ENGLISH PROSODIC MARKING OF INFORMATION STRUCTURE BY L1-JAPANESE SECOND LANGUAGE LEARNERS

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

SECOND LANGUAGE STUDIES

DECEMBER 2018

By

Aya Takeda

Dissertation Committee:

Bonnie D. Schwartz, Chairperson
Theres Grüter
William O’Grady
Amy J. Schafer
Shinichiro Fukuda

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to all of the people who have accompanied me throughout the journey of my graduate studies. First and foremost, I would like to show my highest gratitude and appreciation to my academic advisor and dissertation chair, Prof. Bonnie D. Schwartz, for her unlimited guidance, patience, and encouragement in each and every step of the way. She has been my inspiration for entering this field and continuing to learn and grow as a future scholar. Through challenging questions and stimulating discussions, she taught me how to think critically and develop a strong intellectual foundation for language research. I was always amazed by her expertise, detailed eyes, insightful comments, and passion toward language acquisition research. Without her inspiration, I would not have become what I am now.

I am also thankful to Prof. Amy J. Schafer for teaching me advanced psycholinguistics and phonology. She provided me with numerous valuable comments on the experimental design, data analysis and statistics, and the dissertation drafts, which really was a tremendous help for me. When I was struggling with data analysis and interpretation, she sat with me every week and discussed the statistical procedure that would be most appropriate for my research. I thank her for the many hours that she spent in helping to develop my dissertation despite her tight schedule. I have always admired her because she is always on top of newly published literature and conference presentations and very familiar with the current state of knowledge, ideas, and methods that are being used at the frontier of psycholinguistic research.

My sincere thanks also go to Prof. Theres Grüter for not only helping with my eye-tracking experiments but for giving me so many wonderful opportunities to learn as a graduate research assistant while I was on her research project team. Under her supervision, I was able to get hands-on experience in the whole research procedure from developing research materials to writing manuscripts for publication, which was of great benefit to this dissertation and my future career. With her professionalism, efficiency, and kindness, Prof. Grüter is one of my best examples.

I would also like to express my gratitude to the rest of my committee members, Prof. Shinichiro Fukuda and Prof. William O'Grady for their encouragement, insightful comments, and hard questions during individual meetings, project presentations, and the dissertation defense.

I would also like to thank Prof. Victoria Anderson, who unfortunately had to step down from the committee due to her sabbatical leave, for teaching me the basics of intonational phonology and helping to design, create, and analyze sound stimuli for my experiments. She was always keen to know what I was doing and how I was proceeding, and when things did not go as expected, her enthusiasm and encouragement raised my spirit and kept me going.

Besides my current and former committee members, I owe many thanks to my research assistants who supported me with various aspects of my dissertation. I would like to thank Meagan Dailey, James Grama, Robert Otsubo, Robin Parrish, and Bonnie Windham for their help with experiment construction and data analysis. For their constructive feedback on my research and presentations, special thanks also go to the members of the Language Acquisition Reading Group at the University of Hawaiʻi as well as the audiences at the Boston University Conference on Language Development; the Chuo-UHM Student Conference on Linguistics, Psycholinguistics, and Second Language Acquisition; the CUNY Conference on Human Sentence Processing; the Generative Approaches to Second Language Acquisition Conference; and ProPro 2017: Processing prosody across languages, varieties, and nativeness.

I feel very fortunate to be surrounded by so many fun, supportive, and warm-hearted colleagues: Hyunah Ahn, Amber Camp, Wei Chu, Keiko Hata, Haerim Hwang, Chae-Eun Kim, Hyunwoo Kim, Kiteak Kim, Elaine Lau, On-Soon Lee, Wenyi Ling, Mari Miyao, Kyae-Sung Park, Nozomi Tanaka, Zhijun Wen, Frederick Zenker, and Jing Crystal Zong. The time I spent with them on campus is one of the most precious memories during my years as a graduate student in Hawaiʻi. Big mahalo also goes to Ivan Bondoc, who kindly agreed to take care of administrative paperwork for my dissertation after I left the island.

I am deeply indebted to the staff in the Language Analysis and Experimentation Labs. Whenever I had questions about computers and research devices, I would run to them for advice. I especially enjoyed discussing technologies with the former IT specialist for the labs, Kurt Brunner, who unfortunately passed away in 2013. He was not only a knowledgeable computer support technician but also (to my surprise) a passionate hula dancer. I will forever remember his brightest smile while performing hula at the international competition that he had invited us graduate students to.

Several people helped me with participant recruitment: the ELI director, Kenton Harsh, the HELP director, Joel Weaver, the HELP coordinator, Christine Guro, and ELI, HELP, and SLS instructors. And of course, this dissertation would not have been possible without the help of all of my participants.

This dissertation research was supported by a National Science Foundation Doctoral Dissertation Research Improvement Grant (BCS-1324511), for which I am very grateful. I also appreciate support from the University of Hawai‘i via the Elizabeth Carr-Holmes Scholarship Fund (Department of Second Language Studies), a graduate student scholarship from the Center for Japanese Studies, a scholarship from the East-West Center Graduate Student Affiliate Program, and a Summer Institute Fellowship from the Linguistic Society of America. They provided me with the financial means to complete my graduate studies and this dissertation.

Last but not least, I would like to thank my parents and dear husband and sons, who endured this long process with me, for their constant patience, love, and encouragement. Their support was in the end what made this dissertation possible.

ABSTRACT

This dissertation examines second language (L2) use of prosody to convey and comprehend Information Structure. In English, contrastive information and new information are typically higher in pitch and longer in duration, whereas given information is often deaccented. Contrastive information and new information are often differentiated by distinct types of pitch accents: the former marked with L+H*, the latter with H*. Tokyo Japanese, by contrast, uses pitch accent primarily to express lexical meaning rather than information status, but contrastive information can be indicated by expanded pitch range, similar to English L+H*, without a change in either duration or lexically-determined pitch accent type.

In light of such differences between English and Japanese, this study investigates how L2 learners develop target-language prosodic marking of Information Structure, attending in particular to whether this is initially modulated by native language (L1) properties. L1-Japanese L2 learners of English were tested on three prosodic patterns–deaccentuation, a regular high pitch accent (H*), and a contrastive pitch accent (L+H*)–and their link to Information Structure via three tasks: a prosody-in-context naturalness rating task, an eye-tracking listening comprehension task, and a production task.

It was found that native English speakers associate given information with deaccentuation, contrastive information with L+H*, and new information with accentuation (both H* and L+H*) in all three tasks. Advanced L2 learners could map given information with deaccentuation and contrastive information with L+H* in the rating and production tasks, but only the mapping between given information and deaccentuation was demonstrated in the eye-tracking listening

task. Less proficient L2 learners were able to associate given information with deaccentuation and contrastive information with L+H* in the rating task, but no evidence for prosody-discourse association emerged in the production or eye-tracking listening tasks.

The outcomes of the study suggest that L2 learners do not necessarily transfer their L1 prosody-discourse mappings (for all tasks), and that they can master target-language prosodic marking of Information Structure that is not instantiated in the L1. Furthermore, how successfully they map prosody to information status depends on several factors: learners' language proficiency, processing difficulty of particular mappings, task difficulty, and whether learners' attention is explicitly drawn to prosody.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

In first language (L1) research, it has been repeatedly shown that prosodic information, viz., the rhythm and melody of speech, plays an important role in language acquisition. Infants are sensitive to and prefer the rhythmic and tonal patterns of their native language from their earliest days of life (Christophe, Mehler, & Sebastián-Gallés, 2001), and such prosodic cues guide young infants to segment a stream of speech sounds into identifiable words (Johnson & Jusczyk, 2001; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Corpus studies in the child-directed speech literature have suggested that while segments in strings consisting of background information are underarticulated, segments in focus are more clearly articulated, and that the latter might aid children to segment a stream of speech on the basis of acoustic properties of focus words (Ratner, Rooney, & MacWhinney, 1996).

A number of studies in L1 processing also document the importance of prosody in adults' sentence comprehension and production (see Cole, 2015; Cutler, Dahan, & Van Donselaar, 1997; Dahan, 2015; Speer & Blodgett, 2006; Wagner & Watson, 2010; Warren, 1996, for review). Prosodic phrasing, i.e., grouping of words within an utterance according to their intonational patterns, often corresponds to syntactic boundaries (Kjelgaard & Speer, 1999; Lehiste, 1973; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Shattuck-Hufnagel & Turk, 1996; Speer, Warren, & Schafer, 2011), allowing listeners to efficiently organize and appropriately interpret incoming segments in relation to the previously processed sentence structure. Prosodic prominence, or saliency, has also been recognized as an indicator of sentential focus in discourse (Chafe, 1974; Pierrehumbert & Hirschberg, 1990; Venditti & Hirschberg, 2003).

While second language (L2) learners' acquisition of consonants and vowels has long been a subject of interest to many scholars and language-teaching practitioners (Best, 1995; Flege, 1995; Kuhl & Iverson, 1995), it is only rather recently that theoretical and applied research on the acquisition and use of L2 prosody has made rapid advances. Available L2 studies on prosodic features have revealed difficulties in the perception and production of word stress (Archibald, 1997), lexical tone (Eliasson, 1997; Juffs, 1990; So & Best, 2010, 2014), and pitch accent (Henrichsen, 1984), including problems in using pitch accent to express given vs. new information in discourse (e.g., Gut, Pillai, & Don, 2013; Gut & Pillai, 2014; Verdugo, 2003, 2006; Wennerstrom, 1994, 1998; see Mennen, 2015; Mennen & de Leeuw, 2014, for review).

The goal of this dissertation is to examine one particular aspect of prosody, intonational marking of discourse structure by L2 learners of English whose L1 is Japanese. The linguistic means for marking information status of discourse referents vary from language to language. In English, words conveying new or contrastive information generally bear a pitch accent, thus being perceptually more salient than given information (Arnold, 2008; Birch & Clifton, 1995; Bock & Mazzella, 1983; Dahan, Tanenhaus, & Chambers, 2002; Isaacs & Watson, 2010; Nooteboom & Kruyt, 1987; Terken & Nooteboom, 1987). Furthermore, new information and contrastive information are often marked with distinct pitch accent patterns–H* (new) vs. L+H* (contrastive) (Pierrehumbert & Hirschberg, 1990), although the relationship appears probabilistic rather than a strict one-to-one mapping (Calhoun, 2010b; Im, Cole, & Baumann, 2018). In Tokyo Japanese, on the other hand, information status is marked primarily by morphological and syntactic means (Kuno, 1973; Shibatani, 1990), and only secondarily by prosodic cues (Beckman & Pierrehumbert, 1986; Nakanishi, 2001; Pierrehumbert & Beckman, 1988; Venditti & Hirschberg, 2003; Venditti, Maekawa, & Beckman, 2008; Venditti & Swerts, 1996). Given vs. new information is generally indicated by the thematic marker *-wa* and syntactic scrambling rather than by the presence or absence of pitch accent, which mainly signals lexical meaning in Japanese; however, contrastive information can be marked prosodically with local pitch range expansion (Venditti et al., 2008) similar to English L+H*. In light of the differences between English and Japanese, the present study investigates how Japanese L2 learners of English, who of course have knowledge of Information Structure in their L1, develop the ability to associate information status and its prosodic marking in the target language (TL)–the mappings of given, new, and contrastive information to deaccentuation, H*, and L+H*. Since Japanese speakers are sensitive to contrast-marking pitch expansion in their L1 (K. Ito, Jincho, Minai, Yamane, & Mazuka, 2012), the contrastive–L+H* link is predicted to be easy due to transfer from the L1; by contrast, the given–deaccentuation and new–H* links should be harder to map, as accentuation in Japanese is a lexical property rather than a discourse-status marker.

How (dis)similarities in L1 and TL influence L2 acquisition has been one of the recurring questions in the field of second language research. Our study extends it to the domain of prosody-discourse interface as very little is known about L2 learners' use of prosodic information. To this end, the study, utilizing naturalness rating, comprehension, and production tasks, asks the following questions:

1. Can L2 learners associate a particular TL pitch accent pattern with an appropriate discourse meaning when focusing on intonation?

2. Can L2 learners make immediate use of TL pitch accent cues to identify discourse referents during real-time comprehension?

3. Can L2 learners employ TL pitch accent cues to express information status of discourse referents in naturalistic production?

The research outcomes should contribute to the theoretical and empirical development of this relatively new field of study in several ways. First, since all tasks are simple and do not require advanced lexical or syntactic knowledge, the study can be conducted across L2 proficiency, which provides us with a detailed picture of early L1 prosodic transfer as well as learners' development in the establishment of the L2 relationship between Information Structure and prosodic marking as proficiency increases. Second, analyses reveal not only whether L2 learners can become native-like in the prosodic marking of Information Structure but also whether Interlanguages possess systematic (albeit non-target) prosodic organization principles for expressing information status. Lastly, by comparing performance on naturalness rating, comprehension, and production by the same L2 learners, the study allows us to examine the relationship between comprehension and production.

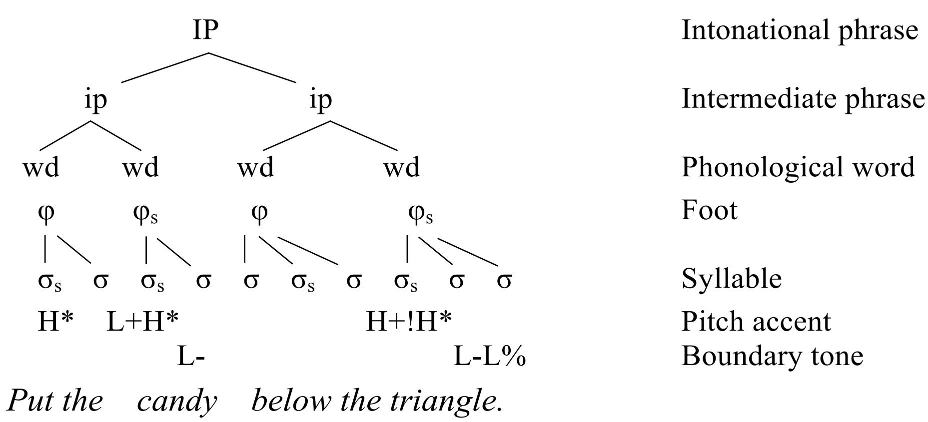
This dissertation is organized as follows: Chapter 2 summarizes the basic intonational systems and prosodic marking of Information Structure in English and in Japanese. Chapter 3 reviews the literature on L1 processing of prosodic marking of given, new, and contrastive information status in Germanic languages such as English and Dutch, and details theoretical proposals and empirical work on the L2 acquisition and processing of intonational marking of information status. Chapter 4 addresses the research questions for the present study and makes predictions about the acquisition of English prosodic marking by L1-Japanese speakers. The chapter also summarizes overall research designs of the three main experiments conducted in this study (i.e., a prosody-in-context naturalness rating task, an eye-tracking listening comprehension task, a semi-spontaneous dyadic production task), and provides participants' background information as well as language proficiency. Chapter 5 covers the prosody-in-context naturalness rating task, the purpose of which was to assess listeners' sensitivity to the link between intonational patterns and information status of discourse referents. Details of the experimental methods as well as the results in terms of L2 proficiency are reported. Chapter 6 is on a visual world eye-tracking listening experiment. In this experiment, participants heard instructions and selected the specified objects while their eye movements were recorded. By tracking listeners' eye movements as they hear sentences, we monitor how quickly they process intonational cues to identify discourse referents. Chapter 7 reports on an interactive production experiment, which is a speaking version of the eye-tracking listening task. Each participant gave instructions to a confederate, directing her to select particular objects on the computer screen. This task examines how speakers' intonation and word durations change according to the information status of particular objects while giving instructions. Chapter 8 summarizes the results of all three experiments in relation to the research questions posed in Chapter 4, discusses and evaluates possible explanations of the results, and compares the current study to the previous research on the L2 acquisition of discourse prosody.

The results of our study shed important light on the issue of L1 transfer in L2 prosody, achievement of native-like performance by L2 learners, role of L2 proficiency, and possible sources of L2 processing difficulty. We also provide some pedagogical implications and instructional strategies for second/foreign language speaking/listening courses in the hope that the study will ultimately benefit not only researchers but also language teachers as well as students.

CHAPTER 2
PHONOLOGICAL BACKGROUND

2.1 Intonation and focus marking in English

This chapter describes the intonation systems of English and Japanese, following the Autosegmental-Metrical (AM) model of intonational phonology (Beckman & Hirschberg, 1994; Ladd, 1996; Pierrehumbert, 1980), and reviews previous findings on intonational marking of information status by native speakers of each language. The AM model is a phonological theory of intonational structure that has been highly influential in both L1 and L2 acquisition and processing research, and is characterized by the hierarchical organization of prosodic prominence and phrasing. Sound patterns are analyzed into different types of elements at several different levels as shown in Figure 2.1. One of the central claims of the AM theory is that the most fundamental segmental features are separated from suprasegmental levels. Intonation is represented phonologically as a string of tones, L(ow) and H(igh), in its own level independent of vowels and consonants (thus, “autosegmental”), and these tones are associated with prominent syllables and phrasal boundaries at the phrase levels such as the phonological word, the intermediate intonational phrase (ip), or the intonational phrase (IP) (thus, “metrical”).



Note. The subscript “s” indicates a strong foot/syllable.

(Adapted from Cole, 2015, p. 3, Figure 1)

Figure 2.1: Prosodic structure of an English sentence

According to the phonological theory developed by Beckman and Pierrehumbert (1986), pitch accents in English are associated with prominent stressed syllables as indicated with an asterisk (*) in Figure 2.1. As mentioned above, intonation is represented in a hierarchical structure: In English, each utterance contains at least one full intonational phrase, which is further divided into intermediate intonational phrases. An IP is a prosodic constituent that contains a perceptually single/coherent intonational contour (Shattuck-Hufnagel & Turk, 1996), and it ends with a strong degree of disjuncture such as a pause. The final disjuncture at the IP level is marked with either a high (H%) or a low (L%) pitch boundary tone. Beckman and Pierrehumbert define an ip as the domain of a single, coherent intonational contour that is smaller than an IP and ends with a level of disjuncture that is less strong than the IP-level boundary. In English, "downstep," a phenomenon that describes gradual lowering of pitch, occurs over the domain of an ip, and therefore pitch is reset to high at the beginning of every ip.[1] An ip must contain at least one pitch accent (e.g., H*, L*, L+H*, L*+H, H+!H*) on a syllable bearing primary stress, and it must end with a phrase accent (e.g., L-, H-).[2] An example of the Tones and Break Indices for Mainstream American English transcription (MAE_ToBI) (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005), which is one of the most commonly adopted prosodic labeling systems for English intonation, is shown in (1). The tonal grammar of English is summarized in Figure 2.2.

(1) a. [[Put the candy]$_{ip}$]$_{IP}$[[below the diamond]$_{ip}$]$_{IP}$.
H* L+H* L-L% H* H-H%

b. [[Now put the bell]$_{ip}$]$_{IP}$[[above the triangle]$_{ip}$]$_{IP}$.
H* L+H* L-H% H+!H* L-L%

(Dahan, Tanenhaus, & Chambers, 2002, pp. 297-298, (1) & (3))

[1] In case of parenthetical expressions (e.g., sentence adverbials such as *however* and *unfortunately*, adverbial clauses, comment clauses such as *I think*, non-restrictive relative clauses, and question tags), which are not part of the main assertion, pitch reset may not occur. Whether or not parentheticals are phrased in an independent intonation domain separate from the host IP is a topic of continuing debate (Dehé, 2009).

[2] H*: high pitch accent; L*: low pitch accent; L+H*: bitonal pitch accent with low tone followed by high tone prominence; L*+H: bitonal pitch accent with low tone prominence followed by high tone; H+!H*: bitonal pitch accent with high tone followed by downstepped (slightly lower) high prominence; L-: low pitch phrase accent; H-: high pitch phrase accent

IP initial boundary tone		Pitch accent		ip final phrase tone		IP final boundary tone
%H %L	+	H* L* L*+H L+H* H+!H*	+	H- L-	+	H% L%

Figure 2.2: Tonal grammar of Mainstream American English

English has multiple types of pitch accents and boundary tones, which are assigned by the speaker at a post-lexical level in order to indicate syntactic, discourse, and pragmatic meanings. Among many discourse functions of English intonation, one is to cue the information status of discourse entities (Pierrehumbert & Hirschberg, 1990). By accenting or deaccenting referring expressions or by varying pitch accent types on accented entities, speakers modulate intonational prominence to express, for example, new vs. given status of referents in discourse (Brown, 1983; Büring, 2006, 2016; Chafe, 1987; Ladd, 2008; Rooth, 1992; Selkirk, 1995).

In English, accentuation is encoded with elevated pitch, longer duration, increased intensity (i.e., loudness), and hyper-articulation of vowels (Breen, Fedorenko, Wagner, & Gibson, 2010; Cho, 2005; Cooper, Eady, & Mueller, 1985; De Jong, 1995; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Katz & Selkirk, 2011; Xu & Xu, 2005), all of which enhance the acoustic clarity of segments and facilitate lexical access. As a result, accented words are processed and recognized faster than unaccented words (e.g., Cutler & Foss, 1977). Hirschberg and Pierrehumbert (1986) suggest that the speaker's decision to accent or deaccent expressions in discourse reflects the intentional and attentional structuring of discourse: Speakers typically deaccent a given entity since its referent is already salient and accessible in the discourse, requiring little phonetic information for the listener to select the intended discourse entity among the restricted number of referents that have been activated. At the same time, speakers accent a new entity in order to introduce the new referent into the hearer's (shared) attentional space (Nakatani, 1997; Terken, 1984). Accentuation speeds up the processing of phonetic information, which in turn promotes faster and more accurate recognition of the word referring to a new discourse referent for the listener.

The mechanism behind this sort of acoustic variation by the speaker has been attributed to a listener-oriented language production process, or so called "audience design" (Clark, 1996;

Clark & Murphy, 1982). It holds that speakers select linguistic forms in each utterance based on their estimates about the knowledge and needs of the listener, and consequently the listener makes use of such cues in comprehension. Contrasting to audience design is the speaker-oriented view of production processes. Repeated words are easy to produce because their semantic and phonological representations have been activated recently in the production system, and words related to the activated discourse entities can also be accessed more quickly, as demonstrated in a number of lexical priming studies (e.g., Swinney et al., 1979). Although whether acoustic prominence in speech is driven by audience design, speaker-internal design, both, or some other processing-oriented mechanisms is beyond the scope of the present study, previous studies suggest that both the needs of the listener and the efficiency for the speaker play roles in the speaker's language production processes (Arnold, Kahn, & Pancani, 2012; Kahn & Arnold, 2015).

In addition to Information Structure, there are other factors affecting accent distribution such as grammatical functions and the surface structure of referring expressions (Terken & Hirschberg, 1994), lexical-semantic weight, or rhythmic alternation in the English metrical system (Calhoun, 2010a). For instance, Terken and Hirschberg's production study (1994) revealed that given information is more likely to be deaccented if it bears the same grammatical role (e.g., subject vs. direct object) and appears in the same surface syntactic position as when it was previously mentioned in the discourse. In addition, words with more semantic weight tend to receive accent as in the example, *I ate an APPLE* vs. *I ATE something*, and accents are associated with metrically strong syllables as can be seen at the syllable level in Figure 2.1 (Calhoun, 2010a). Taking these findings together, deaccentuation vs. accentuation does not perfectly correlate with given vs. new status of Information Structure, but rather, the relationship is probabilistic.

Pierrehumbert and Hirschberg (1990) further portray interpretations of pitch accents in terms of discourse, advocating a compositional analysis of tonal meaning. They hypothesize that each tonal element (a particular pitch accent, phrase accent, or boundary tone) carries specific meanings, and the meaning of an overall intonational sequence is the result of the combination of the meanings conveyed by individual tones. For instance, H* signals that the accented entity is new in the discourse and thus needs to be added to the discourse model. L- indicates that the current phrase is separate from the upcoming phrase, and the L% boundary tone implies that the current phrase/utterance can be interpreted without paying particular attention to the subsequent utterances. When these tonal elements are combined in an utterance such as (2), it forms a neutral declarative intonation pattern for introducing new information into the discourse.

(2) The train leaves at seven.
H* H* H* L-L%

(Pierrehumbert & Hirschberg, 1990, p. 286, (5))

The L+H* accent is claimed to evoke a salience scale among possible referents and to convey that the accented item, not some other candidates, should be activated in the listener's discourse model. The most common example of such a function is to mark a correction or contrast. In example (3), H* marks *Fred* as new information added by Speaker B to the discourse, and the L+H* accent marks the word *beans*, indicating a contrastive meaning such as "As for the beans, Fred ate them. As for the other food, other people may have eaten it."

