually integrated meaning. The authors used a probe recognition as well as a naming task (in their Experiments 1 and 2–3, respectively) to compare the activation level of a noun concept that was expressed in either an affirmative phrase (e.g., some bread) or a negative phrase (e.g., no bread). They used sentences like Almost every weekend, Elizabeth bakes [some cookies but no bread / no bread but only cookies] for children, where the negation word appeared before either the first (N1) or the second (N2) noun in the direct object phrase. The participants’ job was to read the sentence word-by-word in a self-paced manner and then either (a) indicate whether a subsequent probe (i.e., either the
negated or non-negated noun: e.g., *bread, cookies*) had occurred in the original sentence (in Experiment 1), or (b) simply name the probe word (in Experiments 2 and 3) as quickly as possible. Afterwards, participants also had to answer the truth value of the comprehension statements (e.g., *Elizabeth bakes bread for the children*). The results showed that, regardless of the position of the negated noun (N1 or N2), people were slower to react to the probes when the probe word represented the negated noun rather than the non-negated noun, which suggests that negation suppressed the activation level of the following noun’s concept. Interestingly, however, they found that this activation-reducing function of negation does not impact the activation of the negated word’s associates (Experiment 3). That is, even though the negation word *no* reduced the activation level for *bread*, it did not do so for a conceptual associate of bread such as *butter*. With these findings, the authors argued that “the processes involved in encoding negation of a noun change the activation level of the noun’s representation” and “negation effect is relatively long-lived,” since negation effect was found regardless of whether the N1 or the N2 position was probed and even when the probe word appeared more than three words after the negation marker (pp. 638–639).

MacDonald and Just’s (1989) finding clearly suggests that the availability of the negated concept is reduced in the resulting mental representation. However, there was no dependent measure that showed precisely when the negation marker reduced the following word’s activation level. This raises two questions. First, does negation reduce the following word’s activation level at the time the word (e.g., *bread*) is encountered? Or, does this happen sometime after the following word meaning (e.g., bread) is accessed? Because the probe response times were measured at the sentence offset (which was more than three words apart from the negation), it is hard to know when the negation impacted the availability of the following word’s concept. The fact that the negated word’s associate, such as *butter*, was still available suggests the second option as a possibility, although this is an empirical question that needs further research. Assuming that the conceptual effect of negating a noun is no different from that of negating a predicate, the current study further investigated the availability of the negated concept, by using sentences with contradictory predicates (see Section 3.2.1 and 3.2.2).
In summary, the previous research discussed thus far does not present a clear picture for the time course and nature of negated information processing. That is, it is unclear whether negation processing can be explained by two distinct steps, or whether it can be better explained via a more general processing mechanism that takes multiple constraints (e.g., types of negated concept) into account. Moreover, the differences in materials and tasks across the studies make it difficult to infer any conclusion. Therefore, the current dissertation’s Experiment 2 investigated how negative sentences with contradictory predicates are processed, using an end-of-sentence picture naming task similar to Kaup et al.’s (2006), but with four major differences: (a) testing language and (b) modality (spoken English vs. written German), (c) the types of sentences tested (contrastive/affirmative/negative vs. affirmative/negative), and (d) the number of ISIs where meaning representations were probed (six ISIs vs. two ISIs). The focus is to explore the extent to which mental simulation captures subtle differences in meaning that are driven by explicit negation (i.e., \textit{not}) or implicit negation (i.e., contrastive prosody), in the process of constructing an implied meaning. The next section describes the details of Experiment 2.

3.2 Experiment 2

3.2.1 Design

The current experiment selected 30 test sentences from Experiment 1A, each of which was prepared in three forms, contrastive, affirmative, and negative neutral. The goal was to investigate the types of meaning representations evoked by each sentence, as well as the time it takes to arrive at those meanings. The most critical question was whether listeners would mentally simulate the state-contrast implicature evoked by prosody, and if so, through what process and time course. Previous studies have shown that objects’ implied attributes such as shape, color, and orientation appear during mental simulation (e.g., Connell, 2007; Stanfield & Zwann, 2001; Zwaan et al., 2002). However, existing evidence on mental simulation comes mostly from processing lexical cues, and to my knowledge, there is no study investigating the effect of sentence prosody on mental simulation. As spoken language is one of the main modes of communication and prosody
is a vital factor influencing speaker’s meaning, the current study aimed to understand to what degree prosody affects dynamic meaning activation and integration in the course of mental simulation. Moreover, the impact of contrastive prosody was compared in both affirmative and negative sentences to that of neutral prosody, in order to illuminate to what extent similar or different processes are involved in comprehending sentences that differed by only prosody or negation.

The 30 sentences were crossed with two types of images that depicted the subject entities mentioned in the sentences. Table 3.1 presents the resulting six conditions. For contrastive and affirmative sentences, Image 1 (e.g., a full mailbox), or the “mentioned-state image,” represented the sentential-subject entity in a state that was asserted by the predicate adjective, (e.g., *The mailbox was full*). For the negative neutral sentences, this same image type, Image 1 (e.g., a full mailbox), represented the state that was actually negated in the sentence as in *The mailbox was not full*; thus, Image 1 for negative sentences is referred to as the “counterfactual-state image.” On the other hand, Image 2 (e.g., an empty mailbox) depicted a state that is the opposite from what was originally asserted in the contrastive and affirmative sentences, and is therefore called the “opposite-state image”. For the negative neutral sentences, the depiction in Image 2 (e.g., an empty mailbox) matched the inferred factual state, i.e., one possible truthful state when the negation is integrated into the sentence meaning. This was therefore called the “factual-state image.” These six conditions were manipulated within participants and items.
Table 3.1. Six experimental conditions crossing 3 Sentence types with 2 Image types.

<table>
<thead>
<tr>
<th>Cond.</th>
<th>Sentence</th>
<th>Image</th>
<th>State of affairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 Contrastive</td>
<td>L+H* L-H%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>The mailbox WAS full.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 Affirmative Neutral</td>
<td>H* H*L-L%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>The MAILBOX was FULL.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3 Negative Neutral</td>
<td>H* H* H*L-L%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>The MAILBOX was NOT FULL.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2 Neutral</td>
<td>H* H*L-L%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>The MAILBOX was FULL.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3 Neutral</td>
<td>H* H*L-L%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>The MAILBOX was NOT FULL.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2 Neutral</td>
<td>H* H*L-L%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>The MAILBOX was FULL.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For all critical trials, the depicted entities were to be named by the participants. Therefore, lexical properties of the entity names were carefully controlled for any unintended source of variance. For example, all names consisted of one or two syllables at most and always carried an initial lexical stress. In addition, words’ frequencies, as well as naming reaction times (RTs) to those isolated words were closely matched on the basis of the English Picture Naming Database collected as part of the International Picture Naming Project (IPNP; Szekely et al., 2004).

Importantly, the current study probed meaning representation at six different ISIs (i.e., 0 ms, 500 ms, 1000 ms, 1500 ms, 2000 ms, and 2500 ms). As introduced in Section 3.1, ISI (Inter-Stimulus Interval) refers to the predefined time lag (in milliseconds) between the sentence offset and the image onset. Testing six different ISIs was motivated by three reasons. First, there is no precedent study showing exactly when prosody induces the state-contrast implicature. The most relevant reference is the result from Experiment 1A showing that participants who frequently generated this implicature spent nearly 1.5–1.6 seconds to produce continuation sentences. However, this information was taken from the onset of typing, which began after participants heard the sentence twice.
Thus, there is no precise information regarding how long it will take participants to activate and represent the state-contrast meaning.

Second, testing six ISIs was to investigate the ways in which mental representations of meanings change over the time course of meaning activation and integration. Previous studies have suggested that meaning representations at different time points reflect different levels/depths of semantic encoding. For example, it has been claimed that at the time of or right after lexical access, semantic features that are generally related to individual words are activated regardless of discourse/sentential context (Conrad, 1974; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979), whereas after a deep semantic integration, only those semantic features that are contextually relevant remain activated (Merrill, Sperber, & McCauley, 1981). For instance, reading a word *cat* activates both *fur* and *claw*, but a sentence like *The girl touched the cat* activates *fur* more strongly than *claw* while a sentence like *The girl fought the cat* does the opposite. Merrill et al. (1981) found such results, but crucially, the results reflect meaning representations examined at 1000 ms after either the single-word context or the sentential context. That is, given the same amount of meaning integration time, only the sentential context could successfully encode contextually relevant semantic features, whereas the single word context activated any general features associated with the word (Merrill et al., 1981). This finding suggests that the time window of 1000 ms from the sentence offset is when individually activated meanings in the sentence are tightly integrated and meaning enrichment takes place (see also, Mirković & Altmann, 2008). However, some other studies have also shown that selective meaning access is possible when the sentential or discourse context is sufficiently constraining (e.g., Paul, Kellas, Martin, & Clark, 1992; Tabossi, 1988; Tabossi, Colombo, & Job, 1987). For example, a homonym like *port* can immediately activate only its dominant meaning when the preceding sentential context provides biasing information toward that meaning as in *The violent hurricane did not damage the ships which were in the port...* (Tabossi et al., 1987). Therefore, it is important to investigate with the same set of test materials the extent to which semantic representation changes over the time course of processing.
The third reason to test six ISIs was to make data interpolation and interpretation processes more reliable. As previously practiced by Kaup et al. (2005, 2006; Kaup et al., 2007), among others, it is possible to test just two or three ISIs (e.g., one short and one long ISI) and infer meaning representations happening in between those time points. However, such an interpolation process could potentially disguise dynamically changing meaning representations. As mentioned above, the conclusion for the two-stage negation model in Kaup et al. (2006) was grounded on the null effect for negation at 750 ms together with the match-facilitation effect for the factual-state at 1500 ms. However, this finding cannot guarantee that there was actually a dominant representation for a counterfactual state at an ISI earlier than 750 ms. As suggested in various studies discussed in Section 3.1, it is instead possible that people actually represent the factual state as soon as they finish with the sentence, perhaps due to the activation spreading from the contradictory adjective to its antonymic partner, in conjunction with the function of negation that reduces the activation level of the negated word’s concept, but not the associated concept.

In consideration of these points, testing six evenly distributed ISIs will provide more data points, making for more reliable data interpolation. Of course, the best solution is to have a dependent measure that is closely time-clocked to the ongoing meaning processing. In Chapter 4, I present this dissertation’s last experiment, which used such a measure—people’s eye-movements collected during language processing. Also, the ISI variable in the current experiment was tested as a between-subjects factor, meaning that different participants were randomly assigned to one of the six ISIs conditions. This is because previous studies have reported that varying ISIs within an experiment often disrupted participants’ attention for the task (e.g., Kaup et al., 2006). The eye-tracking experiment presented in Chapter 4 will complement the current experiment’s findings by accumulating further evidence for the time course of implicature processing within participants and items.
3.2.2 Predictions

3.2.2.1 Negative neutral sentences

If mental simulation of negative sentence meaning requires a simple two-step sequential processing as suggested in Kaup et al.’s 2006 model, then naming times should be significantly smaller for the pictures that depict the counterfactual state than they are for the factual state at earlier ISIs, whereas the opposite pattern should be found from naming times collected at longer ISIs. This pattern of finding will confirm that people initially focus on the counterfactual state regardless of the type of negated information, before turning their attention onto a factual state (Kaup et al., 2006, 2007).

However, drawing an inference for a possible factual state from the negation processing requires comparisons among possible alternatives, and therefore, people might benefit from simultaneous meaning contrast initially, to infer a factual state. This hypothesis is plausible especially when the contrastive notions expressed by words and phrases are tightly associated, and thus easily accessible. Related to this view, Israel (2004) argues that the contrastive relation innate to the antonymic word pairs is grounded on the conceptual association developed through perceptual experience. For example, as we grow up, we learn to distinguish good behavior from bad behavior, hot weather from cold weather, and long fingernails from short fingernails, and these notions co-occur and develop as part of schematic knowledge for an event, topic, or entity. If this is the case, listening to a word like full might immediately activate its antonymic notion like empty because these notions are part of the same event schema. Indeed, Gross et al.’s (1989) study on lexical access mentioned in Section 3.1 provides initial evidence on this view.

In negative sentences like The mailbox was not full, however, the predicate, full, is preceded by a negation marker not, which is known to reduce the activation level of the negated word’s concept but not its associate’s concept (MacDonald & Just, 1989). If this is the case, the activation level for the predicate that is negated in the sentence, full, will be lower than that of its antonym empty, thus making the concept of the factual state relatively more accessible (i.e., smaller naming times to the picture of an empty mailbox than a full mailbox). While we might see such an effect as early as the sentence offset (i.e., at 0 ms ISI) as seen in MacDonald & Just’s (1989) negated noun results, it is hard to
predict the exact timing information since there is no precedent study that tested negative sentences with contradictory predicates at 0 ms ISI.

Also, how long the reduced activation of the negated concept will persist is an empirical question. However, with these particular materials containing the contradictory predicates, the concept of the negated word’s antonym expresses one of the most plausible factual states. Thus, the pattern of initial automatic activation from individual word meanings and incremental construal of those meanings matches the representation resulting from meaning integration. This suggests that naming times will be smaller when the picture matches the factual state, even at a longer ISI. Moreover, because this is based upon meaning integration, this effect might be stronger than the effect based on meaning activation.

3.2.2.2 Contrastive sentences

If mental simulation of an inferred meaning not only from explicit negation but also from other cues in general requires a two-step sequential processing that involves a transition from the literal meaning to an implied meaning, then sentences spoken with the contrastive prosody L+H* L-H% will lead to an initial access to the described state (e.g., *The mailbox was full*) and then to the simulation of an implied state that contradicts the asserted state (e.g., but the mailman came and now it is empty). If this is the case, people will initially be faster at naming a picture of the mentioned state (Image 1) than a picture of the opposite state (Image 2). People will then construct the state-contrast implicature and thus show shorter naming times for a picture of the opposite state than the mentioned state at a longer ISI.

However, this simple two-step processing of a prosody-induced implicature suggests an interpretive mechanism that proceeds with lexical information initially before making use of prosodic information for meaning enrichment. On the contrary, a growing body of research shows that prosodic information is processed in parallel with segmental information, thus making an impact on meaning activation as each syllable or word comes in (Dahan et al., 2002; Ito & Speer, 2008). The earliest moment that the current task measures meaning representation is at 0 ms ISI, i.e., as soon as the auditory sentence

86
is finished. If mental simulation rapidly integrates meaning evoked by both lexeme and prosody, then by the time participants process the predicate adjective, their mental representation should contain the mentioned entity (e.g., in its canonical shape, see Dahan & Tanenhaus 2005 for individual word processing activating canonical shape information; or in a form that they recently experienced that entity, Zwaan, 2004, see Section 3.3), as well as time and state information from processing the auxiliary was because the auxiliary expresses both of these meanings. While the detailed mechanism of representing abstract information like time in an experiential format requires more research, existing literature suggests that comprehenders keep track of the temporal sequence of an event in their mental representation (Barsalou, 1999; Gernsbacher, 1990; Zwaan, 1996; Zwaan, Madden, & Whitten, 2000), and some possible ways to do so include (a) taking time as a signal to create a mental space or structure (e.g., Fauconnier, 1997; Gernsbacher, 1990), or (b) time can be projected via some concrete domain that utilizes spatiotemporal regions (e.g., Boroditsky, 2001; Zwaan & Radvansky, 1998). In any case, the contrastive pitch accent L+H* on the auxiliary will assist the activation of not only the past time was but also some other alternative time frames (e.g., is, will be, etc.), as well as alternative state information (e.g., wasn’t, isn’t, will not, etc.). Lastly, the sentence finishes with a contradictory predicate, and as mentioned above, the contradictory predicate is likely to activate both the mentioned attribute and its opposite attribute due to the associative link between those opposing attributes (Gross et al., 1989; Israel, 2004). This predicts that naming times at earlier ISIs might not show any dominant pattern of match facilitation to either type of picture, since people will be simultaneously accessing the state mentioned by the predicate and the state suggested by its antonym.

However, there is also the sentence-final continuation rise L-H%, which signals a search for the forward reference or logical consequence of the asserted event (Pierrehumbert & Hirschberg, 1990). Since the current experiment does not provide external information or context supplying such information, people may construct an implicature to fill in that information. In doing so, people will likely make use of previously activated meanings which include different time frames as well as the binary alternative states, because they are more accessible than other meanings. I predict that the
conclusion that people draw for the implicature will reflect their background knowledge, i.e., what they already know of the target entity or an event; for example, a full mailbox becoming empty after a mailman takes the mail away (see Section 3.3 for more discussion on this point). These internal processes related to building an implicature will certainly require time. Therefore, I predict that the evidence for an implicature will emerge only at a later ISI, for example, when all lexical meanings are already encoded and more processing time is allocated for meaning integration prompted by the L-H%.

When compared to negative sentences, I predict that the inference effect might be weaker with contrastive sentences than negative sentences. The meaning of the negation word not is likely to be less invariable in the minds of native English speakers, whereas the use of prosodic cues requires a successful mapping between text and tune information, which could be more susceptible to additional factors such as attention level. Also, the critical cue not in the negative sentences unfolds temporally earlier than the critical inference cue L-H% in the contrastive condition. That is, not is the penultimate word in the sentence, whereas the L-H% cue in the contrastive condition unfolds over the last word of the sentence (i.e., the contradictory predicate). My analysis on the sound materials shows that there is on average 363 milliseconds difference between the onset of the negation marker not and the earliest onset of the rising cue L-H% (i.e., the low pitch dip). This suggests that the inference effect (i.e., faster naming times for the pictures depicting the state opposite from the mentioned state) might show up at least 363 ms earlier in the negative condition than in the contrastive condition.

3.2.2.3 Affirmative neutral sentences

A main function of neutral prosody is to instantiate new topics into the discourse (e.g., Cruttenden, 1997; Ladd, 1996; Pierrehumbert & Hirschberg, 1990). Therefore, comprehending affirmative sentences in neutral prosody should let people accept what was asserted in the sentence. In this case, people’s naming times should be smaller for the pictures depicting the mentioned-state than for those depicting the opposite-state. Previous studies on affirmative sentence reading indicate mental simulation of a described meaning. That is, when the meanings suggested by the sentence and the
subsequent stimulus possessed compatible properties, a match-facilitation effect was found (i.e., shorter response times in the matching condition). And the timing of finding such an effect from those reading studies includes a 0 ms ISI (i.e., right after the offset of sentences like *The elephant is above the giraffe*; Kaup et al., 2005), as well as a 250 ms ISI (e.g., Zwaan et al., 2002 with sentences like *The ranger saw the eagle in the sky*) and a 750 ms ISI (e.g., Kaup et al., 2006 with sentences like *The door was open*).

However, the affirmative sentences also contain the contradictory predicates that could easily activate antonymic alternatives. To my knowledge, no studies on mental simulation have tested affirmative sentences containing contradictory predicates at an ISI earlier than 750 ms. Therefore, no predication can be made precisely when and whether the match facilitation toward the mentioned-state meaning will be found from affirmative sentences with contradictory predicates.

To summarize, I predict that the word *not* that is present only in the negative sentences will create a differential processing pattern for the negative sentences on the one hand and the contrastive and affirmative sentences on the other hand. That is, only the negative sentences contain the word *not* that is known to reduce the activation level of the following word’s concept. As for the negative sentences and contrastive sentences, they may show similar result patterns especially at longer ISIs, due to the linguistic cues that assist in evoking contrast and/or inference (e.g., *not* vs. L+H*/ L-H%). Lastly, when compared to the affirmative sentences, only the contrastive sentences will create the state-contrast implicature due to the combination of the accented auxiliary and the final rising cue. If any, the affirmative sentences might indicate some level of activation of the antonymic concepts due to the presence of the accented contradictory predicates, but in the absence of any cue whose function is designated for contrast or inference, there will be no reliable inference effect.

### 3.2.3 Norming study

Two norming studies checked whether there was any initial bias toward each subtype of the sentences or images among the test items. Since Experiment 1A served as the norming study for the test sentences, this section describes the picture norming study.
only. Thirty-four pairs of pictures (32 test pairs and two filler pairs) depicted two opposite states for an entity (e.g., an old car vs. a new car), whose names were controlled across the set to have limited variation in the number of syllables, lexical stress, frequency, and raw naming times (IPNP; Szekely et al., 2004). These pictures were counterbalanced to create two experimental lists, and 16 native English speakers were each randomly assigned to one of the two lists.

After signing the consent form, each participant was seated in front of a computer and received two short training sessions. First, participants named alphabet letters one at a time to adjust the sensitivity of the voice key, which was operated by an E-prime button box connected to a handheld microphone. Then, participants practiced the main task with four practice trials. Once the main session had started, the E-prime software (version 1.2, Psychology Software Tools, Inc.) randomly presented each image in the center of the screen, one at a time. As participants spoke the name of the depicted entity out loud as quickly as possible, the naming onset time was automatically recorded into the computer. Naming one image removed it from the screen and called up the next one.

Among the 32 test pairs, 30 pairs were named as intended with at most two errors per condition for each participant (88% accuracy). The other two pairs (the images of a gift and of a stocking) were frequently named by other words (e.g., present for gift, sock for stocking) and thus subsequently excluded from the analysis. The naming times for the 30 remaining pairs were submitted to two paired t-tests to compare means in each condition for participants and items analyses. There was a marginal effect by participants ($t_1(1,15) = -1.94, p = .07$) due to mean naming times that were smaller for Image 1 than Image 2. For the items analysis, however, there was no effect of condition on naming times ($t_2(1,29) = -1.2, p = .24$). These 30 pairs were used for the main experiment (see Appendix B).

### 3.2.4 Materials

Besides the 30 sentence-picture pairs for critical items, 70 filler trials were prepared to disguise the purpose of the experiment. First, filler sentences never mentioned the pictured entities to offset the obvious relationship between the sentence
and the picture in critical trials, in which the items in the pictures were always mentioned. Instead, filler pictures were either completely unrelated to the sentences (50 fillers), or related to the sentential subjects or predicates either phonologically or semantically (20 fillers).

In addition, the proportion of neutral versus non-neutral prosody was balanced. While one third of the critical sentences were spoken with the L+H* L-H% tune, this prosodic pattern is reported to be less common than neutral prosody in language corpora such as the Boston University Radio News Corpus (Dainora, 2001, 2002). Therefore, 50 filler sentences were embedded in a range of neutral contours. The remaining 20 filler sentences then included some other non-neutral contours, which placed different types of pitch accents in different locations and also used different types of sentence edge tones. All of the critical and filler sentences were produced by a highly-trained female phonetician who is an expert in the Mainstream American English ToBI system. These sentences were evaluated for the appropriate realization of the intended prosody (see Section 2.2.1 for phonetic and phonological analysis for the test sentences).

Also, half of the filler trials (i.e., 35 fillers) were followed by comprehension questions for two reasons. First, the questions distributed participants’ attention to different parts of the sentence meaning, considering that critical sentences always mentioned the critical pictures in the subject position. Hence, 25 questions asked something about the state of affairs of the entity mentioned (e.g., How was the toothpaste? after the sentence The toothpaste was minty), whereas the remaining 10 questions asked something about the sentential subject (e.g., What was yummy? when the sentence was The donut was yummy). Second, comprehension questions added different levels of difficulty to ensure that participants comprehended sentences for meaning, rather than focusing only on the picture naming task.

3.2.5 Participants and procedures

Three hundred and thirty-three students at the University of Hawai‘i at Mānoa were each randomly assigned to one ISI version of the experiment either for course credit or $5 in compensation. Due to Hawai‘i’s multilingual environment, both pre-experiment
solicitation and a post-experiment language background questionnaire ensured that participants were native English speakers who learned only English before age five.

Upon completing a consent form, each participant was seated in front of a computer to practice how to use the voice key, which was operated by a handheld microphone connected to an E-prime button box, then to the computer. As the speech onset activated the voice key, the onset time was automatically recorded into the computer that was controlled by the E-prime software, version 1.2 (Psychology Software Tools, Inc.). Once the voice key was properly adjusted for the sensitivity level, participants practiced the main task with four practice trials. All items in the practice session as well as the main session were presented in a completely randomized manner for each participant.