(3) A: What about the beans? Who ate them?
B: Fred ate the beans.
H* L- L+H* L-H%

(Adapted from Jackendoff, 1972, p. 260, (14))

Although whether there is a categorical distinction between H* and L+H* is still under debate (Bartels & Kingston, 1994; Calhoun, 2007, 2010b; Dilley, 2010; Ladd & Schepman, 2003), evidence has shown that native English speakers reliably mark new vs. contrastive information using prosody (Breen et al., 2010; Cooper et al., 1985; Katz & Selkirk, 2011) and exploit such cues to aid their comprehension of Information Structure (Breen et al., 2010; Dahan et al., 2002; Isaacs & Watson, 2010; K. Ito & Speer, 2008; Watson, 2010).

In sum, it has been commonly assumed that in English, speakers typically place a pitch accent, or more specifically H* according to Pierrehumbert and Hirschberg (1990), on referents that are new to the discourse and deaccent referents that are given. Previous research has also shown that listeners make use of such prosodic cues to help process given vs. new discourse information. Furthermore, L+H* is said to be a contrastive pitch accent that evokes a set of alternatives at the point where the pitch accent is encountered. As these long-standing claims have received empirical support in past studies (e.g., Arnold, 2008; Dahan et al., 2002; Isaacs

& Watson, 2010; K. Ito & Speer, 2008; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Watson, Tanenhaus, & Gunlogson, 2008; Weber, Braun, & Crocker, 2006), we expect that native speakers of English in our study will take advantage of the close mapping between prosody and information status, although we may also see some variability in the use of prosodic information since more recent studies have put forward evidence that there is no exact one-to-one correspondence between prosodic categories and discourse functions (Breen, Kurumada, Wagner, Watson, & Yu, 2018; Calhoun, 2010b).

The next section reviews the intonational and focus marking systems in Japanese, and summarizes similarities and differences between English and Japanese with respect to the role of pitch accent in relation to Information Structure.

2.2 Intonation and focus marking in Japanese

Applying the AM model of intonational phonology, Beckman and Pierrehumbert (1986) compared intonational structures and prosodic features of English and Tokyo Japanese and proposed a pioneering theory of Japanese tone structure. Further phonological and experimental analyses of Japanese intonation contributed to the development of a more recent model of Japanese intonation, called the Japanese Tone and Break Indices (J_ToBI) (Venditti, 2005), which draws heavily on Beckman and Pierrehumbert's original model. Here we describe the Japanese intonation system based on the J_ToBI model, since the model has been most widely applied to phonological and psycholinguistic research on Japanese.

According to J_ToBI, the Japanese language has a hierarchical intonational structure like English and, in both languages, tonal elements are limited to simple tones (e.g., H-) or bitonal complex tones (e.g., H*+L). Instead of the ip level, however, the Japanese intonation system has the accentual phrase (AP) level just below the IP level. The IP is the highest level in the prosodic hierarchy as well as the basic unit of intonation, as in English. Every IP has to contain at least one AP and end with an IP boundary tone (H% for prominence-lending rise and insisting rise, LH% for incredulity and information question rises, and HL% for explanatory rise-fall boundary movement). APs are defined by their unique pitch movement pattern transcribed as [%L H- L%]: An AP starts with L and rises to H around the second mora, and gradually falls to L at the right edge of the phrase. Weak low boundary tones (%wL or wL%) are used when the following phrase begins with an initial accent or a heavy syllable. The tonal grammar of Japanese is summarized in Figure 2.3.[3]

AP boundary tone		AP tone		AP boundary tone		IP boundary tone
%L (%wL)	+	H- [unaccented] (H-) H*+L [accented]	+	L% (wL%)	+	H% LH% HL%

Figure 2.3: Tonal grammar of Tokyo Japanese

One of the features distinct from English pitch accents is that the presence or absence of accentuation is specified lexically in Japanese. Every Japanese word is either accented or unaccented, and all accented words are realized as H*+L (high tone followed by a sharp fall within the accented mora). Unaccented words do not have lexically specified pitch; therefore at the AP level, they are realized as the default AP pitch contour H- (high plateau without a sharp fall). Another important difference between English and Japanese pitch accent is that in English, phrase- or utterance-level focus-related accentuation is realized by various combinations of pitch movement, elevated pitch, increased loudness, prolonged duration, and full and clear articulation of the accented word/phrase, whereas in Japanese lexical accentuation, duration and clarity of articulation do not differ between accented and unaccented words (Beckman, 1986; Ladd, 1996, 2008; Venditti et al., 2008).

Since pitch accents are lexical properties in Japanese, the manner in which Japanese intonation marks discourse focus differs from that of English. Firstly and most importantly, the H*+L accent in Japanese can never be used to mark lexically unaccented words. Since adding or removing an accent could result in a change in lexical meaning, new vs. given status of referential entities is primarily marked morphologically or syntactically (Kuno, 1973; Shibatani, 1990). Morphologically, the thematic topic marker *-wa* is often associated with given

[3] Other models describe the Japanese intonation system differently. Selkirk's Match Theory (Selkirk, 2009; Selkirk et al., 2011) proposes that each constituent in phonological structure matches up with a corresponding constituent in syntactic structure: An intonational phrase, a phonological phrase, and a prosodic word correspond to a syntactic clause, a phrase, and a word, respectively. Other models hold that instead of the single category of phonological phrase, there should be two separate categories below the intonational phrase level and above the prosodic word level, for example, major vs. minor phrase or intermediate vs. accentual phrase (Kubozono, 1987; McCawley, 1968; Pierrehumbert & Beckman, 1988). Ito and Mester (2013) further claim that the two separate categories are indeed two layers of a single, recursively occurring phonological phrase.

information, and the subject marker *-ga* with new information (Kuno, 1972).[4] Syntactically, the cleft construction and scrambling are among the most commonly adopted focalization strategies in Japanese. However, contrastive information can be expressed prosodically, in addition to other morphological and syntactic means, such as the contrastive marker *-wa* and the cleft construction. When indicating contrast, pitch range expansion typically occurs with a steep rise from the initial AP boundary %L to the target tone H- or H*+L (K. Ito, 2002). In many cases, this focal pitch rise is also accompanied by insertion of an IP boundary at the left edge of the focused word and elimination of all the accentual phrase boundaries as well as pitch range compression in the immediately following phrase within the same IP (processes called "dephrasing" and "prosodic subordination," respectively) (Venditti et al., 2008). One of the analyses explaining the underlying mechanism of these focus effects posits that the insertion of an IP in the pre-focus position resets intonational downstep, causing pitch boost of the focused element, and post-focus dephrasing eliminates any pitch reset in the post-focus region, resulting in continuous downstep that contributes to prosodic subordination (Nagahara, 1994; Tomioka, 2016). This whole process makes the pitch boost on the focused material perceptually more salient to the listener.

The association between expanded pitch range and contrastiveness in Japanese is indeed attested in the processing study by Ito, Jincho, Minai, Yamane, and Mazuka (2012). In their visual world eye-tracking experiment, adults and pre-school children participated in a visual search task while their eye movement was monitored. They were presented with a visual display with various animals in different colors, heard a sequence of instructions such as *Pinku-no neko-wa doko? Jaa, midori-no neko-wa doko?* ("Where is the pink cat? Then, where is the green cat?"), and then selected the target animal on the screen. The results showed that for both adults and older children, fixation to the target contrastive referent was facilitated when the adjective preceding the target noun in the second instruction was produced with expanded pitch range compared with normal pitch range, supporting the hypothesis that L1 Japanese speakers utilize prosodic cues for the processing of contrastive information.

All in all, although both English and Japanese have a hierarchical organization of the intonational system, details of the organization, nature of accentuation, and prosodic

[4]The topic marker *-wa* is said to have two discourse functions: thematic and contrastive (Kuno, 1972). The thematic *-wa*, which means "speaking of...," or "talking about...," marks either a generic or anaphoric (i.e., previously mentioned) theme as opposed to a comment, while the contrastive *-wa* marks the contrasted element of the sentence. The contrastive *-wa* typically bears prosodic emphasis (Nakanishi, 2001).

manifestations of focus are considerably different between the two languages. The English intonational system consists of the intermediate phrase and the intonational phrase levels, whereas the Japanese system comprises the accentual phrase and the intonational phrase levels. In addition, English allows a wider variety of pitch accents and combinations of pitch accents and phrasal-boundary tones than Japanese, which has only one type of pitch accent and a few kinds of boundary tones. And most importantly, English and Japanese contrast in the functions of pitch accents: In English, pitch accents are inserted post-lexically to signal semantic and pragmatic meaning, while in Japanese, pitch accents are specified at the lexical level, and thus the presence or absence of pitch accent indicates lexical contrast. Below is a summary of the principal cross-linguistic differences in prosodic functions between English and Japanese.

- In English, both pitch height (maximum F0, minimum F0, mean F0) and duration serve as cues to distinguish accented vs. unaccented words. In Japanese, pitch height, but not duration, plays a major role in marking accented vs. unaccented words.
- In English, new vs. given information is typically marked by, respectively, accentuation (H*) vs. deaccentuation. In Japanese, on the other hand, presence vs. absence of pitch accents signals lexical contrast.
- English contrastive focus is often realized by L+H* (although whether L+H* is intonationally distinct from H* is still controversial), whereas Japanese contrastive focus can utilize an expanded pitch range. Phonetically, both are characterized as a sharp pitch rise followed by a fall as well as post-focus pitch compression.

Regarding the first point, as documented in a number of auditory speech studies, there is no simple relation between acoustic properties of the speech signal and perceived prosodic prominence. For instance, human speech perception (of any language) is to some degree shaped by the listener's native language: Because the phonetic cues for marking prosodic prominence on certain types of words vary from language to language, listeners learn both to disregard acoustic information irrelevant or less crucial for their native language and to optimize the detection of subtle signal changes that are meaningful in their native language (Carroll, 2006; Carroll & Shea, 2007). Therefore, native speakers of English are sensitive to both pitch and durational cues when processing sentences (Isaacs & Watson, 2010). Japanese speakers, on the other hand, may initially not be sensitive to the durational factor contributing to the perception of accentuation when listening to English. As for the second point, to what degree prosody correlates with information status varies across languages depending partially on the availability of other means, such as morphological and syntactic marking. While accentuation vs. deaccentuation signals different discourse status in English, such a distinction is a lexical property in Japanese. Due to this cross-linguistic difference, English speakers are likely to exploit prosodic prominence to help produce and comprehend discourse structure, whereas Japanese speakers may initially expect the distinction between accented vs. unaccented words in English to be merely a lexical difference, not yet able to relate prosodic prominence to new vs. given information status. With respect to the third point, English and Japanese are similar in that in both languages, a steep rise-fall followed by pitch compression signals contrast. Therefore, for this particular prosody-discourse mapping, L1-Japanese learners of English may be able to detect the pitch cue and quickly interpret it as contrastive in comprehension, and produce contrastive information with a native-like pitch pattern in production.

In light of these cross-linguistic differences, this dissertation aims to address the issue of whether and how L2 learners come to acquire TL prosodic marking of Information Structure. The next chapter reviews the literature on L1 and L2 prosodic marking of given, new, and contrastive information status and presents the theoretical framework in which the current study is embedded.

CHAPTER 3
PSYCHOLINGUISTIC BACKGROUND

3.1 Processing studies on L1 prosodic marking of Information Structure

Since the main issue that this dissertation addresses is whether L2 learners of English come to acquire native-like use of prosody in the processing of discourse information, it is essential to understand how reliably native speakers of English use prosody to comprehend and convey Information Structure. This section reviews key psycholinguistics studies that have investigated the role that prosody plays in the L1 processing of Information Structure in English, Dutch and German, which are said to share similar prosodic functions. We will first go over studies examining the relationship between accentuation and new vs. given information, and then look at studies dealing with H* vs. L+H* marking of new vs. contrastive information status.

3.1.1 Processing studies on L1 prosodic marking of new vs. given information

Many of the earlier studies examining the relationship between pitch accent and Information Structure utilize offline experimental paradigms such as a prosody-context matching task and a naturalness rating task. In Most and Saltz (1979), for example, participants heard the target sentence with pitch accent on the agent or on the patient (e.g., *The PITCHER threw the ball.* vs. *The pitcher threw the BALL.*) and wrote the most appropriate question to which the target sentence might be a reply (*Who threw the ball?* vs. *What did the pitcher throw?*). Most and Saltz found that the types of questions the participants produced were highly dependent on the location of pitch accent in the target answer.

Bock and Mazzella (1983) used a speeded sentence comprehension task in which the target sentence (e.g., *DORIS fixed the radio.*) was presented following the appropriate context sentence (e.g., *ARNOLD didn't fix the radio.*), an inappropriate context sentence (e.g., *Arnold didn't FIX the radio.*), or a neutral context sentence with no special emphasis on any of the words (e.g., *Arnold didn't fix the radio.*), and participants pressed a lever as soon as they understood what the target sentence meant. The participants' comprehension times were shorter when the accent placement in the target sentence was appropriate for the context sentence than when it was not, confirming the facilitative effect of appropriate accentuation on listeners' comprehension. Similar results were found in a study conducted in Dutch (Terken & Nooteboom, 1987). In a picture verification task, participants saw a visual display with alphabetical characters moving from one position to another in a sequence, heard audio stimuli consisting of a context sentence followed by a target sentence such as (the Dutch equivalent of) *The Q is on the left of the P. The P is on the right of the K*, and judged whether the sentences they heard were correct descriptions of the visual changes. Terken and Noteboom manipulated the information status (new vs. given) and accentuation (accented vs. deaccented) of the subject noun and the head noun of the place adverbial in the target sentence. The results showed shorter verification latencies when new information was accented and when given information was not, which led to a conclusion that accented and unaccented expressions are processed in different ways: The presence of an accent drives the listener's attention to acoustic details, while the absence of an accent leads the listener to map the expression onto a set of discourse entities that are currently activated.

Using question-answer pairs similar to those in Most and Saltz (1979), Birch and Clifton (1995) conducted a prosodic naturalness rating task and a speeded meaning judgment task. In the former, participants used a 5-point scale to rate the prosodic naturalness of intonation patterns of the target answer given to a question (e.g., Q: *Isn't Kerry good at math?* – A: *She teaches MATH.* vs. A: *She TEACHES math.*). In the latter task, which was intended to examine how prosody affects meaning comprehension, participants heard the same question-answer pairs with appropriate and inappropriate pitch accent patterns in the target answer, and responded by pushing the "yes" or "no" lever depending on whether the target sentence was a meaningful answer to its question. In both tasks, listeners generally preferred answers in which new information was accented and given information was not, as reflected in higher ratings in the rating task and shorter comprehension time in the meaning judgment task.

Native speakers' production of new vs. given referents has also revealed the close associations between accentuation and new referents, and deaccentuation and given referents. In an L1-Dutch study by Krahmer and Swerts (2001), utterances with an adjective-noun sequence were recorded using a semi-spontaneous interactive card game between pairs of participants.

The two participants took turns in describing geometrically shaped, colored cards to each other, specifying the color and the shape of the objects (e.g., *red circle*). The cards were ordered beforehand as a stack so that the target utterance (e.g., *blue square*) would be elicited in the following four contexts: all new (beginning of game → *blue square*), single contrast in the adjective (*yellow square → blue square*), single contrast in the noun (*blue triangle → blue square*), and double contrast (*red circle → blue square*). It was found that Dutch speakers generally accented new and contrastive information and deaccented given information, and that single contrastive accents on the adjective were especially more prominent than new accents.

These offline studies seem to confirm native Germanic speakers' reliance on accentual cues in the processing of new vs. given information status. However, more recent studies employing online measures such as the eye-tracking paradigm offer mixed results. Dahan, Tanenhaus, and Chambers (2002), for example, used pairs of cohort items such as *candy* and *candle* displayed on a computer screen along with geometric shapes. Participants' eye movements were tracked as they heard sentences such as *Put the candy/candle above the triangle. Now put the CANDLE/candle above the square*. Dahan et al. manipulated (1) the first-mentioned object so that the target noun in the second instruction (*candle*) becomes either anaphoric or non-anaphoric, and (2) the presence vs. absence of pitch accent on the target noun itself. The results showed that during the time the ambiguous segment [kæn] was being heard, accentuation on the target syllable (of the noun) triggered more looks to a new object (one not mentioned in the first instruction), whereas deaccentuation of the syllable (of the noun) evoked increased looks to the object previously mentioned. These results provide additional evidence for the effect of accentuation on L1 speakers' real-time processing of new vs. given referents.

Using more controlled, resynthesized sound stimuli, Isaacs and Watson (2010) replicated Dahan et al. (2002) in order to test what acoustic features listeners use to make decisions about presence and absence of a pitch accent in reference resolution. They manipulated pitch (high vs. low F0 contour) and duration (short vs. long) of the target noun and examined how those acoustic manipulations affect listeners' eye fixations to new vs. given referents. The results showed that high pitch facilitated looks to new referents and low pitch to given referents, but this difference between high and low pitch conditions appeared only when duration was short. The authors interpreted these results as suggesting that the interaction of both pitch and duration, i.e., F0 slope, contributes to listeners' perception of pitch accents. When duration is short, pitch slope becomes steeper. That is, the high pitch condition generates a steep pitch rise and the low pitch condition a steep fall, which enhances the pitch effect on eye fixations to new and given referents. When duration is long, pitch slope is less steep in both high and low pitch conditions, which reduces the pitch effect on fixation preferences for both new and given referents. The study implies that the combination of both pitch and duration guides native English listeners to distinguish accented nouns from unaccented ones, and helps them identify new vs. given discourse referents.

Contrary to Dahan et al. (2002) and Isaacs and Watson (2010), evidence against the accented word bias was found in Arnold (2008). Using the visual world eye-tracking paradigm, she tested adults and four- and five-year-old children to see if presence or absence of accent influences their referential interpretation. Participants listened to and followed the auditory instructions such as *Put the bacon on the star. Now put the bagel/bacon on the square*, in which the target noun in the second instruction (e.g., *bagel/bacon*) was either a new referent or a given referent, presented with or without a pitch accent. For both adults and children, unaccented nouns led to increased looks to previously-mentioned objects, while accented nouns did not trigger a strong bias toward new objects.

Several production studies also provide evidence against the accentuation bias toward new information. As briefly mentioned in Chapter 2, Terken and Hirschberg (1994) conducted a picture description task in which native speakers of English saw objects move on a visual display ("events"), one at a time, and described each event in a sentence (e.g., *The ball touches the cone. The ball touches the cross. The ball touches the diamond. The ball touches the star.*). Crucially, the grammatical role of the critical noun (e.g., *the ball*) in the last event was varied (e.g., *The ball touches the star.* vs. *The star touches the ball.* vs. *The box pushes the star against the ball.*) so as to manipulate the noun phrase (NP) carrying the same or different grammatical role as in the previous utterances. Native speakers showed a tendency, first, to deaccent given information only when the NP's grammatical role remains unchanged through the sequence of events and, second, to accent given information when the NP bears a different grammatical role. In fact, accenting given information appears rather common according to the corpus study by Sityaev (2000). He reports that 86% of given referents were accented in English speech that was read and points out that rhythmicality in the metrical grid and topicality of given information appearing in the subject position contributed to the unexpectedly large incidence of accentuation on given referents.

In sum, the studies reviewed in this section suggest that native speakers make rapid use of accentual cues to process and even anticipate new vs. given information status of discourse referents. However, it seems that there is an asymmetry between accentuation and deaccentuation in their cue reliability: Deaccentuation appears to be a fairly strong cue for given information, whereas accentuation does not reliably indicate new information. In order to examine this asymmetry further, the current study tests both native controls and L2 learners in two different types of comprehension tasks as well as in a semi-spontaneous production task comparing the effect of H* and deaccentuation.

3.1.2 Processing studies on L1 prosodic marking of new vs. contrastive information

The difference between two types of English pitch accents, H* and L+H*, in the identification of discourse referents has also been investigated in both offline and online studies. Cruttenden (1985) tested adult and 10-year-old child native speakers of English on their interpretation of contrastive pitch accent placement. Participants saw a set of three pictures (e.g., (a) a boy with four oranges and a girl with two oranges, (b) a boy with four oranges and a girl with four bananas. or (c) a boy with three oranges and a girl with four oranges), heard the target sentence (e.g., *John's got FOUR oranges.*), and chose the best matching picture out of the three options. The mean accuracy rate was significantly higher for the adult group than for the children, which led the author to conclude that L1 adults are highly sensitive to the location of prosodically encoded contrastive focus in spoken sentence interpretation, while children are not.

More recent studies provide evidence for native sensitivity to contrastive prosody during real-time discourse processing. For example, Ito and Speer (2008) found anticipatory effects of H* and L+H* accents in their eye-tracking study incorporating a naturalistic, interactive task. In their second experiment, participants listened to instructions such as *Hang the green drum. Now hang the BLUE drum*, and decorated four holiday trees with a given set of ornaments. The authors manipulated the color adjective and the noun in the first and second instructions so as to create two discourse conditions: In the adjective contrastive condition, the target noun remained the same while the adjective changed from the first instruction to the second (e.g., *green drum* → *blue drum*). In the no contrast condition, the second-mentioned object differed from the first-mentioned object by both color and type (e.g., *red angel* → *blue drum*). Each condition was presented with either an H* or L+H* pitch accent on the color adjective in the second instruction in order to examine whether participants' eye fixations to the target object would be affected by pitch accent type. The results confirmed the prediction that in the adjective contrastive condition, L+H* on the modifying adjective resulted in faster eye gaze to the target object (e.g., *green drum* → *BLUE drum*) compared to when H* was on the same adjective. Furthermore, in the no contrast condition, contrastive L+H* pitch incorrectly guided, or "garden-pathed", participants to the contrastive referent (e.g., increase in looks to *blue drum* in the trial *green drum* → *BLUE angel*). These results confirmed the distinct functions of H* and L+H* pitch accents and the immediate effect of prosodic cues on discourse processing in L1 English comprehension.

In a similar vein, Watson, Tanenhaus, and Gunlogson (2008) investigated native English speakers' real-time interpretation of H* vs. L+H* also using the eye-tracking paradigm. Listeners heard instructions such as *Click on the camel and the dog. Move the dog to the right of the square. Now, move the camel/candle below the triangle*, while looking at a visual display showing four objects, which included two members of a phonetic cohort set (i.e., items that match in initial phonetic segments), and four geometric shapes. The target object in the third instruction carried either H* or L+H*. Watson et al. predicted that L+H* should trigger a bias toward the contrastive referent (e.g., *camel*) and H* toward the new referent (e.g., *candle*). The results revealed listeners' strong tendency to look at contrastive referents with L+H* but no preferential looks to new referents with H*, which led to the conclusion that L+H* is a strong cue for contrast while H* is compatible with both new and contrastive information.

The role of contrastive pitch accent was also investigated by Weber, Braun, and Crocker (2006) in German. Native German listeners heard two consecutive instructions such as (the equivalent of) *Click on the purple scissors. Click now on the red vase*, at which point they were to click on the specified objects on a computer display while their eye movements were monitored. In the critical stimuli, two factors, referent (contrastive vs. non-contrastive) and prosodic accent (L+H* on the color adjective vs. on the noun) were crossed, yielding four conditions: non-contrastive referent/L+H* on the noun (e.g., *purple scissors* → *red vase*$_{L+H*}$), non-contrastive referent/L+H* on the adjective (e.g., *purple scissors* → *red*$_{L+H*}$ *vase*), contrastive referent /L+H* on the noun (e.g., *purple scissors* → *red scissors*$_{L+H*}$), and contrastive referent/L+H* on the adjective (e.g., *purple scissors* → *red*$_{L+H*}$ *scissors*).

Although there was an overall preference for contrastive referents regardless of the prosodic accent condition, listeners fixated on the contrastive referent–both as the target in the contrastive referent (L+H* on the adjective) condition, and as the competitor in the non-contrastive referent (L+H* on the adjective) condition–more than the non-contrastive referent upon hearing L+H* on the adjective, consistent with the previous finding that L+H* pitch accent facilitates rapid identification of contrastive discourse referents.

These results stand in contrast with the findings in Sedivy, Tanenhaus, and Chambers (1999), which is one of the earliest studies that employed the eye-tracking technique to investigate the effect of contrastive prosody on online reference resolution. They used sentences such as *Touch the pink comb. Now touch the yellow/YELLOW comb*, with either an H* or L+H* pitch accent on the color adjective in the second instruction, and these sentences were aurally presented along with a display containing four objects: the first-mentioned object (e.g., *pink comb*), the target object (e.g., *yellow comb*), a competitor that had the same color as the target object (e.g., *yellow bowl*), and a distractor (e.g., *blue knife*). The authors predicted that if listeners are sensitive to the presence of the contrastive pitch accent, they should fixate on the contrastive target object more quickly when the color term is produced with L+H* than H*. The results, however, did not support this prediction. Regardless of the pitch accent type, the modifier was immediately interpreted as contrastive, showing speeded fixation to the target object.