For each trial, participants listened to a sentence while the computer monitor remained blank. The sentence offset led to a predetermined pause (i.e., ISI), followed by a picture appearing at the center of the computer screen. Participants verbally identified the depicted entity as quickly and accurately as possible by uttering a single noun (e.g., “mailbox”), and as they did so, the speech onset was automatically recorded as the picture simultaneously disappeared. After the naming task, comprehension questions occasionally appeared on the screen (i.e., after 50% of the filler trials), and participants provided verbal answers to those questions. All verbal responses to the pictures and questions were recorded into a digital voice recorder (model WS-311M; Olympus Imaging Corporation). A coder subsequently coded the responses to check the accuracy of both picture naming and answers to the comprehension questions. Each experiment session lasted for less than 25 minutes; the time it took to finish the experiment was positively correlated with the length of the ISI between the sentence and the picture.

3.2.6 Results

Data from 27 people (8% of total data) were eliminated due to: (a) lip smacks and naming errors resulting in less than three observations per test condition (3%), (b) mean naming times exceeding 2.5 standard deviations from participants’ grand mean (2%), and (c) comprehension question accuracy lower than 80% (3%). The remaining data went
through two steps of the outlier cleaning procedures. First, naming times outside the range of 350 ms and 1400 ms were omitted due to the scarcity in data distribution. Then, naming times exceeding 2.5 standard deviations from each participant’s mean were replaced with the respective participant’s mean plus 2.5 standard deviations from that mean (Ratcliff, 1993). This removed a total of 1.45% of the data from the 306 remaining participants.

The dependent measure was the naming time, which was measured from the image onset to the naming onset in each test trial. All valid data were fitted into several linear mixed-effects regression models using the \textit{lmer} function from the statistical software R (R Development Core Team, 2007). For each model, every valid data point was included without any need to aggregate them for either participant or item. Moreover, error variances innate to those participants and items were handled by two random-effect terms, while the test factors of ISI, Image, and Sentence were included as fixed-effect terms. The regression models calculated coefficient values between the dependent variable (i.e., naming times) and each of those testing factors. For each fixed-effect term, the models took the Affirmative Sentence, Image 1, and 0 ms ISI as the default levels; thus, the intercepts in the modeling results represent the estimated group means that considered those default levels (for more details about the modeling, see Baayen 2008). This section reports the results from the fixed-effect terms only. Also, the significance of an intercept estimate is not discussed due to lack of theoretical interest, i.e., it always tests whether the intercept was significantly different from zero.

All valid data were first fitted into a model including ISI and Image as two fixed-effect terms, in order to evaluate whether naming times were influenced by any unintended initial bias toward either Image type. The results confirmed that the Image type itself did not influence the naming times in any significant way, nor did Image type interact with ISI. Based upon this result, I present three mixed-effects models that were built for each Sentence type, as a product of 6 ISIs and 2 Image types. After that, naming times collected for each Image type were separately modeled, as a product of 6 ISIs and 3 Sentence types. Then, a full model compared the effects of Sentence, Image, and ISI types in a single model.
3.2.6.1 Results from the affirmative sentence model

Table 3.2 summarizes participants’ data collected when the affirmative neutral sentences preceded either Image type across the 6 ISI versions of the experiments. The mean naming times (in milliseconds) in each condition are plotted in Figure 3.2, where the unfilled data points indicate mean naming times for the images depicting the mentioned state, i.e., Image 1, as in a full-mailbox image after a sentence like *The mailbox was full.* The filled data points represent naming times for Image 2, which depicted a state that is the opposite of what was stated originally (e.g., an empty mailbox).

Table 3.2. Participants’ mean naming latencies/standard errors/standard deviations (in milliseconds) as a function of 2 Image types and 6 ISI conditions, collected after the Affirmative Sentences.

<table>
<thead>
<tr>
<th>ISI (ms)</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affirmative Neutral sentence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image 1: Mentioned state</td>
<td>Mean</td>
<td>738.28</td>
<td>728.27</td>
<td>763.82</td>
<td>772.82</td>
<td>782.90</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>17.12</td>
<td>14.63</td>
<td>17.74</td>
<td>18.17</td>
<td>16.95</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>125.77</td>
<td>107.48</td>
<td>130.38</td>
<td>133.52</td>
<td>124.53</td>
</tr>
<tr>
<td>Image 2: Opposite state</td>
<td>Mean</td>
<td>741.72</td>
<td>728.67</td>
<td>761.37</td>
<td>754.74</td>
<td>765.15</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>14.77</td>
<td>13.80</td>
<td>18.98</td>
<td>16.21</td>
<td>15.85</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>108.54</td>
<td>101.41</td>
<td>139.48</td>
<td>119.09</td>
<td>116.46</td>
</tr>
</tbody>
</table>

Figure 3.2. Mean naming times in milliseconds (y-axis) collected after the Affirmative sentences as a function of 2 Image types at each of the 6 ISIs (x-axis).
As seen in the figure, mean naming RTs at early ISIs were nearly identical, suggesting no initial match facilitation for either Image type. While the mean naming times to the two Image types indicated numerical differences after 1000 ms ISI, the Affirmative-sentence model that fitted all valid data points did not find this pattern to be statistically valid. In Table 3.3 below, the intercept represents the model’s estimated mean for Image 1, which was calibrated for RTs at 0 ms ISI. The model found a very small main effect of ISI, suggesting that naming RTs were greater in general when they were collected at longer ISIs than shorter ISIs ($p = 0.096$). However, this pattern did not interact with the Image type, nor had the two Image types exerted any differential influence to the naming RTs (i.e., the coefficient of the ISI is shared for both Image types). Subsequent modeling for the data collected just at 1500 ms and 2000 ms found that the numerical RT differences suggested by the averaged values were not statistically valid at either ISI.

Table 3.3. The Affirmative model’s fixed effects; the intercept shows the estimated group mean for Image 1 (i.e., mentioned-state) at the 0 ms ISI.

| Fixed effects | Estimate | Std. Error | $t$ value | Pr(>|t|) |
|---------------|----------|------------|-----------|---------|
| (Intercept)   | 741.565  | 16.234     | 45.68     | 0.000   |
| ISI           | 0.014    | 0.008      | 1.66      | 0.096   |
| Image2        | −0.376   | 10.084     | −0.04     | 0.970   |
| ISI*Image2    | −0.005   | 0.007      | −0.67     | 0.506   |

These results indicate that processing affirmative sentences activated representation of two states of affairs immediately, i.e., one for the mentioned state and the other for the unmentioned opposite state, and there was no initial advantage for the mentioned state even at an early ISI. This suggests activation spreading from the mentioned contradictory predicate to its unmentioned antonym. The results also indicate that neither the unaccented *was* nor the H*-accented predicate could stably induce an inference even at a longer ISI, which allowed sufficient time to integrate meaning from the sentence before the image was presented. An unexpected finding is that there was no match-facilitation effect for the described state at any point in time after the affirmative sentence offset. Section 3.3 discusses some possible reasons for this finding.
3.2.6.2 Results from the contrastive sentence model

Table 3.4 and Figure 3.3 below present data collected from the contrastive sentences.

Table 3.4. Participants’ mean naming latencies/standard errors/standard deviations (in ms) as a function of 2 Image Types and 6 ISI conditions, collected after the Contrastive Sentences.

<table>
<thead>
<tr>
<th>ISI (ms)</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>36</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>15.14</td>
<td>13.09</td>
<td>19.94</td>
<td>18.84</td>
<td>15.36</td>
<td>20.18</td>
</tr>
<tr>
<td>S.D.</td>
<td>111.26</td>
<td>96.20</td>
<td>146.53</td>
<td>138.44</td>
<td>112.88</td>
<td>121.07</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>20.46</td>
<td>15.30</td>
<td>17.55</td>
<td>17.19</td>
<td>15.39</td>
<td>19.94</td>
</tr>
<tr>
<td>S.D.</td>
<td>150.37</td>
<td>112.46</td>
<td>128.98</td>
<td>126.35</td>
<td>113.06</td>
<td>119.65</td>
</tr>
</tbody>
</table>

Figure 3.3. Mean naming times in milliseconds (y-axis) collected after the Contrastive sentences as a function of 2 Image types at each of the 6 ISIs (x-axis).

While the mean plot’s appearance approximated that of the affirmative sentences, only the Contrastive statistical model produced two significant effects (Table 3.5). First, the main effect of ISI suggested that naming RTs were significantly greater when they were collected at longer ISIs than shorter ISIs. More importantly, however, the ISI effect interacted with the Image type, such that the rate of the increase in naming RTs across ISIs was significantly smaller for Image 2 than Image 1 ($t = -2.01$, $p = 0.045$).

Subsequent models testing data at each ISI found that naming RTs were significantly smaller when the depicted images represented the reversed-state (Image 2) than they were
for the mentioned state (Image 1) at a 2000 ms ISI (coefficient: \(-36.34, \text{SE}: 13.24, t = -2.75, p = 0.0063\)). This suggests that the initial competition between two activated meanings led to a later dominance for the implied meaning when 2000 ms had elapsed from the sentence offset. There was no indication of a crossover point from the literal interpretation to the implicature, despite intuition for such a serial processing. This indicates that the state-contrast implicature evoked from the tune L+H* L-H% on sentences with contradictory predicates results from parallel meaning activation initially, followed by an implicature construction at a later point. This finding’s implications are further discussed in the discussion section.

Table 3.5. The contrastive model’s fixed effects; the intercept value shows the estimated mean naming times for Image 1 (i.e., mentioned-state) at the 0 ms ISI.

| Fixed effects | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 734.135  | 16.039     | 45.77   | 0.000    |
| ISI           | 0.018    | 0.008      | 2.13    | 0.034    |
| Image 2       | 7.039    | 10.434     | 0.67    | 0.500    |
| ISI: Image 2  | -0.015   | 0.007      | -2.01   | 0.045    |

3.2.6.3 Results from the negative sentence model

The negative neutral sentences produced somewhat different results. As seen in Table 3.6 as well as in Figure 3.4, naming RTs were always smaller when the image matched the factual state (Image 2) than when it mismatched.

Table 3.6. Participants’ mean naming latencies/standard errors/standard deviations (in ms) as a function of 2 Image Types and 6 ISI conditions, collected after the Negative Sentences.

<table>
<thead>
<tr>
<th>ISI (ms)</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>36</td>
</tr>
<tr>
<td>Pair 3: Negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image 1</td>
<td>Mean</td>
<td>756.65</td>
<td>758.02</td>
<td>772.25</td>
<td>800.98</td>
<td>776.28</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>18.57</td>
<td>15.56</td>
<td>19.64</td>
<td>19.29</td>
<td>17.06</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>136.44</td>
<td>114.31</td>
<td>144.33</td>
<td>141.76</td>
<td>125.33</td>
</tr>
<tr>
<td>Image 2</td>
<td>Mean</td>
<td>723.68</td>
<td>751.86</td>
<td>750.90</td>
<td>748.37</td>
<td>764.37</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>17.60</td>
<td>15.00</td>
<td>16.14</td>
<td>14.06</td>
<td>16.64</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>129.33</td>
<td>110.19</td>
<td>118.58</td>
<td>103.30</td>
<td>122.30</td>
</tr>
</tbody>
</table>
Figure 3.4. Mean naming times in milliseconds (y-axis) collected after the Negative sentences as a function of 2 Image types at each of the 6 ISIs (x-axis).

The statistical model for the negative sentence found only the main effect of the Image type, without any interaction of this effect with ISI. In fact, a likelihood ratio test (Baayen, 2008) suggested that adding the ISI factor to the negative sentence model did not improve a simpler model that included only one fixed-effect term: Image ($\chi^2 = 1.07$, $df = 2$, $p = 0.5856$).

Table 3.7. The Negative model’s fixed effects; the intercept value shows the estimated mean naming times for Image 1 (i.e., mentioned-state) at the 0 ms ISI.

| Fixed effects | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 764.058  | 16.372     | 46.67   | 0.000    |
| ISI           | 0.004    | 0.008      | 0.50    | 0.617    |
| Image2        | -26.136  | 10.238     | -2.55   | 0.011    |
| ISI: Image2   | 0.004    | 0.007      | 0.61    | 0.544    |

To further verify that the Image effect found only from the negative sentences, as well as the interaction of Image and ISI found only from the contrastive sentences, were not carried by outliers, I followed Baayen’s (2008) suggestion and fitted each sentence model to a subset of the data, which removed long RTs constituting 4% of the initially trimmed data. The models produced essentially the same, or even stronger effects: (1) no significant effect from the affirmative sentence model, (2) a main effect of Image (coefficient = 0.021, $SE = 0.009$, $t = 2.41$, $p = 0.016$) as well as an interaction of Image and ISI (coefficient = -0.017, $SE = 0.007$, $t = -2.50$, $p = 0.013$) from the contrastive
sentence model, and (3) a main effect of Image only for the negative sentence model (coefficient = −31.624, SE = 9.637, t = −3.28, p = 0.001).

Further analyses for the negative sentences at each ISI found that naming RTs were significantly smaller for the factual-state images (Image 2) than for the counterfactual-state images (Image 1) at 0 ms ISI (coefficient = −33.82, SE = 14.23, t = −2.38, p = 0.018), 1000 ms ISI (with the reduced data set having 4% data removal; coefficient = −28.88, SE = 14.44, t = −2.00, p = 0.046), and 1500 ms ISI (coefficient = −48.02, SE = 13.39, t = −3.59, p = 0.0004). There was no significance in naming time differences between the two Image types at either 2000 ms ISI or 2500 ms ISI. Naming RTs at 500 ms ISI were numerically smaller for the factual-state images than the counterfactual-state images, although this result didn’t reach statistical significance. This null effect found from the current study’s 500 ms ISI experiment is similar to the null effect from Kaup et al.’s (2006) negative sentence data collected at 750 ms ISI. A possible reason is discussed in Section 3.3.

To summarize the findings on the negative sentence models, the robust main effect of Image suggests that after processing negative sentences, participants in the current study accessed the factual meaning right away. As this result was repeatedly found regardless of several outlier trimming methods used, this result challenges a previous claim that negative sentences with contradictory predicates activate two states of affairs in a sequential manner (i.e., the counterfactual-state first and the factual state second), and that people mentally simulate both of those states by default (Kaup et al., 2006, 2007). Implications of these findings are discussed in Section 3.3.

### 3.2.6.4 Results from two image models

Each Image was separately modeled to evaluate the effect of sentence type across the 6 ISIs. For the model of Image 1 (which depicted the mentioned state for affirmative and contrastive sentences and the counterfactual state for negative sentences), there was a main effect of negative sentences, meaning that naming times to Image 1 were significantly greater across ISIs after participants had processed negative sentences than after they had processed affirmative sentences (coefficient for negative sentence =
22.054, $SE = 10.145, t = 2.17, p = 0.03$). This effect is in line with the finding mentioned above that naming times after negative sentences were significantly smaller for the factual state image (Image 2) than the counterfactual state image (Image 1). With respect to Image 2, however, there was no significant effect of sentence type or interaction between Sentence and ISI, indicating that contradictory predicates presented in all sentence types activated the opposite state to a similar degree. The lack of difference between affirmative and contrastive sentences in both Image models overall is attributable to the fact that both sentences contained the same lexical items.

Table 3.8. The Image 1 model’s fixed effects; the intercept value shows the estimated mean naming times for Affirmative sentences at the 0 ms ISI.

| Fixed effects         | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------------------|----------|------------|---------|----------|
| (Intercept)           | 742.361  | 17.752     | 41.82   | 0.000    |
| ISI                   | 0.014    | 0.009      | 1.59    | 0.113    |
| SentTypeContrastive   | -7.669   | 10.115     | -0.76   | 0.448    |
| SentTypeNegative      | 22.054   | 10.145     | 2.17    | 0.030    |
| ISI:SentTypeContrastive | 0.005   | 0.007      | 0.7     | 0.484    |
| ISI:SentTypeNegative  | -0.009   | 0.007      | -1.26   | 0.207    |

Table 3.9. The Image 2 model’s fixed effects; the intercept value shows the estimated mean naming times for Affirmative sentences at the 0 ms ISI.

| Fixed effects         | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------------------|----------|------------|---------|----------|
| (Intercept)           | 742.900  | 16.550     | 44.88   | 0.000    |
| ISI                   | 0.008    | 0.008      | 0.95    | 0.341    |
| SentTypeContrastive   | -0.936   | 9.950      | -0.09   | 0.925    |
| SentTypeNegative      | -3.481   | 9.931      | -0.35   | 0.726    |
| ISI:SentTypeContrastive | -0.005  | 0.007      | -0.66   | 0.508    |
| ISI:SentTypeNegative  | 0.001    | 0.007      | 0.13    | 0.893    |

3.2.6.5 Results from the full model

Figure 3.5 below plots mean naming times collected in all test conditions (i.e., 3 Sentence Types, 2 Image Types, and 6 ISIs). To compare the effects for Sentence, Image, and ISI in a single model, all valid data were fitted into a full model including those three factors as the fixed-effect terms, and participants and items as two random-effect terms (Table 3.10). The analysis found two effects. First, the significant main effect of negative sentences (coefficient: 21.686, $SE$: 10.261, $t = 2.11, p < 0.035$) suggested that naming RTs collected for Image 1 were significantly greater after processing negative sentences than affirmative sentences. However, a marginal interaction between negative sentences
and Image 2 (coefficient: −24.398, SE: 14.467, t = −1.69, p = .092) suggested that naming times collected after the negative sentences were smaller when the image matched the factual state (Image 2) than when it matched the counterfactual state (Image 1). These results suggest that soon after processing negation, the accessibility to the counterfactual state was substantially reduced, while the accessibility to the factual state was sustained.

The fact that there was no main effect of Image once again confirmed that Image types did not create any initial bias. In fact, the picture norming study showed a marginal advantage for Image 1 (only from the participant analysis), meaning that naming times were smaller for this Image type than for Image 2 when pictures were processed without sentences. Thus, any facilitation effect found toward Image 2 after sentence processing cannot be reduced to mere preferences for Image 1. This point is further supported by an additional analysis that took the difference values between naming times collected in the norming study and in the main study as the dependent measure, which found the essentially same result patterns. Besides, the fact that there was no three-way interaction between Contrastive: Image2: ISI and Negative: Image2: ISI suggests that the way affirmative sentences behaved with respect to Image2 and ISI was globally similar to the way the contrastive and negative sentences behaved to that image. This indicates that the presence of contradictory predicates across all sentence types elevated the availability of the unmentioned opposite state.

Table 3.10. The full model’s fixed effects; the intercept value shows the estimated mean naming times for the mentioned-state Image (Image 1) for affirmative sentences at 0 ms ISI.

| Fixed effects          | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------------|----------|------------|---------|---------|
| (Intercept)            | 742.529  | 16.260     | 45.67   | 0.000   |
| Contrastive            | −8.034   | 10.233     | −0.79   | 0.432   |
| Negative               | 21.686   | 10.261     | 2.11    | 0.035   |
| Image2                 | −0.576   | 10.251     | −0.06   | 0.955   |
| ISI                    | 0.013    | 0.008      | 1.60    | 0.111   |
| Contrastive:Image2     | 7.516    | 14.463     | 0.52    | 0.603   |
| Negative:Image2        | −24.398  | 14.467     | −1.69   | 0.092   |
| Contrastive: ISI       | 0.005    | 0.007      | 0.66    | 0.512   |
| Negative: ISI          | −0.009   | 0.007      | −1.26   | 0.207   |
| Image2: ISI            | −0.005   | 0.007      | −0.71   | 0.478   |
| Contrastive:Image2: ISI| −0.010  | 0.010      | −0.95   | 0.342   |
| Negative: Image2: ISI  | 0.009    | 0.010      | 0.91    | 0.361   |
Figure 3.5. Mean naming times in milliseconds (y-axis) as a function of 3 Sentence Types and 3 Image Types at each of the 6 ISIs (x-axis).

To summarize the results, Figure 3.6 plots mean naming-time differences in milliseconds (i.e., naming times collected after Image 1 minus naming times collected after Image 2) for each sentence type per ISI. The positive values indicate that naming times were smaller when the Image depicted the contradictory state than when it depicted the mentioned state (i.e., match facilitation toward the opposite state). The global patterns of the difference scores indicate the following. First, at earlier ISIs, both affirmative and contrastive sentences did not produce any match-facilitation effect toward either Image type due to the activation spreading from the contradictory adjectives between the mentioned state and the unmentioned contradictory state. On the other hand, the negation marker not in the negative sentences suppressed the activation level of the negated state, thus making the alternative’s activation level relatively greater even at 0 ms ISI. Second, when more time was allocated for meaning integration, both contrastive and negative sentences induced significantly smaller naming times when the images depicted the unmentioned contradictory state rather than the mentioned state, while the affirmative sentences did not produce any significant result. This suggests that the presence/absence as well as the nature of the linguistic cue designated for contrast and inference influenced both the timing and size of the effect. That is, the affirmative sentences with no distinctive marker of contrast and/or inference could not reliably induce match
facilitation for the contradictory state. However, contrastive sentences that contained exactly the same lexical items induced strong and reliable inference effects at 2000 ms due to the distinct function of L-H% along with L+H*. Sentences with an overt negation marker produced the strongest inference effect at 1500 ms, which appeared 500 ms earlier than the effect from the contrastive sentences. This difference in timing and manner of processing is further discussed in the discussion section below.

Figure 3.6. Mean difference in milliseconds (i.e., naming times for Image 1 minus naming times for Image 2) for each Sentence Type at each ISI.

3.3 Discussion

The results of Experiment 2 demonstrated that the L+H* L-H% tune induced the state-contrast implicature in the listeners’ minds, even when there was no production demand for generating such an inference. Although intuition as well as the sentence continuation products in Experiment 1 indicated an serial achievement of the implicature that emerged from the initial assertion, the contrastive tune modulated picture naming times in a manner that suggests parallel processing of multiple meanings initially, followed by a later resolution of the competition toward the implicature. There was no point in time where a match-facilitation effect was found for the mentioned-state image, suggesting that the literal assertion never became the dominant interpretation in listeners’ minds. On the contrary, a match facilitation was found for the implied state most significantly at 2000 ms ISI, suggesting that at this point, the implicature became the dominant interpretation.
The affirmative neutral sentences, however, didn’t produce any significant effect, despite the fact that these sentences contained exactly the same lexical items as the contrastive sentences. Naming times collected after short ISIs were statistically identical for both Image types, suggesting that processing the contradictory predicates also activated the asserted meaning together with its opposite meaning. At longer ISIs, while mean naming times were numerically smaller for the opposite-state image than for the mentioned-state image, this pattern did not reach statistical significance, indicating that sentences with an unaccented *was* as well as the H* accented predicate could not reliably induce the state-contrast implicature. The differences in findings between affirmative and contrastive sentences underscore the distinctive functions of prosodic elements such as L+H* and L-H% in higher meaning perception.

One unexpected finding from the affirmative sentences was that there was no point in time that produced the match-facilitation effect for the described state meaning. This is clearly different from the affirmative result found in Kaup et al.’s (2006) 750 ms ISI condition, where people’s naming times were smaller when the picture matched the mentioned state rather than the opposite state. This effect, however, disappeared at 1500 ms condition in Kaup et al. (2006). One thing to note is that, besides the language difference (i.e., German vs. English), Kaup et al.’s study and the current study differed in the modality of sentence processing. That is, whereas Kaup et al. (2006) presented sentences in text, the current study presented sentences auditorily, where a high peak accent H* was placed on both the subject and the contradictory predicates as in *The MAILBOX H* was FULL H* L-L%. Moreover, the pitch accent on the predicate represented the nuclear pitch accent (i.e., the last pitch accent within the intermediate phrase, which in this case coincides with the intonation phrase). It is known that a nuclear pitch accent receives the most prominence within the phrase (e.g., Beckman & Ayers, 1997; Veilleux et al., 2006), and the analysis of the current sound materials confirms this pattern (see Table 2.3). That is, within affirmative sentences, the F0 maximum as well as the pitch excursion value (F0 maximum minus minimum) was greater in the predicate than in the subject. Moreover, the F0 maximum value in the affirmative sentence condition was higher than the corresponding values in the other two conditions.
Interestingly, Michael, Keller, Carpenter, and Just (2001) presented brain imaging evidence suggesting that more semantic processing takes place when sentences are heard rather than read. This finding makes sense when we consider other studies claiming that accented materials induce deeper semantic processing (Sanford, 2002). As mentioned in Section 1.1, Chevallier et al. (2008) found that the disjunctive connector *or* evokes its secondary meaning as in “A and B but not both” more often when the connector is accented than when it is not. Also, Bartels and Kingston’s (1994) perceptual study, among others, showed that heightened pitch peak contributes to the perception of additional contrast. Lastly, Watson, Tanenhaus, and Gunlogson (2008) have recently found that the domain of interpretation between H* and L+H* overlaps, such that high peak accent H* can convey contrast despite its main function of introducing a new topic into the discourse. Thus, it seems that, in the current study, the binary lexical properties innate to the contradictory predicates, in conjunction with the accentual pattern placing a nuclear pitch accent on the predicates, created a unique effect. That is, some participants in some trials might have perceived contrast from the H* accent, which appeared on a lexical item that denoted binary attributes. Follow-up studies are underway to further investigate the extent to which sentence presentation modality contributes to the affirmative sentence processing.

The results from the negative sentences differed from those of both affirmative and contrastive sentences. The only significant effect was that of the Image type, where people’s naming times were always smaller when the picture depiction matched the factual state than the counterfactual state, regardless of the length of the processing time allowed after the sentence offset, even at the earliest ISI (i.e., 0 ms ISI). The fact that there was no ISI producing a match facilitation for the counterfactual-state image challenges previous studies supporting a two-stage model of negative sentence processing (Kaup et al., 2006; Kaup et al., 2007). Such studies have argued that representing the negated or counterfactual state is a necessary step to understanding (all types of) negative sentences; once participants have that representation, they can later turn their attention to the factual state by comparing the deviation between the negated world and the actual world (Kaup et al., 2006, 2007). Instead, the current study’s results suggest that negative
sentences with contradictory predicates can allow a rapid access to the factual-state meaning as soon as the sentence is finished. This was possible particularly due to the explicit negation word not reducing the availability of the negated concept. This point is further supported by the finding that, at early ISIs, naming times to both mentioned and opposite states were nearly identical when sentences lacked the negation word not, i.e., the contrastive and affirmative sentences.

Before discussing the processing mechanism that might be responsible for the current findings, one thing to note is that these overall results reflect meticulous data examination. First, to ensure that these results are not carried by certain participants or items, all valid data points from each condition per ISI were arranged into histograms to evaluate the data distribution patterns. Data in all conditions showed unimodal distribution patterns, without any hint of bimodal or multimodal patterns. Second, the same patterns of results were repeatedly found from data sets with a few different outlier trimming procedures, e.g., fixed cutoffs at 1400 ms or 1800 ms, with longer RTs replaced via 2.5 SD from the respective participant’s mean (Ratcliff, 1993), or outlier removals from each participant’s sorted data (blind to conditions) without any replacement. Third, after each data modeling, the distributions of residuals were examined to evaluate the fitness of the model (Baayen, 2008).

While these procedures confirm that the results reflect a general processing tendency from all participants and items tested in this experiment, this raises a question about the individual variation found in Experiment 1, where the rate at which continuations encoded the state-contrast implicature in the L+H* L-H% condition differed across participants and items. However, this could simply reflect that the dependent measure used in Experiment 1 (i.e., the rate of producing continuations expressing the state-contrast implicature) was insensitive to detect on-line meaning activation. Instead, producing continuations represented some level of conscious and effortful production processes aiming at naturally continuing discourse. In contrast, the dependent measure used in the current experiment (i.e., picture naming time) reflects a subconscious access to meaning representation formed during sentence processing, which in turn influenced processing time for the subsequent picture stimulus. These differences
highlight the importance of using different tasks and/or dependent measures to allow multidimensional understanding of a target phenomenon (MacDonald & Just, 1989).

In the current experiment, picture naming times were collected at six different ISIs. This aspect of the design proved vital for two reasons, each of which is discussed in turn. First, data points collected at equally distributed time points along a continuous scale, especially with the same materials and task, permitted data interpolation processes that are reliable and modest. This enabled immediate evaluation of the two-stage model for negative sentence processing, which was proposed based upon the combined findings at selective ISIs from separate experiments (Kaup et al., 2005, 2006; Kaup et al., 2007; see Section 3.1 for a review). Noting the gap that negative sentences with contradictory predicates were never tested at ISIs earlier than 750 ms in those studies, the current study offered such a test and found that with contradictory predicates, participants accessed the meaning of the unmentioned opposite state (i.e., the factual state) right after the sentence offset. This result cannot be explained by the two-stage model and suggests that any proposed model must come from data interpolation and interpretation based on direct evidence.

Related to this issue, testing six different ISIs allowed evaluation of different components of semantic processing happening at different points in time. Although the importance of examining semantic interpretation at different levels of comprehension processes—e.g., initial meaning access at the level of words, semantic integration at the level of phrase, sentence, or discourse, or high-level (conscious) processes for computing truth value—have been continuously emphasized (e.g., MacDonald & Just, 1989, Merrill et al., 1981), to my knowledge, there haven’t been many studies examining multiple processes within a study. In the current study, meaning representations probed at different ISIs conceivably reflect different semantic processes. Below I present some possibilities, focusing on how different components of semantic processing might constitute an experiential-based mental representation of meaning. Zwaan (2004) offers insights on this view, where three general components—activation, construal, and integration—that are not discrete but continuous processes underlie experiential-based simulation of meaning through an interactive and incremental processing mechanism.
The current study’s results found at a 0 ms ISI (i.e., right after the sentence offset) likely indicate the process of automatic meaning *activation*, where the patterns of neural activities formed and stored during previous perceptual experience (i.e., “functional webs” of neurons located throughout the cortex; Pulvermüller, 1999, 2002) are diffusely stimulated during initial lexical access. According to Zwaan (2004), the characteristics of initially accessed meaning take the form of diffusely activated multiple functional webs, since in the absence of a specific semantic context (e.g., as the first word in the sentence), the activated functional webs represent the entirety of our experiences with the entity or state of affairs under consideration. There are constraints, however, that the degree of diffuseness in representation will change, depending on various factors such as frequency, primacy, or recency in the experience with the target. Also, incoming words will activate new functional webs, which will influence/adjust the level of previous activation. Likewise, the previous level of activation also constrains the degree of activation of the next word. This suggests that at the point of the sentence offset, participants in the current study had already activated functional webs for a subject entity like *door* in multiple possible forms but have just encountered a contradictory predicate like *open*, which will adjust the activated functional webs for the door to represent an open door more strongly than others. Importantly, however, the literature on associative memory and lexicon suggests that when one part of a schema is activated, the associated parts are also activated (e.g., Griffiths, Dickinson, & Clayton, 1999; Marshall, 1995; Moss, Hare, Day, & Tyler, 1994) and that opposition relations are based upon lexical as well as conceptual associations (Gross et al., 1989; Israel, 2004). This suggests that an open door and a closed door are simultaneously activated. The naming times collected at 0 ms ISI suggest exactly this pattern. That is, the meanings of both the mentioned state and the unmentioned alternative state were simultaneously activated in all sentence types, except that the presence of the negation marker in the negative sentences suppressed the activation level of the mentioned state, producing a significant match-facilitation effect toward the factual information only in negative sentences.

However, results at medium ISIs, such as 500 ms and 1000 ms, likely indicate *construal*, which is an integrated meaning representation, of which referential unit is an
event. With respect to negative sentences, both the current study’s 500 ms ISI condition and Kaup et al.’s (2006) 750 ms ISI condition produced naming times that were numerically shorter for the factual state images than for the counterfactual state images, but this pattern did not reach statistical significance. Also, in the current study, naming times collected at 500 ms ISI and 1000 ms ISI after the contrastive and affirmative sentences did not indicate any dominance toward the mentioned state or the implied opposite state. One possibility for these null effects is that this time window between 500 ms and 1000 ms after the sentence offset is where language comprehenders actively evaluate the activated meanings for further integration, which perhaps involve comparisons among alternative meanings. The contradictory predicates make this aspect even more relevant due to the automatic activation of the binary alternatives. This differs from sentences like The girl touched the cat, which do not necessarily evoke a clear alternative meaning that can be compared. Therefore, by the time 1000 ms has passed from the sentence offset, people’s mental representation of the sentence The girl touched the cat can contain only those contextually relevant semantic features like fur (Merrill et al., 1981).

Lastly, results at longer ISIs, such as 1500 ms and 2000 ms, perhaps provide evidence for semantic integration, which refers to the transition process from one construal to the next, where the referential unit of the process is the event sequence (Zwaan, 2004). In the current study, both negative sentences and contrastive sentences included linguistic markers that are designated for indicating contrast (e.g., lexical negation, prosody), whose concept presupposes the existence of multiple alternatives. One interesting aspect of the integration process is that people anticipate what is coming next, and the predictability is largely influenced by the event sequence that we experience in the real world. That is, “if a sequence of events is experienced frequently, we tend to anticipate the second event when presented with the first” (Zwaan, 2004, p. 50) and “event sequences that are not consistent with experience are more difficult [to predict]” (Zwaan, 2004, p. 57). This suggests that the knowledge about objects and event sequences that we have gained from experience is the plausible ground for making inferences (Barsalou, 1999).
One way to test this idea is to create experimental items with the contrastive prosody \( L+H^* L-H\% \), where the implied contrast suggests either plausible consequences or less frequent (or less plausible) consequences as shown in (3.1) and (3.2), respectively. Intuition suggests that constructing an implicature (as exemplified in the continuation sentences in the columns on the right) is easier with the target sentences in (3.1) than with those in (3.2), perhaps because the event sequences suggested in (3.1) are more easily perceived in our experiences. Indeed, target examples in (3.2) may be used to indicate other pragmatic meanings such as jokes, irony, or sarcasm.

(3.1) Target sentence implying a plausible event sequence

<table>
<thead>
<tr>
<th>Target</th>
<th>Continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L+H^* L-H% ) The tomato WAS green…</td>
<td>( But \ now \ it's \ red \ and \ ready \ to \ eat. )</td>
</tr>
<tr>
<td>( L+H^* L-H% ) The dog WAS young…</td>
<td>( But \ it's \ been \ getting \ old \ recently. )</td>
</tr>
</tbody>
</table>

(3.2) Target sentence implying a less frequent or less plausible event sequence

<table>
<thead>
<tr>
<th>Target</th>
<th>Continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L+H^* L-H% ) The tomato WAS red…</td>
<td>( But \ now \ it \ is \ black \ because \ it \ is \ rotten. )</td>
</tr>
<tr>
<td>( L+H^* L-H% ) The dog WAS old…</td>
<td>( But \ it \ looks \ like \ it \ has \ become \ young \ again! )</td>
</tr>
</tbody>
</table>

Assuming that longer ISIs probed meaning integration processes, another discussion point is the difference between contrastive and negative sentences regarding the time at which the opposite meaning became the most dominant. This difference could actually result from the mere difference in the acoustic cue onsets. Acoustic analysis of the test sentences indicated that, on average, the onset of the negation word \( not \) was 363 ms earlier than the onset of the continuation rise (because the low pitch dip in \( L-H\% \) can be realized only on the final syllable of the predicate, whereas the negation marker is the penultimate word). Then, just due to this durational difference in the acoustic cues, the match facilitation for the factual-state in the negative sentences could emerge 363 ms earlier than the match advantage for the implied meaning in the contrastive sentences. Thus, the fact that the strongest match facilitation to the implied factual state was found 500 ms earlier with the negative sentences than with the contrastive sentences doesn’t
mean that constructing implicature is more difficult or slower with prosody than lexical
negation. Rather, the results suggest that inference processes involved in both sentence
types exhibit similarities as well as differences. To further evaluate this point, the next
chapter presents an eye-tracking experiment that compares the time course of inference
processing involved in both sentence types at a millisecond-grain detail.
CHAPTER 4
EXPERIMENT 3: VISUAL SEARCH

Experiments 1 and 2 have so far established that processing contrastive prosody (e.g., L+H* L-H%) enables people to access implied meaning that contrasts with the state of affairs asserted in the sentence. The results from the end-of-sentence picture naming paradigm in Experiment 2 suggested that one mechanism that people utilize to construct such meaning is to run perceptual simulation not only of what was described but also of what was implied in the sentence. Despite its usefulness, however, the picture naming task probed meaning representations only at the sentence offset and after, thus not permitting observation of ways in which meaning is constructed incrementally. Moreover, all materials in Experiments 1 and 2 included contradictory predicates that denoted binary attributes. Although this manipulation was practical for confirming that people do access state-contrast meaning when prompted by nonlexical cues such as prosody, a question remains regarding whether prosody can still evoke the state-contrast implicature when the lexical items themselves do not denote binary attributes. Previous studies on negation, reviewed in Section 3.1, suggest that, in the absence of an easily accessible opposite schema, negating a concept doesn’t necessarily constrain what might be a possible state for the factual situation (e.g., Kaup et al., 2007; Mayo et al., 2004).

This chapter describes Experiment 3, which used an eye-tracking paradigm with a new task and materials, in order to provide more detailed analyses on implicature processing. Because eye movement is a motor response that is fast and largely resistant to strategic control, its usefulness in language processing research has been frequently exploited (e.g., Henderson & Ferreira, 2004; Spivey, 2007; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Tanenhaus & Trueswell, 2006). In the current study, participants’ eye movements were recorded while they were engaged in an active visual search task, where they had to find and click on an object mentioned in the sentence. Crucially, the visual environment displayed more than 20 distinct objects that were changed and moved around for every trial, thus providing participants with an opportunity to be immersed in a challenging visual search task. Moreover, four of the
objects in each trial shared the initial syllable of their names (i.e., cohort objects with names such as *can*, *candy*, *candle*, and *camel*), in order to add more challenge to the task and to allow observation of competitive meaning construction processes happening in response to the temporarily ambiguous syllable.

The details of the experiment design and results are offered in Section 4.2. However, this chapter begins by reviewing some recent eye-tracking studies suggesting that language processing engages highly interactive processes among multiple factors at different levels of linguistic structure. I then discuss several studies that used an eye-tracking paradigm to explore the extent to which prosody can influence on-line meaning construction, especially when the meaning involves a notion of contrast. The last section of this chapter (Section 4.3) will compare the findings of the previous studies to those of the current study.

4.1 Background

Language processing involves multiple sources of information that interact at multiple levels of linguistic structure simultaneously (e.g., Altmann & Steedman, 1988; Crain & Steedman, 1985; Dahan et al., 2002; Sedivy, 2002; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). One line of findings supporting this view demonstrates that people’s knowledge of language and the world influences the ways in which low-level linguistic cues are processed. More importantly, such top-down influences impact language processing immediately, at the moment when a linguistic cue is encountered. For instance, Spivey et al. (2002) monitored eye movements of participants while they followed verbal instructions to move a certain object to a particular location. Crucially, the instructions included a sentence fragment like *Put the apple on the napkin...* where the prepositional phrase *on the napkin* was temporarily ambiguous between two roles: the goal location of the apple or the modifier that delimits the identity of the apple. The eye-movement results showed that when people encountered the ambiguous phrase, their decision on its function was influenced by the number and uniqueness of apples existing in the visual context. That is, when there was only one apple in the context, people interpreted the prepositional phrase as identifying the goal location of the apple.
However, when the context showed two or more apples, people used *on the napkin* to identify which of the apples was being referred to. These findings are striking because they show that diverse linguistic and nonlinguistic knowledge that we possess can determine or even override the ways in which bottom-up information is processed, thus compelling us to recognize that language processing is something other than a mere intake and decoding of linguistic information; rather, it is a highly complex and interactive enterprise that calculates multiple factors.

Recently, several studies have shown that this interactive and parallel processing mechanism extends to cover the manner in which even “suprasegmental” or prosodic information is processed. Key findings on prosodic processing include the finding that accentual information is processed separately from segmental information, in a way that reflects people’s knowledge on the relationship between acoustic prominence and its influence on discourse structure. For example, Dahan et al. (2002) recorded eye movements of people whose task was similar to that of participants in Spivey et al.’s (2002) study; that is, to find an object that was mentioned in the sentence and follow instructions to move it to a specific location. An important element of this experiment was that the researchers manipulated the accentual status of a referent noun that contained an ambiguous syllable like /kæ/, which was temporarily compatible with multiple lexical candidates such as *candy* and *candle* (known as cohort items). At the same time, the authors also manipulated the discourse context, such that the target referent could refer to either a new entity (i.e., previously unmentioned item) or an old entity (i.e., previously mentioned item that is either focal or nonfocal). Eye movement patterns collected during participants’ visual searches showed that, when the target syllable was spoken with a pitch accent, people took it as a signal to identify a previously unmentioned entity or a previously mentioned but nonfocal (or background) entity. However, when the target syllable was spoken without any accent, people treated it as a signal for an already mentioned entity. This is a convincing demonstration of an interactive parallel processing mechanism, because acoustic prominence on a single syllable, which lasted for less than a few hundred milliseconds, affected lexical selection processes at the moment when the syllable was encountered. Moreover, these processes
were modulated by people’s pragmatic knowledge on how accent information is used to indicate information structure.

Along this line, Ito and Speer (2008), among others (e.g., Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006), have demonstrated that not only the presence of accents but also the shape of accents immediately influences the way a referential domain is constrained. In particular, they showed that listeners take the rising peak accent L+H* as a cue to access unmentioned but visually present contrast sets, even before the segmental information of the accented word is fully disclosed. Important evidence comes from Ito and Speer’s (2008) Experiment 2, where eye-movement patterns were recorded while participants followed prerecorded verbal instructions like *Hang the red star. Now hang the GREEN L+H* angel L-L% to select the mentioned ornaments and decorate Christmas trees. The authors found that upon hearing the contrastively accented adjective, GREEN L+H*, people immediately looked at the cell containing the set of stars, not the one containing angels, which reflects their expectation that the accent functions to highlight the contrastive property, in this case, color. These fixations on the incorrect cell began before segmental information of the correct noun angel became available (i.e., during the processing of the color word), and they even persisted for 300 ms into the target noun. This finding demonstrates a highly dynamic interaction among multiple factors during real-time language processing; lexical and syntactic knowledge (i.e., function of a prenominal modifier in a definite noun phrase), pragmatic knowledge on how contrastive information is used (i.e., contrastively focused information is evaluated against background alternatives), and prosody (i.e., the L+H* marks contrastive focus).

The findings from these studies underscore that prosody in the speech signal is tightly related to the type or nature of the meaning that people construct. The immediate impact of L+H* in particular supports the notion that prosody activates meaning beyond what was mentioned in the sentence, by initiating interactions among lexical meaning, pragmatic knowledge, and surrounding contextual information. However, the precise mechanism by which these distinct sources of information are integrated during language processing is still largely unknown. One account offered by Sedivy and colleagues (1995), for example, suggested that contrastive focus expressed by pitch accent (and
other cues) allows people to access relevant contextual presupposition, where the focused item is compared to a set of alternative items (see Rooth, 1992, for a formalization of this notion). Also, this function of contrastive focus is not simply due to the accent that raises the saliency of the focused property, but rather is due to the interaction between contrast sets activated by the accent and the contextual information; it is this interaction that supports a unique identification of the focused set against the alternative sets. Despite this insightful account, however, the definition of contrast as well as the type of materials used to test the notion were extremely constraining, both in the original study of Sedivy et al. (1995) and in subsequent studies (Ito & Speer, 2008; Weber et al., 2006). For example, experiments have used a visual context comprised of one or more contrast sets differing by object properties such as size or color (e.g., purple scissors and red scissors presented with a green clock and a red vase). In addition, a contrastive accent L+H* was placed on language material that provided highly constraining information (e.g., on the prenominal adjective, such as the BLUE comb or the LARGE red square). Therefore, it is not clear whether the immediate effect of the alternative-set evoking function of L+H* will extend to the processing of other types of contrastive meaning. Moreover, existing eye-tracking studies have investigated the role of pitch accents when they were combined with only one type of end contour, i.e., the falling tone (L-L%), thus ignoring the influence of phrasal and boundary tones on meaning construction. 11 As mentioned in previous chapters, the influence of edge-tone type on meaning perception has been the topic of many theoretical as well as empirical studies in intonation phonology. In English, for instance, yes-no questions are usually spoken with a high rising end contour (H-H%), whereas a low falling tone is often used for regular statements. In addition, a low rising end contour (L-H%) can signal pragmatic meanings such as uncertainty, contradiction, or salience of background information, especially when it is combined with certain types of pitch accents such as L*+H or L+H* (e.g., Gussenhoven, 1983; Jackendoff, 1972; Lee, 2007; Liberman & Sag, 1974; Pierrehumbert & Hirschberg, 1990; Ward & Hirschberg, 1985, 1986). To my knowledge, however, there has not been any study that

11 In their Experiment 3, Ito and Speer (2008) placed either L+H* or H* on discourse markers (e.g., And now, After that) in the context of the rising end contour L-H%; however, this end-contour type was not part of their experimental manipulations.
systematically investigated the role of edge tones in computing such high-level meanings during on-line language processing. The goal of Experiment 3, described in Section 4.2, is to fill this gap and advance the current understanding of the relationship between prosody, pragmatics, and processing.

4.2 Experiment 3

This experiment, the last covered in this dissertation, investigates the ways in which the type and location of pitch accents, as well as the type of end contour, affect the on-line comprehension of implied meaning. An example item from the experimental material is *Lisa HAD L+H* the bell *L-H%*, where the contrastive accent on the verb *had* and the sentence-final rising tone work together to support the implicature that “Lisa no longer has the bell,” which is also a type of state-contrast implicature (see Section 2.2 for more details). Sentences similar to this example were presented as part of a visual search task, where participants were required to use the computer mouse to click on depictions of the mentioned objects (e.g., the bell) as quickly as possible while their eye movements were recorded during the comprehension of critical sentences.

One goal was to examine the compositionality hypothesis for meaning construction driven by sentence intonation (Pierrehumbert & Hirschberg, 1990). Experiment 1B offered initial evidence that each prosodic unit (e.g., pitch accents, edge tones) carries particular meanings, and that these are combined to contribute to the interpretation of the whole tune. However, this evidence came from an off-line study that measured the frequency with which participants (deliberately) produced the target meaning after listening to the target intonations twice. The design of the experiment hinges on the theoretical proposal as well as empirical evidence that the contrastive pitch accent *L+H* evokes a salient scale that promotes the accented item against some alternative items (Ito & Speer, 2008; Pierrehumbert & Hirschberg, 1990; Weber et al., 2008). I prepared three types of tunes—contrastive, emphatic, and neutral tunes (see Table 4.1 below for experimental conditions)—in order to examine whether an *L+H* accent on the critical word is sufficient to create the state-contrast meaning, or whether
the implicature is dependent on the combination of the contrastive accent and a subsequent rising end contour, L-H%.

In addition, the experiment design included a prosodically matched set of negative sentences, such as *Bart didn’t have the bee* (see Table 4.1). As reviewed in Section 3.1, researchers have argued that interpreting negative sentences often involves the processing of both the negated or counterfactual meaning and (an inferred) factual meaning (e.g., Kaup et al., 2006; Kaup et al., 2007). Therefore, a second goal of this study was to compare the time course of processing an inferred meaning when it is signaled by a lexical cue (i.e., *not*) versus a prosodic cue (i.e., L+H* and/or L-H%).

In this study, I collected two kinds of data: (a) response times taken by participants to click on the objects mentioned in the sentences, and (b) participants’ eye movements during their visual searches. The results on these data are reported separately in Section 4.2.5.

### 4.2.1 Visual materials

The visual field presented on a computer monitor was split in half to depict two cartoon characters’ bedrooms: Lisa Simpson’s room on the left and Bart Simpson’s room on the right. Figure 4.1 shows an example display. Each room contained 10 to 12 distinct objects, only one of which (e.g., a bell) matched an item mentioned in the test sentence (e.g., *Lisa had the bell*). Multiple objects were displayed to provide an active visual search environment and avoid the “closed-set” issue addressed by researchers who are concerned that any effects from simple displays may be artifacts of task-specific strategies (e.g., Ito & Speer, 2008; Tanenhaus & Trueswell, 2006).