One of the reasons for the lack of prosodic effect in Sedivy et al. (1999) could be under-controlled test stimuli, as pointed out in Ito and Speer (2008). In Sedivy et al., the experimenter read aloud the instructions off a written script rather than using pre-recorded and pre-analyzed audio as in the three other eye-tracking studies described above. Since the instructions were not recorded during the experimental session and no acoustic or ToBI analysis was conducted, it is unclear how exactly the two prosodic conditions were produced and whether they differed significantly. In semi-spontaneous speech like this, speech rate, loudness, pitch range, and the overall pitch contour could vary from utterance to utterance even for the same prosodic condition, and therefore the L+H* pitch accent in Sedivy et al. may not have been produced as consistently and prominently as in the other eye-tracking studies. It is critical to control phonetic and phonological details of test materials in prosodic processing experiments since a number of studies investigating acoustic realization and interpretation of L+H* indicate that some acoustic cues are more important than others to the distinction between L+H* and H*.

For instance, Bartels and Kingston (1994) investigated which acoustic feature or features best discriminate H* from L+H* for native English listeners by using the target sentence *Amanda had a banana*, with acoustic manipulations on the critical word *banana* in four different dimensions: height of the pitch peak, depth of the low dip at the onset of the stressed syllable *na*, timing of the pitch rise, and exact location of the pitch peak. Participants heard a dialogue such as Q: *So, did Amanda eat anything today?* – A: *Yes, she ate her apple*, followed by the target sentence uttered by a different person, B: *Amanda had a banana*, with various intonation patterns. For each target utterance given, participants judged whether the target sentence meant "she ate a banana instead of her apple" (contrastive), or "she ate a banana in addition to her apple" (non-contrastive). Bartels and Kingston found that the strongest acoustic cue that distinguishes contrastive from non-contrastive interpretations of the target utterance was pitch peak height. In fact, in all of the three eye-tracking studies described above that revealed the immediate contrastive interpretation of L+H*, the authors explicitly manipulated their test stimuli so that L+H* condition was realized with significantly higher pitch peak than the H* condition.

Results from the offline English focus interpretation task by Welby (2003) further provide indirect support for the proposal that pitch height is the most reliable cue to the contrastive pitch accent. In her study, native English speakers listened to question-answer pairs marked with an H*, L+H* or no pitch accent on the object NP in the answer (e.g., Q: *How do you keep up with the news?* – A: *I read the DISPATCH*$_{H*}$ vs. A: *I read the DISPATCH*$_{L+H*}$ vs. A: *I READ*$_{H*}$ *the Dispatch*$_{\emptyset}$), and rated the appropriateness of intonation for each pair. It has been claimed that a pitch accent on the head noun of the object NP projects focus to the entire verb phrase (VP), whereas a pitch accent on the verb marks focus only on the verb itself, which makes the former, but not the latter, appropriate for the answer to a broad focus question (Selkirk, 1995). Native listeners gave similar ratings to dialogues with H* and those with L+H* on the object while they dispreferred dialogues in which the the object NP in the answer was not marked with pitch accent. Based on these results, the author concluded that native English speakers are sensitive to presence or absence of pitch accent but less so to different types of pitch accent (H* vs. L+H*) when they interpret focus structure. In their experimental stimuli, the timing of the pitch rise was earlier for H* stimuli than for L+H* stimuli but the height of the pitch peak was kept equivalent between the two types. The closely matched pitch peak between the two conditions may be one of the reasons why their stimuli did not create an interpretative bias between H* and L+H*.

Taken together, the previous studies examining the contrastive pitch accent indicate that pitch height has substantial influence on the interpretation of prosodic emphasis: Listeners perceive pitch accents with relatively higher pitch as more emphatic, and the increased prosodic emphasis is more likely to trigger contrastive interpretations. This is not to say other acoustic cues are irrelevant to contrastive prosody. In addition to pitch height, duration was a distinguishing feature between L+H* and H* pitch accent conditions in Ito and Speer (2008), Watson, Tanenhaus, and Gunlogson (2008), and Weber, Braun, and Crocker (2006), all of which have shown significantly different eye fixation patterns for L+H* and H*. Contrary to these studies, Sedivy, Tanenhaus, and Chambers (1999) and Welby (2003), in which no distinct effect of L+H* and H* was found, did not control duration in a systematic way. These studies thus imply that duration may be another important acoustic cue to the distinction between L+H* and H*.

Others have argued that the strongest cue to the distinction is intensity. Breen, Fedorenko, Wagner, and Gibson (2010) investigated English speakers' prosodic marking of broad (new) vs. narrow (contrastive) focus in both production and comprehension. In a dyadic task, one participant serving as a speaker saw a written question on a computer screen (e.g., *What did Damon fry this morning?* or *Did Damon fry a chicken this morning?*) and answered the question based on pictorial cues presented on the screen (e.g., *I think Damon fried an omelet this morning.*). The other participant serving as a listener heard the answer produced by the speaker, and among a list of seven written questions presented on his/her own computer screen, chose the one that he/she thought the speaker was answering. As for production, contrastive information was produced with longer duration and greater maximum intensity than new information, but new information was produced with higher mean and maximum pitch than contrastive information. The latter finding about new referents is inconsistent with the observation from the previous literature that contrastive focus is produced with higher maximum pitch and larger pitch range than non-contrastive focus (Bartels & Kingston, 1994; Ladd & Morton, 1997). Based on their production results, Breen et al. support the claim by Kochanski, Grabe, Coleman and Rosner (2006) that higher intensity serves as a stronger cue than higher pitch to acoustic prominence. In comprehension, however, only six out of ten listeners were able to distinguish between new and contrastive information above the chance level, implying that intensity and duration may not actually be strongly associated with listeners' perception of acoustic prominence. In addition, Breen et al. do not conduct phonological analysis by annotating the type of pitch accent used in production, which leaves it unclear whether acoustic signals and discourse meaning are associated directly or mediated by intonational categories such as H* and L+H*.

Although which acoustic features differentiate contrastively and non-contrastively focused elements is still an unresolved issue, the studies reviewed in this section provide reasonable evidence that relatively higher pitch is the primary cue to distinguish L+H* from H*, and that L+H* is more likely to be interpreted as contrastive than new, whereas H* may be compatible with either new or contrastive. Apparently, native speakers of English can quickly integrate prosodic cues into the interpretation of discourse meaning and even form expectations about upcoming linguistic elements based on the available prosodic information. For non-native speakers of English whose L1 prosodically differs from English, however, it is extremely difficult to acquire target-like prosodic marking of discourse status especially because the relationship between prosody and discourse interpretation in English is not as simple as a one-to-one mapping (Calhoun, 2010a; Katz & Selkirk, 2011; Terken & Hirschberg, 1994). In addition, not all languages use pitch accents to convey focus or information status, which adds an additional layer of difficulty for learners of English whose L1 does not mark discourse information prosodically or does mark it prosodically but in a different way. Jun (2005) proposes that languages can be classified into head-prominence and edge-prominence languages. Head-prominence languages include English and other Germanic languages such as Dutch and German, whereas edge-prominence languages include Korean and Japanese. In head-prominence languages, focus is signaled by placing a pitch accent on a stressed syllable of the focused word regardless of its syntactic position. In edge-prominence languages, Information Structure is typically realized by flexible word order, and speakers insert a prosodic boundary immediately before the focused element and delete boundaries in the post-focus phrase. Valluduví (1992) proposes a different way of classifying languages depending on whether information status influences accent distribution. In plastic languages like English and Dutch, accentuation patterns are rather flexible while word order is generally fixed, and therefore the location of pitch accents is by and large determined by discourse and pragmatic factors. In non-plastic languages such as French and Spanish, accents are usually placed in fixed positions, while discourse and pragmatic information is conveyed via flexible word order.

The next section reviews the previous literature investigating the acquisition and processing of English (and Dutch) prosodic marking of information status by L2 learners who come from a typologically different language group such as Mandarin and Spanish and those from the same group like Dutch, and it introduces acquisition theories and models that have been applied to the research on L2 prosody-discourse mappings.

3.2 Previous studies on L2 prosodic marking of Information Structure

The investigation of L2 speech perception and production at the suprasegmental level falls into several different areas: the use of prosodic rhythm for word segmentation and recognition (Cutler, 2000; Cutler, Mehler, Norris, & Segui, 1983, 1986; Cutler, Mehler, Otake, & Hatano, 1993; Cutler & Otake, 1994; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Murty, Otake, & Cutler, 2007; Tremblay, 2008; Tremblay, Broersma, Coughlin, & Choi, 2016), prosodic word structure and morphological marking (Goad & White, 2004, 2008), lexical prosody in tone languages (Y.-S. Lee, Vakoch, & Wurm, 1996; So & Best, 2010; Wang, Spence, Jongman, & Sereno, 1999; Wayland & Guion, 2004), the use of prosody for phrasing and syntactic analysis (Dekydtspotter, Donaldson, Edmonds, Fultz, & Petrush, 2008; Harley, Howard, & Hart, 1995; Hwang, 2007; Ma, 2007; Nibert, 2005, 2006; O'Brien, Jackson, & Gardner, 2014; Ying, 1996), sentence-level word prominence (Backman, 1979; Jenner, 1976), prosodic marking of Information Structure (Akker & Cutler, 2003; Braun & Tagliapietra, 2011; Chen & Lai, 2011; Kelm, 1987; Turco, Dimroth, & Braun, 2015; Wennerstrom, 1994, 1998), the role of intonation in foreign accent (van Maastricht, Zee, Krahmer, & Swerts, 2017), age, experience and other factors affecting L2 prosody (Harley et al., 1995; Huang & Jun, 2011; Piske, 2012; Trofimovich & Baker, 2007), and the use of paralinguistic intonation (Chen, 2009).

Although L2 prosody has received increasing attention in recent years (Mennen & de Leeuw, 2014; Trouvain & Gut, 2007), it still lacks established theories or models as to whether suprasegmental characteristics of learners' L1 transfer to L2, and how perception and production of non-native prosody develop over the course of L2 acquisition. The lack of well-developed theories on the L2 acquisition of discourse prosody is probably one of the main reasons why the majority of L2 studies on prosodic marking of Information Structure has been conducted without specifying any acquisition model, but simply comparing L1 and TL intonational systems and examining L1 traces in L2 learners' use of prosody. The following sections summarize the previous L2 studies on prosodic marking of Information Structure that make no reference to L2 acquisition theories or models. We present L2 production studies first as they are more common, then turn to the relatively sparse comprehension/perception studies.

3.2.1 Production studies on L2 prosodic marking of Information Structure

One of the earlier production studies in L2 discourse prosody is Wennerstrom (1994). She investigated how Japanese, Spanish, and Thai native speakers, who were in-training international teaching assistants in an American university, employ English intonation to signal discourse structure (given vs. new information, contrastive vs. non-contrastive, pitch reset in a new paragraph, and sentence continuation vs. end) in oral paragraph-reading and spontaneous picture-description tasks conducted both in English and in the participants' native language. At the time of the experiment, the participants were enrolled in an intermediate class in the intensive English program and their length of residence in the US ranged from less than three weeks to two years. For data analyses, new, contrastive, and given information (among others) were identified in the recorded speech, then pitch and intensity were measured. The results revealed native speakers' consistent use of clear pitch contrast especially for contrastive and given information as opposed to the non-native speakers' underuse of pitch signals for information status. The author explains that Thai speakers in particular performed poorly due to insufficient amount of TL exposure and the L1 effect of not using tones for post-lexical discourse marking. Their length of residency in the US (less than three weeks) was indeed much shorter than that of the other non-native groups (four months - two years for the Japanese group, and less than nine months for the Spanish group). However, the effect of L1 transfer is rather speculative, as the study does not provide detailed descriptions of cross-linguistic differences in intonation nor does it present data from the tasks conducted in the participants' native language.

Lack of detailed descriptions of participants' L1 intonation system is in fact not uncommon in L2 prosody studies. A longitudinal corpus study by Verdugo (2002) also reports discrepancies between native English speakers and L1-Spanish L2 learners of English in their prosodic marking of new vs. given information, but without referring to the cross-linguistic differences in discourse prosody between the two languages. Their English speakers placed tonal prominence on new information and not on given information, whereas the upper-intermediate L2 learners marked both new and given information with prosodic prominence. In addition, the L2 learners in general used narrower pitch range compared to the native group, making it difficult to identify prosodically focused elements in the L2 data, and they often placed accent on the last word of an utterance regardless of Information Structure. Since the author does not provide interpretations of the data in terms of the learners' L1, the exact role of L1 transfer in this study remains unclear. In addition, although this longitudinal study was conducted over a period of three years, the data are analyzed as a whole rather than by different time points. This makes it impossible to see whether and how the L2 learners developed their ability to use prosody over the course of development.

Another study by Verdugo (2006), investigating L1-Spanish upper-intermediate English learners' production of broad (new) and narrow (contrastive) focus in a scripted dialogue, further reveals important differences between native and non-native speakers' prosodic patterns. At the phonological level, English native speakers used nuclear pitch accent on the last lexical word in the broad focus sentences and on the focused word in the narrow focus sentences, while the L2 learners often placed prominence on the last word of an utterance for both broad and narrow focus conditions. At the phonetic level, the native speakers used expanded pitch range with a steep rising contour followed by a fall (L+H* L-) to indicate contrastive focus, while the Spanish learners used narrower pitch range with an H plateau (H*) or a low rising contour (L*+H L- or L*+H%). The study, however, again does not describe the differences between Spanish and English intonation systems, and therefore it is unknown to what extent the distinct prosodic performance found in the task is due to L1 transfer.

Gut, Pillai, and Don (2013) is another such study that lacks a systematic cross-linguistic comparison between learners' L1 and the TL. They investigated L1-Malay learners' marking of new and given information in Malaysian English. While very little research has been conducted on prosodic marking of information status in Malay, Gut, Pillai, and Don refer to one study demonstrating that Malay speakers tend to employ demonstrative pronouns and particles (such as *lah*) to mark new information rather than using pitch accents or intonation. Based on this report, the authors predicted that Malay learners of English would not utilize prosody to indicate information status in their L2. The study consisted of three parts: a dyadic semi-spontaneous card game, scripted story reading, and a perception task. In the dyadic card game, each participant had a pile of cards with pictures of different objects in different colors (e.g., *green moon*), and they took turns to read out and move the card on top of their pile. The cards had been ordered in such a way that either the color or the type of object constituted new or given information. The utterances from the card game were marked by two independent raters as to whether the color adjective, or the noun, or both had accent. In the scripted reading task, participants read a short story that contained new and repeated words. In the perception task, another group of Malaysian speakers listened to 15 recorded utterances from the new–new condition, the new–given condition, and the given–new condition in the card game and indicated which one of the three questions the utterance they heard was the most likely answer. For example, they heard *blue moon* and selected either *What is this?* (implying a new–new structure), *Was it a blue sun?* (implying a given–new structure), or *Was it a green moon?* (implying a new–given structure).

The results revealed that the L1-Malay L2 learners of English placed accent on the noun in all conditions, suggesting that they did not differentiate new vs. given information by using distinct types of pitch accents. Furthermore, in a perception experiment, listeners were not able to correctly identify new vs. given information status in the L2 participants' utterances. The acoustic analysis of the reading task showed a tendency for new information to be marked with an earlier and steeper pitch rise than given information. However, in their follow-up study (Gut & Pillai, 2014) comparing L1-Malay and L2-English production using the same scripted story reading task from Gut et al. (2013), the authors found no difference in types of pitch accents, peak alignment, and steepness of rises for new vs. given information in both L1 Malay and L2 English, confirming direct L1-Malay influence on prosodic marking of information status in L2 Malaysian English. Although the additional L1 data in Malay provide some insights into the impact of L1 prosody on L2 production, the L2 learner groups tested in the studies included mixed populations and proficiency levels, from ones who spoke English as one of their first languages to those who rarely used English outside the university. This could be a reason why no prosodic differences between new and given information were found in either the semi-spontaneous production task or the perception task. Or it could simply be that prosodic marking of Information Structure in Malaysian English is quite different from that of Mainstream American English. To tease apart those two possibilities, an additional control group consisting of native Malaysian English speakers needs to be tested in the same experimental design.

Nava (2008) is more informative in that she illustrates the differences between English and Spanish in the placement of nuclear pitch accents in relation to focus. The two languages differ with respect to prosodic marking of broad (new) focus: In English, a nuclear pitch accent on the subject is allowed in unergative or unaccusative constructions (where the predicate contains no object), while for the same constructions in Spanish, a VS order is typically used and the nuclear pitch accent is placed at the right edge of the sentence because a nuclear pitch accent has to fall in the position preceding a phrase boundary in Spanish. This is illustrated in the examples below.

(4) English
Q: Why did they scream?
A: The BUILDING just shook.

(5) Spanish
Q: ¿Por qué está molesto tu papá?
why is-3SG upset your father
"Why is your father upset?"
A: Se rompió la VENTANA!
REFL break-PST-3SG the window
"The window broke."

L1-Spanish learners of English at intermediate and high proficiency levels completed two main tasks and a cloze-test as an independent proficiency measure. The main tasks involved scripted dialogue reading and story description designed to elicit all new information in intransitive, transitive, and VP focus constructions. Though the study does not describe how the recorded utterances were annotated and analyzed, the results for the intransitive construction showed that in the reading task, some highly proficient L2 learners demonstrated native-like stress patterns by placing the nuclear accent on the subject. The other learners often placed the nuclear accent on the sentence-final word regardless of Information Structure. In the story description task, the L2 learners often used the "there is" construction when producing unaccusative sentences (e.g., *There is a SHARK in the water!* instead of *A shark comes out of the water!*), leaving little data on whether the learners allowed accentuation on the subject NP, but the author does not report how this tendency relates to the participants' proficiency in this particular task. Based on the results from the two tasks, it was concluded that while advanced learners of English are able to acquire target-like nuclear pitch accent placement, they also tend to rely on syntactic means to avoid placing an accent-bearing subject conveying new information in the utterance-initial position perhaps due to the L1 constraint.

In contrast to the above-mentioned studies that only allude to possible influence of L1 on the intonational characteristics of L2 learners' English production, Swerts and Zerbian (2010) is more explicit with respect to the interplay between conceivable L1 influence and language proficiency. Using a semi-spontaneous picture description task, their study investigated L1 effects on English prosodic marking of boundaries and focus by L1-Zulu speakers at various proficiency levels measured by an independent English proficiency test. Zulu is similar to English in that high boundary tones signal continuity and low boundary tones indicate finality. As for focus marking, English marks focus by prosodic means while Zulu lacks such prosodic marking of focus structure. In the picture description task, participants described a series of colored pictures (e.g., *blue flower* → *yellow tree* → *red star* → *red house* → *red cow*) in which the target picture (e.g., *red cow*) appeared in the middle of the sequence or at the end of it (non-final vs, final). In addition, by manipulating the color and type of objects, the task elicited three types of focus: weak contrastive focus, strong contrastive focus, and corrective focus. In the weak contrastive focus condition, the target object built contrast with the two previous pictures only (e.g., *blue flower* → *yellow tree* → *red star* → *red house* → *red cow*), while in the strong contrastive focus condition, contrast was built over all the preceding pictures (e.g., *red flower* → *red house* → *red star* → *red tree* → *red cow*). In the corrective focus condition, participants were presented with only the target picture, followed by a slide which said *Please correct if necessary: Is this a X?* where X gave an inappropriate description of the target object. The recorded speech samples were coded by native English speakers based on their perceived prominence (color adjective, noun, or neither) and continuity (rising, falling, or level pitch) of the target NP. The results showed that L2 learners of all proficiency levels used prosodic boundaries to signal finality or continuity in the same way as the native English controls did. As for prosodic marking of focus, highly proficient learners were comparable to the native English controls in weak and strong contrastive focus, while low proficient learners did not prosodically mark Information Structure in a systematic way. The study suggests that novice L2 learners are susceptible to L1 transfer but they may gradually approximate the native-like level over the course of L2 development.

Using a similar semi-spontaneous picture naming task from Swerts and Zerbian (2010), van Maastricht, Krahmer, and Swerts (2016) investigated to what extent L1 prosodic transfer is constrained by learners' proficiency level. L1-Dutch learners of Spanish and L1-Spanish learners of Dutch as well as control groups of Dutch natives and Spanish natives participated in the study. The L2 participants in each language were further split into a more proficient group and a less proficient group based on the level of the university language course they had completed before the study. As mentioned earlier, Dutch is a plastic language: New and contrastive information are usually accented, while given information is often deaccented. Spanish, on the other hand, is a non-plastic language: Information Structure is signaled via syntactic movement rather than accent patterns. In the picture naming task, participants saw a sequence of pictures of different objects in different colors, and named each picture (e.g., *pink broom* → *blue donkey*). The pictures were ordered so as to create four types of Information Structure: contrastive/contrastive (e.g., *pink broom* → *blue donkey*), given/contrastive (e.g., *blue broom* → *blue donkey*), contrastive/given (e.g., *red donkey* → *blue donkey*), given/given (e.g., *blue donkey* → *blue donkey*). Acoustic analyses of pitch values on the critical color adjectives and the noun revealed that the native Spanish speakers consistently placed prosodic prominence on the noun across all four conditions, whereas native Dutch speakers made distinctions between given and contrastive information prosodically. As for the L2 learners', L1-Spanish learners of Dutch performed more like Spanish natives than Dutch natives, and L1-Dutch learners of Spanish were more similar to Dutch natives than Spanish natives, implying traces of L1 influence in both L2 groups. Furthermore, the degree of L1 prosodic transfer in L2 prosodic focus marking seemed to diminish as proficiency increased, which lends further support for proficiency effects on L1 prosodic transfer in L2 focus marking.

While both Swerts and Zerbian (2010) and van Maastricht et al. (2016) found L1 transfer and proficiency effects, other researchers have shown that factors other than those may also be at play. He, Hanssen, van Heuven, and Gussenhoven (2001) examined how L1-Mandarin speakers mark broad (new) focus and two types of narrow focus, informational (new) focus and corrective focus, in L2 Dutch. Both Mandarin Chinese and Dutch employ longer duration and expanded pitch range on focused elements followed by post-focal pitch range compression, which is realized as an H*L pitch accent. Since the two languages share comparable phonological categories for realizing prosodic focus enhancement, the authors predicted that both Mandarin L2 learners of Dutch and Dutch natives would mark the three types of focus in similar ways. L1-Chinese participants were divided into higher and lower proficiency groups based on their mean segmental and prosodic proficiency scores measured in a prior experiment conducted by the same authors. Participants performed a scripted question-answer dialogue reading task where they read out Dutch sentences that contained the three types of focus (e.g., Q: *What are your plans for tomorrow?* – A: *I would like to cycle to Momberen* [broad focus]; Q: *Where would Karel like to take you?* – A: *He would like to take me to Momberen* [informational (new) narrow focus]; and Q: *Did your mother want you to go to Zaltbommel?* – A: *No, she wanted to send me to Momberen* [corrective narrow focus]). Contrary to the prediction, the L2 group showed irregular performance in prominence marking, and their performance diverged from that of the native Dutch group in phonetic details of the critical word (e.g., *Momberen*): Neither high nor low proficiency L2 group demonstrated native-like patterns in their use of pitch range, word duration, and post-focal pitch fall for the marking of the three types of focus. This suggests that L2 leaners may not always transfer their L1 prosodic properties to L2 even when there is no difference in prosodic meaning between the two languages.

Another example comes from O'Brien and Gut (2010), which investigated whether L1-German intermediate-advanced L2 learners of English (judged by the Oxford Online Placement Test) exhibit L1 transfer in pitch accent placement and phonetic details of prosodic marking of broad, narrow, and contrastive focus during a semi-spontaneous picture question-answering task and a reading task conducted in both L1 German and L2 English. While no description of German pitch accent and focus marking is provided in the study, the authors cite a previous study showing that although pre-nuclear rising pitch accents exist in both languages, the phonetic realization of those pitch accents differed between L1-German learners of English and native English speakers. For the former group, the pitch peak was delayed in a similar way to the pattern found in L1 German. On the basis of this report, O'Brien and Gut predicted that L1 prosodic transfer would occur in the phonetic realization of phonological categories rather than at the phonological level. The produced utterances were analyzed for pitch accent placement and pre-nuclear pitch rises in the three focus conditions (e.g., Q: *What's happening?* – A: *A HOUSE is burning* [broad focus]; Q: *Who is hitting the ball?* – A: *The MAN is hitting the ball* [narrow focus]; Q: *Who is eating the apple and who is eating the banana?* – A: *The WOMAN is eating the apple and the MAN is eating the banana* [contrastive focus]). While most of the learners were able to correctly place pitch accents on the focused words both in L1 German and L2 English, phonetic details of pre-nuclear pitch rises in broad and narrow focus marking differed between when they spoke L1 German and L2 English, and when they performed the semi-spontaneous task and the highly monitored reading task. The authors concluded that those L2 learners were using strategies that are not a simple reflection of L1 prosodic transfer.