Further, the current task compared conditions in which the mentioned object was found or was not found in the room of the referent of the sentential subject. The visual search task thus increased the difficulty of locating the mentioned object, which in turn allowed an evaluation of the extent to which a prosodically conditioned implicature versus a lexically expressed negation meaning induced a shift in attention from the sentential subject’s room to the alternate room, or the whole display.
The entire experiment included 32 critical TARGET objects, all of which had monosyllabic names. This was to ensure that participants had opportunities to perceive prosodic signals, especially the final rising or falling edge tones, before the critical segmental information was fully disclosed. To further prevent object search purely based on segmental information, each target item (e.g., bell) was displayed with three cohort items (e.g., belt, bed, bench)—one located in the same room as the target, hence called SAME ROOM COHORT, and two positioned in the other room, hence called OTHER ROOM COHORTS.

To participants, the displays appeared as if objects were added, removed, and/or moved for every trial. There were six designated locations, however, where the target and cohort objects appeared. The six squares marked on Figure 4.1 represent the critical cells (the squares themselves were not shown to participants). Objects for filler trials (i.e., 128 objects) appeared all over the screen including in the six critical cells, hence preventing any expectations for locating a target object.

4.2.2 Participants and procedures

Sixty-seven native English speakers participated in this study and received either course credit or $10. After signing the consent form, participants sat down in front of a desktop computer and wore a lightweight head-mounted eye-tracker with eye-head integration (ASL E5000). The eye-tracker consisted of a scene camera, along with a 60 Hz eye-camera and a small magnetic receiver, which was connected to a wooden pole with a magnetic transmitter. Each participant’s eye position was calibrated before the
experiment began, but it was recalibrated whenever necessary to prevent track loss throughout the experiment.

The experiment was run by E-prime, version 1.2 (Psychology Software Tools, Inc.). Participants were first presented with a cover story that provided a clear behavioral goal: to find and click on the objects mentioned in the sentences. The story stated that some creature moves the belongings of Lisa and Bart around, and this upsets the kids. The mother in the story, Marge Simpson, constantly recites “who had what” from her memory to aid the return of objects to the correct owner’s room. This story line ensured a felicitous use of the past tense verb had for both ongoing possession of an object and a switch of possession. Participants in a pilot study confirmed the naturalness of this set-up.

Each trial consisted of two visual searches directed by two separate sentences. Target sentences, however, appeared only as the first sentence in a trial, and all of the second sentences were filler items. In each trial, the onset of the display was synchronized with the onset of the first sound file, which included about 100 ms of silence before the sentence (thus providing a 100 ms preview). Participants listened to a sentence (e.g., Lisa had the bell) and clicked on the object mentioned in the sentence (e.g., bell) as quickly as possible. Mouse click response times were measured from the offset of each sound file. Upon the mouse click (i.e., the completion of the first search), another sentence was played while the visual display remained the same. Overall, each participant completed a total of 80 randomly presented pairs of visual searches (i.e., a total of 160 searches), directed by 32 target and 128 filler sentences. In most cases, the experiment lasted about 30 minutes including the eye-tracker set-up and calibration procedures. Participants in the debriefing sessions stated that they had fun with the task and did not consider intonation as a possible focus of the study. A few of them mentioned that sometimes the objects were not found where they had expected them, so they had to look for them more thoroughly. A few participants also mentioned that a small number of objects they were looking for were different from their mental images of the objects. This comment indicates, even if post hoc, that during language processing people visually simulate what is being talked about, which supports the mental simulation hypothesis tested in Experiment 2.
### 4.2.3 Sentence materials and experimental conditions

All sentences used the past tense; this was to induce one of the dominant implications that evoke the state-contrast implicature (i.e., “…but not anymore”). Although the L+H* L-H% tune on a sentence like Lisa had the bell can imply “…but now she has something else,” the experimental set-up biased participants to perceive the meaning that “Lisa had the bell…but now Bart has it.” The specific layout of the experiment suggested that the possessor of an object can only be either Lisa or Bart (i.e., a limitation on the number of alternative possessors), whereas there were too many possible alternative objects for one to be selected by the target sentence alone. This aspect of the experimental manipulation was never stated in the instructions. The binary set of possessors was simply presented as background in the room display, and it never became the center of attention because the background picture remained the same throughout the experiment (i.e., Lisa’s room on the left, Bart’s room on the right). In this sense, the contrast set manipulation in the current experiment is much subtler than those in the previous studies, where the display included visually contrasting objects—e.g., a colored pair of the same category of object such as a pink comb versus a yellow comb among a total of only four visual objects (e.g., Sedivy et al., 1995; Weber et al., 2006), or 11 sets of colored quadrates of the same object such as blue angel, red angel, brown angel, green angel (Ito & Speer, 2008).

Table 4.1 below presents detailed information about the testing conditions, which consisted of two sets of four comparable conditions. The first set included four affirmative sentences, whereas the second set contained four negative sentences. These two sets used different items for both the sentential subjects and the objects. The subject of the affirmative set was always Lisa, while that of the negative set was always Bart. This was to maintain the situation that each prosodic condition, from both sets, indicated the same room as the target location (e.g., the contrastive tune in both affirmative and negative sets was accompanied by the target object in Bart’s room), thus allowing qualitative comparisons for the pattern and speed of finding the objects. Filler sentences disguised this manipulation.
Table 4.1. Two sets of conditions comprised of three types of tunes and two types of truth for the object location in the display.

<table>
<thead>
<tr>
<th>Cond.</th>
<th>Tune</th>
<th>Sentence</th>
<th>Object location</th>
<th>Truth of the object location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lisa’s</td>
<td>Bart’s</td>
</tr>
<tr>
<td>Affirmative set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>contrastive</td>
<td>C1 L+H* L-H% Lisa HAD the bell.</td>
<td>bell</td>
<td>True</td>
</tr>
<tr>
<td>C2</td>
<td>emphatic</td>
<td>C2 L+H* L-L% Lisa HAD the bell.</td>
<td></td>
<td>False</td>
</tr>
<tr>
<td>C3</td>
<td>emphatic</td>
<td>C3 L+H* L-L% Lisa HAD the bell.</td>
<td>bell</td>
<td>True</td>
</tr>
<tr>
<td>C4</td>
<td>neutral</td>
<td>C4 H* H* L-L% LISA had the BELL.</td>
<td>bell</td>
<td>True</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>contrastive</td>
<td>C5 L+H* L-H% Bart DIDN’T have the bee.</td>
<td>bee</td>
<td>True</td>
</tr>
<tr>
<td>C6</td>
<td>emphatic</td>
<td>C6 L+H* L-L% Bart DIDN’T have the bee.</td>
<td></td>
<td>False</td>
</tr>
<tr>
<td>C7</td>
<td>emphatic</td>
<td>C7 L+H* L-L% Bart DIDN’T have the bee.</td>
<td></td>
<td>True</td>
</tr>
<tr>
<td>C8</td>
<td>neutral</td>
<td>C8 H* H* H* L-L% BARt DIDN’T have the BEE.</td>
<td></td>
<td>True</td>
</tr>
</tbody>
</table>

Both the affirmative and negative sets employed three types of intonational tunes. The first was the Contrastive Tune, where the contrastive accent L+H* on the verb had was followed by the continuation rise L-H%, resulting in the L+H* L- H% tune. The second type of tune used the contrastive accent (again on the verb) with a falling end contour, as in L+H* L-L%. This tune was named the Emphatic Tune, referring to previous research suggesting that, besides the contrastive role, this accent delivers some feeling of “emphasis” on what was stated (e.g., Bartels & Kingston, 1994; Ladd & Morton, 1997), which may in turn enhance the level of general attention (e.g., Ito and Speer’s Experiment 3, 2008). Thus, listening to Lisa HAD L+H* the bell L-L% can assure the listener that Lisa indeed had the bell (versus had not). The last type of tune was the Neutral Tune that placed the presentational pitch accent H* on the subject and the object, as well as the negation word not in the negative set (H* (H*) H* L-L%). Note that the acoustic signal of H* on the object was stronger than that on the subject phrase, because the last pitch accent in an intermediate phrase serves as the nuclear pitch accent (e.g., Beckman & Ayers, 1997; Ladd, 1996).
These sentences were carefully recorded in each of the tune types at 44 kHz by the same female phonetician/intonation phonologist who recorded the materials for Experiments 1 and 2. To maintain the greatest naturalness of the materials, the cross-splicing technique was not used this time. Instead, materials were re-recorded until both ToBI analysis and acoustic analysis confirmed that the sentences had similar overall durations (mean sentence duration, 1.09 seconds for affirmatives and 1.15 seconds for negatives, without including the pre-silence) and were produced with the intended intonation. Tables 4.2 and 4.3 present mean durations, as well as F0 minimum and maximum values for each word region across the test conditions. Statistical analyses on these values are provided at the end of this section.

Table 4.2. Mean duration (in ms) for test items; the bold font marks values from the accented syllables.

<table>
<thead>
<tr>
<th>Affirmative set</th>
<th>Present silence</th>
<th>Lisa</th>
<th>had</th>
<th>the</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
</tr>
<tr>
<td>Contrastive</td>
<td>105</td>
<td>264</td>
<td><strong>332</strong></td>
<td>–</td>
<td>48</td>
</tr>
<tr>
<td>Emphatic</td>
<td>106</td>
<td>269</td>
<td><strong>333</strong></td>
<td>–</td>
<td>51</td>
</tr>
<tr>
<td>Neutral</td>
<td>107</td>
<td><strong>274</strong></td>
<td>247</td>
<td>–</td>
<td>45</td>
</tr>
<tr>
<td>Mean</td>
<td>106</td>
<td>269</td>
<td>304</td>
<td>48</td>
<td>473</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative set</th>
<th>Present silence</th>
<th>Bart</th>
<th>didn’t</th>
<th>have</th>
<th>the</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
<td>ms</td>
</tr>
<tr>
<td>Contrastive</td>
<td>139</td>
<td>264</td>
<td><strong>235</strong></td>
<td>195</td>
<td>44</td>
<td>414</td>
</tr>
<tr>
<td>Emphatic</td>
<td>126</td>
<td>262</td>
<td><strong>210</strong></td>
<td>211</td>
<td>44</td>
<td>392</td>
</tr>
<tr>
<td>Neutral</td>
<td>128</td>
<td><strong>254</strong></td>
<td>194</td>
<td>231</td>
<td>39</td>
<td><strong>454</strong></td>
</tr>
<tr>
<td>Mean</td>
<td>131</td>
<td>260</td>
<td>213</td>
<td>212</td>
<td>42</td>
<td>420</td>
</tr>
</tbody>
</table>

Table 4.3. F0 minimum and maximum (in Hz) for test items; “a” and “b” mark F0 maximum on the accented/non-accented syllable and at the rising boundary, respectively.

<table>
<thead>
<tr>
<th>Affirmative set</th>
<th>Lisa</th>
<th>had</th>
<th>the</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Contrastive</td>
<td>162</td>
<td>190</td>
<td><strong>164</strong></td>
<td>242</td>
</tr>
<tr>
<td>Emphatic</td>
<td>159</td>
<td>191</td>
<td><strong>163</strong></td>
<td>254</td>
</tr>
<tr>
<td>Neutral</td>
<td><strong>163</strong></td>
<td><strong>194</strong></td>
<td>149</td>
<td>198</td>
</tr>
<tr>
<td>Mean</td>
<td>161</td>
<td>192</td>
<td>159</td>
<td>232</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative set</th>
<th>Bart</th>
<th>didn’t</th>
<th>have</th>
<th>the</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
</tr>
<tr>
<td>Contrastive</td>
<td>136</td>
<td>192</td>
<td><strong>183</strong></td>
<td><strong>275</strong></td>
<td>126</td>
</tr>
<tr>
<td>Emphatic</td>
<td>143</td>
<td>200</td>
<td><strong>200</strong></td>
<td><strong>265</strong></td>
<td>131</td>
</tr>
<tr>
<td>Neutral</td>
<td><strong>145</strong></td>
<td><strong>193</strong></td>
<td><strong>164</strong></td>
<td><strong>200</strong></td>
<td>158</td>
</tr>
<tr>
<td>Mean</td>
<td>141</td>
<td>195</td>
<td>182</td>
<td>247</td>
<td>139</td>
</tr>
</tbody>
</table>
In the Contrastive Tune and Neutral Tune conditions, the object location was always faithful to the sentence meaning, under the assumption that the Contrastive Tune dominantly evokes the state-contrast implicature, and the Neutral Tune conveys the literal meaning. The Emphatic Tune, however, generated two different experimental conditions, where the locations of the target object differed. Here, the “true” condition was considered as the one satisfying the general role of the L+H* pitch accent, which emphatically affirms the mentioned state (e.g., *Lisa HAD L+H* the bell L L% to mean “Yes, Lisa indeed had the bell,” and the target object is found in Lisa’s room). Accordingly, the other condition was called the “false” condition for convenience, although this false condition could serve truth if the L+H* alone could induce the state-contrast implicature. These two Emphatic conditions provided critical comparisons for themselves, as well as against the Contrastive condition. I explain this further in the discussion of the predictions in Section 4.2.4.

Overall, the filler materials limited the proportion of the false trials to just 7% and normalized the range of tunes. To encourage an active object search even further, a small portion of fillers (constituting 12% of the sentences) used some underspecified descriptions such as *Bart had a kind of hat* and *Lisa might have the bone.*

**Statistical analysis on acoustic measures**

Duration and F0 maximum data at each sentence region were fitted into several linear mixed-effects models in R, in order to test the influence of tune type on acoustic measures. Thus, each model included Tune as one fixed-effect term and Item as one random-effect term while treating the Contrastive condition as the default one. In the affirmative set, the subject duration was marginally longer in the Neutral condition than in the Contrastive condition due to the H* accent present only in the former condition (Neut. coefficient = 10.38, Standard Error = 6.08, \( t = 1.71, p = 0.095 \)), whereas no difference was found between the Contrastive and Emphatic conditions. As for the verb *had*, the duration was shorter in the Neutral condition than in the Contrastive condition due to the L+H* that was present only in the latter condition (Neut. coefficient = −85.19, \( SE = 4.77, t = −17.87, p < 0.00 \)). While the duration of *the* was significantly longer in the Emphatic condition than in the Contrastive condition (Emph. coefficient = 3.31, \( SE = 1.6, t = 2.14, p = 0.038 \)).
\( t = 2.08, p = 0.044 \), the pattern was the opposite for the target object duration (Emph. coefficient = -48.385, \( SE = 8.35, t = -5.79, p < 0.00 \)). The longer object duration in the Contrastive condition is due to the L-H\% contour leading to syllable lengthening.

The subject phrase in the negative set didn’t generate any significant difference. However, the duration of didn’t was significantly longer in the Contrastive condition than in both the Emphatic (Emph. coefficient = -25.19, \( SE = 5.64, t = -4.47, p = 0.0001 \)) and Neutral conditions (Neut. coefficient = -40.94, \( SE = 5.64, t = -7.26, p < 0.00 \)), whereas have was significantly shorter in the Contrastive condition than in the other two conditions (Emph. coefficient = 15.5, \( SE = 3.43, t = 4.52, p < 0.00 \); Neut. coefficient = 35.88, \( SE = 3.43, t = 10.46, p < 0.00 \)). The duration of the was shorter in the Neutral condition than in the Contrastive condition (Neut. coefficient = -4.81, \( SE = 1.97, t = -2.44, p = 0.02 \)). The target object duration was shorter in the Emphatic condition than in the Contrastive condition that elongated the last syllable with L-H\% (Emph. coefficient = -22.50, \( SE = 6.83, t = -3.29, p = 0.002 \)); however, the object noun was in turn longer in the Neutral condition than in the Contrastive condition, due to the former condition containing the H* accent (Neut. coefficient = 40.31, \( SE = 6.83, t = 5.90, p < 0.00 \)).

As for the F0 maximum value in the affirmative set, no difference was found in the subject name, whereas F0 maximum on had in the Contrastive condition was significantly smaller than the value in the Emphatic condition (Emph. coefficient for had = 12.0, \( SE = 2.23, t = 5.39, p < 0.00 \)) but greater than the value in the Neutral condition (Neut. coefficient for had = -44.1, \( SE = 2.23, t = -19.80, p < 0.00 \)). For the target object noun, however, the trend was the opposite (Emph. coefficient for object = -52.97, \( SE = 4.53, t = -11.68, p < 0.00 \); Neut. coefficient for object = 11.05, \( SE = 4.53, t = 2.44, p < 0.029 \)). F0 of the was significantly smaller in the Contrastive condition than in the other two conditions (Emph. coefficient = -24.67, \( SE = 6.94, t = -3.55, p = 0.001 \); Neut. coefficient = -41.84, \( SE = 6.94, t = -6.03, p < 0.00 \)).

In the negative set, the F0 in the subject phrase was greater in the Emphatic condition than in the Contrastive condition (Emph. coefficient = 7.83, \( SE = 3.94, t = 1.99, p = 0.05 \)). F0 for didn’t in the Contrastive condition was significantly smaller than those values in the other two conditions (Emph. coefficient = -9.46, \( SE = 3.50, t = -2.71, p =\).
0.01; Neut. coefficient = −74.60, \( SE = 3.50, t = −21.34, p < 0.00 \). F0 of *have* in the Contrastive condition was significantly smaller than that in the Emphatic (Emph. coefficient = 4.34, \( SE = 4.71, t = 0.92, p = 0.36 \)) but significantly greater than the value in the Neutral condition (Neut. coefficient = −18.08, \( SE = 4.71, t = −3.84, p < 0.00 \)). F0 of both *the* and the object noun in the Contrastive condition was greater than in the Emphatic condition (Emph. coefficient of *the* = −4.67, \( SE = 4.58, t = −1.02, p = 0.31 \); Emph. coefficient of object = −49.71, \( SE = 4.46, t = −11.15, p < 0.00 \)) and smaller than in the Neutral condition (Neut. coefficient of *the* = 34.51, \( SE = 4.58, t = 7.54, p < 0.00 \); Neut. coefficient of object = 16.09, \( SE = 4.46, t = 3.61, p < 0.00 \)).

### 4.2.4 Predictions

#### 4.2.4.1 The affirmative set

If semantic processing is highly incremental (e.g., Altmann & Steedman, 1988; Dahan et al., 2002; Sedivy, 2002; Sedivy et al., 1999) and there is a tight relationship between attention and visual behavior (e.g., Henderson & Ferreira, 2004; Tanenhaus & Trueswell, 2006; Tanenhaus et al., 1995), then the subject name *Lisa* in the auditory stimuli should immediately direct participants’ looks to Lisa’s room in the display, hence providing a quick preview of the objects in that room while the sentence progresses. Therefore, looks to and mouse clicks on the target objects will be faster in the Neutral True condition than in any other conditions, because the rest of the linguistic information following the subject name unambiguously supports the sentential subject referent as the possessor of the mentioned object. Moreover, only the Neutral True condition places a pitch accent (\( H^* \)) on the direct object noun itself, and proper accentuation is known to facilitate information processing (e.g., Cutler et al., 1997; Ito & Speer, 2008).

Besides the Neutral True condition, the remaining three conditions include sentences with the L+H* accent on the verb *had*. Assuming that the impact of accentual cues on language processing is immediate (e.g., Dahan et al., 2002; Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006), the eye movement patterns in these conditions will differ substantially depending on the role that the L+H* accent serves.
As reviewed in Section 4.1, previous eye-tracking studies have demonstrated that the L+H* accent evokes alternative sets in listeners’ minds, thus leading to incorrect fixations on the competing alternatives to the target in a visual world paradigm (Ito & Speer, 2008; Sedivy et al., 1995; Watson, Tanenhaus, & Gunlogson, 2008; Weber et al., 2006). If the alternative-set evoking function of L+H* itself suffices to systematically induce an implied meaning that suggests a state contrast (i.e., “…but now Bart has it”), the three conditions containing the L+H* accent should reveal the following object search patterns. First, participants will start looking into Lisa’s room upon hearing Lisa, but they will switch their looks to Bart’s room as soon as they hear the contrastively accented verb, HAD \( L+H^* \). In the Contrastive True condition as well as the Emphatic False condition, the target object is located in Bart’s room. Hence, these two conditions will allow object searches at a similar pace. However, the Emphatic True condition presents the target object in Lisa’s room. Therefore, upon realizing that the target object is not in Bart’s room, people will have to switch their looks back to Lisa’s room in this Emphatic True condition, and this additional switch in the looking pattern will result in slower response times for the looks as well as the mouse clicks, when this condition is compared to the Contrastive True and Emphatic False conditions.

However, Experiment 1B found that, in the absence of the final rising tone L-H%, the probability that participants would indicate the perception of the state-contrast implicature substantially decreased. This suggests that L+H* alone may not be sufficient to systematically induce the implicature, and that the clause-final rising tone L-H% is needed. If this is the case, then the earliest time that participants will change their looks to Bart’s room will only be when they encounter the last component of the critical linguistic cue, i.e., the rising edge tone, L-H%, or when they realize a search failure (in the absence of any supporting linguistic signal). Under this hypothesis, the Emphatic True condition will generate the quickest object search among the three conditions containing L+H*.

This is because the accent is followed by a simple falling tone L-L%, and the target object in this condition is located in Lisa’s room. In contrast, both the Contrastive True condition and the Emphatic False condition place the target object in Bart’s room. Because participants will start from Lisa’s room and will have to switch their looks to
Bart’s room, response times in these two conditions will generally be slower than both the Emphatic True condition and the Neutral True condition.

This hypothesis that L+H* alone is not a sufficient cue for computing the state-contrast implicature leads to a further prediction for the comparison between the Contrastive True and Emphatic False conditions. That is, I predict that latencies for visually locating the target object for the first time in a given trial, as well as response times to click on the object to complete the search, will be shorter in the Contrastive True condition than in the Emphatic False condition. This is because the Contrastive True condition contains a linguistic cue (i.e., the rising tone L-H%) that suggests a shift in the looking pattern from Lisa’s room to Bart’s room, whereas the eye-movement shift in the Emphatic False condition will be driven by the realization of the search failure.

Moreover, the emphatic tune in the Emphatic False condition may actually support Lisa’s room as the target room due to the tune’s meaning that affirms what was stated in the sentence. Thus, the earliest possible moment of the eye-movement shift between the two rooms will be the end of the target object name in the Emphatic False condition (i.e., the moment that the segmental information is fully disclosed and the visual search confirms no matching object candidate), whereas the low pitch dip within the object name may serve as the earliest signal for a room switch in the Contrastive True condition. Or, depending on the cue strength (influenced by the combination of the segmental properties and L-H% in a given item), people may switch their looks to the contrasting room anytime between the F0 dip and the F0 maximum on the final rise, or right after the final rise. Material analysis in these two conditions suggests that switching looks to Bart’s room in the Emphatic False condition should be slower by as much as the durational difference between the F0 dip on the vowel of the object in the Contrastive True condition and the offset of the object noun in the Emphatic True condition (i.e., mean difference of 217 ms). Note, however, that the mean duration of the target words is actually shorter in the Emphatic False condition than in the Contrastive True condition by 48 ms. This suggests that the moment when the segmental identity of the target word is fully disclosed will come faster in the Emphatic False condition than in the Contrastive True condition, which could potentially reduce differences between these conditions.
Thus, if found, any effect of the rising end contour in the Contrastive True condition will serve as a strong demonstration of the function of L-H%.

Last, under the hypothesis that L+H* itself is not a sufficient cue for a state-contrast implicature, the Emphatic True condition might generate quicker object search times when this condition is compared to the Neutral True condition. As mentioned before, previous research suggests a general role for L+H* in enhancing people’s attention level (see Ito & Speer, 2008, regarding L+H* accent on discourse markers in their Experiment 3). Hence, the L+H* on the verb might increase people’s search attention, which may assist a quicker object search. It is conceivable, however, that the increased attention may not have a direct advantage for object identification. Rather, the H* on the target object name in the Neutral True condition could be useful for a faster search.

4.2.4.2 The negative set

As in the affirmative set, the sentential subject name in the negative set will initially direct participants’ looks to the corresponding subject’s room (Bart, in this case). Upon listening to the negation word didn’t, participants must now process the negation. I consider two possibilities for negation processing and resulting looking patterns in a visual display.

First, negative sentence processing might be fundamentally similar to affirmative sentence processing. That is, listeners will take each word as it comes in and integrate its meaning with the existing meaning, while also making predictions about the upcoming words (e.g., Altmann & Kamide, 1999). In that case, participants’ behavior with respect to the negative conditions will resemble their behavior in the affirmative conditions, except that participants will begin by looking in the contrasting room (which is the mentioned subject’s room, i.e., Bart’s room). Upon listening to the negation word didn’t, participants will shift their looks to Lisa’s room and look around in that room with the expectation of receiving an object name, unless they receive an additional cue for a room switch (e.g., L+H* or the final rise L-H%, depending on the hypothesis explained in the section above). This suggests that response times to visually locate and click on the target
objects will be smaller in the Emphatic True and Neutral True conditions than in the other two conditions, because the rest of the information after the word didn’t will be processed in a similar way as in the affirmative set, and the visual display indeed supports Lisa’s room as the target room.

Alternatively, processing negative sentences could be fundamentally different from processing affirmative sentences. Affirmative sentences mostly evoke the described meaning (as long as there is no lexical item that spreads its activation to its antonym, e.g., the contradictory predicates, as discussed in Chapter 3). Therefore, these sentences should direct visual attention to just the sentential subject’s room (unless there is an additional cue for a switch). However, negative sentences refer to two states of affairs, i.e., one state that is being negated and the other state that is actually the case (e.g., Glenberg et al., 1999; Kaup et al., 2006; Mayo et al., 2004). Thus, upon hearing the word didn’t, participants may widen their visual attention to include both the alternative room and the current (mentioned possessor’s) room, given that the background picture implies only two possessor candidates. That is, there will be an expansion of the search fields from Bart’s room alone to the whole space. This visual search field expansion upon hearing the word didn’t would then greatly reduce differences in object search times across the four conditions. Participants will already have gotten a preview of the objects in Bart’s room upon processing the subject name, and the search field expansion upon processing didn’t will assist a rigorous visual scanning for the target objects in both rooms regardless of the prosodic pattern.

4.2.4.3 Comparison between affirmative and negative sets

Taking the affirmative and negative sets together, I predict three types of object search times overall—short, medium, and long—that are reflected in the latencies of the first fixations to the target objects as well as the mouse click times. On the one hand, the shortest response times will result from the Neutral True and Emphatic True conditions in the affirmative set, since these conditions warrant successful object searches within one room (i.e., no need to switch rooms, assuming the insufficiency of the L+H* in evoking implicature). Therefore, I call them the fast conditions.
On the other hand, the longest response times will come from the Contrastive True condition and the Emphatic False condition in the affirmative set. Participants will initially search for objects in the mentioned sentential subject’s room (i.e., Lisa’s room), but upon receiving a prosodic cue (in the Contrastive True condition) or realizing a search failure (in the Emphatic False condition), participants will shift their searches to the alternative room (i.e., Bart’s room). Since this eye-movement shift is motivated only with a reason, response times to find the targets in Bart’s room will be much slower in these conditions. Thus, I call these conditions the slow conditions.

The response times from negation conditions will fall in between (i.e., the medium conditions) on either view of how negation is processed. Participants will start their searches in Bart’s room, but the negation word didn’t will lead to a quick shift of participants’ attention from Bart’s to Lisa’s room, or to the whole space. Because of this necessity to change the search field upon processing the negation word, object searches in the negative set will be slower than those for the fast conditions in the affirmative set. However, both the initial looks to Bart’s room and the expansion of the search field will reduce the overall search times in the negative conditions to be less than in the Contrastive True and Emphatic False conditions in the affirmative set.

Finally, there might be a durational difference between computing an implicature based on the prosodic cue versus lexical negation, simply because the onset of negation (approximately as the onset of didn’t in the negative baseline conditions, i.e., the Emphatic True and Neutral True conditions) is realized temporally earlier than the onset of the rise in the Contrastive True condition in the affirmative set by 606 ms on average.

4.2.5 Results

Data were first evaluated for mouse click accuracy as well as fixation calibration accuracy. Data from seven participants were excluded from the analysis: two for experimenter error (not recording the eye movements), one for a participant’s misunderstanding of the task (resulting in clicking anywhere in the room rather than on the target objects), three for a calibration difficulty generating less than 75% of fixation observations for critical trials, and one for mean click times that exceeded 2.5 standard
deviations from the overall participants’ mean. Thus, analyses included data from the remaining 60 participants.

4.2.5.1 Mouse click results

Mouse click data showed overall mean accuracies of 99.96% from the participant analysis and 96.9% from the item analysis when both critical and filler items were included, suggesting that the task was straightforward. When only the critical items were analyzed, mouse click accuracy decreased to 95.4%, where most of the incorrect clicks happened on one of the three cohort objects. However, there were no significant differences in click accuracy across test conditions.

Mouse click times were analyzed from accurate trials only. Click times were first trimmed with a fixed cutoff at 6000 milliseconds based upon the overall click time distributions blind to the experimental conditions. This removed 1.81% of total data. Then, long latencies exceeding 2.5 standard deviations from each participant’s mean were replaced with the respective cutoff value, i.e., the participant’s mean plus 2.5 standard deviations from that mean (Ratcliff, 1993). Figure 4.2 below presents mean mouse click times collected in each test condition from the participant analysis. The left panel displays the results from the affirmative set and the right panel includes the results from the negative set.

![Figure 4.2. Mean mouse click times averaged across participants (in milliseconds, measured from the sentence offset); the four bars on the left represent data from the affirmative set; the four on the right represent data from the negative set.](image-url)
As for the affirmative set, mean click times were substantially different across the four conditions. Two one-way repeated-measures ANOVA tests crossing the test condition as a by-participant and by-item factor and the experimental list as a by-participant factor found a strong main effect of condition: $F_1(3,168) = 101.6, p < .01$; $F_2(3,45) = 29.16, p < .01$.\(^{12}\) The difference pattern among the conditions indicated strong support for the hypothesis that perceiving the state-contrast implicature from prosody depends on the presence of the sentence-final continuation rise L-H%, and not just the presence of the contrastive pitch accent L+H*. That is, click times in the Emphatic True condition (C3) were similar to click times in the Neutral True condition (C4) and not to those in the Emphatic False condition, suggesting that the contrastive pitch accent L+H* itself could not induce eye-movement shifts between the rooms. Planned contrasts supported this pattern. That is, correct mouse click times in both the Contrastive True (C1) and Emphatic False (C2) conditions were significantly greater than mouse clicks in the Emphatic True (C3) and Neutral True (C4) conditions, all at $p < .01$: C1 vs. C3 ($F_1(1,56) = 92.23, F_2(1,15) = 23.35$), C1 vs. C4 ($F_1(1,56) = 107.4, F_2(1,15) = 30.57$), C2 vs. C3 ($F_1(1,56) = 121.54, F_2(1,15) = 42.36$), and C2 vs. C4 ($F_1(1,56) = 164.56, F_2(1,15) = 53.87$). Note that this includes the critical comparison of the Emphatic False and Emphatic True conditions; click times were significantly shorter in the Emphatic True condition than in the Emphatic False condition, not the reverse. This pattern suggests that participants perceived the L+H* on the verb as an emphatic cue, affirming what was stated in the sentence, rather than taking it as a contrastive cue for a switch of looks to the alternative location. Further explanations of the difference between an emphatic cue and a contrastive cue are found in the discussion in Section 4.3.

Mouse click times in the two slow conditions also showed a significant difference. That is, the click times in the Contrastive True condition (C1) were significantly smaller than those in the Emphatic False condition (C2), highlighting the facilitatory role of linguistic cues (in this case, prosody) in visual search: C1 vs. C2 ($F_1(1,56) = 13.63, p < .01, F_2(1,15) = 2.66, p = .124$).\(^{13}\) Also, the difference between click

---

\(^{12}\) ANOVA analysis will be replaced with a regression analysis in subsequent publication.

\(^{13}\) Two out of 16 affirmative items produced much lower accuracy in one of their conditions (bell: 67% in the Emphatic False condition; lime: 73% in the Contrastive True condition), due to cohort competition (e.g.,
times in the Emphatic True condition (C3) and the Neutral True condition (C4) was marginally significant by participants, although not by items: C3 vs. C4 ($F_1(1,56) = 3.93, p = .052, F_2(1,15) = 2.34, p = .15$).\footnote{After excluding the two lowest accuracy items, the $p$ value is .09.}

As for the negative set, however, one-way ANOVA tests indicated no meaningful differences across the four conditions: $F_1(3,168) = 1.79, p = .15, F_2(3,45) = 1.18, p = .33$. The click times in the Contrastive True and Emphatic False conditions were numerically greater than the click times in the Emphatic True and Neutral True conditions, but these differences did not reach statistical significance.

Post hoc paired $t$-tests showed that mouse click times from the comparable conditions in the affirmative and negative sets were significantly different. For the Contrastive True pair (C1 vs. C5), for example, click times in the affirmative set were significantly greater than those in the negative set from the participant analysis: $t_1(1,59) = 2.00, p < .01, t_2(1,30) = 2.04, p = .21$. For the Emphatic False pair (C2 vs. C6), click times in the affirmative set were significantly greater than those in the negative set from both participant and item analyses: $t_1(1,59) = 2.00, p < .01, t_2(1,30) = 2.04, p < .01$. In contrast, click times for the Emphatic True pair (C3 vs. C7) and the Neutral True pair (C4 vs. C8) were significantly smaller in the affirmative set than in the negative set: $t_1 = 2.00, p < .01$ for both pairs; for C3 vs. C7, $t_2 = 2.04, p < .05$; for C4 vs. C8, $t_2 = 2.04, p < .01$.

Overall, there was a difference of nearly 1000 ms in click times between the fast affirmative conditions on the one hand (i.e., Emphatic True and Neutral True conditions) and the slow affirmative conditions on the other hand (i.e., Contrastive True and Emphatic False conditions). Furthermore, the 447 ms difference on mean click times between the Contrastive True condition in the affirmative set (C1) and the baseline negative conditions (i.e., C7 and C8) was smaller than the durational difference of 606 ms in the acoustic signals between the onset of the rise in C1 (i.e., the F0 minimum in the syllable) and the onset of the negation didn’t. This suggests that computing an implicature from a prosodic cue is as fast as doing it from a lexical negation. Section 4.3 will compare and contrast the current findings with the findings from Experiment 2. The

\textit{belt} and \textit{lion}, respectively). After removing these two items, the item analysis comparing these two conditions showed marginal significance at $p = .06$.\footnote{After removing these two items, the item analysis comparing these two conditions showed marginal significance at $p = .06$.}
next section presents the results of the analyses of eye movements, first for the affirmative set, and then for the negative set.

4.2.5.2 Eye movement results

Based upon the eye movements recorded at 59.94 Hz (i.e., at approximately every 17 ms), fixations were automatically generated by the eye-tracker software, ASL E5000. Evaluation of fixation accuracy revealed that, among the valid trials collected from 60 participants, 27 trials induced correction fixations on the targets without the targets being clicked on (i.e., incorrect clicks). These trials were removed from the analysis because they suggest that participants had other objects in mind during their visual searches (e.g., likely the cohort items). In summary, 96.8% of the critical trials that induced correct mouse clicks also generated correct fixations on the target objects, especially after the onset of the target object noun was heard. The distributions of trial losses were not different across the conditions in the affirmative set. However, there were additional trial losses for the negative set due to the experimenter’s mistake, where one particular item (the whisk) in one particular condition (C7, the Emphatic True condition) was placed in an unintended critical cell for half of the participants’ data collected within a particular list. The mouse click data showed that this mistake didn’t generate any noticeable difference in the participants’ task performance, because they still successfully found and clicked on the target objects from that new location. Although these trials could have been tracked down and included for the fixation analyses, they were omitted due to the complications they would bring to handling the large amount of data in a consistent manner.

4.2.5.2.1 The onset of the signal-driven eye gazes

Valid fixations on the TARGET (i.e., that happened after the target word onset) were first evaluated for signal-driven fixations. According to Matin, Shao, & Boff (1993), among others, 180–200 ms after the critical word onset is the earliest possible time point at which one can observe eye gazes that are driven by processing a linguistic cue (due to the time needed to program an eye movement). However, it is also known that
complexity in the visual and/or language materials could influence the amount of time needed for fixations to represent signal processing (e.g., Barr, 2008; Henderson & Ferreira, 2004; Tanenhaus & Trueswell, 2006). Therefore, to visually assess the onset of the signal-driven fixations blind to the specific experimental conditions in the current study, I plotted mean fixations on the target collapsed across the affirmative conditions (Figure 4.3) and, separately, across the negative conditions (Figure 4.4). In both figures, the x-axis presents time that is aligned to the target object noun onset at the 0 ms point. Therefore, negative times represent the moments when pre-object materials were presented to participants. Fixations were then aggregated into 50 ms time bins, and the y-axis represents these fixations on a log odds scale.\textsuperscript{15}

![Affirmative set: grand mean fixation on the target](image)

Figure 4.3. Mean fixations on the target objects collapsed over four affirmative conditions and aggregated into 50 ms bins that were time-locked at the onset of the target object noun; the value 0 on a log odds scale (y-axis) corresponds to 0.5 on a proportion scale (Johnson, 2008).

\textsuperscript{15} The choice of a log odds scale over a proportion scale is based upon previous reports that it is “the appropriate scale for assessing effects on a categorical dependent variable” (Barr, 2008, p. 462; also see Agresti, 2002; Cohen et al., 2003; and Jaeger, 2008).
Both figures show two increasing trends. As for the affirmative set, one trend starts at 350 ms before the target word onset and the other starts at 500 ms after the object noun onset. In the negative set, the onsets of two rising trends are 600 ms before and 400 ms after the target word onset. The slopes before the target object noun onsets in both figures indicate fixation increases on the target that happened by chance once the sentence presentation had begun. In the current experiment, the beginning of the visual display was synchronized with the beginning of the sound file as well as the eye-movement recording; hence, as participants were looking around the display while listening to pre-object materials, target objects were sometimes fixated by chance. For the affirmative sound files, the grand mean duration of the pre-object materials consisted of 106 ms silence plus 620 ms for the sentence fragment: Lisa (269 ms) had (304 ms) the (48 ms). For the negative sound files, the grand mean duration of the pre-object materials includes 131 ms silence plus 727 ms for the sentence fragment: Bart (260 ms) didn’t (213 ms) have (212 ms) the (42 ms). Aligning these values on the x-axis (the blue dotted lines) suggests that the chance fixations on the TARGET (i.e., fixations that happened without clear intention of finding the object) started increasing as participants processed the name of the sentential subject, Lisa or Bart, which indicates a highly incremental processing mechanism.

Figure 4.4. Mean fixations on the target objects collapsed over four negative conditions.
In both figures, fixations to the TARGET start sharply increasing again, once 500 ms (in the affirmative set) or 400 ms (in the negative set) have passed from the critical object noun onset. To my knowledge, these values represent one of the slowest onsets for a signal-driven eye gaze (i.e., visually locating an object due to processing a linguistic signal), when compared to similar studies dealing with visual behavior and language processing. While the majority of studies have assumed 180–200 ms from the critical sound onset as the starting point for critical fixations, Barr (2008) observed a signal-driven increase of eye gaze happening at 300 ms after the critical word onset, when the sentences used unconventional descriptions to refer to unfamiliar objects even when only four objects were displayed on the screen. Ito and Speer (2008)—who displayed 11 cells, each of which contained three to four objects differing only in color—stated that they assumed that 200–300 ms were needed for a signal-driven eye fixation after the critical noun onset, referring to Viviani (1990). However, their data showed that the average first fixation latencies to the target cell in their felicitous L+H* conditions from their Experiments 1 and 2 were 407 ms and 388 ms, respectively.

In the current study, the whole visual field displayed 20–24 distinct objects that differed in color, shape, and identity. Moreover, while the total number of objects on the screen remained constant, the objects kept changing, and they were moved around for every trial. To my knowledge, this set-up represents one of the highest visual complexities reported in the eye-tracking studies investigating language processing; participants in the current study were engaged in a truly challenging visual search task. The unusually slow latency for the signal-driven gaze behavior is attributable to this visual complexity in the display, which confirms that researchers of a visual world paradigm must not assume any normative value (e.g., 180–200 ms) for a signal-driven gaze, and instead carefully assess the fixation pattern under the scope of the complexities in the audio and visual materials for that individual study (e.g., Barr, 2008; Henderson & Ferreira, 2004).
4.2.5.2.2 Results from the affirmative set

Based upon the signal-driven fixation information, this section begins by summarizing general patterns observed in the fixation data collected in the affirmative set. I first present fixation data on a proportion scale to make data interpretation more accessible. For statistical analysis, however, I convert proportion data into a log odds scale, as explained later.

The two figures (Figures 4.5 and 4.6) below show the proportion of fixations collected on the target objects averaged across participants in the four affirmative conditions. The x-axis represents time, with the averaged onset of the target words in the sound files aligned to the 0 ms point. Negative time values thus indicate the moments when the pre-object materials were played to participants. As mentioned above, the data include only those trials where participants actually looked and clicked on the target objects. This is to ensure that the analysis includes eye movements collected when participants were searching for the correct target objects (and not for cohort competitors). The patterns of fixations in the two figures are essentially the same. However, Figure 4.5 uses a new method that graphs the proportion of fixations at a millisecond grain, thus providing a way to represent the continuous measure of eye movement.\footnote{I thank Kurt Brunner for providing me with this solution.} In contrast, Figure 4.6 follows a traditional method that aggregates fixations on a visual category (e.g., target) gathered at every 16.66 milliseconds from a given participant in a given condition into a series of predefined time bins (here, 50 ms bins). This aggregation is to handle the non-independence in the eye data within a trial, due to the inherent logistics of how eyes move (i.e., consecutive fixations can not be completely discrete; see Barr, 2008).
Figure 4.5. Proportion of fixations to the target objects at a millisecond grain in the four affirmative conditions; time on the x-axis is aligned to the target word onset.

Figure 4.6. Proportion of fixations to the target object averaged across participants, aggregated into 50 ms bins in four affirmative conditions.

For both graph types, proportions give the number of fixations to the target objects divided by the number of possible fixations, 240 (i.e., 60 participants multiplied by 4, the maximum number of trials per condition). Using the possible number of
fixations as the denominator, rather than the total number of actual fixations in the corresponding condition (i.e., C1: 228, C2: 230, C3: 234, C4: 237), is to consider previous researchers’ concerns regarding the validity of removing non-target fixation events. Barr (2008), for example, noted that two things are hard to know: (a) the kinds of events that constitute non-target-looking instances (e.g., blinks, fixations on other regions, transitions between fixations), and (b) the distribution of those noise events across the experimental conditions. Considering a possible influence of these elements on the inferential statistics, I consistently use the total number of possible fixations as the denominator for calculating fixation proportions.

These figures confirm the validity of the signal-driven fixation onset determined earlier. More importantly, there were two twin patterns of fixation curves. On the one hand, both the Emphatic True and Neutral True conditions generated fixation proportions on the TARGET that increased quickly and peaked earlier (at about 1200–1400 ms from the target word onset). On the other hand, the fixation proportions to the TARGET in the Contrastive True and Emphatic False conditions increased at a slower rate, reaching the peak substantially later, roughly at around 1700 ms–1900 ms after the target noun onset. This split in the fixation pattern converges with the pattern found from the mouse click data; mouse click times were much smaller in the Emphatic True and Neutral True conditions than in the Contrastive True and Emphatic False conditions. This likely indicates that only the two slow conditions required an eye gaze shift from the mentioned subject’s room (i.e., Lisa’s room) to the alternate subject’s room (i.e., Bart’s room), whereas the two fast conditions aided participants to find objects from the mentioned subject’s room without any fixation shift between the rooms.\(^\text{17}\)

This observation was verified by examining the patterns of fixations not only on the TARGET but also on the cohort objects, especially the OTHER ROOM COHORTS. As mentioned above, the visual display for each critical trial presented the TARGET together with three cohort items in two subtypes: one SAME ROOM COHORT (which was always displayed with the TARGET in the same room) and two OTHER ROOM COHORTS (which

\(^{17}\)The experiment gathered fixation data with respect to another area of interest: Lisa’s room vs. Bart’s room. These data are expected to provide direct evidence on how often participants shifted their eye movements back and forth between the rooms. They will be analyzed in a future report.
always appeared in the contrasting room). Therefore, fixation latencies and proportions on the OTHER ROOM COHORTS along with those on the TARGET can serve as the measure of when and how frequently participants switched their looks between two rooms. In the figures below, fixations to the two OTHER ROOM COHORTS were collapsed and represented simply as proportions of fixations to OTHER ROOM COHORT.

*Proportion of fixations in the Emphatic True and Neutral True conditions*

Figures 4.7 and 4.8 below present fixation proportions (averaged across participants) on each of three categories—TARGET, SAME ROOM COHORT, and OTHER ROOM COHORT—in 50 ms time bins, collected from the Neutral True and Emphatic True conditions. The x-axis represents time (in milliseconds) where 0 ms indicates the target word onset averaged across the relevant sound files (i.e., the black solid line), and the y-axis indicates fixation proportions. Three blue dotted lines mark the average offset of each content word: *Lisa, had/HAD*, and the target object name. Figures showed that participants rarely looked at the OTHER ROOM COHORTS in these two conditions. Instead, fixation proportions on both the TARGET and SAME ROOM COHORT started increasing at around −150 ms, that is, before the target word onset, until 500 ms had passed from it, at an approximately similar rate. Aligning sentence materials with the graphs and adding 500 ms needed for programming a signal-driven eye-fixation to those values indicate that the initial increase of the looks to TARGET and SAME ROOM COHORT is due to the processing of the sentential subject *Lisa*, which suggests that visual behavior is tightly linked to the incoming linguistic cue. Once 500 ms has passed from the target word onset, fixations to TARGET increased sharply, while fixations to the competitor, SAME ROOM COHORT, started decreasing quickly. This divergence point corresponds to the time at which the target word’s segmental information started being processed and lasted for an additional 495 ms on average in the Neutral True condition and 439 ms on average in the Emphatic True condition (which correspond to 995 ms and 939 ms on the x-axis, respectively).
This confirms that participants found the \textsc{target} in Lisa’s room without any consideration of objects in the contrasting room, Bart’s room, thus verifying that these two conditions behaved as the fast-search conditions (see the summary in Figure 4.9).
This finding, especially in the Emphatic True condition, is crucial, given that this condition placed the contrastive pitch accent L+H* on the verb *had*. The fact that the L+H* accent itself did not induce a shift of looks between two rooms is the first piece of evidence suggesting that L+H* alone was not sufficient to convey the state-contrast implicature.

![Proportions of fixations to TARGET and OTHER ROOM COHORT](image)

Figure 4.9. Proportions of fixations to TARGET and OTHER ROOM COHORT in the Emphatic True and Neutral True conditions.

### Proportions of fixations in the Emphatic False condition

The fixation proportion patterns in the Emphatic False condition (Figure 4.10) were completely different from the Emphatic True condition, even though both conditions used exactly the same sentence materials that contained L+H* on the verb *had* followed by the low falling tone at the end of the target object word. Recall that the Emphatic False condition (as well as the Contrastive True condition) placed the TARGET and the SAME ROOM COHORT in the room of the non-mentioned alternate possessor, Bart, while placing the OTHER ROOM COHORTS in the room of the mentioned subject, Lisa.
The patterns shown in the figure indicate that participants were looking only at the OTHER ROOM COHORT during the time window between the sentence beginning and to approximately the 600 ms point after the target word onset, shown on the x-axis. There was no fixation on the TARGET or SAME ROOM COHORT during this time window, indicating that participants were searching for the target objects in Lisa’s room as they processed the pre-object materials, Lisa HAD the. This is the second piece of evidence confirming that L+H* on the verb had itself did not induce eye-movement shifts between the rooms, perhaps because it was not a sufficient cue to support the implied meaning that there is a change in the mentioned state.