To summarize, the production studies reviewed thus far seem to support L1 prosodic transfer and the effect of proficiency on L2 prosodic marking of discourse status. It is also possible that other factors such as the amount of exposure to the TL, age of onset, and task type may of course play important roles, although these issues have rarely been investigated in the available literature. As we reviewed the previous research, two major methodological problems were identified. First, the majority of the L2 production studies do not make systematic cross-linguistic comparisons between L2 learners' native language and the TL with respect to intonational phonology and prosodic marking of discourse, and therefore it is often difficult to draw definite conclusions as to whether the observed L2 performance derives in part from the learners' L1 or something else. In the present study, we purposefully chose languages whose intonational systems are well documented (i.e., English and Japanese) so that we can make clear predictions about expected L2 outcomes on the basis of the linguistic comparisons between the two languages.

The second problem was that some of the previous studies lack objective measures of language proficiency and analyze L2 learners of mixed proficiency as one group without considering the effect of proficiency on task performance. This could obscure potentially interesting behavioral patterns rooted in the proficiency levels and lead to null findings as in the case of Gut et al. (2013). In order to avoid such problem, our study employed two proficiency measures and analyzed L2 data both as a whole group and by proficiency. In the next section, we turn our attention to perception research and present an overview of the previous studies investigating the interpretation of prosodic cues in L2 discourse processing.

3.2.2 Comprehension/Perception studies on L2 prosodic marking of Information Structure

Very few studies have been conducted on the L2 perception of prosodic marking of information status. Among these is an offline study investigating L2 learners' interpretation of ambiguous pronouns (Schafer, Takeda, Rohde, & Grüter, 2015). In the first experiment of the study, native speakers of English and L1-Korean/Japanese learners of English heard an aural context sentence such as *David served/was serving Paul a pint of beer*, with an contrastive pitch accent (L+H*) on either the Source (e.g., *David*) or the Goal (e.g., *Paul*), and saw a written pronoun prompt which was ambiguous with respect to its referent (e.g., *He*...). Their task was to complete the story by writing the continuation following the ambiguous pronoun. Both L1 and L2 participants chose the contrastively-accented antecedent as the referent of the ambiguous pronoun.

In the second experiment, accentuation on the ambiguous pronoun rather than the possible antecedents was manipulated so that the pronoun was either contrastively-accented or unaccented in the aural prompt (e.g., *David served/was serving Paul a pint of beer. He/HE obviously...*). This time, the L1 group, but not the L2 group, demonstrated the effect of contrastive intonation on reference choice. The native speakers chose the Source interpretation about 65% of the times with unaccented pronouns, whereas with accented pronouns the Goal interpretation increased, which suggests that L+H* on the pronoun triggered selection of the less salient referent as the antecedent. The study concludes that L2 learners' successful use of prosody in discourse depends on how complex the relevant mapping is: In Experiment 1, listeners perceive L+H* on a referent, the referent becomes salient in their discourse model, and when the ambiguous pronoun is encountered, the listeners simply consider the salient referent as the antecedent of that pronoun. Reference resolution in Experiment 2 is more complex in that after the listeners hear L+H* on the pronoun, they need to establish two possible referents, identify which one of the two referents is more salient, and select the other referent as the antecedent of the ambiguous pronoun. Though the study provides evidence for L2 sensitivity to contrastive intonation, how mapping complexity influences L2 processing of discourse prosody could have been clearer if the authors had analyzed the L2 data in Experiment 2 in terms of proficiency, as advanced L2 learners are likely to better handle complex L2 processing and thus perform more like the native speakers.

Akker and Cutler's (2003) online phoneme detection study also shows that L2 processing of discourse prosody is not native-like even with advanced L2 proficiency and similar intonational systems between learners' L1 and TL. The main purpose of their study was to test whether L1-Dutch advanced learners of English can process prosodic cues to discourse structure as efficiently as native English speakers do. English and Dutch are similar in their lexical stress assignment rules, prosodic structures, and prosodic focus marking, based on which the authors predicted that L1-Dutch learners of English should rely on the same processing routines for English accentual marking of focus structure as native speakers of English. In the phoneme detection task, participants first saw the target phoneme on the computer screen (e.g., /d/), heard a question-answer pair sentences (e.g., Q: *Which bones were found by the archaeologist?* – A: *The bones of the dinosaur were found by the Cuban archaeologist.*) and pressed a button as soon as they heard the word beginning with the target phoneme. The question and the accentuation in the answer were manipulated to create four conditions: accented and focused target condition (e.g., Q: *Which bones were found by the archaeologist?* – A: *The bones of the DINOSAUR were found by the Cuban archaeologist.*), accented but unfocused target condition (e.g., Q: *Which archeologist found the bones?* – A: *The bones of the DINOSAUR were found by the Cuban archaeologist.*), unaccented but focused target condition (e.g., Q: *Which bones were found by the archaeologist?* – A: *The bones of the dinosaur were found by the CUBAN archaeologist.*), and unaccented and unfocused target condition (e.g., Q: *Which archeologist found the bones?* – A: *The bones of the dinosaur were found by the CUBAN archaeologist.*).

In the L1 experiments, both English and Dutch control groups showed an interaction effect of accentuation and focus position: Listeners detected the target phoneme faster when it was stressed and in a focused position, and when it was deaccented and in a non-focus position. L2 learners, on the other hand, lacked such an interaction effect: Regardless of whether the target phoneme was in a focus position, learners detected the phoneme faster when it was accented. The authors concluded that the interplay between accentuation and focus is what makes native discourse processing fast and efficient, and that the observed divergence between the native and non-native groups implies L2 learners' reduced ability to integrate various pieces of linguistic information even at the advanced proficiency level, attesting to the difficulty of successful acquisition of L2 discourse prosody. Whether those learners were really at the advanced proficiency level is uncertain since no proficiency test in English was administered in the study. It is nevertheless intriguing that L2 learners do not apply their L1 processing strategies to L2 comprehension even when the two languages are comparable with respect to the phonological structure and focus marking.

In contrast to Akker and Cutler (2003), Braun and Tagliapietra (2011) report L1 influence on L2 learners' online processing of information status. They employed a lexical decision task to examine the perception of a particular pitch contour, the hat pattern (rising accent followed by high pitch and then falling accent), by L1-German L2 learners of Dutch at various proficiency levels. In Dutch, the hat pattern carries a neutral meaning, whereas in German the same contour indicates contrastive meaning. During the experiment, participants heard sentences like (the equivalent of) *In Florida he photographed a flamingo* in Dutch, and upon seeing a visual target contrastively related to the prime (e.g., *pelican* for the prime *flamingo*), they decided whether or not the visual target was a real word. The L2 learners' proficiency was measured on the basis of the mean accuracy scores in filler trials. The authors hypothesized that when the prime was presented with a hat pattern, L1-German speakers would be likely to interpret the prime word contrastively due to their L1, which would then speed up the recognition of the contrastively related visual target word. If, on the other hand, German speakers have learned that the hat pattern signals a neutral rather than contrastive meaning in Dutch, such priming would not occur. The results confirmed this prediction. When the prime was embedded in a sentence with the hat pattern, only more proficient L2 learners, not Dutch natives or less proficient L2 learners, showed shorter latencies in lexical decision, which suggests that the hat pattern triggered contrastive interpretation only for the L2 learners because of L1 influence. The authors attribute the lack of a priming effect among less proficient learners to their insufficient L2 ability to understand the prime sentences accurately and to complete the task successfully. The study suggests that the interpretation of L2 prosodic meaning is automatic and guided by the prosody-discourse mapping in the L1, as the authors conclude, and that task difficulty can affect the degree to which L1 influence emerges, especially for low proficiency learners.

Less conclusive results on L2 comprehension of discourse prosody are found in Chen and Lai (2011), which investigated L2 online integration of accentual cues into discourse referential processing using the visual world eye-tracking paradigm. Following the experimental design in Dahan et al. (2002), intermediate-advanced L1-Dutch learners of British English heard sentences such as *Put the comb/coat below the triangle. Now put the comb*$_{(H*L/L*H/deaccentuation)}$ *below the diamond*, and moved pictures displayed on a computer screen according to the directions they had heard. The visual display contained the target object (e.g., *comb*), the competitor with a phonetically similar onset (e.g., *coat*), and two unrelated objects, plus four geometric shapes. Crucially, the object in the first sentence was manipulated so that the target object in the second sentence (e.g., *comb*) was either a new referent or a given referent. In addition, the target referent was produced with H*L (fall), L*H (rise), or no accent.[5] A prior eye-tracking listening study using the same experimental paradigm (Chen et al., 2007) had revealed that native speakers of British English fixated on new referents more when they were produced with H*L than with L*H and no accent, and given referents more with L*H and no accent than with H*L. An analogous production study conducted by Braun and Chen (2010) had further shown that while both British English and Dutch native speakers produced H*L on a new target object, only the Dutch speakers used L*H on a new object when the geometric shape that the new object was moved to was also new. Based on these previous findings, Chen and Lai predicted that native English speakers would associate H*L with new information and both L*H and deaccentuation with given information, whereas Dutch natives would associate both H*L and L*H with new information and deaccentuation with given information. If L1-Dutch learners of English rely on their L1 knowledge of prosody-discourse associations, both H*L and L*H on the ambiguous segment (e.g., *co-*) should trigger more fixations to the new referent, and deaccentuation should facilitate looks to the given referent. If, on the other hand, L2 learners have acquired the target-like mapping between H*L and new information in English and dissociate L*H from new information, their fixation to the new referent should be facilitated with H*L only, and L*H and deaccentuation should trigger looks to the given referent.

Overall, the L2 group was inclined to fixate on given referents more than new referents regardless of the pitch conditions, implying that learners tend to delay their gaze shift to new referents until disambiguating information becomes fully available. However, further analyses of their fixation patterns for given referents revealed a curious finding: The intermediate L2 group performed, more than the advanced group, like the native English speakers tested in their previous study (Chen et al., 2007). The intermediate group linked both L*H and deaccentuation to given information, while the advanced group linked only deaccentuation with given information reflecting their L1 prosody-discourse mapping. The authors' explanation for these puzzling results is that intermediate learners may be more alert to L1 transfer or they may benefit more from exposure to naturalistic use of intonation in spoken English than advanced

[5]The authors adopted the Transcription of Dutch Intonation notation (ToDI) instead of ToBI in their study based on the assumption that ToDI reflects tones in British English more closely than ToBI (Gussenhoven, 2005). They explain that the intonational systems of British English and Dutch are similar, and the phonological categories defined in ToDI also exist in British English. In addition, ToDI pitch accents do not have leading tones characterizing the pitch of a pre-accentual syllable (e.g., L+ in L+H*), and there is no intermediate phrase level, both of which correspond to the British English tradition (Chen, Den Os, & De Ruiter, 2007).

learners, but these scenarios are rather speculative with no supporting data or empirical evidence from their study.

L2 learners' non-native performance in the processing of contrastive pitch accents has also been reported in Lee and Fraundorf (2017), whose primary research objective was to investigate how pitch accents are encoded and represented in learners' discourse memory. Their earlier study (Fraundorf, Watson, & Benjamin, 2010) had revealed that for native speakers of English, contrastive pitch accents improve memory by enhancing the semantic representation of not only the accented entity itself but also its contrastive entities. During the study phase in the memory task, native speakers heard stories consisting of a context sentence followed by a continuation such as *Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys. Finally, the British spotted one of the monkeys in Malaysia and planted a radio tag on it.* The context sentence contained two contrast sets (e.g., British vs. French and Malaysia vs. Indonesia), and the continuation included one item from each set (e.g., British and Malaysia). In addition, the target word (e.g., British) in the continuation was marked with either an H* or L+H* pitch accent. Later at the test phase, participants were given a statement about the continuation that named either the correct entity (e.g., *The British scientists spotted the endangered monkey and tagged it.*), the contrast entity (e.g., *The French scientists spotted the endangered monkey and tagged it.*) or an unmentioned entity (e.g., *The Portuguese scientists spotted the endangered monkey and tagged it.*), and answered whether the given statement was true or false. When L+H* was on the target word as opposed to H* in the continuation, native speakers of English were more likely to accept the test statement naming the correct entity (e.g., *British*) and reject the one with the contrastive entity (e.g., *French*), while they showed no pitch difference in rejecting the unmentioned entity (e.g., *Portuguese*). The results suggest that L+H* strengthens memory representations of the contrast set for native speakers.

Lee and Fraundorf replicated the experiment with L2 learners in order to examine whether L1-Korean low, mid, and high proficiency learners of English, grouped based on a cloze-test, can come to process English contrastive pitch accents in the same way as native speakers, given that contrastive focus in Korean is marked by prosodic phrase boundaries rather than pitch accents. The results showed that for the high proficiency L2 learners, L+H* led to correct rejection of the statement including the unmentioned entity (e.g., *Portuguese*) but did not discriminate the correct entity from the contrastive entity (e.g., *British* vs. *French*). As for the low and mid proficiency learners, no effect of L+H* was observed. Based on the results, the authors propose that L2 learners, compared to native speakers, have a shallower representation of discourse information encoded by pitch accents. Even though advanced L2 learners can master the relationship between L+H* and contrast meaning which is not part of their L1 and establish a contrast set, they suffer from confusions in memory due to cognitive resource limitations rather than L1 transfer, failing to accurately encode in their discourse model which one of the entities is the target and which is the contrastive alternative. Lower proficiency learners, on the other hand, are more susceptible to L1 effects and thus unsuccessful in the L2-specific intonation-meaning mapping between L+H* and contrastive information.

In sum, perception studies on L2 prosodic marking of information status are still rather rare. The few available studies that specifically look at L2 interpretation of prosodic marking of focus or information status seem to suggest that even for advanced L2 learners, native-like processing of the mappings between pitch accents and Information Structure is challenging due to L1 interference, limitations in memory, or difficulty in integrating various pieces of linguistic information. However, these studies are so diverse in their experimental designs that it is difficult to compare results from those studies and draw any obvious conclusions about the effects of L1 transfer, language proficiency, task type, and cognitive resource limitations on L2 sensitivity to discourse prosody.

In the present study, we conducted two different kinds of comprehension tasks with the same group of L2 participants in an attempt to fill in the existing research literature on this topic. One is a prosody-in-context naturalness rating task in which we asked participants to rate the naturalness of different versions of intonation presented in different discourse contexts. Learners were instructed to focus on intonation so that their attention was explicitly drawn to the prosodic aspect of experimental stimuli in relation to discourse, and no time limit was imposed during the task. The purpose of including this task in our study is to examine whether L2 learners (come to) possess native-like knowledge of prosody-discourse associations in the TL. By using a task that deliberately draws learners' attention to prosody and gives sufficient time to think about the meaning of the sentence and make a conscious decision using their metalinguistic abilities, we aim to reveal if L2 learners are capable of exercising their knowledge of L2 discourse prosody when the linguistic property being investigated in the task is transparent to the learners, and the processing demand due to time pressure is minimized. The other task used an online, visual world eye-tracking paradigm in which participants heard instructions and clicked on specified objects on a computer screen. There was no mentioning of intonational manipulation in the task directions so that the participants would concentrate on completing the task rather than paying special attention to prosodic features of the recorded stimuli. The purpose of this task is to see whether learners make use of their knowledge of L2 prosody-discourse associations during real-time processing. All of the L2 comprehension studies described above employed an online processing task that measures learners' behavior in real time to millisecond precision. In the phoneme detection task by Akker and Cutler (2003) and the eye-tracking task by Chen and Lai (2011), participants had to respond as they listened to the sentences unfold rather than at the end of the sentences. In the lexical-decision task in Braun and Tagliapietra (2011), participants had to indicate as quickly as possible whether a given target was a real word or not.

Although online methodologies are excellent means to investigate participants' automatic and unconscious processes while minimizing their reliance on metalinguistic knowledge (Marinis, 2003), they could pose a great challenge to L2 learners at relatively low proficiency level if the task contains unfamiliar vocabulary or if the learners' lexical access is so slow that they cannot comprehend spoken utterances in real time. In order to avoid such possibilities, our eye-tracking task only included simple vocabulary and syntactic structure that would not cause any interference with learners' online comprehension of sentence meaning. Results from the two types of comprehension tasks should give us an insight into whether the level of prosodic awareness modulated by task type influences L2 learners' interpretation of prosodic cues to discourse structure.

In addition to the two comprehension tasks, a production task was also administered to the same group of L2 learners in our study. Although the relationship between production and perception of L2 discourse prosody has not been investigated much, some research has shown that pedagogical training on L2 intonational perception abilities can help learners improve their L2 production skills (Bot & Mailfert, 1982; 't Hart & Collier, 1975), which suggests that there may be some link between production and perception in L2 intonation. On the other hand, Baker (2010) found no correlation between pitch accent perception and realization (production) tasks in her study. In this dissertation research, both production and comprehension tasks were conducted with the same L2 population during a one-time experiment session for a better understanding of this issue. The next section provides an overview of theories and models that have been previously used in studying L2 acquisition of prosodic marking of discourse, and describes the framework in which the present study is couched.

3.2.3 Theories on the L2 acquisition of discourse prosody

Despite the amount of published work on L2 prosody and information status, only a limited number of studies have adopted or developed a particular framework in their studies. Some of the existing prosody studies have extended the major models proposed for the L2 acquisition of segments, such as the Speech Learning Model (SLM) (Flege, 1995) and the Perceptual Assimilation Model (PAM)/PAM L2 (Best, 1995; Best & Tyler, 2007), to suprasegmentals, and used these models as a framework to explain how L2 learners come to comprehend and produce prosodic patterns in the TL.

In these two most-often cited theoretical models of L2 phonological acquisition, the notions of transfer and contrast play a central part. They both share the fundamental assumption that L2 speech is acquired through the phonological representations of their L1. SLM predicts that the greater the perceived distance between a TL sound and the closest L1 sound, the more likely a separate category will be created for the TL sound. Using SLM as a framework, Mennen (1999), for example, investigated L1-Dutch L2-Greek speakers' nucleus placement in the production of Greek yes/no questions. In Greek, yes/no questions are prosodically described as L* H L%, with a low nuclear pitch accent followed by a rise-fall boundary tone at the end of the sentence. The exact location of H (pitch peak) of this boundary tone depends on the location of the low nuclear pitch accent. If the nuclear pitch accent is on the utterance-final word, H occurs on the last syllable of that word even when the syllable is unstressed. If, on the other hand, the nuclear pitch accent appears somewhere before the utterance-final word, H occurs on the stressed syllable of the last word. A pitch contour similar to the latter, i.e., a rise-fall on a lexically stressed syllable, exists in Dutch, but it normally appears in statements rather than yes/no questions. The former pattern, i.e., a rise-fall on an unstressed syllable, is not possible in Dutch. Based on these cross-linguistic observations, Mennen predicted that a rise-fall pitch contour on a stressed syllable in Greek yes/no questions would be more difficult to produce than a rise-fall contour on an unstressed syllable for L1-Dutch L2 learners of Greek. The experiment using scripted dialogues asked six near-native L2 speakers of Greek to produce yes/no questions reflecting different nuclear accent placement, and the results revealed that, in accordance with SLM's main claim that L2 learners have less difficulty with a new sound than a similar TL sound, a rise-fall in an unstressed syllable (i.e., the TL pitch contour that does not have a counterpart in the L1, and therefore a new contour) was produced more accurately than a rise-fall in a stressed syllable (i.e., one which is similar to an existing L1 pitch contour).

PAM proposes that similarities in articulatory gestures between TL sounds and L1 sounds determine listeners' perceptual assimilation of the TL phones to L1 categories. It predicts that two TL speech sounds that can be assimilated equally well to a single L1 phoneme will be difficult to distinguish and hence difficult to acquire, whereas two TL sounds that can be assimilated into separate L1 phonemes or TL sounds that are unlike any L1 sounds are easy to acquire. So and Best (2010) extended PAM to the identification of L2 Mandarin lexical tones by speakers of Cantonese, Japanese and Canadian English. Both Mandarin and Cantonese are lexical tone languages: Variations in pitch indicate different word meanings. While Mandarin has four main tones, Cantonese has six distinct tones, and some of these tones are phonetically similar. For instance, both Mandarin and Cantonese have a high level tone (Tone 1), a rising tone (Tone 2), and a tone with an initial fall followed by a rise (Tone 3). Japanese is a pitch-accent language in which accentuation is a lexical property, and tonal patterns are determined based on presence vs. absence of pitch accent as well as the location of the accented mora. So and Best propose that Mandarin tones and Japanese pitch accent are both phonemic in nature, indicating lexical contrasts. In addition, Japanese LH (rise) and HL (fall) sequences are phonetically similar to Mandarin mid rising tone (Tone 2) and high falling tone (Tone 4), respectively. English is a stress-accent language: The role of pitch at the lexical level is very limited (e.g., pitch difference in the accented syllable in homophones such as *SUBject* (noun) vs. *subJECT* (verb)). At the post-lexical level, pitch is used primarily to indicate grammatical and discourse functions such as rising pitch for questions vs. falling pitch for statements and high pitch for focused words/phrases. It was predicted that Cantonese speakers would have difficulty distinguishing two Mandarin tones when both are phonetically similar to a single Cantonese tone, and that Japanese speakers would be better than English speakers at distinguishing rising and falling tones. Overall, the Cantonese and Japanese groups were able to identify Mandarin tones better than the English group, possibly because of the tone (Cantonese and Japanese) vs. non-tone (English) language difference. In addition, further analyses of language-specific errors revealed assimilation of non-native Mandarin tones to Cantonese speakers' native prosodic categories. The authors concluded that the phonological system of listeners' native language constrains L2 perception of tones, consistent with PAM.

PAM has also been applied to L2 discourse prosody. In Gili Fivela (2012), L1-Italian low-intermediate learners' perception of English marking of broad focus and corrective focus was tested in various tasks, including a discrimination task and a category goodness-rating task. In English, corrective focus has higher pitch peak than broad focus while they do not differ in pitch peak alignment, whereas in Italian, corrective focus is realized with lower pitch without a peak delay compared to broad focus. It was hypothesized that Italian speakers would assimilate English broad focus and corrective focus pitch accents into the Italian corrective focus category because Italian corrective focus is similar to English broad and corrective focus with respect to pitch peak alignment. Furthermore, the author predicted that if linguistic functions are taken into consideration, English broad focus would be more difficult to discriminate from Italian corrective focus since they are phonetically similar but convey different focus meanings. The results showed that as predicted, the Italian learners had most difficulties in acquiring English broad focus pitch accent due to its similarity to Italian corrective focus pitch accent.

Although SLM and PAM provide an essential framework for predicting and interpreting learner performance for the L2 acquisition of segments, they may not be suitable for TL suprasegmentals, especially discourse prosody, because of the complexity of prosodic sound systems and the multifaceted functions of prosody in context. Prosodic features such as H and L tones are not always easy to define and categorize due to speaker and context variability, and also prosody has been described and analyzed using different frameworks, which makes it hard to identify cross-linguistic similarities and differences of prosodic features (Jun, 2005). In addition, since prosody signals both linguistic (e.g., word segmentation, syntactic structure, discourse structure, focus and information status) and paralinguistic information (e.g., intentions, attitudes, and emotions), there is never a strict one-to-one relationship between a particular prosodic feature and its function/meaning. Both Mennen (1999) and Gili Fivela (2012) assumed that a TL pitch contour that is phonetically similar to but functionally different from an existing L1 contour is most difficult to acquire, but neither SLM nor PAM gives a clear indication as to how learners come to acquire the mappings between intonation and functions.