Once 600 ms had passed from the target word onset, fixations to the TARGET began increasing. However, until about the 800 ms point on the x-axis, fixation proportions on the OTHER ROOM COHORT also continued increasing. This concurrent pattern of fixation increase on the TARGET and on the OTHER ROOM COHORT during the same time window (600–800 ms from the target word onset) could suggest that (a) there was some competition between the two rooms, at least for some participants in some trials, or (b) some participants in some trials had already shifted their eye movements to
the alternate subject’s room, whereas some other participants were still looking into the mentioned subject’s room during this time window, or (c) a combination of (a) and (b). In any case, the overlapping increase pattern was soon terminated, and the proportion of fixations to the OTHER ROOM COHORT began decreasing as fixation proportion on the TARGET continued to increase. Finally, there was a crossover between the proportion of fixations on the TARGET and on the OTHER ROOM COHORT at around the 1150 ms point on the x-axis. This crossover point represents a value that is 211 ms greater than the earliest possible crossover point, 939 ms (i.e., the mean offset of the target word, 439 ms, plus 500 ms needed for eye-movement programming in the current study’s complex visual display). This suggests that the eye-movement shift in this condition was fairly delayed, since participants needed to confirm the failure of their initial searches in Lisa’s room, which was actually a good potential location based upon the meaning of the sentence intonation. Figure 4.11 shows how the search pattern differed as the function of the object location in two otherwise identical conditions, Emphatic True and False conditions.

![Figure 4.11. Comparison between Emphatic True and False conditions.](image)

**Proportions of fixations in the Contrastive True condition**

The fixation proportion patterns in the Contrastive True condition approximated the pattern found in the Emphatic False condition, as the TARGET and the SAME ROOM COHORT in this condition were also placed in the unmentioned alternate possessor’s room, whereas the OTHER ROOM COHORTS were present in the mentioned subject’s room (Figure 146).
4.12. Sentences in this condition also contained L+H* on the verb had, but the accompanying end contour was the rising one, L-H%, which could potentially trigger the meaning of contrastive implicature. In the figure, the red solid line indicates the earliest potential onset of the rising cue L-H% on average (i.e., the F0 minimum on the vowel of the target word across items). Three blue dotted lines mark the average offset of each content word: Lisa, HAD, and the target object name.

Figure 4.12. Proportions of fixations to TARGET, SAME ROOM COHORT, and OTHER ROOM COHORT in the Contrastive True condition.

Despite the globally similar patterns in this condition and in the Emphatic False condition, close inspection shows several differences. For a visual aid on the detailed comparison between these two conditions, Figure 4.13 plots fixation proportions only on the TARGET and OTHER ROOM COHORT. Fixation proportions on the SAME ROOM COHORT are left out because they were nearly identical in both conditions.
Figure 4.13. Proportion of fixations to TARGET and OTHER ROOM COHORT in the Contrastive True and Emphatic False conditions.

The first difference is that fixation increase on the TARGET began about 100 ms earlier (i.e., after the 500 ms point on the x-axis), and the proportion of fixations to the OTHER ROOM COHORT continued to increase for a shorter amount of time in the Contrastive True condition (i.e., between 500 ms and 650 ms on the x-axis) than in the Emphatic False condition. Moreover, the crossover point (where the fixation proportions on the TARGET catch up with the fixation proportions on the OTHER ROOM COHORT) also occurred about 100 ms earlier in the Contrastive True condition (i.e., about 1000 ms on the x-axis) than in the Emphatic False condition. Crucially, this particular crossover point suggests that the shift of attention between the two rooms mostly happened while the rising cue L-H% was being processed. Adding 500 ms (needed for the signal-driven gaze in the affirmative set) to the mean onset value for the rise (222 ms) and the mean offset value for the target word (487 ms) in the Contrastive True condition supports this observation. After the crossover point, the fixation proportions on the OTHER ROOM COHORT decreased more slowly and remained higher in the Emphatic False condition than in the Contrastive True condition, whereas fixation proportions on the TARGET increased more quickly and remained higher in the Contrastive True condition than in the
Emphatic False condition. This suggests that participants might have been looking back and forth between the two rooms more often in the Emphatic False condition to confirm if their initial searches had indeed failed.

An important question is whether the subtle timing differences between the Contrastive True and Emphatic False conditions are statistically reliable. As mentioned in the prediction section, the speed of finding the TARGET (in the alternate subject’s room) could be different due to the nature of the cue: a linguistic cue L-H% that implies a state change in the Contrastive True condition, versus a search failure from the mentioned subject’s room in the Emphatic False condition. Also, it is important to statistically validate the object search timing difference observed between the Emphatic True and False conditions, as well as the similarity between Emphatic True and Neutral True conditions. One dependent measure that allows such comparisons is the latency of the very first fixation made to the TARGET in each of the trials collected from all participants per test condition. The next section reports results of this analysis.

**Latencies of the very first fixations to the TARGET**

For data preparation, initial fixations to the target objects were collected in each test condition, whereas fixations made before the target word were eliminated. Then, outlying values were determined and replaced with different fixed-cutoff values: 3 seconds for the Neutral True and Emphatic True conditions (i.e., fast conditions) and 6 seconds for the Emphatic False and Contrastive True conditions (i.e., slow conditions) in the affirmative set, and 5 seconds for the four conditions in the negative set (i.e., medium conditions; for negation results, see Section 4.2.5.2.3). While a convention is to use one common fixed-cutoff value across all experimental conditions, the following two reasons necessitated using different fixed-cutoff values to protect against unnecessary data loss or data skew in only some of the conditions. First, the descriptive statistic tests found that there was a huge difference between the valid fixation onset latencies in the fast conditions and those in the slow conditions. For example, the median values for the Emphatic True and Neutral True conditions were 0.82 and 0.85 seconds respectively, whereas the median values were 1.68 and 1.9 seconds for the Contrastive True and
Emphatic False conditions (Note that these values are measured from the target word onset; therefore, 500 milliseconds needs to be deducted from these values when discussing the actual latencies taken from the signal-driven fixation onset). Second, there is a theoretical reason that explains the split pattern in the fixation latency data. As explained in the predictions section and seen in the fixation proportion figures above, the nature of the experimental conditions was very different, especially in the affirmative set. In particular, a need to shift looks between two rooms resulted in much slower fixation latencies to the target in Contrastive True and Emphatic False conditions. Due to these reasons, all instances of fixation onset times were organized for the fast, slow, and medium (i.e., negative) conditions, and were examined in sorted line graphs. The percentage of data replacement (with those fixed-cutoff values mentioned above) are: 3.28% in the fast conditions, 3.60% in the slow conditions, and 1.61% in the medium conditions (i.e., all negative conditions).

All remaining data were then trimmed once more to include only those fixations that were driven by processing the sound signal (i.e., only those fixations that occurred after 500 ms from the target word onset; about 11% of the data was removed in the affirmative set). These valid fixations from all individual trials (without aggregation) were fitted into linear mixed-effects regression models in R (R Development Core Team, 2007), to determine the best model for each of the following comparisons in the affirmative set: (a) all four conditions, (b) two slow conditions, and (c) two fast conditions. All models that are reported here included Experimental Condition as one fixed-effect term, and Items and Subjects as two random-effect terms, since likelihood ratio tests using the function anova (Baayen, 2008; Johnson, 2008) verified that more complicated models including another random factor, Experimental Lists, produced the same pattern of significant results.

Table 4.4 below presents the results from the four-condition model, which treated the Contrastive True condition as the default condition. As seen in the table, the positive coefficient for the Emphatic False condition suggests that the estimated first fixation latency in this condition was 0.372 seconds greater than that in the Contrastive True condition, and this difference was significant at $\alpha = 0.001$ level. This verifies that
participants shifted their looks to the other room significantly faster in the Contrastive True condition than in the Emphatic False condition. On the other hand, the negative coefficients for both Emphatic True and Neutral True conditions suggest that the estimated first fixation onset latencies were −0.84 and −0.85 smaller in these two conditions than in the Contrastive True condition at α = 0.0001. In other words, when compared to the Contrastive True condition, first fixations on the TARGET object occurred significantly faster in the Emphatic True and Neutral True conditions, whereas the first fixations on the TARGET were significantly slower in the Emphatic False condition. This also confirms that the difference in object search times between the Emphatic False and Emphatic True conditions was the greatest.

Table 4.4. Fixed effects from the four condition model in the affirmative set; the intercept estimate shows the predicted mean first-fixation latency on the TARGET in the Contrastive True condition (measurement starting from the target word onset).

|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| (Intercept)| 2.0422   | 0.1429     | 14.2960 | 0.0000   |
| EmFalse    | 0.3722   | 0.0924     | 4.0280  | 0.0001   |
| EmTrue     | −0.8364  | 0.0984     | −8.4960 | 0.0000   |
| NeuTrue    | −0.8531  | 0.0972     | −8.7780 | 0.0000   |

To verify if the difference between the Contrastive True and Emphatic False conditions results from data driven by a room-switch cue (e.g., L-H% or search failure), the next model compared first fixation latencies that occurred 722 ms after the target object word onset in these two conditions. The reason to use 722 ms as the analysis onset is because this value represents, on average, the earliest possible time when people receive a meaningful signal for a room switch by processing the low pitch dip in the rising cue L-H% in the Contrastive True condition (i.e., average onset of the rise L-H% 222 ms + 500 ms needed for signal-driven fixation rise). On the other hand, the earliest possible time to confirm a search failure in the Emphatic False condition is on average 939 ms, the time point when the target word’s segmental information is fully processed (i.e., average offset of the target word 439 ms + 500 ms). The model’s results in Table 4.5 show that, when compared to the Contrastive condition, the estimated coefficient for the Emphatic False condition is now even 0.418 seconds greater at the significance level of α = 0.0001.
Table 4.5. Fixed effects from the two slow condition model in the affirmative set; the intercept estimate shows the predicted mean first-fixation latency on the TARGET in the Contrastive True condition (measurement starting from the target word onset).

|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| Intercept  | 2.0686   | 0.217      | 9.531   | 0.0000   |
| EmFalse    | 0.4183   | 0.1086     | 3.853   | 0.0001   |

The last analysis compared just the Emphatic True and Neutral True conditions, using the latencies for the first fixations on TARGET that happened 500 ms after the target noun onset. The results in Table 4.6 indicate that both the Emphatic True and Neutral True conditions were equally beneficial for finding the TARGET quickly, and there was no significant difference between these conditions. This perhaps suggests a comparable effect from L+H* (on the verb had in the Emphatic True condition) and H* (on the object in the Neutral True condition) as a general attention-increase cue.

Table 4.6. Fixed effects from the fast condition model in the affirmative set; the intercept estimate shows the mean first-fixation latency predicted for the Emphatic True condition (measured from the target word onset).

|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| Intercept  | 1.1509   | 0.0666     | 17.2800 | 0.0000   |
| NeuTrue    | 0.0385   | 0.0626     | 0.5160  | 0.5386   |

In summary, the analysis on the latencies of the first fixations to the TARGET verified that participants were able to visually locate the TARGET significantly faster in the Emphatic True and Neutral True conditions than in the Contrastive True and Emphatic False conditions. Between these slow conditions, latencies were significantly smaller in the Contrastive True condition than in the Emphatic False condition, which confirms the usefulness of a linguistic cue (L-H%) in inducing eye-movement shifts between the rooms. This suggests that the L-H% cue enabled participants to construct an inference for a new state based upon the mentioned state.
**Statistical analysis on the likelihood of fixations on TARGET vs. OTHER ROOM COHORT**

The next analysis concerned the extent to which experimental conditions modulated the way time influenced the proportion of looks to TARGET and/or OTHER ROOM COHORT. That is, the dependent measure is a categorical variable (i.e., eye fixations: fixated or not), while the independent variable includes a continuous measure (i.e., time). Related to this point, there is an increasing awareness in the psycholinguistics field that an analysis method like ANOVA that was traditionally used for the fixation proportion analysis is suitable for a completely opposite situation, where a researcher investigates the influence of a categorical variable on a continuous measure such as reaction times (e.g., Barr, 2008; Jaeger, 2008). Thus, I provide an alternative analysis suggested by Barr (2008), where a mixed-effects linear regression modeling is adjusted to handle fixation data on a weighted log-odds scale, instead of a regular proportion scale; so called the mixed-effects logistic regression (for detailed discussion about this method, see Barr, 2008).

For data preparation, fixations on the TARGET (or OTHER ROOM COHORT) gathered from each subject in each condition were aggregated into 50 ms time bins. As mentioned above, this aggregation handles the dependence in eye data created by the same participant in the same condition (because consecutive fixations cannot be completely discrete; see Barr, 2008). Then, the number of fixations in each bin for the relevant category (e.g., TARGET or OTHER ROOM COHORT) was transformed into empirical logit using the following function, where “y” represents the number of fixations within a bin and “N” represents the total possible number of fixations in that bin (Barr, 2008):

\[ \eta' = \ln((y + .5) / (N - y + .5)) \]

This function takes care of the problem that, in log odds, the value becomes infinite when the exponent value approaches to 0 or 1. Figure 4.14 below plots the transformed log of the odds for looking at the TARGET in four affirmative conditions.
Figure 4.14. Time course of the fixations to the TARGET objects in the four affirmative conditions; log odds scale.

The pattern in Figure 4.14 is essentially the same as the one that was shown earlier with the proportion scale (Figure 4.6). For ease of understanding, the values on the log odds scale were translated into the proportion scale in Table 4.7 (adapted from Johnson, 2008). The fact that the likelihood of looking at the TARGET was overall less than 0.5 in the probability term is simply because there was a maximum of 12 objects per room. For example, the probability of looking at the TARGET given 12 options is only 0.0833. In terms of odds (i.e., likelihood of an event occurrence over non-occurrence), the odds of looking at the TARGET is 0.0909. If objects in both rooms are considered as possibilities, then the odds of looking at the TARGET decrease remarkably to 0.0435, which is less than 0.1 in the probability scale.

Table 4.7. Scale comparison.

<table>
<thead>
<tr>
<th>Log odds</th>
<th>Odds</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.386</td>
<td>4.00</td>
<td>0.8</td>
</tr>
<tr>
<td>0.847</td>
<td>2.33</td>
<td>0.7</td>
</tr>
<tr>
<td>0.405</td>
<td>1.50</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>0.000</strong></td>
<td><strong>1.00</strong></td>
<td><strong>0.5</strong></td>
</tr>
<tr>
<td>-0.405</td>
<td>0.67</td>
<td>0.4</td>
</tr>
<tr>
<td>-0.847</td>
<td>0.43</td>
<td>0.3</td>
</tr>
<tr>
<td>-1.386</td>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td>-2.197</td>
<td>0.11</td>
<td>0.1</td>
</tr>
<tr>
<td>-4.595</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Before fitting data into a model, an effect coding scheme was created, where a combination of three numeric variables could represent each of the four test conditions in a way that the parameter estimates in the model (i.e., the regression coefficients) would be equivalent to the mean difference in log odds between conditions (Barr, 2008). The values in those variables were set up in a way that the model could treat the Contrastive True condition as the baseline condition.

Among the three models that I report on here, the first model tested the likelihood of looking at the TARGET across the four conditions in the affirmative set. The analysis window covered the fixations that occurred during 501–3000 ms after the target object noun onset. The analysis onset 501 ms was chosen in consideration of the 500 ms needed to plan and execute eye movements as a result of processing a linguistic signal in the complex visual display. Before modeling the data, each empirical logit in the data was given regression weights (to fulfill the requirement of using a transformed variable on a linear regression model) via a formula ‘1/v’, where $v = (1/y + .5) + (1/(N − y + .5))$ (Barr, 2008; McCullagh & Nelder, 1989). Table 4.8 presents the parameter coefficients from this model. The model included seven fixed effect terms: the three variables in the effect coding scheme, of which combinations uniquely represented each condition (i.e., to test the main effect of each test condition at the analysis onset, 501 ms), time (which translates the bin’s identity into seconds information to test the slope in the function), and the time combined with each of the three variables (i.e., to test the effect of the test factor on the slope, time). The model also included two random-effect terms to handle the dependency as well as variability resulting from experiment conditions and participants, on both the intercept and the slope (for more detailed information about how a model treats these factors, see Barr, 2008).

\[^{18}\text{I thank Dr. Dale Barr for a personal communication regarding a solution for an effect coding scheme for the current study.}\]
Table 4.8. The affirmative set model; the Contrastive True condition as the base level; Analysis window = 501–3000 ms (from the target object noun onset).

<table>
<thead>
<tr>
<th></th>
<th>Estimate&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Estimate&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Std. Error</th>
<th>t value</th>
<th>Sig. level (α)&lt;sup&gt;19&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>−1.95248</td>
<td>−7.05</td>
<td>0.02935</td>
<td>−66.52</td>
<td>0.001</td>
</tr>
<tr>
<td>C2(EmFalse)</td>
<td>−0.04166</td>
<td>−1.04</td>
<td>0.08372</td>
<td>−0.50</td>
<td>ns</td>
</tr>
<tr>
<td>C3(EmTrue)</td>
<td>0.97026</td>
<td>2.64</td>
<td>0.08066</td>
<td>12.03</td>
<td>0.001</td>
</tr>
<tr>
<td>C4(NeuTrue)</td>
<td>1.09153</td>
<td>2.98</td>
<td>0.08095</td>
<td>13.48</td>
<td>0.001</td>
</tr>
<tr>
<td>t1 (slope)</td>
<td>0.10986</td>
<td>1.01</td>
<td>0.02142</td>
<td>5.13</td>
<td>0.001</td>
</tr>
<tr>
<td>t1C2(EmFalse)</td>
<td>−0.05819</td>
<td>−1.00</td>
<td>0.05699</td>
<td>−1.02</td>
<td>ns</td>
</tr>
<tr>
<td>t1C3(EmTrue)</td>
<td>−0.51681</td>
<td>−1.03</td>
<td>0.05941</td>
<td>−8.70</td>
<td>0.001</td>
</tr>
<tr>
<td>t1C4(NeuTrue)</td>
<td>−0.59391</td>
<td>−1.03</td>
<td>0.05923</td>
<td>−10.03</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>: Estimate values in log odds.

<sup>b</sup>: values in Estimate<sup>a</sup> translated into odds.

The intercept represents the estimated mean likelihood of looking at the TARGET in the Contrastive True condition at the onset of the analysis window: 501 ms after the target object noun onset. The negative sign suggests that at this time, participants were 7.05 times more likely to be looking somewhere other than at the TARGET in the Contrastive True condition (exp(1.95) = 7.05). This is expected, given that participants were still looking in Lisa’s room during this time, while the TARGET object was in Bart’s room in this Contrastive True condition. The coefficient of the slope term (time), however, suggests that, as every 50 ms passed from the analysis onset, the likelihood of looking at the TARGET in the Contrastive True condition increased 1.01 times (exp(0.10986*0.05) = 1.01), and this increase was significant at α = 0.001.

As for the Emphatic False condition, the likelihood of looking at the TARGET at the analysis onset was 1.04 times smaller in this condition than in the Contrastive True condition, although this difference didn’t reach significance. This is not surprising, given that the analysis onset (i.e., 501 ms) represents the time where participants’ fixations started showing the evidence of processing the initial segment of the target word, which is before the time point at which cues are differentiated (i.e., L-H% vs. search failure) in these two conditions. Also, the rate of fixation increase over time was not statistically different in these two conditions, although the likelihood of fixation increase to the

<sup>19</sup> α values are extracted from a table with percentage points of the two-tailed t distribution (Myers & Well, 2003), because MCMC sampling is not yet implemented in R for this type of modeling (Barr, 2008; R Development Core Team, 2007).
TARGET was numerically smaller in the Emphatic False condition than in the Contrastive True condition.

As for the Emphatic True and Neutral True conditions, the coefficients at the onset of the analysis window suggest that participants were 2.64 and 2.98 times more likely to be looking at the TARGET in these two conditions than in the Contrastive True condition, which converges with the findings from the first fixation latency and the mouse click data. Over time, however, the likelihood of fixation increase to the TARGET was 1.03 times smaller in these two conditions than in the Contrastive True condition, which suggests that, after reaching the peak earlier, fixations in the Emphatic and Neutral True conditions decreased quickly.

Given these global results, the next analysis reduced the analysis time window to just 1050 ms–1350 ms after the target object noun onset. The analysis onset at 1050 ms represents the mean crossover point between the Contrastive True and Emphatic False conditions, where the likelihood of looking at the TARGET started dominating the likelihood of looking at the OTHER ROOM COHORT. A question evaluated here was whether the fixation shift guided by a linguistic signal like the final rising tone (L-H%) in the Contrastive True condition had a reliable processing advantage over the fixation shift that happened merely because of a search failure. The model contained the same fixed and random effect terms as the other model explained above, except that the size of the analysis window was reduced. Table 4.9 shows the results of the model.

Table 4.9. The affirmative set model; the Contrastive True condition as the base level; Analysis window = 1050–1350ms (from the target object noun onset).

<table>
<thead>
<tr>
<th></th>
<th>Estimate(^a)</th>
<th>Estimate(^b)</th>
<th>Std.Error</th>
<th>t value</th>
<th>Sig. level ((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.1006</td>
<td>-8.17</td>
<td>0.1201</td>
<td>-17.491</td>
<td>0.001</td>
</tr>
<tr>
<td>C2(EmFalse)</td>
<td>-0.7911</td>
<td>-2.21</td>
<td>0.3487</td>
<td>-2.269</td>
<td>0.05</td>
</tr>
<tr>
<td>C3(EmTrue)</td>
<td>1.1902</td>
<td>3.29</td>
<td>0.3224</td>
<td>3.692</td>
<td>0.001</td>
</tr>
<tr>
<td>C4(NeuTrue)</td>
<td>0.5954</td>
<td>1.81</td>
<td>0.3312</td>
<td>1.798</td>
<td>0.1</td>
</tr>
<tr>
<td>t1 (slope)</td>
<td>0.5603</td>
<td>1.03</td>
<td>0.1628</td>
<td>3.441</td>
<td>0.001</td>
</tr>
<tr>
<td>t1C2(EmFalse)</td>
<td>0.9049</td>
<td>1.05</td>
<td>0.4727</td>
<td>1.914</td>
<td>0.1</td>
</tr>
<tr>
<td>t1C3(EmTrue)</td>
<td>-0.6094</td>
<td>-1.03</td>
<td>0.4432</td>
<td>-1.375</td>
<td>n.s.</td>
</tr>
<tr>
<td>t1C4(NeuTrue)</td>
<td>0.4015</td>
<td>1.02</td>
<td>0.4506</td>
<td>0.891</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

\(a\): Estimate values in log odds.
\(b\): values in Estimate\(^a\) translated into odds.
The coefficient estimates suggest that, at the analysis onset, participants were still 8.17 times more likely to be looking somewhere other than the TARGET in the Contrastive True condition. It is important, however, that the likelihood of looking at the TARGET was significantly smaller by 2.21 times in the Emphatic False condition than in the Contrastive True condition. This difference contrasts with the results from the Emphatic True condition, where the likelihood of looking at the TARGET was significantly greater, by 3.29 times, in the Emphatic True condition than in the Contrastive True condition. This again confirms that, despite having exactly the same sentence materials, the effect of the Emphatic False and Emphatic True conditions is the greatest. While the Emphatic True condition was as useful as the Neutral True condition for finding the TARGET quickly, the Emphatic False condition slowed object search times even more than another condition with an equivalent object location but with a tune that contained a meaningful signal for eye-movement shift between the rooms (i.e., the Contrastive True condition).

The rate of fixation increase on the TARGET over time (as every 50 ms passed) was not significantly different among these conditions. Although the coefficient on the time effect on the Emphatic False condition suggests a marginally faster increase rate in that condition than in the Contrastive True condition (at $\alpha = 0.1$), this pattern does not bear any theoretical interest because once the looks are switched between rooms, the matter becomes simply to find the TARGET quickly.