Rasier and Hiligsmann (2009) used the Markedness Differential Hypothesis (Eckman, 1977) to explore prosodic signals of information status by both L1 and L2 speakers of Dutch and English. The Markedness Differential Hypothesis asserts that in the areas where learners' L1 and TL differ, marked structures are more difficult to learn than the corresponding unmarked structures (Eckman, 2008). Drawing on Valluduví's (1991) typological distinction between plastic and non-plastic languages, Rasier and Hiligsmann describe Dutch as a plastic language where word order is relatively fixed and prosodic structure is rather flexible, and French as a non-plastic language where word order can change more freely but variability in accent distribution is more restricted. Furthermore, they make an additional point that in many languages, accent placement at the utterance level relies on both structural information and pragmatic information (although each language has a preference for one over the other), and there is no language that completely lacks structural constraints on accent distribution, characterized by a perfect correlation between pragmatic information and accentual patterns. Based on this observation, Rasier and Hiligsmann hypothesize that pragmatic accentuation rules are more marked than structural accentuation rules, and therefore acquiring the pragmatically determined accentuation rules in Dutch is more difficult than acquiring the structurally determined accentuation rules in French. In other words, L1-French L2-Dutch learners will have more difficulty with Dutch accentuation than L1-Dutch L2-French learners do with French accentuation rules. Using a picture description task similar to Swerts, Krahmer, and Avesani (2002), the authors asked participants to describe different geometrical shapes appearing in different colors (e.g., *red circle* → *blue triangle*) on a computer screen. The shapes and colors were systematically varied in order to elicit new/new information (e.g., beginning of the task → *blue triangle*), given/contrastive information (e.g., *red circle* → *red triangle*), contrastive/given information (e.g., *red circle* → *blue circle*), and contrastive/contrastive information (e.g., *red circle* → *blue triangle*). There is no mention of the learners' L2 proficiency in the study.

The results showed some discrepancies between L2-Dutch learners' accentual patterns and those of Dutch native speakers. One of the main differences was that while Dutch natives deaccented given information, learners tended to accent both the adjective and the noun in all experimental conditions. L2-French learners, on the other hand, performed similarly to French native speakers. These findings suggest that it is easier for speakers of a plastic language (e.g., Dutch) to acquire the accentuation rules of a non-plastic language (e.g., French) than the other way round, supporting Eckman's claim that there is a direct link between the degree of markedness and the degree of learning difficulty. The methodological approach that Rasier and Hiligsmann adopt may suffice if the languages being investigated can be clearly identified as either plastic or non-plastic as in the case of Germanic and Romance languages such as English, Dutch, French, Italian, and Catalan. Vallduví's classification may not be applicable to other languages like Japanese, in which word order is highly flexible, yet intonational prominence does not always fall on a particular location since phrase-level accentual patterns are by and large determined by lexical pitch accents.

Klassen's (2013) study is framed within the Interface Hypothesis (Sorace, 2011; Sorace & Filiaci, 2006; Sorace & Serratrice, 2009). The Interface Hypothesis makes a distinction between grammar internal interfaces and grammar external interfaces. Grammar internal interfaces are those within the linguistic system such as a syntax-semantics interface or a syntax-phonology interface. Grammar external interfaces are those between the linguistic system and cognitive domains, such as a syntax-discourse interface. The central claim of the hypothesis is that L2 learners have greater difficulties in acquiring linguistic properties pertaining to grammar external interfaces than grammar internal interfaces. Klassen investigated whether L1-Spanish L2 intermediate learners of English experience difficulty in mastering target-like prosodic focus, given the cross-linguistic difference that English marks focus prosodically while Spanish focus marking is realized via word order change. In a perception task, participants listened to question-answer dialogues (e.g., Q: *Who met David?* – A: *LISA met David.* vs. A: *Lisa met DAVID.*) and rated the intonational naturalness of the answer. The results showed that the L2 learners as a whole performed like the native control, giving high ratings for correct focus and low ratings for incorrect focus, which provides evidence against the Interface Hypothesis. Furthermore, when the data were analyzed by proficiency (measured via an independent proficiency test), the high intermediate group showed a significant difference between the focus-correct and focus-incorrect conditions, while the low intermediate group did not. These findings confirm the effect of proficiency on the L2 interpretation of focus prosody.

The Interface Hypothesis is also adopted in Yoshimura, Fujimori, and Shirahata (2015), which examines whether L1-Japanese learners face problems in producing and comprehending prosodic focus marking in English using three tasks: a focus identification task, a prominence identification task, and a production task. Participants in the advanced group were third-year

English majors and those in the novice group were first-year science majors at a Japanese university. In the focus identification task, participants were given written question-answer dialogues such as Q: *Where did you go last Sunday?* – A: *I went fishing with my friend in the river*, and marked a single word in the answer that they thought should be emphasized. The prominence identification task was a listening version of the focus identification task: Participants heard question-answer pairs and identified the most prominent word in the answer. In these tasks, Japanese EFL learners, both advanced and novice, could correctly identify focused elements. In the read-aloud production task, however, the learners had difficulty in placing prosodic prominence on the focused words regardless of their proficiency. Based on the results, they conclude that the difficulty at the prosody-discourse interface lingers in production but it can be overcome in comprehension. Although the authors attribute the production difficulty to L1 prosodic transfer, it could also be due to lack of experience in producing English as it is a common issue among EFL learners.

In these two studies described above, the Interface Hypothesis seems insufficient to account for the observed data, and thus may be less applicable to the L2 acquisition of focus prosody. Rather than relying on existing L2 acquisition theories, Baker (2010) proposes a new framework specific to L2 prosodic marking of discourse structure based on the results of her own study testing English broad and narrow focus marking by L1-Korean and L1-Mandarin learners. She claims that the L2 acquisition of prosodic focus marking can be explained by a combination of the Transfer Model and the L2 Challenge Model. The Transfer Model states that when learners' L1 and TL both mark focus prosodically, that helps them acquire focus marking in the TL even if there are some differences between the L1 and TL in how exactly the focus prosody is realized. This was supported by the results from a prominence perception task, where participants heard question-answer pairs and identified the most prominent word in the answer, and from a prominence placement task, where participants were given written question-answer dialogues and identified the most appropriate pitch accent location in the answer. In both tasks, Korean speakers outperformed Chinese speakers because, according to Baker, Korean uses pitch at the post-lexical level for marking broad and narrow focus as in English, whereas in Chinese pitch is primarily a lexical property.[6] The second model, the L2 Challenge Model, holds that regardless of learners'

[6]Note that while pitch in Mandarin primarily indicates lexical meaning, new and corrective focus can be signaled by expanded pitch range followed by post-focal pitch range compression as described in He et al. (2001).

proficiency, some types of pitch accents are easier to perceive and produce than the others, and this also depends on task difficulty. In hard tasks such as a question-answer production task, in which participants had to handle both segmental and suprasegmental aspects of speech at the same time, both Korean and Chinese speakers found broad focus marking easier than narrow focus marking, possibly because the former is more frequent. In easy tasks such as a prosody naturalness judgment task where learners' attention was directed to prosody and they made conscious decisions about the naturalness of intonation without time constraints, narrow focus marking was easier than broad focus because of the direct relationship between accent location and the focused word. Though this new framework certainly achieves important progress in the area of L2 discourse prosody, Baker does not lay out clear and detailed explanations as to what constitutes "easy" or "hard" tasks for learners, which makes it difficult for other researchers to apply the models to further studies. In addition, the author does not provide the underlying acquisition/processing mechanism that allows the frequency factor to be more influential than the accent-focus relationship factor in "hard" tasks, and vice versa in "easy" tasks.

Another theory, which has been developed recently to account for learners' abilities or inabilities to attain native-like discourse prosody, is the L2 Intonation Learning theory (LILt) (Mennen & de Leeuw, 2014; Mennen, 2015). On the basis of cross-linguistic differences in intonation systems between learners' L1 and the TL, the theory aims to predict and explain difficulties in speech production that learners are likely to encounter in the course of L2 intonation acquisition. Built on Ladd's (1996) classification of intonational differences, the LILt proposes four dimensions in which intonation of two languages can differ:

1. Systemic dimension: the inventory and distribution of categorical phonological elements
2. Realizational dimension: the phonetic implementation of these categorical elements
3. Semantic dimension: the functionality of the categorical elements or tunes
4. Frequency dimension: the frequency of use of the categorical elements

(Mennen, 2015, p. 173)

The systematic dimension concerns differences in the inventory of tonal categories between languages, including differences in types of pitch accents, accentual phrases, and boundary tones. Language-specific differences in possible combinations/sequences of tonal categories are also included in this dimension. Mennen uses an example from Ladd (1996) that a variety of English spoken in Belfast and Glasgow has rising intonation L*HL% on ordinary statements, while this intonation pattern does not occur in statements in North American English. The realizational dimension refers to the ways in which the same tonal category is realized phonetically in two languages, such as precise alignment of pitch accents with the segmental string (tonal alignment), peak height and dip depth (scaling), timing of rise or fall onset (steepness), etc. For example, Dutch and Greek share the phonologically identical, prenuclear rising intonation LH* for declarative sentences, but they differ in the exact timing of rise: The tonal peak appears earlier in Dutch than in Greek (Mennen, 2004). The semantic dimension is concerned with differences in meaning or function of identical tonal patters or differences in tonal patterns for indicating the same meaning/function in between two languages. For instance, a high-rising intonation conveys questions in most varieties of English, while the same tune signals statements in Belfast English. Use of different intonational patterns for focus marking is also included in this dimension. Frequency dimension, which is not part of Ladd's original classification but was added to the LILt by Mennen, has to do with how frequently tonal categories and patterns in the inventory of a language are used. For instance, Mennen mentions that rising intonation is more frequent among female speakers of Northern Standard German than female speakers of Southern Standard British English (Mennen, Schaeffler, & Docherty, 2012).

As Mennen proposes these four intonational dimensions in the LILt, she incorporates some of the theoretical assumptions underlying SLM and PAM L2. First, SLM and PAM L2 assume that TL phones are automatically perceived in relation to learners' existing L1 phonological categories. Though this assumption seems to apply to TL intonation as evidenced in previous studies showing L2 learners' difficulty in perceiving TL tonal patterns that do not exactly map onto any L1 category, Mennen warns that identifying perceptual differences and similarities of intonation is far more challenging because of various forms and functions it can take. Therefore, it is crucial to take into consideration the semantic dimension when it comes to determining perceptual differences/similarities between L1 and TL intonation. Secondly, the LILt takes on the assumption in SLM and PAM L2 that L1 influence on TL speech segmental perception occurs both at the phonological and phonetic levels and that phonetic details of L1 and TL categories are important in TL speech perception. For this reason, the LILt posits that both the systematic and realization dimensions may have an effect on perception and production of TL tonal categories. Third, SLM and PAM L2 consider age of onset to be one of the determining factors for successful L2 acquisition of sounds. Mennen, too, hypothesizes that the age at which an L2 learner is first exposed to the TL in a naturalistic environment can be a reliable predictor of success in TL intonation learning, even though she notes that age of onset may have a variable degree of effects in each of the four dimensions of intonation. Fourth, SLM and PAM L2 assume that perceptual mechanisms are intact and retained throughout the lifespan and, therefore, adult L2 learners continue to learn and adjust their perception as they gain more experience in the TL. Following this assumption, the LILt holds that L2 production of intonation may initially reflect L1 traces due to limited exposure to the TL. However, learners' production becomes fine-tuned gradually to approach the native norm. Lastly, SLM and PAM L2 maintain that L1 and L2 phonological categories co-exist in a shared perceptual space, causing bi-directional interaction between the two languages. The LILt agrees with SLM and PAM L2 and posits that the same mechanism applies to L2 intonation, drawing on the previous research findings that L2 learners' performance is often positioned somewhere in between their L1 and the TL (Mennen, Schaeffler, & Dickie, 2014), and that L2 development can cause reverse changes in their L1 production (de Leeuw, Mennen, & Scobbie, 2012).

To summarize, the LILt is a useful framework for identifying cross-language differences in intonation and predicting non-native deviations from the native norm. It allows us to compare intonation of two languages systematically and to incorporate factors such as L1 background, learning experience, age of onset, and proficiency levels into research on the L2 acquisition of intonation. For this reason, we framed our study within the LILt to examine L1-Japanese learners' acquisition and processing of English prosodic marking of Information Structure. The next chapter discusses our research questions and using the LILt as a guide, we identify cross-linguistic differences between English and Japanese discourse prosody and make predictions about whether and how L1-Japanese speakers come to acquire prosodic marking of new, contrastive, and given information status in English.

CHAPTER 4
RESEARCH QUESTIONS AND OVERALL EXPERIMENTAL DESIGN

4.1 Research questions

Acquisition of prosody is a relatively new topic in L2 research; especially prosodic marking of Information Structure has largely been underexplored. As research interests in the field of L2 acquisition expand from the analysis of linguistic structure such as sounds, words, and sentences to more context-oriented approaches like discourse and pragmatics, prosody has gained recognition as an essential part of spoken language and human communication. Prosody is abundant in language input: It exists in virtually every utterance we hear, and children master adult-like use of intonation effortlessly and unconsciously in the course of language development. Despite this ease, the L2 literature reviewed in Chapter 3 clearly indicates that establishing a reliable and meaningful relationship between particular prosodic patterns and discourse meaning imposes a great challenge to adult L2 learners. One reason could be the multifunctional properties of prosody, serving as cues for both linguistic (lexical stress, structural boundaries, illocutionary force, discourse structure) and nonlinguistic (emotions, attitudes) information. During language comprehension, learners are exposed to various intonation patters signaling a wide range of information all at the same time, which makes it extremely difficult to pinpoint the function or meaning that a particular tonal contour conveys. Another reason for the difficulty could be that many languages exploit not only prosody but also lexical, morphological, or syntactic means to mark focus, so in language production, learners may be inclined to rely on those other means in order to express focus to the listener in a more transparent way. Moreover, discourse intonation is an unlikely topic to be dealt with in language classes. While most teachers acknowledge that intonation is crucial in listening and speaking, they tend to avoid teaching it explicitly in the classroom perhaps because they are unsure how to apply theoretical accounts of intonation to actual language teaching, or because teaching materials that focus on intonation components, especially discourse-level functions of prosody, are still sparse (Chun, 2002). Thus, the complex nature of discourse prosody combined with lack of explicit instructions in the classroom presents a real learning problem in L2 acquisition. This suggests that understanding the learning difficulty in L2 prosody is a significant issue not only to acquisition researchers but also to language learners, teachers, and other practitioners such as materials writers.

To gain a better understanding of this issue, the current study investigates whether and, if so, how L2 learners of English come to acquire the relationship between a part of discourse and prosody in the TL. In particular, it focuses on specific types of English accentuation choices: deaccentuation, H*, and L+H*, and their discourse functions, i.e., given, new, and contrastive focus marking. We chose Japanese as the L1 of the learner group for three reasons. First, the intonational system of Japanese and its prosodic annotation system (J_ToBI) are well-documented, allowing us to make a systematic comparison between English and Japanese intonation structures. Second, as described in Chapter 2, English uses pitch primarily for conveying pragmatics/discourse meaning such as information status, whereas Japanese uses pitch mainly for lexical semantics and secondarily for indicating contrast. On the grounds that English and Japanese are somewhat similar with respect to prosodic marking of contrastive information but different in the ways they mark given vs. new information, we can examine whether L2 learners transfer the mapping that exists in the L1, and subsequently acquire new prosody-discourse mappings in the TL that are not instantiated in their L1. Third, L2 leaners are reported to transfer their L1 prosody-discourse mappings to L2 processing when pitch accents are post-lexical properties in both languages (e.g., L1-German L2-Dutch; Braun and Tagliapietra, 2011). By comparing English and Japanese, it will become clear whether the acquisition of target-like prosody-discourse associations is even more difficult when transitioning from L1 lexical (Japanese) to L2 post-lexical (English) use of pitch accents.

To this end, we address the following research questions:

1. Can L2 learners associate a particular TL pitch accent pattern with an appropriate discourse meaning when focusing on intonation?

2. Can L2 learners make immediate use of TL pitch accent cues to identify discourse referents during real-time comprehension?

3. Can L2 learners employ TL pitch accent cues to express information status of discourse referents in naturalistic production?

The first part of research question (1) is concerned with the perception of acoustic cues for (de)accentuation. Realization of English focus is encoded with a wider pitch range, prolonged duration, increased intensity, and hyper-articulation of the stressed syllable within the focused element. Therefore, English speakers are sensitive to these phonetic cues when processing sentences. In Japanese, pitch range expansion, insertion of a pre-focus boundary, and post-focus pitch suppressions are among the properties of prosodic focus (Venditti et al., 2008), but unlike in English, duration of the syllable/word in the focused element does not signal accentuation. Due to this cross-linguistic difference in focus marking, Japanese-speaking learners of English may not be sensitive to the durational factor contributing to the perception of accentuation when listening to English. Though the L2 Intonation Learning theory does not make a specific prediction about L2 acquisition of duration, it takes on one of the assumptions underlying the Speech Learning Model and the Perceptual Assimilation Model that the perception of L2 speech is filtered, at least initially, through the automatic L1 perceptual strategies. If L2 learners indeed transfer their L1 speech processing strategies, as the LILt holds, they will experience difficulties in perceiving changes in duration but not in pitch at the early stage of L2 acquisition. This prediction was tested in the current study via a prosody-in-context naturalness rating task that manipulates both pitch height and durational value of the target word.

The second part of research question (1) has to do with the association between prosody and discourse meaning. While L2 learners learn to exploit both pitch and durational cues for perceiving deaccentuation, H*, and L+H*, they also need to acquire the target mappings between various prosodic features and their functions, in this case, deaccentuation, H*, and L+H* with, respectively, given, new, and contrastive information marking. The prosody-in-context naturalness rating task investigates whether our L1-Japanese L2 participants are able to do so when their attention is drawn specifically to intonation in order to make conscious judgments about the appropriateness of a certain intonational pattern presented in various discourse contexts.

Research question (2) further asks if L2 learners do possess knowledge of the target-like prosody-discourse mappings, are they able to exercise such knowledge in real time, while they focus on comprehending utterances for meaning? As seen in Chapter 3, Baker's (2010) L2 Challenge Model holds that learners' perception and production of pitch accents is partially affected by task difficulty. By conducting two listening experiments that involve different levels of time pressure and of listeners' attention to prosody, we aim to examine if task type influences L2 learners' interpretation of prosodic cues to discourse structure.

Research question (3) asks whether L2 learners employ prosodic means to indicate Information Structure in a semi-spontaneous production task. We are interested to see whether L2 learners' ability to mark a particular information structure with appropriate pitch accents depends on their ability to accurately perceive TL pitch accents. To this end, we tested both comprehension and production tasks with the same group of native English controls and L2 learners.

On the basis of the LILt, we predict that it will be relatively easy for Japanese learners of English to associate L+H* with contrastive information. Even though the L+H* pitch accent does not exist in the inventory of the Japanese intonation system, English L+H* and pitch range expansion in Japanese contrastive marking share phonetically similar characteristics: They both start off low and show a steep rise to the target H followed by a sudden fall. As for the frequency dimension, Hedberg and Sosa's (2008) intonationally annotated corpora of English spontaneous speech revealed that about 50% of L+H* pitch accents indicated contrastive information in naturally occurring utterances, and Speer and Ito's (2011) production experiment showed that contrastive information was most likely to be produced with L+H* (58% of contrastive cases was marked with L+H* and 29% with H*). These previous findings imply that TL input provides at least some indication that L+H* often signals contrast and that contrast in many cases is expressed via L+H* marking.

We predict the English link between deaccentuation and given information to be rather challenging for Japanese speakers. English deaccentuation is determined at the post-lexical level reflecting pragmatic and discourse functions and has no fixed pitch contour associated with it: An intonation contour on a deaccented word is determined via linear pitch interpolation connecting the preceding and following pitch accent targets. In Japanese, on the other hand, deaccentuation is determined lexically, with no associated pragmatic/discourse meaning, and it is realized as H- at the phrase level. English deaccentuation and Japanese H- are therefore quite different in systematic, realizational, and semantic dimensions. As for frequency, Hedberg and Sosa report that 76% of all the instances of deaccentuation indicated given information, while Terken and Hirschberg (1994) and Sityaev (2000) found that given information was accented frequently (86% of given referents in Sityaev, 2000). In other words, deaccentuation is a fairly reliable cue for given information, whereas given information is not necessarily realized with deaccentuation. Due to this asymmetry, learners are likely to have a hard time in detecting the association between deaccentuation and given information, but with ample exposure to the TL, they may be able to learn that deaccentuation signals given information.

Similar to the deaccentuation–given mapping, the English H*–new association may also be difficult to acquire for L2 learners. Japanese has no pitch accent corresponding to English H*, and even though the pitch contour of H* may look somewhat similar to that of H- in Japanese (shallow rise to high plateau), they carry distinct functions: English H* is assigned at the post-lexical level for pragmatic/discourse functions, whereas Japanese H- is assigned to an unaccented phrase determined lexically. According to Hedberg and Sosa, H* is most frequently used in spontaneous speech compared to other types of pitch accents, but only 26% of H* pitch accents signal new information. However, new information is almost always accented (95% of new referents in Sityaev, 2000). Thus, H* is an unreliable cue for new information, but new information in most cases bears some sort of prosodic prominence such as elevated pitch height. The asymmetry may again pose a great difficulty for learners in realizing the link between H* (or accentuation) and new discourse status in English, but it is possible that ample language exposure would allow them to realize that new information should be emphasized prosodically.

Based on the assumptions underlying the LILt, to the extent that L2 learners are able to establish a link between particular pitch patterns and meanings in the process of learning, they are expected to show more target-like performance as proficiency increases. That is to say, low proficiency learners are more likely to rely on their L1 focus marking strategies than advanced learners, so the former should be able to link L+H* to contrastive information due to L1 transfer but not H* to new information nor deaccentuation to given information. With more and more exposure to the TL, advanced learners may be able to acquire new mappings that do not exist in their L1, namely the deaccentuation–given mapping and the H*–new mapping, based on the prosodic correlates of Information Structure available in the language input.

Our predictions based on the LILt are summarized below:

1. The L+H*–contrastive association will be easiest to master for L1-Japanese learners of English due to L1 transfer.

2. The associations between H* and new information and between deaccentuation and given information will be challenging for Japanese speakers.

3. L2 learners are expected to acquire a new link between a particular pitch pattern and its discourse meaning in the course of learning, showing more target-like performance as their proficiency increases.

These predictions were tested via three main experiments in the present study. Since all tasks are simple and do not require advanced lexical or syntactic knowledge, the study was conducted across L2 proficiency, providing us with a detailed picture of any early L1 prosodic transfer as well as gradual L2 development in the establishment of the TL relationships between prosodic marking and Information Structure as proficiency increases. The next section outlines the experimental design; it also provides proficiency and demographic information about our participants.

4.2 Experimental design

In the present study, 70 English native speakers and 64 L1-Japanese L2-English speakers performed all three main tasks conducted in English (a prosody-in-context naturalness rating task, an eye-tracking listening comprehension task, and a semi-spontaneous production task) during their one-time, 1-hour visit to the Language Analysis and Experimentation Labs at the University of Hawai‘i. In addition, participants completed two English proficiency tests, a C-test (Appendix A) and a read-aloud task (Appendix B), as well as a background questionnaire (Appendix C). These tasks were administered in the following order:

1. Production task
2. Background questionnaire
3. Eye-tracking listening comprehension task (Block 1)
4. C-test
5. Eye-tracking listening comprehension task (Block 2)
6. Read-aloud task
7. Prosody-in-context naturalness rating task

A C-test is a fill-in-the-blanks test similar to a cloze test.[7] Although both C-tests and cloze tests have been developed to assess learners' overall language proficiency, especially lexical and morpho-syntactic competence, we chose to administer a C-test because it is less time-consuming for participants, and scoring is easy and objective as there is, in most cases, only one acceptable answer for each blank. The particular version of the C-test used in our study (see Appendix A) was adopted from Schulz (2006). The maximum possible score was 40. Each C-test score was multiplied by 1.25 and then divided by 10 in order to convert the raw scores to a 5.0 scale.

The intended purpose of implementing a read-aloud task was to measure participants' phonological competence. Participants were audio-recorded as they read a short paragraph (Appendix B) taken from the *Speech Accent Archive* (Weinberger, 2015), an online database containing a large number of speech samples by a variety of speakers. The paragraph was constructed in such a way that it contained all of the vowel and consonant sounds of English. Participants were instructed to read aloud the paragraph twice, and the recordings from the second reading were independently rated by two trained native speakers of English on a 5-point scale (1 = strong foreign accent, 2 = noticeable foreign accent, 3 = mild foreign accent, 4 = very little foreign accent, 5 = no foreign accent). Since the weighted Kappa statistic confirmed a significant positive correlation between the two raters [K = .89, $p < .001$], the two ratings were averaged for each participant.

[7] The difference between a C-test and a cloze test is that in a C-test the first half of a word is provided so that a test taker does not need to restore a complete word for each blank as in a cloze test.

Next, we examined the relationship between the read-aloud ratings and the C-test scores using Kendall's non-parametric statistic.[8] Since there was a significant positive correlation between read-aloud and C-test scores [τ = .66, $p < .001$], we decided to include both variables into participants' language proficiency measure. Each participant's total proficiency score was calculated by adding their converted C-test score to read-aloud rating so that out of 10 possible points, half came from the C-test and half from the read-aloud task.

Based on the total scores, the L2 learners were split at the median of 4.5 to form an upper proficiency group (n = 30) and a lower proficiency group (n = 34) in such a way that there were approximately the same number of participants in each proficiency group. Mean score, range, and standard deviation of the C-test, read-aloud task, and total proficiency scores are listed for each group in Table 4.1. The mean proficiency score for the L1 English group was 9.11 out of 10, and the mean score for the L2 group as a whole was 4.93 with a wide range from 1.75 to 8.88. When the L2 participants were divided into two proficiency groups, the mean score for the upper group was 6.16, while the mean score for the lower group was 3.85.

Table 4.1: Participant groups and their proficiency scores

Participant group	C-test (max = 40)			Read-aloud (max = 5)			Total (max = 10)		
	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD
70 L1 English	34.01	25-39	3.04	4.94	4-5	0.21	9.19	7.5-9.88	0.51
64 L1-Japanese L2 learners	22.28	2-37	6.46	2.15	1-4.5	0.85	4.93	1.75-8.88	1.44
30 Upper	26.87	15-37	4.90	2.80	1.5-4.5	0.78	6.16	4.88-8.88	1.07
34 Lower	18.24	2-27	5.22	1.57	1-2.5	0.33	3.85	1.75-4.5	0.60

Note. Mean C-test scores converted to the 5-point scale: L1 English–4.25, L2 upper–3.36, L2 lower–2.28

Table 4.2 summarizes the L2 learners' language learning background information, including their age at the time of the study, age of onset of learning English, and length of residence in an English-speaking country. All of the L2 participants were born in Japan and grew up in a family where Japanese was the only language spoken. Though some L2 learners started weekly English lessons in Japan at age six, none of them had a substantial amount of exposure to English in an English-speaking environment until at least age eleven. Although the mean age for the upper group (28.20) was approximately four years older than that for the lower group (24.68), the two

[8]We chose Kendall's non-parametric statistic over Pearson's or Spearman's correlation because read-aloud ratings were ordinal rather than interval data, and many rating scores had the same rank.

groups were comparable with respect to the age of onset of English learning (11.23 for the upper group vs. 12.03 for the lower group). For length of residency, the L2 learners in the upper group had lived in an English-speaking country more than three years on average, whereas the amount of time for those in the lower group was about one and a half years, which suggests our advanced L2 learners had much more naturalistic exposure to English than the less advanced learners.

Table 4.2: L2 learners' demographic information

L2 group	Age			Age of onset			Length of residency		
	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD
L2 upper & lower groups	26.33	18-56	8.64	11.66	6-14	1.83	2.46	0-14	3.53
Upper group	28.20	18-56	9.28	11.23	6-14	2.08	3.35	0-14	4.29
Lower group	24.68	19-53	7.80	12.03	7-14	1.51	1.67	0-9	2.49

In the following chapters, we present each of the three main experiments in turn. After describing the experimental methods and predictions, results for the L1 English group are presented first, followed by L2 results as a whole group. Then the results analyzed by L2 proficiency are discussed in each chapter.

CHAPTER 5
EXPERIMENT 1: PROSODY-IN-CONTEXT NATURALNESS RATING TASK

5.1 Introduction

The purpose of Experiment 1 is to examine whether native and non-native English listeners are sensitive to pitch and durational cues that help distinguish words produced without a pitch accent (hereafter: null accent), with H*, and with L+H*, and whether they can associate each type of pitch accent to its appropriate pragmatic meaning. The prosody-in-context naturalness rating task utilizes question-answer pairs: Participants listened to a pre-recorded answer in different prosodic patterns preceded by different types of questions. For each question-answer pair, they rated the prosodic naturalness of the target answer given the question. Prosodic manipulation was conducted using speech synthesis techniques as precise control over pitch and duration was crucial in this experiment.

Since accentual prominence in Japanese is realized via pitch height rather than duration whereas English accentuation involves both pitch and duration, if L2 learners indeed impose their L1 speech perception strategies onto TL prosodic perception, we would expect sensitivity to pitch changes but underuse of durational cues by Japanese-speaking learners of English. With respect to the mappings between pitch accents and discourse, we predict that based on the L2 Intonation Learning theory, the association between English L+H* and contrastive information would be easiest for Japanese speakers since contrastive information can be marked by enhanced pitch range in both languages. The mappings between null accent and given information and between H* and new information are predicted be more challenging because accentuation signals lexical distinctions rather than information status in Japanese.

5.2 Methods

5.2.1 Participants

Seventy native speakers of English and 64 Japanese-speaking L2 learners of English with no visual or hearing impairment were recruited from the University of Hawai‘i community for course credit or small monetary compensation.

5.2.2 Audio stimuli

The target sentence used in this experiment is *Mariana made the banana bread*. The sentence was chosen because it contains no difficult vocabulary for L2 learners, the sentence structure is simple, and it mostly consists of sonorants. It is crucial to avoid non-sonorant sounds so as to elicit the clearest pitch contours possible in acoustic analyses of the target stimuli.

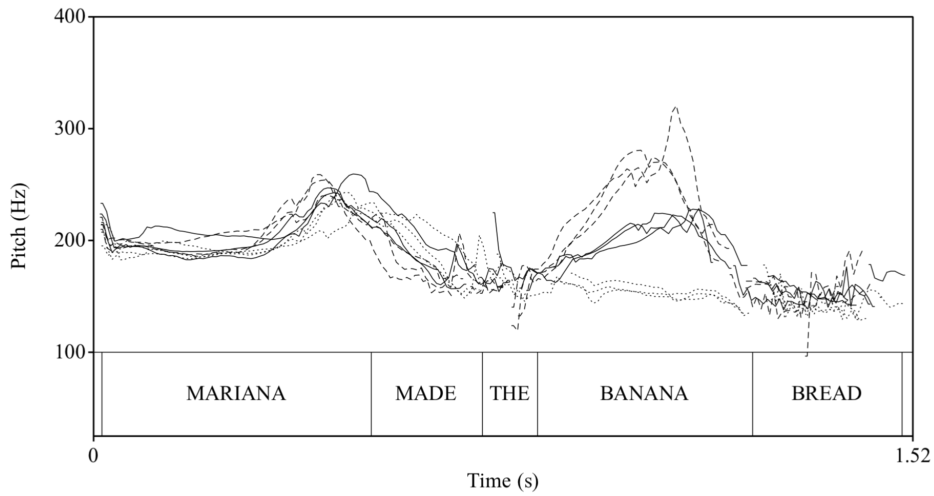
A female native speaker of English with professional training in prosody read the target sentence with null accent, H*, and L+H* on *banana* three times each (Table 5.1).[9]

Table 5.1: English target sentence

Condition	Sentence
Null accent	Mariana made the *banana* bread.
	H* ∅ L-L%
H*	Mariana made the *banana* bread.
	H* L- H* L-L%
L+H*	Mariana made the *banana* bread.
	H* L- L+H* L-L%

[9]Note that in the ToBI transcription in Table 5.1, the phrasal boundary L- appears in the H* and L+H* conditions but not in the null accent condition even though phonetically these three conditions are similar up to the critical region *banana* as shown in the figure. This is because the ToBI transcription scheme (Veilleux, Shattuck-Hufnagel, & Brugos, 2006) specifies that each intermediate intonational phrase must contain at least one pitch accented word. This rule is violated if *made* is marked with L- in the null condition since L- creates two intermediate intonational phrases within the sentence, and the second intermediate intonational phrase will be left with no pitch accented word.

Overall pitch contours of the base recordings are shown in Figure 5.1.[10] Table 5.2 lists the means of duration, mean pitch, and maximum and minimum pitch, for each word by pitch accent types, taken from the base recordings.



Note. Null accent–dotted lines, H*–solid lines, L+H*–dashed lines

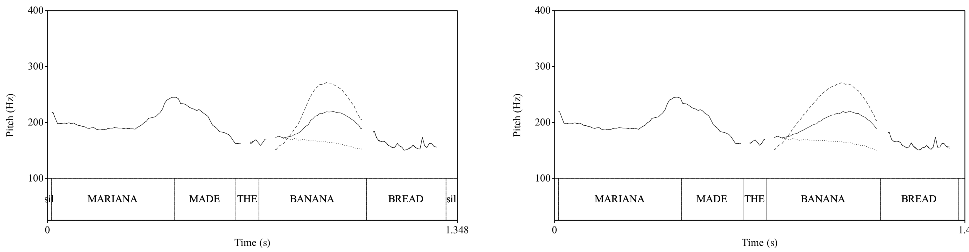
Figure 5.1: Pitch contours of the base recordings

[10]It could be argued that the L+H* contour in Figure 5.1 is an H* with expanded pitch range rather than L+H*. The L+H* pitch tracks in the figure show a slight fall during *the*, which cannot be accounted for by the preceding L-, and then a sharp rise into the prominent syllable *na* in *banana*, whereas the H* pitch tracks show a more gradual rise into prominent H* from the onset of *banana*. For this reason, we considered these base recordings to be appropriate samples for the H* and L+H* tonal conditions.

Table 5.2: Acoustic analysis of the base recordings

		Mariana	*made*	*the*	*banana*	*bread*
Duration						
(ms)	Null	449.74	203.22	120.28	366.59	240.75
	H*	455.50	203.60	95.69	394.26	275.91
	L+H*	426.80	209.90	94.64	389.67	303.34
F0 mean						
(Hz)	Null	202.12	203.28	168.41	158.00	148.11
	H*	208.50	201.57	175.34	199.49	157.69
	L+H*	211.02	194.95	167.27	224.03	160.54
F0 max						
(Hz)	Null	240.84	234.46	197.93	159.81	152.19
	H*	251.89	238.56	203.45	203.50	161.27
	L+H*	254.17	237.04	189.95	230.93	167.33
F0 min						
(Hz)	Null	187.60	158.57	152.20	156.39	141.31
	H*	193.29	166.07	162.34	195.07	151.07
	L+H*	193.47	161.29	156.95	216.29	154.59

All the experimental stimuli were derived from these base recordings. For each of the utterances recorded, the critical word *banana* was divided into 20 segments of equal duration (Isaacs & Watson, 2010). Then the mean pitch for each segment was calculated. For each segment, mean pitch values were plotted in a figure and connected by a four-term polynominal equation line that best fits the 20 pitch points. The best fitted lines for null accent, H*, and L+H* base sentences are shown in Figure 5.2 below. Experiment audio files were created using the pitch-synchronous overlap-and-add algorism (PSOLA) in Praat (Boersma & Weenink, 2016). One of the original H* utterances was used as a base for resynthesis. First, the critical word *banana* was spliced out of the original utterance. The word was divided into 20 segments, and for each segment the F0 value was replaced with a new F0 value calculated by the polynominal equations to create three pitch versions of the same word.



(a) short condition (b) long condition

Note. Null accent–dotted lines, H*–solid lines, L+H*–dashed lines

Figure 5.2: Pitch contours of the manipulated sound files

In addition to the three-level pitch values, the duration of the stressed syllable in the critical word *banana* was also manipulated to create long and short conditions (Isaacs & Watson, 2010). Duration for the long condition was determined by taking the mean duration of the H* base utterances (stressed vowel duration for the long condition = 131.30 ms). Duration of the stressed vowel for the short condition (65.65 ms) was 50% shorter than the vowel for the long condition.[11] The three pitch values were crossed with the two durational conditions, creating six stimulus types in total. The six types created by this procedure were spliced back into the original carrier utterance. Pitch and durational values of the resynthesized target stimuli are listed in Table 5.3

[11] In Isaacs and Watson (2010), the mean ratio of the word duration between the accented (long) target word and the unaccented (short) target word was 0.84. Our study manipulated duration of the stressed syllable in the critical word rather than the entire word duration because prosodic changes due to accentuation typically occur in the accented syllable. When we applied Isaacs and Watson's 1:0.84 ratio to our long and short stimuli, the durational difference was barely noticeable by native speakers of English. We therefore increased the ratio to 1:0.5, which was the point where the durational difference became noticeable while short stimuli did not sound too unnatural to native speakers.

Table 5.3: Pitch and durational values of the target stimuli

		Mariana	*made*	*the*	*banana*	*bread*
Duration						
(ms)	Short	419.44	210.51	78.78	324.76	265.02
	Long				390.41	
F0 mean						
(Hz)	Null	201.95	201.39	164.93	165.62	159.84
	H*				198.65	
	L+H*				228.23	
F0 max						
(Hz)	Null	245.32	239.24	169.60	175.50	182.64
	H*				220.27	
	L+H*				271.08	
F0 min						
(Hz)	Null	186.59	161.85	159.23	150.58	150.53
	H*				172.45	
	L+H*				151.70	

5.2.3 Procedure

In the prosody-in-context naturalness rating task, participants were presented with three pictures depicting three different contexts (Figure 5.3) on a computer screen. In each trial, participants first saw one of the pictures and heard its accompanying aural context (e.g., new context picture with *What did Mariana do?*). Then they listened to a resynthesized target stimulus (*Mariana made the banana bread* with manipulated pitch and duration), and rated the prosodic naturalness of the target stimulus on a three-point Likert scale, with 1 being "unnatural", 2 being "somewhat natural", and 3 being "completely natural". Each of the six target stimuli was presented in each of the three contexts only once, yielding 18 items in total. These 18 items were randomized for each participant in order to eliminate any trial order effects. At the beginning of the task, participants were specifically asked to pay attention to intonation and evaluate its naturalness within a given context (see Appendix D for the actual directions given to participants for this task). There was no response time limit, and no feedback of any kind was provided during the task. Participants' ratings were recorded using E-prime 2.0.

(a) Given context
Who made the banana bread?
Mariana made the banana bread.

(b) New context
What did Mariana do?
Mariana made the banana bread.

(c) Contrastive context
Did Mariana make the apple bread?
Mariana made the banana bread.

Figure 5.3: Visual display and accompanying aural contexts

5.3 Predictions

For native English speakers, we predict that in the given context, null accent will be preferred over accentuation, i.e., H* or L+H*. In the new context, if native speakers make a distinction between H* and L+H*, H* will be rated higher than null accent and L+H*. In the contrastive context, ratings for L+H* will be higher than those for null accent or H*. These predictions on pitch effects (collapsing over duration) are illustrated in Figure 5.4. If, on the other hand, native speakers do not make a clear difference between H* and L+H*, H* and L+H* in the new context and in the contrastive contest should result in similar ratings.

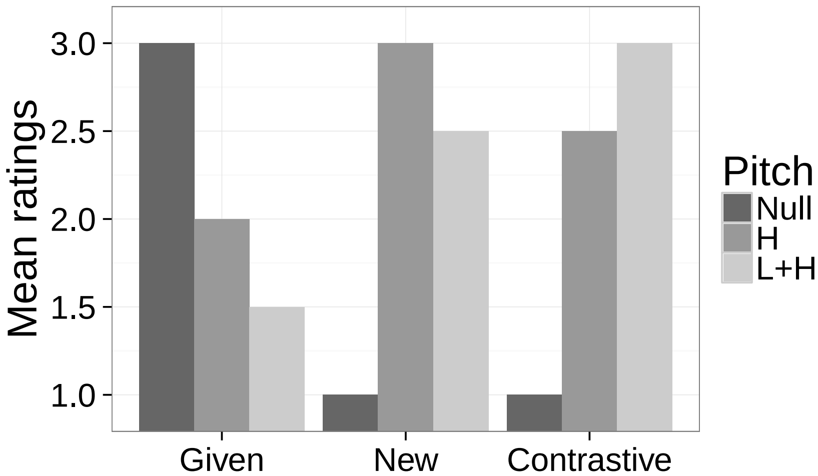


Figure 5.4: Naturalness ratings: Predicted results for pitch

For duration, if native English listeners use duration in addition to pitch as a cue to pitch accents as claimed in the previous research (Beckman & Pierrehumbert, 1986), short stimuli will be rated higher than long stimuli in the given context, whereas long stimuli will be rated higher than short stimuli in the new and contrastive contexts, as shown in Figure 5.5 (collapsing over pitch).

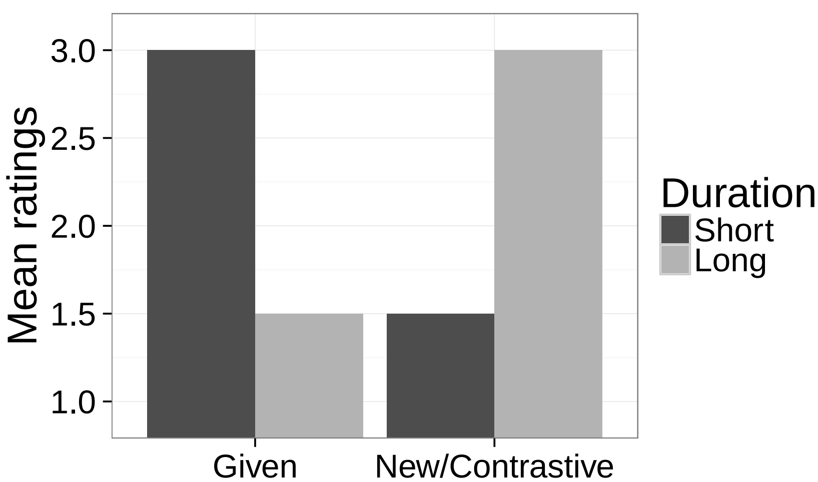


Figure 5.5: Naturalness ratings: Predicted results for duration

As for Japanese listeners, the LILt predicts that if they perceive L+H*, they will be able to evoke a contrastive interpretation since contrastiveness in Japanese can be expressed with expanded pitch range similar to English L+H*. When they hear H* or null accent, on the other hand, they may interpret it as a lexical rather than a discourse cue since the presence or absence of a pitch accent is determined at the lexical level in Japanese. If the L2 intonational meaning is driven solely by L1, we expect higher ratings for L+H* than for H* or null accent in the contrastive context. In the new context and the given context, L+H* will be rated low while H* and null accent will be rated neither low or high. As for duration, no difference between short and long conditions is expected in any context because unlike English, syllable duration does not signal word accentuation or discourse status in Japanese (Beckman & Pierrehumbert, 1986). However, as L2 proficiency increases, learners may show more target-like ratings of null accent and H* in, respectively, the given and new context conditions (high ratings for null accent in the given context and for H* in the new context) as well as durational effects (high ratings for short duration in the given context and for long duration in the new/contrastive contexts).

5.4 Results

In this section, we present data separately for pitch and for duration since no significant interaction between pitch and duration was found for either language group in data analyses.[12] For the rating graph and statistical results by discourse context, pitch, and duration, see Appendix E. Below we discuss the pitch results first, followed by the duration results, since the former shows clearer effects on listeners' intonational naturalness judgments.

Figure 5.6 shows mean naturalness ratings by pitch and discourse context (collapsing over duration) for the L1 and L2 groups with error bars representing the 95% confidence interval (CI). In the native English group, null accent was rated much higher than H* or L+H* in the given context, and L+H* was rated higher than H* or null accent in the contrastive context. In the new context, ratings for H* and L+H* were comparably high, while null accent was rated much lower than H* and L+H*. The L2 group showed similar rating patterns for the given context and the contrastive context. In the given context, null accent was higher than H* or L+H*, and in the

[12]We used the following formula with the *clmm* function in the *ordinal* package: clmm(ratings ∼ pitch * discourse context * duration + (1 | participant)). Pitch, discourse context, and duration were Helmert coded: For the pitch factor, null accent vs. H* & L+H* was coded as -2/3, 1/3, 1/3, and then H* vs. L+H* was coded as 0, -1/2, 1/2. For the discourse context factor, the given context vs. the new & contrastive contexts was coded as -2/3, 1/3, 1/3, and the new context vs. the contrastive context was coded as 0, -1/2, 1/2. Short vs. long duration was coded as -1/2, 1/2.

contrastive context, L+H* was rated higher than H* and null accent. In the new context, null accent was only slightly lower than H* and L+H* unlike for the native group.

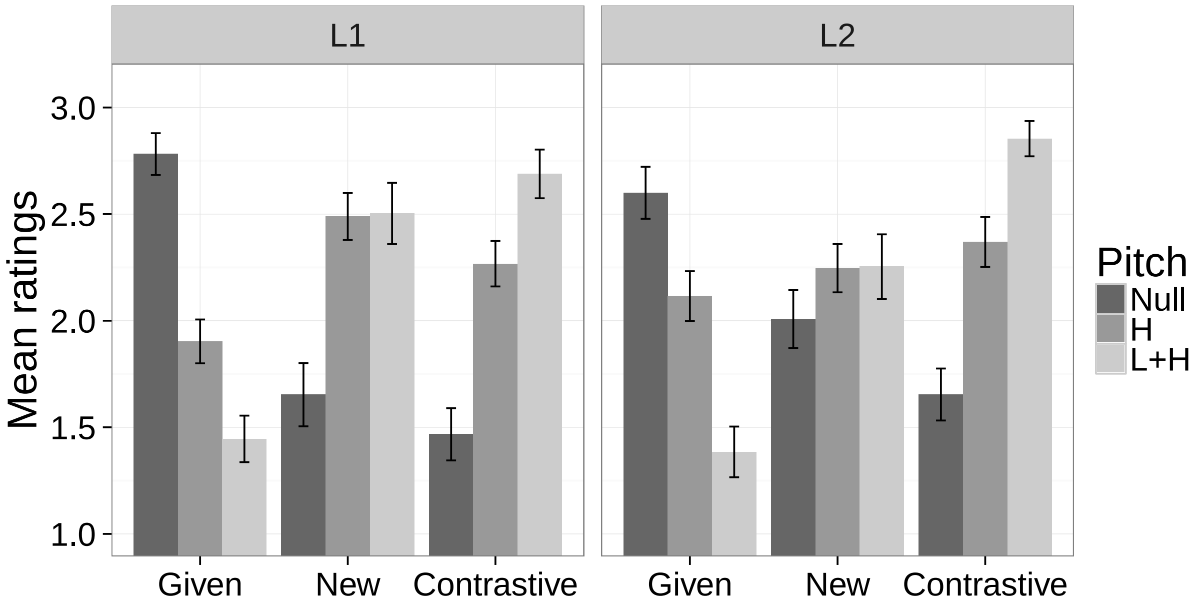


Figure 5.6: L1 and L2 naturalness ratings

Percentage breakdowns of the three response categories for each experimental condition is shown below. In Figure 5.7 and Figure 5.8, the y-axis marks the nine experimental conditions, and the x-axis indicates response proportions on a 200% scale, ranging from -1.0 to 1.0. Each stacked bar is 100% wide, and is broken down into the three response categories ("unnatural," "somewhat natural," and "very natural") according to the percentage of participants in each category. The median of the neutral category "somewhat natural" is aligned with 0 on the x-axis so that the right half of the bar indicates stronger preference, and the left half indicates dispreference in the naturalness ratings.

Figure 5.7 shows that for the L1 group, there were far more “very natural” responses for the given–null accent condition than for the given–H* or the given–L+H* conditions. For the new context, the new–H* condition and the new–L+H* condition were similar in their breakdown, while the new–null accent condition received more “unnatural” responses. In the contrastive context condition, the contrastive–L+H* condition resulted in most “very natural” responses, and the contrastive–H* condition received “very natural” or “somewhat natural” responses from the majority of the native participants. The contrastive–null accent condition was dispreferred by native speakers as reflected in the increased number of “unnatural” responses.