To complement the findings so far, the last model compared fixations to the OTHER ROOM COHORT in the Contrastive True and Emphatic False conditions, to evaluate how quickly looks to the OTHER ROOM COHORT decreased after the fixation crossover point. The model was built for the analysis window starting from 1050 ms until 3000 ms after the target object noun onset. The model included three fixed effect terms—experimental condition, time, and the effect of experimental condition on time—aalong with two random effect terms, one that controls variance caused by the participant-per-condition aggregation and the other for the participant variance (see Barr 2008). The model considered the Contrastive True condition as the baseline condition. Table 4.10 shows the parameter estimates from the model.
Table 4.10. The slow set model; the Contrastive True condition as the base level; Analysis window = 1050–3000 ms (from the target object noun onset).

<table>
<thead>
<tr>
<th></th>
<th>Estimate&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Estimate&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Std.Error</th>
<th>t value</th>
<th>Sig. level (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.26453</td>
<td>-9.627</td>
<td>0.10083</td>
<td>-22.459</td>
<td>0.001</td>
</tr>
<tr>
<td>C2(EmFalse)</td>
<td>0.32737</td>
<td>1.387</td>
<td>0.14798</td>
<td>2.212</td>
<td>0.05</td>
</tr>
<tr>
<td>t1</td>
<td>-0.30632</td>
<td>-1.015</td>
<td>0.05993</td>
<td>-5.111</td>
<td>0.001</td>
</tr>
<tr>
<td>t1C2(EmFalse)</td>
<td>-0.08515</td>
<td>-1.004</td>
<td>0.09388</td>
<td>-0.907</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

<sup>a</sup>: Estimate values in log odds.
<sup>b</sup>: values in Estimate<sup>a</sup> translated into odds.

At the analysis onset, participants were looking at the OTHER ROOM COHORT 1.39 times more in the Emphatic False condition than in the Contrastive True condition, which was significant at α = 0.05. However, the likelihood of looking at the OTHER ROOM COHORT decreased at approximately the same rate in both conditions, perhaps because of the small durational difference between the onset of the rise in the Contrastive condition and the offset of the target word in the Emphatic False condition (i.e., 65 ms on average), which quickly disclosed the entire segmental information of the target word.

Overall, the results from three mixed-effects logistic regression models on the likelihood of looking at the TARGET or the OTHER ROOM COHORT verified the following points. First, fixations on the TARGET peaked and decreased earlier in the two fast conditions (Emphatic True and Neutral True conditions) than in the two slow conditions (Emphatic False and Contrastive True conditions). This confirms that L+H* on the verb itself didn’t induce a shift in looks between rooms. Second, the likelihood of fixations to the TARGET was significantly greater in the Contrastive True condition than in the Emphatic False condition, whereas the trend was the opposite with respect to the likelihood of looking at the OTHER ROOM COHORT, when the analysis onset represented the mean fixation crossover time point in these two conditions. This pattern is in part due to the fact that the linguistic signal (i.e., prosody) in the Emphatic False condition itself was in conflict with the actual location of the TARGET object in the visual display. In contrast, the continuation rise L-H% in the Contrastive True condition aided faster eye-movement shift because participants constructed the implicature that the object possessor must be changed. The overall findings converge with what was found from the first fixation latency analysis as well as the mouse click results.
4.2.5.2.3 Results from the negative set

Figure 4.15 below shows proportions of fixations to the TARGET objects in four negative conditions, as the function of time. Again, the target word onset was synchronized at the 0 ms point on the x-axis. As explained above, TARGETS were located in Bart’s room in the Contrastive True and Emphatic False conditions, whereas they were present in Lisa’s room in the Emphatic True and Neutral True conditions. This experiment set-up reflected a hypothesis that participants who hear sentences like Bart didn’t have the X can draw an inference that the object then should belong to Lisa’s room instead, given the binary option for the object possessors, which was suggested in the background display (i.e., only two rooms: Lisa’s and Bart’s).

![Time course of fixation to the target object in the negative set](image)

Figure 4.15. Proportion of fixations to the TARGET object averaged across participants, aggregated into 50 ms bins in four negative conditions.

The looking patterns in the negative set were substantially different from those in the affirmative set in a number of ways. First, starting from 400 ms after the target word onset (i.e., from the initial moment of the signal-driven gazes), the proportion of looking at the TARGET increased more quickly in the Contrastive True and Emphatic False conditions than in the Emphatic True and Neutral True conditions and not vice versa, suggesting that listening to Bart’s name directed participants’ eye movements to Bart’s room initially, thus providing a preview of objects in that room, which potentially assisted a quicker object identification in these two conditions.
To verify this pattern, fixation proportions on the TARGET were converted into a log odds scale to analyze the likelihood of looking at the TARGET in the time window between 400 ms and 1200 ms from the target word onset. The method of building a mixed-effects logistic regression model was similar to that in the affirmative set, except that the data included only the negative set. Table 4.11 shows that, at the analysis onset, the likelihood of looking at the TARGET was not different between the Contrastive True and Emphatic False conditions. However, the likelihood of looking at the TARGET was significantly smaller in the Emphatic True and Neutral True conditions than in the Contrastive True condition. Over time, fixations on the TARGET in the Contrastive True condition increased 1.08 times more as every 50 ms passed by, and the increase rate was similar in the Emphatic False condition. However, looks to the TARGET increased 1.03 times faster (as every 50 ms passed by from the analysis onset) in the Emphatic True and Neutral True conditions than in the Contrastive True condition.

Table 4.11. Fixed effects from a weighted empirical logit model analyzing the likelihood of fixating the TARGET in the negative set, as a function of experimental condition and time between 400 ms and 1200 ms from the target word onset.

|             | Estimate<sup>a</sup> | Estimate<sup>b</sup> | Std. Error | t value | Pr(>|t|) |
|-------------|----------------------|----------------------|------------|---------|---------|
| (Intercept) | -2.7920              | -16.3138             | 0.0473     | -59.03  | 0.001   |
| C2(EmFalse)| -0.0910              | -1.0952              | 0.1247     | -0.73   | n.s.    |
| C3(EmTrue)| -0.5695              | -1.7673              | 0.1287     | -4.42   | 0.002   |
| C4(NeuTrue)| -0.5860              | -1.7968              | 0.1275     | -4.60   | 0.002   |
| t1 (slope) | 1.5085               | 1.0783               | 0.0695     | 21.70   | 0.001   |
| t1C2(EmFalse)| 0.2383              | 1.0120              | 0.1868     | 1.28    | n.s.    |
| t1C3(EmTrue)| 0.6137              | 1.0312              | 0.1948     | 3.15    | 0.020   |
| t1C4(NeuTrue)| 0.6263              | 1.0318              | 0.1890     | 3.31    | 0.010   |

<sup>a</sup>: Estimate values in log odds
<sup>b</sup>: values in Estimate<sup>a</sup> translated into odds.

Despite this difference found in the fixation likelihood data, the analysis on the latencies for the very first fixations to the TARGET objects revealed that response times to visually locate the TARGET objects initially were not different across the conditions (Table 4.12). The dependent measure was the first fixation onset times, which happened once 400 ms had passed from the target object noun onset (i.e., signal-driven first fixations). Data were fitted into a linear mixed-effects regression model that included Condition as
one fixed-effect term and Subjects and Items as two random-effect terms. The intercept value shows the estimated mean in the Contrastive True condition. As seen in the table, there was no significant difference among conditions, which converges with the mouse click results; response times to click on the TARGET objects were not statistically different across the negative conditions. These results are clearly different from the results found in the affirmative set, where both first-fixation onset times and mouse click times were significantly faster in the Contrastive True condition than in the Emphatic False condition but slower in the Contrastive True condition than in the Emphatic True and Neutral True conditions.

Table 4.12. First fixation latency analysis for the negative set.

| Negative set          | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------------------|----------|------------|---------|---------|
| (Intercept)           | 1.5858   | 0.1054     | 15.0440 | 0.0000  |
| C2(EmFalse)          | 0.0394   | 0.0983     | 0.4010  | 0.6886  |
| C3(EmTrue)           | −0.0922  | 0.0957     | −0.9640 | 0.3354  |
| C4(NeuTrue)          | −0.1334  | 0.0952     | −1.4010 | 0.1615  |

To evaluate exactly how negative sentences were processed, I examined the looks to the TARGET as well as to the OTHER ROOM COHORT, as an indication of the eye-movement shift between the two rooms. Figure 4.16 provides relevant data in the Contrastive True and Emphatic False conditions and Figure 4.17 provides the data in the Emphatic True and Neutral True conditions. The red dotted line indicates the mean onset and the red solid line indicates the offset of the negation word didn’t, which was spoken with a contrastive accent L+H* in all conditions except the Neutral True condition. The onset of the target object noun is indicated by the black solid line anchored at the 0 ms point on the x-axis, and the blue solid line indicates the mean offset of the target object noun. The blue dotted line in Figure 4.16 indicates the onset of the rise L-H% in the Contrastive True condition. Note that it is necessary to add 400 ms to these values, in order to evaluate fixation changes that occurred due to the processing of linguistic signals in the current study’s complex visual display.
In Figure 4.16, fixations to the TARGET (in Bart’s room) sharply increased in both Contrastive True and Emphatic False conditions once the object noun started being processed (i.e., 400 ms from the target noun onset on the x-axis), and because of the room preview upon processing Bart, any potential differences in these two conditions were greatly reduced. However, a more subtle and yet important pattern was that by the time participants finished processing negation (i.e., that corresponds to 153 ms on the x-axis), looks to the OTHER ROOM COHORT (in Lisa’s room) slowly increased and were sustained until participants found the TARGET objects, which suggests that negation expanded people’s attention to include the contrasting room (i.e., Lisa’s room) as part of their search field. In fact, the proportion of looks to the OTHER ROOM COHORT matched the probabilities of 0.0833–0.1 (i.e., two OTHER ROOM COHORTS, among 20 to 24 objects in the entire display). A more direct measure for testing when and how quickly negation evoked shift of attention back and forth between the rooms will be to analyze the proportion of fixations counted on Lisa’s room versus Bart’s room. This analysis is planned for the near future.
The fixation patterns in the Emphatic True and Neutral True conditions (Figure 4.17) reflected the differences in the object location; that is, these conditions placed the OTHER ROOM COHORT in the mentioned subject’s room (Bart’s room) and the TARGET in the contrasting room (Lisa’s room). The figure suggests that participants started looking into the mentioned subject’s room (as evidenced by fixations on OTHER ROOM COHORT) upon processing the sentential subject. While these fixations continued to increase, once participants finished processing didn’t, their looks to the opposite room (Lisa’s room) also started increasing slowly (as evidenced by looks to TARGET at around 150 ms on the x-axis). Then, upon the processing of the target word’s onset (at around 400 ms on the x-axis), fixation proportions on the TARGET began sharply increasing, and by the time participants finished processing the target word (a little after 800 ms on the x-axis), there was a cross-over of fixations on the two rooms. Even after this point, however, looks to the OTHER ROOM COHORT were sustained for another 500 ms or more. These patterns together suggest that negation expands people’s attention to include not only what is being negated but also what could be a potential factual situation. The implications of these findings are further discussed in Section 4.3.
4.3 Discussion

Previous studies examining online reference resolution guided by affirmative sentences have shown that the contrastive pitch accent (L+H*) on a prenominal adjective, for example, is used immediately to guide the processes of finding a correct referent (e.g., Ito & Speer, 2008; Sedivy et al., 1995; Weber et al., 2008). One mechanism that has been discussed for these processes is that the contrastive pitch accent evokes people’s pragmatic knowledge of presupposition, where compatible alternative items must exist in the world under consideration, and the contrastive focus is applied to single out the target item against those alternatives. Integration of such knowledge during language processing is found to be strikingly rapid, such that people immediately look at the visually presented alternative sets as they predict the intended referent at the moment of encountering the contrastive cue, even before they hear the actual segmental information of the target referent (Ito & Speer, 2008; Weber et al., 2008).

The current study, however, has demonstrated that the type of accompanying boundary tones affects how and when the contrastive pitch accent information influences online processing of affirmative sentences. The results suggest that the pitch accent L+H* alone was not strong enough to evoke an implied meaning suggesting a change in the state of affairs mentioned in the sentence. Instead, the rising boundary tone L-H% was needed to integrate the contrast evoked by the pitch accent to build an implicature from it. Key evidence for this finding was that upon the processing of L+H*, there was a lack of eye-movement shift from the mentioned subject’s room (Lisa’s room) to the alternative subject’s room (Bart’s room). One might question whether the lack of such effects was perhaps due to the presence of the Emphatic True condition, where L+H* together with L-L% signaled emphatic affirmation of what was asserted in the sentence, and this in turn weakened the alternative-set evoking function of L+H*. However, experimental conditions also included the Emphatic False condition, which placed the target objects in the alternate subject’s room. This condition could have supported the alternative-set evoking function of L+H*. In fact, out of three conditions containing L+H*, two placed the target objects in the alternate subject’s room. However, the results showed that participants switched their looks to the alternative subject’s room (indicating the
perception of the implied contradiction), only when they encountered the rising end contour L-H% along with L+H*. This finding was further supported by the immediate shift of looks when the L-H% cue was processed, which happened significantly faster than the shift of looks to the alternative subject’s room based upon a purely behavioral reason, i.e., search failure in the initial room. That is, when the target object was located in the alternate subject’s room in the absence of supporting linguistic signals (i.e., the Emphatic False condition), both eye-movement data and the mouse click data showed a significant delay in the object search speed.

On the one hand, the finding that L-H% together with L+H* supported the state-contrast implicature suggests the importance of the whole tune (L+H* L-H%) and its distinctive meaning (e.g., Cruttenden, 1997; Dainora, 2001, 2002; Jackendoff, 1972; Lee, 2007; Liberman & Sag, 1974). On the other hand, this finding does not speak against the contrastive function of L+H*. Rather, it suggests a clear separation between promoting the accented item against some potential alternative sets (e.g., asserting that Lisa HAD the bell) and making use of an alternative set in drawing an implicature (e.g., the implied state-contrast that Bart now has the bell). The L+H* accent on the verb had promotes the state information “had” while contrasting it to the alternative state “had not.” However, it is the rising end contour L-H% that leads participants to use this contrastive information to construct an implicature.

On this view, the results of Experiment 3 provide empirical evidence for the compositionality hypothesis for tune meaning (Pierrehumbert & Hirschberg, 1990). That is, the target implicature discussed in the current study (Lisa had the bell…“but now Bart has it”) results from the distinct functions of each prosodic element of the tune, L+H* and L-H%. The eye-fixation data, closely time-locked to the processing of incremental speech input, provided on-line evidence for the separate functions driven by each prosodic unit. The L+H* accent always guided participants’ eye movements to the mentioned subject’s room, whereas it was L-H% that initiated a shift of looks to the alternative subject’s room.

This differential effect of prosody, however, was greatly reduced when the sentence materials included negation, as in Bart didn’t have the pants. Because the sentential subject name directed participants’ attention to the mentioned subject’s room,
participants had an initial opportunity to preview objects in that room. Then, upon receiving the negation cue, participants expanded their search window to include the contrasting room, while still maintaining their attention on the initially mentioned subject’s room. Because this search-expansion pattern was not found in the results from the affirmative set, this cannot be reduced to the fact that the contrasting room contained phonologically similar items (i.e., OTHER ROOM COHORT) or that TARGET objects were sometimes presented in the “incorrect” room (i.e., the Emphatic False condition). Analysis of the first fixation latencies and mouse click times on the target objects showed that there was no significant advantage for the object search speed across four negative conditions.

At first glance, this negation result seems to be quite different from the negation results found in Experiment 2, where negation reduced the activation level of the following predicate adjective, thus resulting in stronger activation of the unmentioned contradictory (factual) state right away. Negation data in the current experiment, however, suggest that people maintained their attention to the negated or counterfactual state even more strongly than to the contrasting state until they finally reached the ultimate factual state confidently. This difference, however, does not speak against the general role of lexical negation, i.e., reducing the activation level of the negated concept, while leaving the activation level of its associated concept intact. Instead, the difference suggests that the nature of the negated concept as well as the surrounding contextual information that embeds negative sentences (including the end goal) determines the extent to which either the negated/counterfactual state or the inferred factual state can maintain its activation level. For example, predicate adjectives that were negated in Experiment 2 all expressed clear binary attributes (i.e., contradictory predicates like \textit{open}), which easily spread activations to their counterpart attributes (e.g., \textit{closed}). Also, the task in Experiment 2 did not impose any behavioral goal for retaining the negated information. Thus, the reduced activation level of the negated concept remained low while the contradictory state remained highly activated. In the current experiment, the lexical item that follows the negation marker \textit{not} was the verb of possession \textit{have}, which could potentially activate an antonymically associated concept \textit{not have}. This suggests
that, even when negation processing reduces the activation level of the negated word *have*, the notion of *not have* becomes doubly highlighted because it coincides with the concept of the negated word’s associate. More important, participants’ job was to find and click on the object mentioned in the sentence, which in this case was actually the object that was the negated possession. Considering the experiment setting (i.e., objects belong to either of the two possessors), participants could have potentially inferred that something that Bart didn’t have must belong to Lisa, which should have reduced participants’ attention to the negated possessor right away. However, making such an inference was perhaps not natural or meaningful for the search task, as a variety of other sentences truthfully indicated where the objects must belong. Also, the alternative possessor was out of the negation scope and was subtly indicated only in the background, whereas the visual display presented multiple objects to choose from. This aspect might have increased the sense of contrast with respect to the “implicit” alternative objects more so than the contrastive sense with respect to the alternative possessors. Another possibility related to this point is that participants might have created an inference *Bart didn’t have X, but now he does*, in order to pick out exactly those objects that were moved by the creature in the story.

Negation results from the current study together with the results in Experiment 2 thus suggest that the way people process lexical negation reflects a probabilistic approach, where all surrounding factors (e.g., lexical property, contextual information, behavioral goal) simultaneously constrain the level of meaning activation influenced by negation processing. In this sense, processing negative sentences is different from processing affirmative sentences, which normally guides incremental integration of the incoming meaning. Moreover, this study suggests differences in the processing time for a contradictory meaning that is signaled by a lexical cue (i.e., *not*) versus a prosodic cue (i.e., L+H* and/or L-H%). The first fixation latencies as well as mouse click times obtained from the affirmative data suggest that the contradictory meaning perceived from the prosodic tune reflects a clear shift of attention from the literal meaning to the implied meaning (as long as there is no lexical item that automatically activates an opposing concept, as in Experiment 2). The contradictory meaning constructed by the explicit
negation, however, allowed lingering activation of the literal meaning while the ultimate contradictory meaning was being constructed, especially when the negated information (e.g., negated possession) was needed to complete the behavioral goal (e.g., finding that object) as in Experiment 3. Future research will be needed to clarify to what extent the persistence of the residual activation of the literal (negated) meaning is based upon task demands rather than being an integral part of the processing of negative sentences.
CHAPTER 5
CONCLUSION

This dissertation has presented three sets of distinct experiments that investigated how listeners construct implied meaning by creating conceptual relations among lexical cues, prosodic cues, and experience with the world. This section begins by summarizing the behavioral tasks and findings from each experiment. Then, the findings are discussed in terms of their implications for a range of fields to which the issues of intonation, implied meaning, and language processing matter. The dissertation concludes by offering future research questions.

5.1 Summary of tasks and findings

Chapter 2 presented two perception experiments to investigate whether the form of sentence intonation is responsible for the type and nature of the meaning that listeners construct, particularly those meanings that are not expressed in the words themselves. Under the assumption that sentence understanding engages people to construct mental representations of the discourse, both Experiment 1A and Experiment 1B used a continuation paradigm, where participants heard target sentences spoken in various intonations and produced follow-up sentences that would naturally continue the discourse.

The results from Experiment 1A showed that, when a simple lexical string like The jar was full was spoken with a strong contrastive accent (L+H*) on the auxiliary was, followed by a slightly rising tone (L-H%) at the end of full, listeners perceived that a change has occurred with respect to the asserted state, thereby creating a follow-up sentence like “But now it is empty.” Producing such state-contrast continuations was an otherwise unusual response that was induced by this particular intonation, as indicated by the extremely low rate of participants’ producing similar continuations after processing either affirmative neutral sentences or sentences containing the sentential negation not, which is an established linguistic marker for contrast. Subsequent analysis showed that continuations following negative sentences mainly provided a reason for why the state did not hold, e.g., The candle was not lit...“because it was against the fire code,” or discussed a consequence of the negated state, e.g., The cage was not locked...“so my dog
ran away,” rather than focusing on any perceived contrast. This is likely due to the nature of the task, which required a natural continuation of the discourse. Similarly, conveying state-contrast implicature was not the only function of the L+H* L-H% tune, as it readily permitted other types of meanings as well, confirming the one-to-many relationship between intonation form and meaning.

Experiment 1B further investigated whether the state-contrast implicature results from the meaning of the intonation contour (L+H* L-H%) as a whole, or if the meaning reflects separate functions of individual prosodic units (L+H* vs. L-H%). The results indicated a strikingly compositional mechanism, where the rate at which continuations expressed the state-contrast implicature was proportional to the presence of both cues, just the rising cue, or the contrastive pitch accent, in that order. There was no evidence for a super additive effect resulting from the entire intonation contour, which indicated that intonational meaning is composed incrementally as each cue arrives at the interpretation system.

Experiment 1, however, gathered sentence interpretation from an off-line task, where participants deliberately typed out continuations that would naturally extend the discourse. Thus, conscious and effortful production processes might have complicated the nature of the meanings that were initially accessed before the continuations were created. To probe listeners’ on-line meaning representation at a more subconscious level, Experiment 2 (presented in Chapter 3) employed a picture naming paradigm, where meaning activation was indirectly and implicitly measured through response times to the pictures that followed sentence presentation at six different inter-stimulus intervals (0–2500 ms, with 500 ms intervals between the sentence offset and the image onset).

Research questions included whether the forms of mental representation that are accessed from lexical and nonlexical cues (e.g., negation vs. contrastive prosody) share similar features, and also, whether both described and implied meanings take similar representational formats. Moreover, the time course through which people arrive at an implied meaning was investigated, in order to evaluate whether different types of linguistic cues require different manners and amounts of processing during on-line meaning construction.
The picture naming response times indicated that one mechanism people utilize to construct meaning—it being explicitly described or merely implied—is by evoking and integrating experience-based knowledge that is tightly associated with the concepts expressed in the words and phrases. Crucially, successful simulation of implied meaning (indicating a change in the state of affairs) was possible only when the sentence included a linguistic cue whose function is designated for marking contrast or inference (as in the negation *not*, or the contrastive pitch accent L+H*, as well as the continuation rise L-H%). However, the detailed processes involved in constructing implied meaning were different depending on the cue type. When the sentences explicitly negated words that denote binary attributes (as in *not full* or *not open*), the negation reduced or suppressed the activation level of the mentioned state, thereby allowing people to access the contradictory factual state (as in *empty* or *closed*) most dominantly even at the sentence offset. While this relatively enhanced availability of the factual concept was sustained throughout the ISIs, the strongest match effect was found at 1500 ms ISI, which suggests that people successfully selected the interpretation as it represents one of the most plausible factual states under consideration. However, in the absence of the lexical item *not*, processing the binary predicates activated both of the states initially, until a change in state was successfully inferred from contrastive prosody (i.e., the contrastive sentences). The strongest match-facilitation effect toward the contradictory state appeared 500 ms later in the contrastive sentences than in the negative sentences, although there was also a temporal difference of 363 ms in the acoustic signal: the onset of the critical element L-H% unfolded temporally later than the onset of the negation cue *not*. These findings together highlight differences as well as similarities between the cue type in evoking an implied meaning, which is further discussed in the section on implications below. The results also underscore the important role of intonation on higher meaning perception. Sentences with exactly the same lexical contents that lacked the contrastive intonation (i.e., affirmative neutral sentences) did not generate such an inference effect, even though both mentioned and opposite meanings were initially available upon processing the contradictory predicate (in the absence of the negation *not*).
Chapter 4 presented the last experiment of this dissertation research, which provided more fine-grained analysis on how implied meaning is constructed at each moment of sentence processing. To do so, I used a visual search paradigm, where participants’ eye movements were being tracked while they were actively finding various objects mentioned in sentences. This experiment aimed to obtain detailed evidence on precisely which subunit of sentence intonation is responsible for people perceiving state-contrast and further constructing an inference about a possible factual state. For this question, a novel task environment and materials were created, where two cartoon characters’ belongings were moved around, and the participants’ job was to find and click on the mentioned item as quickly as possible in order to return it to the correct possessor’s room. The notion of contrast set was manipulated subtly, with the alternate possessors presented as part of the fixed background information. Combining three intonational tunes in test sentences together with two possible object locations (Lisa’s room vs. Bart’s room) allowed the speed and proportion of eye-movement shifts between the rooms to indicate how quickly and reliably participants perceived a state-contrast implicature that suggests a change in the object possessor as in *Lisa HAD the bell...“but now Bart has it.”*

Both the mouse click data and eye-movement data showed that the contrastive pitch accent L+H* alone was not enough to induce a shift in eye movements between the rooms. Instead, the rising cue L-H% was needed to integrate what was asserted in the sentence together with an evoked contrast by L+H*, in order for participants to infer a factual situation under consideration. In the absence of L-H%, participants were able to use the focused concept (i.e., *Lisa HAD*) against the contrasted concept (e.g., *Lisa had not*) to quickly constrain their search domain and find the target object in the mentioned subject’s room, especially when the visual display supported the meaning of the emphatic assertion (i.e., the Emphatic True condition). When the visual environment failed to support such an interpretation, the elevated accessibility of the focused concept did not produce any observable effect (i.e., the Emphatic False condition). Also, the felicitous effect of the Emphatic True condition was similar to that of the Neutral True condition, where the target object name was spoken with the H* accent. This suggests that listeners
are flexible in utilizing different types of felicitous cues to benefit their sentence interpretation to the greatest extent.

Experiment 3 also compared negative sentences that were matched for intonation form and object location with the affirmative sentences. The negation results showed that processing negation expanded people’s search attention to the whole visual field, which included not only the mentioned possessor’s room but also the alternate possessor’s room, thus reducing any observable effect of sentence intonation. Furthermore, the eye movement measure that is closely time-locked to the unfolding speech input disclosed detailed information on precisely when the negation reduces or suppresses the availability of the negated concept. Taking the default case (i.e., the Neutral True condition), for example, the proportion of fixations on the target as well as on the other room cohort over time showed that, when participants finished processing negation (as in Bart didn’t), they still focused more on the mentioned subject’s room (i.e., Bart’s room), producing a greater proportion of fixations on the target (at that point, these fixations on the target reflected chance fixations). However, as soon as they finished sentence processing (i.e., at the sentence offset), participants immediately shifted their looks to the alternative subject’s room (i.e., Lisa’s room), which suggests that the meaning of negation was successfully integrated with the sentence meaning to reduce the availability of the negated concept by the sentence offset. While this finding converges with what was found in Experiment 2, it will be useful to look further into the new fixations in the period following negation.

5.2 Implications of the findings

The set of findings presented thus far implies several notable points that can benefit our understanding of the connection between three distinct topics—sentence intonation, implied meaning, and spoken language processing. Moreover, the investigation of the time course of processing in Experiments 2 and 3 discovered an important piece of missing information in the current literature: precisely when the negation meaning is integrated with sentence interpretation. I take each of these issues as
the center of a discussion to highlight the integral aspects of these topics on sentence processing from different angles.

First, sentence intonation has substantial impact on the meaning that people construct and represent in their minds. The current study extends this point further by establishing that individual subunits of intonation—not just pitch accents but also the shape of the phrase and boundary tones—generate measurable effects on meaning construction both off-line and on-line. The results of Experiments 1B and 3 together highlight this compositional mechanism, through which, at the time when each prosodic unit is encountered, its function is evaluated and integrated with the meaning that is being concurrently constructed through the segmental information channel. This suggests that a seemingly detachable meaning from an intonational contour actually emerges from each prosodic unit’s independent function, thereby validating the autosegmental approach to intonation theory (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980).

Second, the incremental composition of a tune’s meaning is possible due to the sentence processing mechanism that takes each cue’s function probabilistically in the context of multiple factors influencing the interpretive decision at a given moment. This claim is in line with a growing body of research showing that multiple information sources are simultaneously communicated among different channels or levels of linguistic structure. Also, a few recent studies have added that such interaction occurs even between two far ends of the information spectrum, namely phonetics and phonology on one end and pragmatics on the other (Dahan et al., 2002; Ito & Speer, 2008; Rohde & Ettlinger, 2010; Watson, Tanenhaus, & Gunlogson, 2008). The current study (Experiment 3 in particular) advances these claims by confirming that people’s knowledge about the discourse situation together with their knowledge on intonation inventories can give rise to differential effects of a single element like L+H*. Existing experimental evidence has uniformly supported the understanding that the L+H* accent evokes contrast sets or alternative sets, which means that it induces immediate or even predictive looks to the alternative sets when they are visually present. The current study, however, provided evidence that L+H* in fact serves two separate functions simultaneously. On the one hand, it promotes the accented concept to the discourse foreground. On the other hand, it
creates a sense of contrast, thus evoking alternatives to the accented notion. I expect that different acoustic components of the accent (i.e., the high prominence vs. the rising movement) are perhaps responsible for each of these functions. What is more important is that the discourse environment and the rest of the unfolding tonal context constrain the ways in which these information pieces are integrated into the ongoing meaning construction and influence subsequent behavioral decisions. For example, the foregrounded status of the accented concept was able to influence behavioral outcomes (e.g., people completed their object search faster) only when the promoted concept continued to be supported by both (a) the low falling tone L-L%, affirming what was just stated, and (b) the visual environment supporting the evoked meaning (e.g., the object was placed in the expected room); i.e., the Emphatic True condition. When these constraints were violated, the foregrounded concept failed to exert any noticeable contribution (i.e., the Emphatic False condition). However, when the L+H* accent was followed by another meaningful cue (L-H%) signaling a link between what was stated and what will (or could) be stated, the alternative sets that were previously evoked by L+H* were evaluated against the accented concept to create meaningful inferences about a changed state that is contradictory to what was asserted. This suggests that people’s holistic knowledge of the discourse environment (e.g., an object being searched for must belong to either Lisa or Bart) is rapidly integrated with the incoming speech input, where acoustic prominence at a given moment (e.g., HAD) is evaluated against the background information. This in turn allows a prosodic unit like L+H* to generate differential effects depending on the contextual situation, in terms of both tonal context and visual environment. Thus, the fact that there was no immediate eye-movement shift upon L+H* does not mean that there was a delay in the use of the prosodic information.

These findings also highlight the fact that the contrastive function of L+H* is somewhat different from that of the negation marker not. Negation functions as an explicit cue that reduces the availability of the negated concept, which in turn can boost the relative saliency of the alternative state by the time sentence processing is finished. This effect of negation can be observed especially when the alternative concept is accessible (e.g., through associative activation from the contradictory predicate as in
Experiment 2, or discourse context limiting the number of alternative sets as in Experiment 3). However, the contrastive pitch accent L+H* does not necessarily bias people to pay attention only to the alternative state even after the sentence offset, especially in the absence of visually presented alternative sets (Experiment 2). This suggests that contrastive prosody is a form of “pragmatic” negation indeed. A speaker’s choice of contrastive prosody over negation perhaps reflects that, while indicating the existence of contrast, they want to leave room for listeners to entertain the alternative possibilities to a greater extent.

As mentioned above, the current study found that the moment when negation suppresses the negated word’s concept is not right at the negation offset but when time is given for meaning integration (e.g., at the sentence offset). This finding bears significance regarding a long-standing debate in the lexical processing literature: to what extent initial lexical access is constrained by prior (sentential) contextual information. Based upon the results in Experiments 2 and 3, I suggest that initial activation of word meaning (e.g., open) is not completely constrained by the prior sentential context. Thus, even if the word is preceded by negation as in not open, the concept of the negated word open can still spread its activation to its conceptual associates such as closed. Then, as the meanings of the words not and open are quickly integrated to form a construal of the state or event, negation can effectively reduce the negated word’s concept without influencing the associated word’s concept. While this is in line with MacDonald and Just’s (1989) finding that no bread reduced the availability of bread without influencing the availability of butter, the study did not provide direct evidence on precisely when the negation impacts meaning integration. The results from the eye-tracking experiment provide direct evidence.

The last implication of this study is that sentence prosody plays a functional role during mental simulation of meaning (Experiment 2). An intriguing aspect of this point is that prosodic cues that trigger changes in mental simulation do not (and can not) specify the form or type of the simulated meaning. In fact, those cues (e.g., the contrastive cue L+H* or the continuation cue L-H%) are mere acoustic properties that change the ways segmental information is pronounced and heard. However, listeners are still able to use
this acoustic information to construct simulation of meaning that goes beyond what was described in the sentence. The current study showed that this effect is not simply due to particular properties of the lexical items used in the sentence. When exactly the same lexical contents were spoken with different intonations (e.g., a neutral one), people could not successfully simulate an implied factual situation. This finding underscores that language comprehension (here, viewed as mental simulation of meaning) emerges from dynamic cognitive processes that listeners carry out while engaging with both lexical and nonlexical information sources. During these processes, perceptual traces that are tightly associated with people's experience with entities and events are subconsciously activated, and crucially, intonation modulates the activation level of these associated concepts as perceptual traces.

As mentioned in previous chapters, acoustic salience is closely related to listeners' attention level in a structured way. That is to say, what is important is not the mere strength in the individual acoustic signal but the differing degrees of strength that create the prominence relationship among those signals. From the mental simulation perspective, I propose that this prominence relationship among the acoustic signals is closely related to what aspects of perceptual experiences are simulated in greater or lesser detail. Future research is needed to systematically investigate this issue, but this study showed that contrastive accent on the verb expressing both time and state highlighted perceptual traces related to event sequences, which were then employed to create an inference upon another discourse linking cue (L-H%).

In summary, this study underscores the highly interactive nature of processing mechanisms that utilize multiple information sources arising from a wide range of channels rapidly and coherently. Section 5.3, the final section, discusses some remaining questions.

5.3 Remaining questions

Using three completely different experiments, the current research delineated some of the underlying cognitive processes that are responsible for constructing and representing implied meanings cued by sentence intonation. While the results provide
important insights, this research represents only the first step toward an understanding of the dynamic relationships among intonation, implied meaning, and processing. In this section, I discuss some of the questions that can further advance this interdisciplinary endeavor.

First, the current study expanded the notion of contrast by exploring *state-contrast implicature*, which taps into conceptual relations among discourse entities, states, and event sequences over time. Nevertheless, experimental investigations necessitated the use of materials that could easily constrain the set of alternatives activated in the mind. This suggests that further experimental delineation is needed to show in a less-constraining or non-constraining setting (e.g., natural conversation): (a) what alternative sets are evoked and why, (b) which of the alternatives is ultimately selected and why, and (c) what additional factors are needed to systematically guide those processes. Another question to explore is to what extent evoked alternatives have direct implications on the activation of perceptual traces. The current study suggests that perception of alternative notions is reflected in the resulting visual simulation. Does this imply that people construct contrastive notions by activating perceptual traces?

Second, a spoken language processing model that can incorporate the set of findings described in this study is needed. The accumulating evidence for interactive processing for language has led processing modelers to propose various ways in which multiple information sources can be combined at a given level of linguistic structure during incremental processing (e.g., Goldinger, 1996; Jurafsky, 1996; MacDonald et al., 1994; McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2000; Spivey & Tanenhaus, 1998). However, these existing models do not articulate how prosodic information that operates at the suprasegmental level can be incorporated with the other structures at a given moment of processing. Furthermore, these models have yet to explain the precise mechanism that can account for the ways in which high-level information like pragmatic knowledge can be rapidly combined with prosodic information that comes through a low-level acoustic channel. What needs to be addressed is exactly how and at what point these sets of acoustic features (e.g., pitch, loudness, duration) are translated into a meaningful intonation unit (e.g., L+H*, L-H%), questions
that have direct implications for not just lexical access but also meaning computation. One possible way to envision this is to build upon something similar to the TRACE model (McClelland & Elman, 1986), where multidimensional acoustic and suprasegmental features are represented for distributed parallel activation at each given time slice. Then, the system needs to include a way that these diffusely activated sound features directly communicate with high-level information such as discourse or information structure that is incrementally constructed.

Third, the current study addressed the relationship between intonation form and meaning only from the listener’s perspective. For a complete discussion, however, we need to look at how production systems incorporate the details of intonational forms into their calculation of meaning. At what point of production do speakers plan their prosodic forms? Bock & Levelt’s (1994) production model, for example, treats pronunciation and articulation at the phonological encoding level, which is at the bottom of the production conveyer belt. However, the accumulating evidence suggests that the production mechanism must allow more direct connections between message level and phonological encoding level, because the prosodic form that speakers choose reflects what message they want to convey (e.g., the acoustic prominence is influenced by the importance and predictability of the message that speakers encode; Watson, Arnold, & Tanenhaus, 2008). With respect to the particular intonation form (L+H* L-H%) and meaning (state-contrast implicature) that was explored in this dissertation, a complementary study would provide speakers with the meaning and have them produce sentences for the meaning. Analysis of the prosodic forms produced by the speakers will illuminate our understanding of the target tune’s capacity for the proposed meaning, as well as any other prosodic forms that suit the meaning, thereby enriching our discussion on the form-meaning interface.

Fourth, the current study limited its initial investigation to only native English speakers. However, similar questions can be asked for second language learners of English. Both anecdotal experience and research reports in second language learning suggest that learning pragmatic conventions in a second language is difficult. In particular, when the means to express those conventions is a nonlexical cue like prosody, it is conceivable that the difficulty is even greater. It needs to be seen whether
development in language proficiency automatically enables learners to construct the type of implied meaning investigated in this study. That is, do advanced English learners perceive state-contrast from sentences like *The shoe was $L^+H^*$ tied $L-H^*$? If they can utilize a compositional processing mechanism that is similar to that of native speakers, it is possible that they can do it. That is, if learners know the function of $L+H^*$ and $L-H^*$, they might be able to infer the implied message by combining that information with their world knowledge about the object. If the learners have yet to learn the independent functions of intonational morphemes, they may not perceive the targeted meaning. A range of questions await answers.

Finally, similar developmental questions can be asked about children who learn English as their mother tongue. It has been documented that children are late in their development of certain constructions not because they lack syntactic knowledge but because they lack pragmatic knowledge (e.g., children’s difficulties in learning focus constructions related to the particle *only* due to their inability to construct the alternative set: Paterson et al., 2003). In terms of the construction and meaning investigated in this dissertation, one needs to have knowledge on (a) lexical items, as well as the tonal units and functions, (b) world knowledge on different states of affairs that an entity can have, and (c) a pragmatic distinction between focus and contrast. Furthermore, one must be able to integrate these different kinds of knowledge in a coherent way to infer the contrast relationship between the mentioned target and unmentioned alternatives. This is certainly a cognitively challenging task, and it needs to be seen at what point in development children can successfully accommodate all these needs to understand the speaker’s implicit message. Investigating children at different developmental stages is expected to illuminate some of the key understandings on which of the multiple factors carries more weight than others. Obtaining knowledge in this area will advance our knowledge in the field of language acquisition, as well as other related fields such as psychology and cognitive science.
APPENDIX A

Critical sentences used in Experiments 1A, 1B, and 2

<table>
<thead>
<tr>
<th>Item</th>
<th>Sentence</th>
<th>Exp.1A</th>
<th>Exp.1B</th>
<th>Exp.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The baby was dressed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>His beard was long.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>The bed was made.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>The belt was buckled.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>The cage was locked.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>The candle was lit.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>The car was new.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>The cat was fat.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>The coat was buttoned.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>The curtain was closed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>The dog was groomed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>The door was ajar.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>The drawer was shut.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>The dress was loose.</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>The gift was wrapped.</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>The jar was full.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>The leg was hairy.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>The mailbox was full.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>The necklace was latched.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>The pants were folded.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>The pencil was sharp.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>The plate was clean.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>The road was curvy.</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>The rope was coiled.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>The rug was rolled up.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>The shirt was wrinkly.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>The shoe was tied.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>The stocking was hung.</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>The suitcase was open.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>The tire was flat.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>The tree was leafy.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>The window was broken.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>The zipper was done.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total number of critical items used 33 32 30

Note: The X marked items were not included in the respective experiment due to the reasons explained in the relevant chapter.
# APPENDIX B

Critical pictures used in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Mentioned state</th>
<th>Opposite state</th>
<th>Mentioned state</th>
<th>Opposite state</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td>7</td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
</tr>
<tr>
<td>8</td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
</tr>
<tr>
<td>9</td>
<td><img src="image33.png" alt="Image" /></td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
<td><img src="image36.png" alt="Image" /></td>
</tr>
<tr>
<td>10</td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
<td><img src="image40.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td>Mentioned state</td>
<td>Opposite state</td>
<td>Mentioned state</td>
<td>Opposite state</td>
</tr>
<tr>
<td>---</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>11</td>
<td><img src="image1" alt="Dog" /></td>
<td><img src="image2" alt="Dog" /></td>
<td><img src="image3" alt="Door" /></td>
<td><img src="image4" alt="Door" /></td>
</tr>
<tr>
<td>13</td>
<td><img src="image5" alt="Drawer" /></td>
<td><img src="image6" alt="Drawer" /></td>
<td><img src="image7" alt="Jar" /></td>
<td><img src="image8" alt="Jar" /></td>
</tr>
<tr>
<td>15</td>
<td><img src="image9" alt="Leg" /></td>
<td><img src="image10" alt="Leg" /></td>
<td><img src="image11" alt="Mailbox" /></td>
<td><img src="image12" alt="Mailbox" /></td>
</tr>
<tr>
<td>17</td>
<td><img src="image13" alt="Necklace" /></td>
<td><img src="image14" alt="Necklace" /></td>
<td><img src="image15" alt="Shorts" /></td>
<td><img src="image16" alt="Pants" /></td>
</tr>
<tr>
<td>19</td>
<td><img src="image17" alt="Pencil" /></td>
<td><img src="image18" alt="Pencil" /></td>
<td><img src="image19" alt="Plate" /></td>
<td><img src="image20" alt="Plate" /></td>
</tr>
<tr>
<td>Mentioned state</td>
<td>Opposite state</td>
<td>Mentioned state</td>
<td>Opposite state</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td><img src="image1" alt="Image" /></td>
<td>22</td>
<td><img src="image2" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td><img src="image3" alt="Image" /></td>
<td>24</td>
<td><img src="image4" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td><img src="image5" alt="Image" /></td>
<td>26</td>
<td><img src="image6" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td><img src="image7" alt="Image" /></td>
<td>28</td>
<td><img src="image8" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td><img src="image9" alt="Image" /></td>
<td>30</td>
<td><img src="image10" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>
## APPENDIX C

Critical items and three cohorts used in Experiment 3

### Affirmative set

<table>
<thead>
<tr>
<th>Item</th>
<th>Sentence</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lisa had the bell.</td>
<td>belt</td>
<td>bed</td>
<td>bench</td>
</tr>
<tr>
<td>2</td>
<td>Lisa had the bone.</td>
<td>bowl</td>
<td>bow tie</td>
<td>boat</td>
</tr>
<tr>
<td>3</td>
<td>Lisa had the bus.</td>
<td>button</td>
<td>bucket</td>
<td>butterfly</td>
</tr>
<tr>
<td>4</td>
<td>Lisa had the cane.</td>
<td>cake</td>
<td>cage</td>
<td>cape</td>
</tr>
<tr>
<td>5</td>
<td>Lisa had the cap.</td>
<td>camel</td>
<td>cactus</td>
<td>castle</td>
</tr>
<tr>
<td>6</td>
<td>Lisa had the chick.</td>
<td>chips</td>
<td>chisel</td>
<td>chimpanzee</td>
</tr>
<tr>
<td>7</td>
<td>Lisa had the dice.</td>
<td>diaper</td>
<td>dinosaur</td>
<td>diamond</td>
</tr>
<tr>
<td>8</td>
<td>Lisa had the ham.</td>
<td>hamburger</td>
<td>hammer</td>
<td>hamster</td>
</tr>
<tr>
<td>9</td>
<td>Lisa had the lime.</td>
<td>light bulb</td>
<td>lion</td>
<td>lighter</td>
</tr>
<tr>
<td>10</td>
<td>Lisa had the mask.</td>
<td>match</td>
<td>map</td>
<td>mattress</td>
</tr>
<tr>
<td>11</td>
<td>Lisa had the pen.</td>
<td>pencil</td>
<td>penny</td>
<td>penguin</td>
</tr>
<tr>
<td>12</td>
<td>Lisa had the pig.</td>
<td>pillow</td>
<td>pitchfork</td>
<td>pickle</td>
</tr>
<tr>
<td>13</td>
<td>Lisa had the rose.</td>
<td>rope</td>
<td>roller skates</td>
<td>rolling pin</td>
</tr>
<tr>
<td>14</td>
<td>Lisa had the tent.</td>
<td>tennis racket</td>
<td>television</td>
<td>telescope</td>
</tr>
<tr>
<td>15</td>
<td>Lisa had the trap.</td>
<td>tractor</td>
<td>trashcan</td>
<td>traffic light</td>
</tr>
<tr>
<td>16</td>
<td>Lisa had the watch.</td>
<td>watermelon</td>
<td>washing machine</td>
<td>waffle</td>
</tr>
</tbody>
</table>

### Negative set

<table>
<thead>
<tr>
<th>Item</th>
<th>Sentence</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bart didn't have the bee.</td>
<td>beard</td>
<td>beaver</td>
<td>beet</td>
</tr>
<tr>
<td>2</td>
<td>Bart didn't have the box.</td>
<td>bomb</td>
<td>bonnet</td>
<td>bottle</td>
</tr>
<tr>
<td>3</td>
<td>Bart didn't have the can.</td>
<td>candle</td>
<td>candy</td>
<td>canteen</td>
</tr>
<tr>
<td>4</td>
<td>Bart didn't have the car.</td>
<td>carpet</td>
<td>cards</td>
<td>cart</td>
</tr>
<tr>
<td>5</td>
<td>Bart didn't have the cat.</td>
<td>camera</td>
<td>can opener</td>
<td>cannon</td>
</tr>
<tr>
<td>6</td>
<td>Bart didn't have the comb.</td>
<td>cone</td>
<td>coat</td>
<td>coconut</td>
</tr>
<tr>
<td>7</td>
<td>Bart didn't have the duck.</td>
<td>dustpan</td>
<td>dumplings</td>
<td>dumpster</td>
</tr>
<tr>
<td>8</td>
<td>Bart didn't have the lamb.</td>
<td>lamp</td>
<td>laptop</td>
<td>ladder</td>
</tr>
<tr>
<td>9</td>
<td>Bart didn't have the lock.</td>
<td>lobster</td>
<td>lollipop</td>
<td>lava lamp</td>
</tr>
<tr>
<td>10</td>
<td>Bart didn't have the pants.</td>
<td>pancake</td>
<td>panda</td>
<td>parrot</td>
</tr>
<tr>
<td>11</td>
<td>Bart didn't have the peach.</td>
<td>peanut</td>
<td>pizza</td>
<td>peacock</td>
</tr>
<tr>
<td>12</td>
<td>Bart didn't have the rake.</td>
<td>radio</td>
<td>rain jacket</td>
<td>razor</td>
</tr>
<tr>
<td>13</td>
<td>Bart didn't have the seal.</td>
<td>seahorse</td>
<td>seesaw</td>
<td>cereal</td>
</tr>
<tr>
<td>14</td>
<td>Bart didn't have the tie.</td>
<td>tire</td>
<td>tiger</td>
<td>typewriter</td>
</tr>
<tr>
<td>15</td>
<td>Bart didn't have the truck.</td>
<td>trunk</td>
<td>trumpet</td>
<td>trombone</td>
</tr>
<tr>
<td>16</td>
<td>Bart didn't have the whisk.</td>
<td>whistle</td>
<td>whip</td>
<td>wig</td>
</tr>
</tbody>
</table>
References