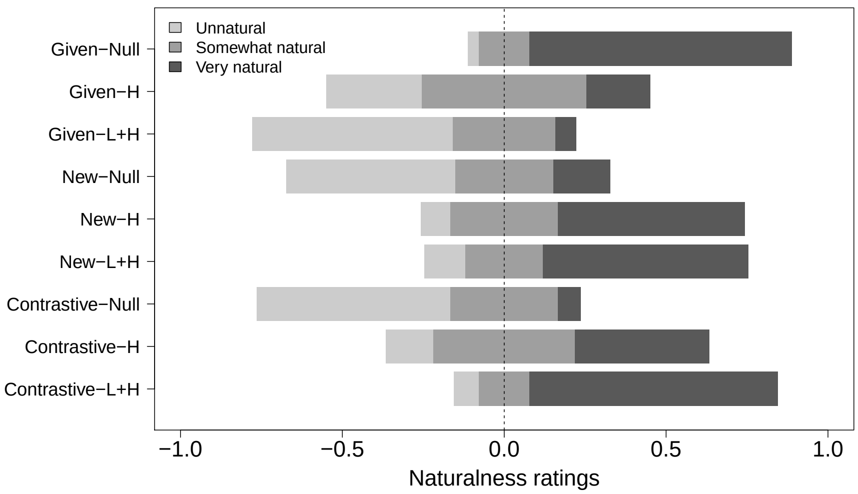


Figure 5.7: L1 response proportions for each experimental condition

As for the L2 group (Figure 5.8), the response pattern for the given context is similar to that of the L1 group: The given–null accent condition received far more “very natural” responses than the given–H* condition or the given–L+H* condition. The new context resulted in similar response patterns among the three pitch accent conditions, although the new–null accent condition received slightly fewer “very natural” responses and more “unnatural” responses than the new–H* and the new–L+H* conditions. The contrastive context condition, on the other hand, yielded far more “very natural” responses for the contrastive–L+H* condition than for the contrastive–H* and the contrastive–null accent condition.

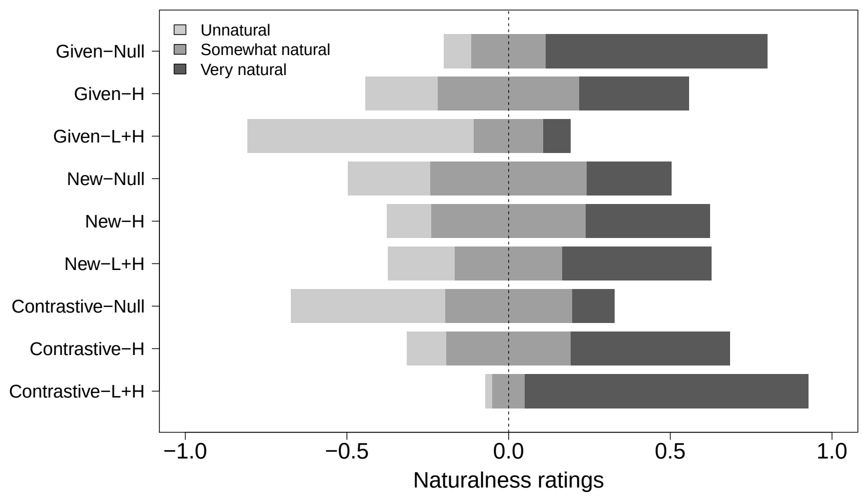


Figure 5.8: L2 response proportions for each experimental condition

Statistical analysis was performed using a cumulative link mixed model in the R package *ordinal* in order to treat the 3-point Likert scale as ordinal data.[13] The model specified the naturalness ratings as the dependent measure, pitch accent and discourse context as fixed factors, participant as a random factor on the intercept, and pitch accent, discourse context, and the interaction between pitch accent and discourse context as random participant factors on the slope.[14] As for the L1 group, there were significant pitch-by-context interaction effects (Table 5.4). Tukey's post-hoc pairwise comparisons using the *lsmeans* package in R revealed that in the given context, null accent was rated significantly higher than H* [b = 5.26, z = 5.37, $p < .001$] and than L+H* [b = 7.07, z = 6.91, $p < .001$]. In the new context, H* and L+H* were rated significantly higher than null accent [b = -3.33, z = -6.53, $p < .001$ for null accent vs. H*;

[13]We used the following formula with the *clmm* function in the R package *ordinal*: clmm(ratings ∼ pitch * discourse context + (1 + pitch * discourse context | participant)). For the pitch factor, null accent vs. H* & L+H* was coded as -2/3, 1/3, 1/3, and then H* vs. L+H* was coded as 0, -1/2, 1/2. The former coding compares the null accent condition with the accent conditions (mean of null accent vs. mean of H* & L+H* combined), and the latter coding compares H* with L+H* when null accent is ignored. Similar coding was done with the discourse context factor: The given context vs. the new & contrastive contexts was coded as -2/3, 1/3, 1/3, and the new context vs. the contrastive context was coded as 0, -1/2, 1/2.

[14]We also ran a statistical analysis including language group as a fixed factor, using the following formula: clmm(ratings ∼ pitch * discourse context * language group + (1 | participant)). For the pitch factor, null accent vs. H* & L+H* was coded as -2/3, 1/3, 1/3, and then H* vs. L+H* was coded as 0, -1/2, 1/2. For the discourse context factor, the given context vs. the new & contrastive contexts was coded as -2/3, 1/3, 1/3, and the new context vs. the contrastive context was coded as 0, -1/2, 1/2. L1 vs. L2 group was coded as -1/2, 1/2. The results revealed no main effect of language group, but a 2-way interaction between discourse and language group as well as a 3-way interaction among pitch, discourse context, and language group. We therefore conducted further statistical analysis for each language group separately.

b = -3.87, z = -5.89, $p < .001$ for null accent vs. L+H*], but no significant difference was found between H* and L+H*. In the contrastive context, L+H* was rated significantly higher than H* [b = -3.50, z = -3.80, $p < .01$], which was in turn rated higher than null accent [b = -3.03, z = -7.38, $p < .001$ for null accent vs. H*; b = -6.52, z = -6.54, $p < .001$ for null accent vs. L+H*].

Table 5.4: Cumulative link mixed models for L1 and L2 groups

Fixed factor	L1 group				L2 group			
	Est.	SE	z	p	Est.	SE	z	p
Null vs. H/L+H	0.74	0.39	1.88	0.06	0.42	0.23	1.82	0.07
H vs. L+H	0.74	0.36	2.06	<.05*	0.12	0.28	0.41	0.68
Given vs. New/Contrastive	0.16	0.38	0.41	0.68	0.87	0.26	3.39	<.001***
New vs. Contrastive	0.15	0.39	0.38	0.70	0.97	0.34	2.90	<.01**
Null vs. H/L+H : Given vs. New/Contrastive	**10.35**	**1.20**	**8.66**	**<.001***	**5.79**	**0.54**	**10.70**	**<.001***
H vs. L+H : Given vs. New/Contrastive	3.84	0.62	6.16	<.001***	4.46	0.54	8.17	<.001***
Null vs. H/L+H : New vs. Contrastive	1.18	0.66	1.77	0.08	3.10	0.54	5.76	<.001***
H vs. L+H : New vs. Contrastive	**2.95**	**1.04**	**2.85**	**<.001***	**3.04**	**0.82**	**3.72**	**<.001***

Significant pitch-by-context interaction effects were observed in the L2 group as well: As in the native group, null accent in the given context was rated significantly higher than H* [b = 2.02, z = 4.90, $p < .001$] and than L+H* [b = 4.87, z = 0.01, $p < .001$]. Unlike L1 speakers, however, ratings for the new context were not statistically different among the three pitch conditions. In the contrastive context, L+H* was rated significantly higher than H* [b = -3.13, z = -4.36, $p < .001$] and null accent [b = -5.47, z = -7.16, $p < .001$], which parallels the L1 ratings.

Rating results analyzed by L2 proficiency are shown in Figure 5.9, Figure 5.10, and Figure 5.11. The rating patterns in the lower and the upper groups look similar except the new context condition, in which the lower group gave similar ratings to the three pitch conditions, while the upper group gave lower ratings to null accent compared to H* and L+H*.

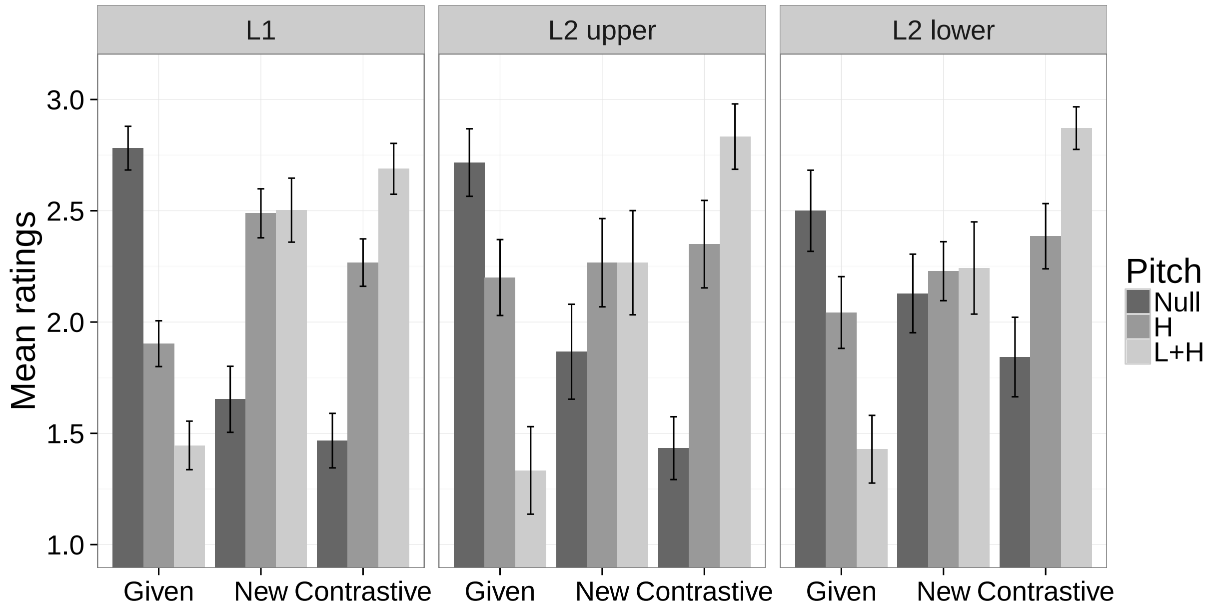


Figure 5.9: L1 and L2 naturalness ratings by proficiency

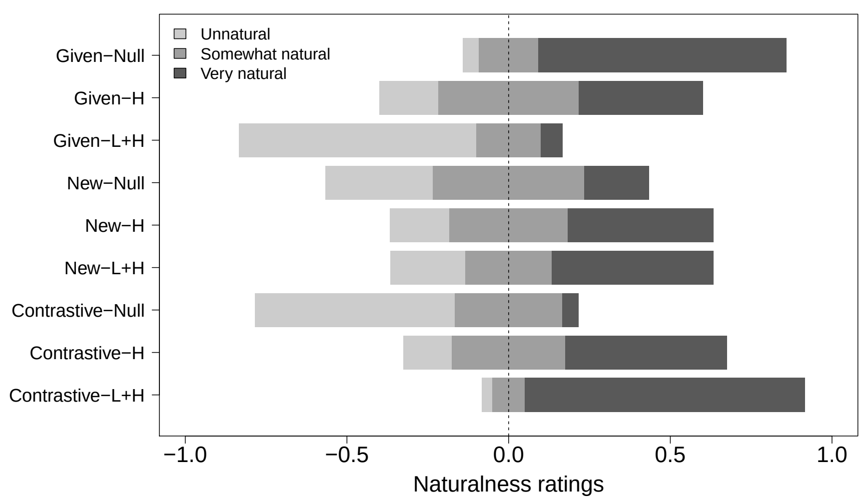


Figure 5.10: L2 upper group response proportions for each experimental condition

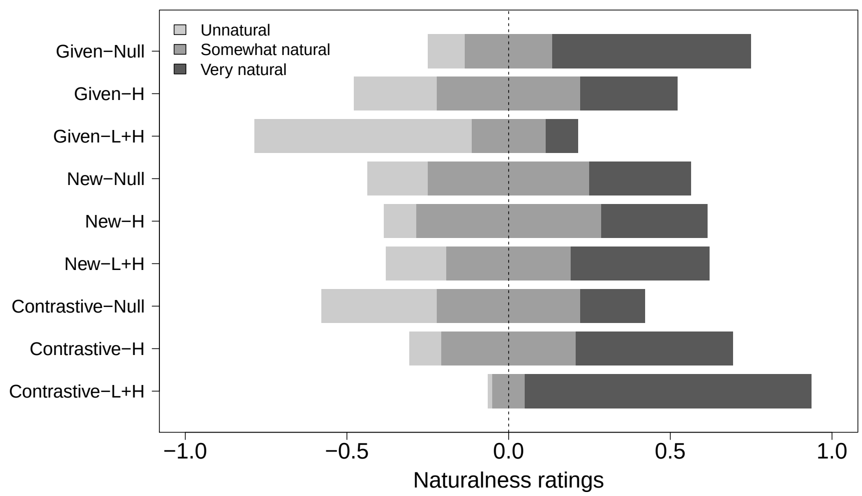


Figure 5.11: L2 lower group response proportions for each experimental condition

Statistical analysis revealed a three-way interaction between pitch, discourse context, and proficiency group [$b = 2.25$, $z = 2.68$, $p < .01$], confirming that the lower proficiency group, in comparison to the upper proficiency group, gave higher ratings for the null accent in the new context and the contrastive context (see Appendix F, Table F.1 for the statistical table).[15]

To summarize the effect of pitch contour on prosodic naturalness ratings, native English speakers demonstrated a strong link between given information and null accent. They also preferred L+H* over H* in the contrastive context, confirming the association between contrastive information and L+H*. For new information, both H* and L+H* were equally natural to native speakers. As for L2 learners, both lower and upper proficiency groups showed a strong preference for null accent in the given context and L+H* in the contrastive context, which mirrors the native speaker patterns. In the new context, the upper proficiency L2 group preferred both H* and L+H* over null accent like the native group, while the lower L2 group did not associate new information with any particular pitch accent type.

Next, we examined the durational effect on prosodic naturalness ratings. Our prediction with respect to duration was that the L1 group would show higher ratings for short stimuli over long

[15]The clmm function was used once again in the formula submitted to the statistical analysis of L2 proficiency: clmm(L2 ratings ∼ pitch * discourse context * L2 proficiency group + (1 + pitch * discourse context | participant)). For the pitch factor, null accent vs. H* & L+H* was coded as -2/3, 1/3, 1/3, and then H* vs. L+H* was coded as 0, -1/2, 1/2. For the discourse context factor, the given context vs. the new & contrastive contexts was coded as -2/3, 1/3, 1/3, and the new context vs. the contrastive context was coded as 0, -1/2, 1/2. For the L2 proficiency factor, the lower group vs. the upper group was coded as -1/2, 1/2.

stimuli in the given context, and higher ratings for long stimuli over short stimuli in the new and contrastive contexts.

Figure 5.12 represents the L1 rating data in which the three pitch accent conditions were collapsed and the three context conditions were divided in a binary fashion into given context vs. new/contrastive contexts.

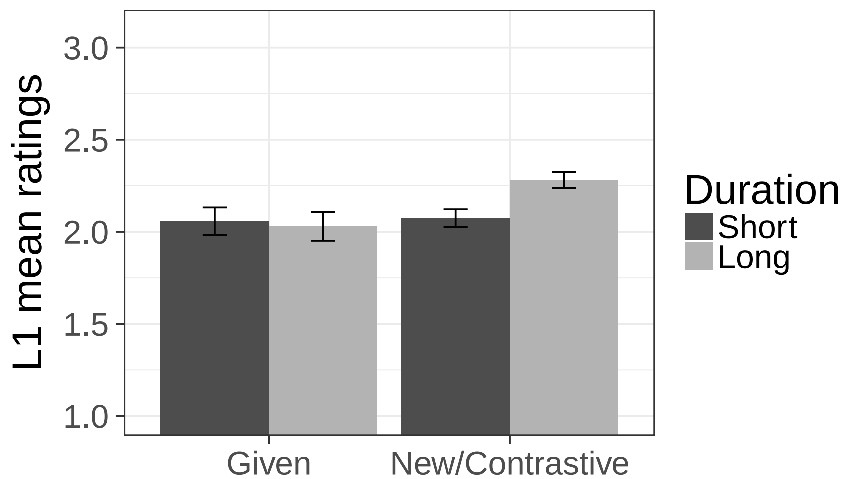


Figure 5.12: L1 naturalness ratings for duration

An ordinal mixed effects regression analysis revealed a significant main effect of discourse context [b = 0.34, z = 3.10, $p < .01$], a marginal effect of duration [b = 0.21, z = 1.92, $p = .06$], and a significant context-by-duration interaction [b = 0.57, z = 2.59, $p < .01$] for the L1 group.[16] Tukey's post-hoc analysis indicated that long stimuli marked significantly higher ratings than short stimuli in the new/contrastive contexts [b = -0.50, z = -3.82, $p < .001$], while short and long stimuli were rated similarly in the given context. This suggests that duration seems to play some role in L1-English speakers' perception of prosodic marking of information status, although its effect is not as clear as the pitch effect.

Figure 5.13 shows rating data by duration for the L2 group. There were main effects of discourse context [b = 0.44, z = 3.46, $p < .001$] and duration [b = 0.29, z = 2.44, $p < .05$], which suggests higher ratings for the new/contrastive contexts than the given context and for long stimuli than short stimuli, but no significant interaction between context and duration was found.

[16]The formula used for the duration analysis: clmm(ratings $\sim$ discourse context * duration + (1 + discourse context * duration | participant)). For the discourse context factor, given context vs. the new & contrastive contexts was coded as -1/2, 1/2, and short vs. long duration was coded as -1/2, 1/2.

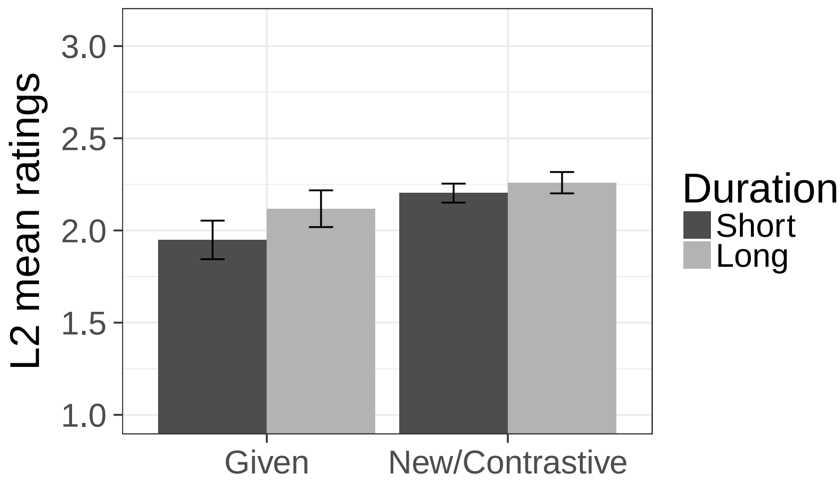


Figure 5.13: L2 naturalness ratings for duration

Figure 5.14 compares durational effects in the L2 upper and lower proficiency groups. For statistical analysis, we included proficiency group as an additional fixed factor in the ordinal mixed effects model.[17]

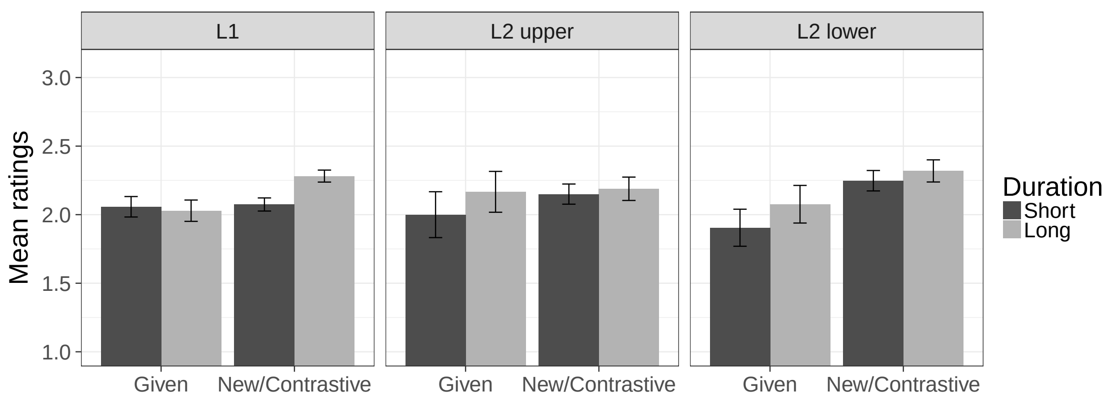


Figure 5.14: L1 and L2 naturalness ratings for duration

In addition to significant main effects of discourse context [b = 0.42, z = 3.40, $p < .001$] and duration [b = 0.29, z = 2.42, $p < .05$], there was also a marginal interaction between context and proficiency group [b = -0.48, z = 0.05, $p = .05$] due to the lower L2 group, but not the upper L2 group, giving higher ratings for the new/contrastive contexts than the given context regardless of duration. No interaction effect of duration in either group suggests that duration does not seem

[17]The formula used for the model: clmm(L2 ratings $\sim$ context * duration * L2 proficiency group + (1 + context * duration | participant)), Base condition = given context–short duration–lower group.

to play a significant role in Japanese L2 learners' association between prosodic saliency and information status of discourse referents in English.

Overall, durational cues, to some extent, affect L1 English speakers' perception of prosodic marking of discourse status, while Japanese-speaking L2 learners of English are insensitive to such cues during the prosody-in-context naturalness rating task.

5.5 Discussion

The primary objective of the prosody-in-context naturalness rating task was to investigate whether L1-Japanese learners of English can perceive differences in pitch and durational cues and associate particular pitch accent patterns–null accent, H*, and L+H*–with given, new, and contrastive information status when their attention is drawn to intonation. When pitch peak was manipulated to create three pitch contours corresponding to null accent, H*, and L+H*, L1 English speakers linked null accent to given information and L+H* to contrastive information, but both H* and L+H* were considered appropriate for new information. These results are in line with previous findings that accented words signal new information, while unaccented words signal given information or information that is available in the discourse context (Birch & Clifton, 1995; Bock & Mazzella, 1983; Most & Saltz, 1979; Nooteboom & Kruyt, 1987; Terken & Nooteboom, 1987). The difference between H* and L+H* emerged only when the accented NP refers to a contrastive referent, not a new referent.

According to Selkirk's focus projection theory (1995), when the direct object of a transitive verb bears a pitch accent, the accent projects focus marking to the VP as well as the NP itself so that the focus of the utterance can be either the direct object only or the entire VP. Our test sentence, *Mariana made banana bread* had acoustic prominence on the stressed syllable of the direct object *banana bread*, making it a feasible answer to both *What did Mariana do?*, which calls for broad focus (focus on the entire VP), and *Did Mariana make apple bread?*, which calls for narrow focus (focus on the direct object only). In addition, acoustic manipulation was performed on the rightmost content word, i.e., *banana bread*, which is the default nuclear pitch accent position for non-contrastive, declarative sentences. Since *banana bread* is a noun-noun compound which requires lexical stress on the first noun, nuclear pitch accent falls on *banana* rather than the final word of the sentence *bread*. In the new context condition, our native English speakers may have interpreted the increased acoustic prominence with L+H* as the nuclear pitch accent rather than a contrastive cue, resulting in reduced sensitivity to the difference between H* and L+H*. In the contrastive context condition, on the other hand, when listeners heard the context sentence *Did Mariana make apple bread?*, they must have expected the target test sentence to convey contrastive meaning. This expectation may have made L+H* on *banana bread* sound more natural than H* to native ears.

If prosodic prominence had appeared in a non-default nuclear pitch accent position, as in the case of Ito and Speer's (2008) test stimuli involving a color adjective-noun sequence (e.g., *green drum* → *BLUE drum*), listeners might be more likely to exploit the prosodic cues for discourse interpretations, resulting in greater sensitivity to the different degrees of acoustic prominence imposed by H* and L+H*. The offline English focus interpretation task by Welby (2003), which was described briefly in Chapter 3, is similar to the current experiment in that native English speakers listened to a question-answer pair with H*, L+H*, or no accent on the last content word of the response sentence, and rated the appropriateness of its intonation. Their naturalness ratings were similar between the H* and L+H* conditions, but significantly lower when there was no pitch accent. This finding provides additional evidence that native English speakers are sensitive to the presence or absence of pitch accent but not to different types of pitch accent (H* vs. L+H*) when the pitch is manipulated within the default nuclear pitch accent position.

Our results of L1 durational sensitivity were much less clear than those of pitch effects: Long stimuli were preferred over short stimuli in the new/contrastive contexts, but long and short stimuli did not yield a significant difference in the given context. Clearer effects of pitch than duration may suggest that L1 listeners rely more on pitch than durational cues in determining prosodic prominence. Watson (2010) has proposed the Multiple Source view, which assumes that multiple acoustic cues contributing to prominence, such as pitch, intensity, and duration, are influenced by various factors. For instance, pitch and intensity are used by the speaker primarily to mark important information for the listener, while increased duration is associated with the speaker's difficulty in the production process, serving as a weak cue for the marking of important information. More reliable correlates between pitch and information status in language input could be a reason why our L1 English listeners were more sensitive to pitch cues than durational cues in the naturalness rating task.

As for the L2 results, advanced L1-Japanese learners of English were able to associate null accent with the given context and L+H* with the contrastive context. They also preferred accentuation, i.e., H* and L+H*, over null accent for the new context. These patterns parallel the results of the L1 group. Less advanced learners could map null accent to the given context and L+H* to the contrastive context, but no difference among null accent, H*, and L+H* was found for the new context. Our prediction for L2 learners was that if they transfer their L1 prosodic marking of discourse, they would give higher ratings for L+H* than H* or null accent in the contrastive context because in Japanese, contrast is marked with expanded pitch range, similar to English L+H*. For H* and null accent, we predicted that there would be no significant difference between H* and null accent in the new and given contexts because presence or absence of pitch accent does not indicate discourse information status in Japanese. These predictions were partially supported in that both upper and lower proficiency learners demonstrated the L+H*–contrastive mapping. Against our predictions, however, both groups have also mastered a new mapping, namely the null accent–given association, in their L2. In addition, the upper proficiency group, but not the lower proficiency group, demonstrated the link between accentuation and new information like the native speaker group did. Why were the lower proficiency L2 learners able to acquire the null accent–given mapping but not the accentuation–new mapping? As discussed in Chapter 3, lack of accentuation usually indicates that the entity has already been activated and is accessible in the discourse, while H* or accentuation can sometimes mark given information in addition to new information depending on grammatical function, the surface structure of the phrase, or rhythmic alternation (Calhoun, 2010a; Sityaev, 2000; Terken & Hirschberg, 1994). Therefore, it may be easier for L2 learners to establish the link between null accent and given information than H* and new information.

Analysis of durational effects revealed no duration-by-context interaction regardless of proficiency level, suggesting that L2 learners do not seem to make use of durational cues when detecting English prosodic prominence and associating accentual signals with discourse structure in the TL. Since in Japanese syllable duration is not a reliable indicator of information status while pitch is (for contrastive marking), L1-Japanese learners are more sensitive to pitch cues than durational cues when it comes to discourse comprehension.

Experiment 1 was a rating task in which participants made a conscious judgment after listening to the whole dialogue. They were told explicitly to pay attention to intonation, and there was no time limit when rating appropriateness in each trial. The results of this task confirmed L2 learners' ability to make use of prosody while interpreting discourse meaning. In Experiment 2, we utilize the eye-tracking paradigm in order to investigate whether or not the same group of L2 learners are able to rapidly and subconsciously integrate prosodic cues and use such cues to facilitate the processing of upcoming discourse information when the learners' focus is directed to the task itself rather than prosody.

CHAPTER 6
EXPERIMENT 2: EYE-TRACKING LISTENING COMPREHENSION TASK

6.1 Introduction

Prior research investigating the use of prominence by native English listeners revealed that pitch accents are closely related to online processing of Information Structure in discourse. As explained in Chapter 3, listeners tend to expect new information to be accented and given information to be deaccented. Furthermore, new information is generally marked with H*, and contrastive information with L+H*. Although a growing number of L1 studies support native speakers' real-time use of pitch accent cues in discourse interpretations, only a few L2 studies have explored such an issue.

The present study adopts the experimental designs from Dahan et al. (2002) and Ito and Speer (2008), in an attempt to examine whether L2 listeners can integrate the target pitch accent patterns (null accent, H*, and L+H*) in a timely manner, and interpret such prosodic cues so as to quickly identify discourse referents while listening to spoken utterances. Utilizing the eye-tracking paradigm as an online measure, the experiment asked participants to perform a computer-based animal-coloring task: They heard pre-recorded instructions and clicked on the specified drawing instruments and animals on the computer screen. The pitch accent patterns and the information status of the target instruments were manipulated to create conditions with appropriate and inappropriate prosodic marking of discourse referents. In order to avoid testing too many experimental conditions in one experiment, we conducted two smaller-scale eye-tracking experiments: One experiment compared the H*–new association with the L+H*–contrastive association, and the other experiment compared H*–new association with the null accent–given association.

6.1.1 Participants

The same 134 participants from Experiment 1 were randomly assigned to either one of the two eye-tracking experiments. The new–contrastive experiment had 45 L1-English participants and 38 L2-English learners, and the new–given experiment had 25 L1-English and 26 L2-English participants. We ran more participants in the new–contrastive experiment than in the new–given experiment based on our L1 pilot results in Experiment 1, which showed a less clear difference between H* and L+H* than between null accent and H*/L+H*. The L2 participants were further grouped into a lower or an upper proficiency group based on their C-test and read-aloud proficiency scores. Among the 38 L2-English participants in the new–contrastive experiment, 21 were lower-level learners and 17 were upper-level learners. Of the 26 L2-English participants in the new–given experiment, 13 were lower-level learners and the other 13 were upper-level learners. Since proficiency scores were calculated after participants had completed the experiment session, the number of participants in each proficiency group is somewhat unbalanced. However, L2 proficiency was comparable between the two experiments. In the new–contrastive experiment, the mean proficiency score (on a scale of 0-10) was 3.75 for the lower group (range: 1.75 - 4.5) and 6.35 for the upper group (range: 5 - 8.88). In the new–given experiment, the mean proficiency score was 4.01 for the lower group (range: 2.88 - 4.5) and 6.01 for the upper group (range: 4.88 - 8.63). Independent-samples t-tests confirmed that there was no statistical difference in proficiency scores between the two upper proficiency groups [$t = .66$, $p = .52$] and the two lower proficiency groups [$t = -1.20$, $p = .24$].

Below we present the new–contrastive experiment first, since its design closely resembles the study by Ito and Speer (2008), which demonstrated the immediate contrastive interpretation of L+H* by native English speakers. We then turn to the new–given experiment, which used the same experimental design as the new–contrastive experiment but with different target stimuli.

6.2 Experiment 2a: New–contrastive experiment

6.2.1 Methods

Task

Experiment 2 was an online listening task: Participants were engaged in a picture-coloring task as a listener, where they were presented with pictures of uncolored animals and a set of drawing tools (as illustrated and discussed in detail below), and were asked to follow pre-recorded instructions to select the specified drawing tools and the animals. Instructions consisted of simple sentences such as *Use the green crayon to color the cow. Now, use the blue paintbrush to color the dolphin*. As participants listened to auditory instructions, they clicked on a drawing tool and an animal on the computer screen. While participants performed the task, their eye-movements and mouse click responses were monitored using a remote eye-tracking device.

Auditory and visual stimuli

Each trial consisted of two auditory instructions (e.g., *Use the green crayon to color the cow. Now, use the blue paintbrush to color the dolphin*). In the first instruction, as shown in Figure 6.1, the adjective-noun sequence was pronounced with a common H* H* pattern, followed by a sentence-final rising contour L-H%, which implies continuation of the current discourse (Pierrehumbert & Hirschberg, 1990).

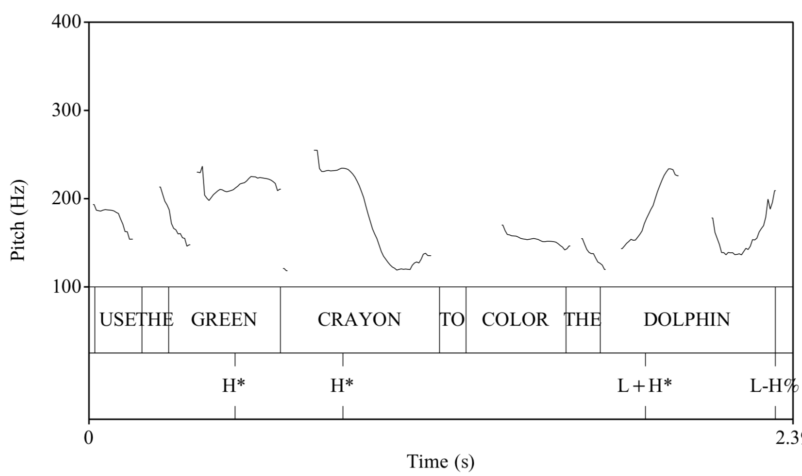
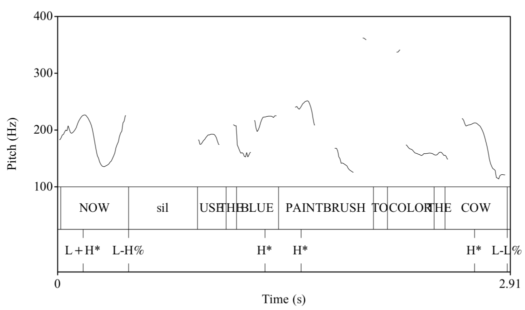


Figure 6.1: Example pitch contour of Instruction 1

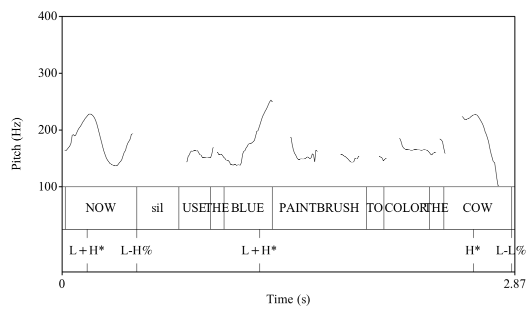
The second instruction contained the critical adjective-noun sequence: The target object was either new (different adjective and different noun) or adjective contrastive (different adjective and same noun). Within each of these two types of context, the pitch accent on the adjective was pronounced with either an H* or L+H* pitch accent (Table 6.1). When H* was assigned to the adjective, H* also occurred on the following noun; when L+H* was on the adjective, the noun was deaccented. Example sentences and their tonal contours are shown in Figure 6.2. All auditory stimuli were recorded by a female native speaker of English who had professional knowledge and training in phonetics and phonology including ToBI.

Table 6.1: Example new vs. contrastive experimental conditions

	Context	Pitch accent	Instruction 1	Instruction 2
(a)	New	H*	green crayon	blue paintbrush H* H*
(b)	New	L+H*	green crayon	blue paintbrush L+H* ∅
(c)	Contrastive	H*	green paintbrush	blue paintbrush H* H*
(d)	Contrastive	L+H*	green paintbrush	blue paintbrush L+H* ∅



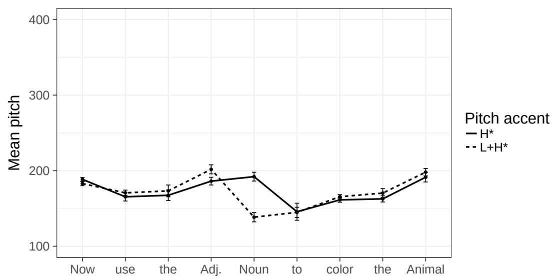
(a) H* condition



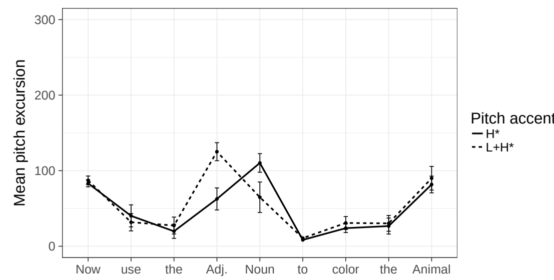
(b) L+H* condition

Figure 6.2: Example pitch contour of Instruction 2

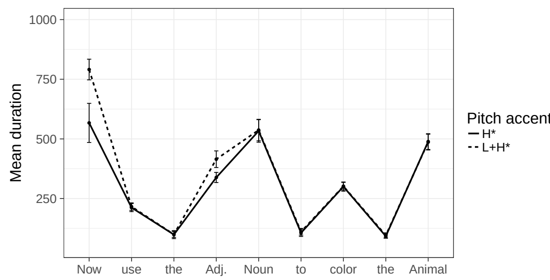
Mean pitch, pitch excursion (maximum minus minimum pitch), and mean duration in each word region are shown in Figure 6.3a, Figure 6.3b, and Figure 6.3c, respectively. Table 6.2 summarizes mean pitch, minimum and maximum pitch, pitch excursion, and duration of each region in the target sentences.



(a) Pitch



(b) Pitch excursion



(c) Duration

Figure 6.3: Mean pitch, pitch excursion, and duration of Instruction 2

Table 6.2: Acoustic analysis of Instruction 2

	Pitch accent	*Now* (& silence)	*use*	*the*$_1$	adjective	noun	*to color*	*the*$_2$	animal
F0 mean									
(Hz)	H*	188.46	165.43	167.46	186.16	192.10	145.65	162.70	191.37
	L+H*	182.92	170.66	173.26	201.80	138.47	155.20	170.34	198.02
F0 max									
(Hz)	H*	230.29	182.16	177.80	208.90	239.77	161.19	175.79	223.92
	L+H*	226.66	180.12	186.20	270.61	176.22	164.81	184.05	231.12
F0 min									
(Hz)	H*	147.02	141.95	158.00	146.27	129.51	145.03	149.10	142.20
	L+H*	139.42	148.34	158.79	145.45	111.39	144.04	153.77	141.08
F0 excursion									
(Hz)	H*	83.28	40.22	19.80	62.63	110.26	16.16	26.69	81.73
	L+H*	87.24	31.78	27.41	125.15	64.83	20.77	30.28	90.04
Duration									
(ms)	H*	567.08	212.08	98.75	338.33	533.33	203.13	92.08	488.75
	L+H*	790.83	216.25	99.58	415.00	537.08	206.25	95.83	486.67

Mean pitch, pitch excursion, and durational values for each word of the experimental sentences were submitted to mixed effects regression models for acoustic analysis.[18] Mean pitch differed significantly between H* and L+H* in *Now* [b = -5.53, $t(23)$ = -3.48, $p < .01$], adjective [b = 15.65, $t(23)$ = 4.09, $p < .001$], noun [b = -53.63, $t(23)$ = -13.90, $p < .001$], *color* [b = 4.21, $t(23)$ = 2.28, $p < .05$], and the_2 [b = 7.64, $t(23)$ = 2.39, $p < .05$]. Significant differences in pitch excursion appeared in adjective [b = 62.52, $t(23)$ = 6.90, $p < .001$] and noun [b = -45.43, $t(23)$ = -4.27, $p < .001$]. Differences in mean duration were found in *Now* [b = 223.75, $t(23)$ = 5.31, $p < .001$] and adjective [b = 76.67, $t(23)$ = 4.68, $p < .001$] (see Appendix G, Table G.1 for the full regression table). This confirmed that the adjective in the L+H* condition was prosodically more prominent than the adjective in the H* condition, marked with higher pitch, larger pitch excursion, and longer duration. The noun region, on the other hand, was more prominent in the H* condition than in the L+H* condition with higher mean pitch and larger pitch excursion.

The auditory instructions were played while participants saw a display with three uncolored animals in the center and four sets of drawing tools in different colors placed at the four corners, as shown in Figure 6.4. There were four types of drawing tools (paintbrush, crayon, roller, spray bottle) in eight colors (blue, green, orange, purple, red, silver, white, yellow). The location of each type of drawing tools was fixed throughout the experiment, while the colors were rotated for each trial to prevent participants from associating a particular color with a specific location. In all the experimental trials, the visual display contained the target object and one competitor which had the same color as the target object but was a different type of tool. For instance, the visual display for the trial *Use the green crayon to color the cow. Now, use the blue paintbrush to color the dolphin* contained a blue paintbrush as the target object and a blue crayon as the competitor.

[18] With *lme* function in the *nlme* package, we entered the following formula for, for example, the analysis of mean pitch: lme(mean pitch $\sim$ pitch condition, random = $\sim$ 1 | item / pitch condition, method = "ML"). Pitch condition was contrast coded: H* = -0.5, L+H* = 0.5.

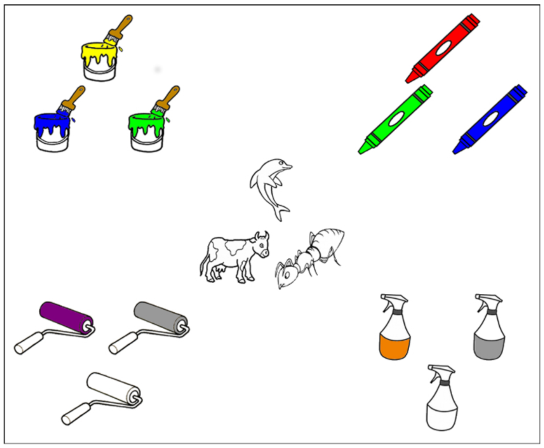


Figure 6.4: Example visual display

In visual world eye-tracking research, it is well known that not only the spoken discourse context but also the visual context alone can evoke contrast and affect native listeners' use of pitch accent. For instance, when the visual scene contains a pair of contrastive objects (e.g., a large blue square, a small blue square; a large yellow circle; a small red triangle), pitch prominence on the size adjective in the target sentence (e.g., *Touch the LARGE blue square.*) triggers faster fixations to the target object than when the display contains two pairs of contrastive objects (e.g., a large blue square, a small blue square; a large yellow circle, a small yellow circle), which implies that even without a preamble in the spoken discourse, listeners establish a contrast set and interpret the pitch prominence contrastively based on the visual cue (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). In the present study, we used four sets of drawing tools in various colors so that (i) the contrastively colored objects in the visual scene would not be too obvious to participants, and (ii) the contrastive relationship would be established according to the auditory stimuli rather than the visual stimuli alone.

For each experimental condition, six sets of experimental stimuli were created and distributed into four lists of 24 items each so that one participant would see only one version of the four conditions for each item. Within a list, 24 target items and 24 fillers were pseudo-randomized and then separated into two blocks of 24 trials each, yielding 48 trials in total (see Appendix H, Table H.1 for a sample item list). In Instruction 2 of the filler trials, the adjective and the noun were both new, or the adjective was given and the noun was contrastive. In both cases, the adjective-noun sequence was marked with a natural prosodic contour of H* H* or H* !H* in order to prevent participants from becoming used to and hence insensitive to unnatural pitch contours in some of the experimental conditions.[19] Five practice items were given at the beginning of the experiment session to make sure all participants were comfortable with the task procedure.

Procedure

Participants were seated in front of a computer, and their eyes were calibrated after the practice trials and immediately before the experimental trials. At the beginning of each trial, the visual display was first presented to the participants as a preview. Following the 1000 ms preview time, they heard the first instruction and clicked on the specified drawing object and the animal. The second instruction began 1750 ms after the offset of the first instruction, and the participant selected a drawing object and an animal again, according to the second instruction. They then pressed the space key to move on to the next trial. No feedback about correct or incorrect mouse responses was provided throughout the experiment. The eye-movements were recorded using the SMI RED250 remote eye-tracker at the sampling rate of 250 Hz (every 4 ms). Participants' mouse clicks were also recorded in order to monitor if they were paying attention to the task during the experiment session.

6.2.2 Predictions

Assuming that the visual display does not create any obvious bias during the preview time, participants are likely to encode the target tool and animal in Instruction 1 (e.g., *green crayon, cow*) as new entities in their discourse model based on the auditory preamble. These entities should remain highly activated in the model as Instruction 2 begins. During Instruction 2, if they primarily associate H* with new discourse entities and L+H* with contrastive entities, we expect more fixations to the new referent target when the prenominal adjective is marked with H* than with L+H* (condition (a) > condition (b) in Table 6.3), and similarly more fixations to

[19]The production studies by Krahmer and Swerts (2001) and Speer and Ito (2011) report that noun contrastive utterances were often produced with a pitch accent on both adjective and noun. For this reason, we assigned H* H* or H* !H* to the adjective-noun sequence in the noun contrastive filler items.

the contrastive referent when the adjective is produced with L+H* than with H* (condition (d) > condition (c)).

In the H*–new condition (condition (a)), when listeners hear H* on the new color, they should interpret it as a prosodic signal for a non-contrastive, new object and quickly shift their eye gaze to the target object on the display (e.g., *green crayon* → *blue paintbrush* not *blue crayon*). In the L+H*–new condition (condition (b)), where a new referent is the target, as soon as listeners hear L+H* on the color adjective (e.g., *blue...*), they should identify the most salient drawing tool in their discourse model (e.g., *green crayon*), establish a contrast set based on the prosodic prominence on the color (e.g., *green crayon* vs. *blue crayon*), and try to search for the other member in the set in the visual display (e.g., *blue crayon*), which guides them (wrongly) to the incorrect, discourse-contrastive object (i.e., competitor) until the disambiguating noun becomes available. This should result in brief fixations to the competitor and a considerable delay in fixating to the actual target object (*green crayon* → *blue crayon* before *blue paintbrush*). In the H*–contrastive condition (condition (c)), where the target is a contrastive referent, H* on the adjective should trigger a new referent interpretation, drawing listeners' attention to the competitor before the segmental information on the noun disambiguates the correct target object (e.g., *green paintbrush* → *blue crayon* before *blue paintbrush*). In the L+H*–contrastive condition (condition (d)), L+H* on the color should trigger a contrast set (e.g., *green paintbrush* vs. *blue paintbrush*) and facilitate the identification of the correct target object on the display (e.g., *green paintbrush* → *blue paintbrush*).

As for the L2 participants, if they are able to utilize their knowledge of the L+H*–contrastive mapping demonstrated in Experiment 1, both lower and upper proficiency groups should be able to detect the L+H* pitch cue and interpret the cue as a signal for a contrastive discourse referent. If L2 learners' ability to exercise their knowledge of the L+H*–contrastive mapping during an online task depends on proficiency, only upper-level learners should display the native-like pattern. If online integration of prosodic cues is difficult for learners at any level, we should expect no difference between the H*–new condition and the L+H*–new condition, or between the L+H*–contrastive condition and the H*–contrastive condition in both lower and upper proficiency groups.

Table 6.3: Predictions for the trial *blue paintbrush*

	Context	Pitch accent	Instruction 1	Instruction 2	Expected eye gaze
(a)	New	H*	green crayon	blue paintbrush H* H*	blue paintbrush (felicitous)
(b)	New	L+H*	green crayon	blue paintbrush L+H* ∅	blue crayon (infelicitous)
(c)	Contrastive	H*	green paintbrush	blue paintbrush H* H*	blue crayon (infelicitous)
(d)	Contrastive	L+H*	green paintbrush	blue paintbrush L+H* ∅	blue paintbrush (felicitous)

6.2.3 Results

Mouse click mean accuracy in the target trials was 99.51% (*SD*: 3.33, range: 95.56-100%) for the L1 group and 98.64% (*SD*: 7.87, range: 92.71-100%) for the L2 group. Mean accuracy in the filler trials was 98.84% (*SD*: 8.05, range: 94.73-100%) for the native group and 96.96% (*SD*: 14.46, range: 91.66-100%) for the L2 group. Since all of the 83 participants had a mean accuracy of 90% or above, they were all included in the subsequent gaze data analyses.

For gaze analyses, we first removed all the target trials in which participants failed to select the correct target objects. We then reframed the gaze data from 4 ms to 50 ms time bins, and for each time bin, fixation on a given area of interest was coded as 1 or 0.[20] Following Ito and Speer (2008), the data were aligned from the onset of the critical noun, as it is the point where segmental information begins to disambiguate the target object. Fixations during the modifying adjective were aligned backward with negative time values, and fixations after the noun onset had positive time values. Assuming that the time to plan and launch an eye movement is between 175 and 200 ms (Rayner, 1995), we expected the effects of pitch accent to emerge approximately 200 ms from the onset of the prenominal color adjective.

Figure 6.5 shows time-course fixation proportions to the target and to the competitor in the new and the contrastive contexts for the L1 group (see Appendix I, Figure I.1, for fixation proportions to all the areas of interest). Remember that in the new context conditions, the target

[20]Eye blinks and saccades were coded as 0 rather than being excluded from the data set because those noise events were not equally distributed across conditions (Barr, 2008).

was a new referent (e.g., *green crayon* → *blue paintbrush*) and its competitor shared the same color as the target (e.g., *blue crayon*). In the contrastive context conditions, the target was a contrastive referent (e.g., *green paintbrush* → *blue paintbrush*) and its competitor had the same color as the target (e.g., *blue crayon*). Inspection of the graph indicates an effect of pitch accent on target fixations in the contrastive context beginning approximately 100 ms after the onset of the noun, while no pitch difference can be seen in the new context. It seems that L+H* facilitated the identification of the contrastive target, but H* did not facilitate looks to the new target. Fixation proportions for the competitor showed very similar patterns between the two pitch conditions in both new and contrastive contexts. However, fixation lines for the target and for the competitor began to diverge slightly earlier for H* than L+H* (around 300 ms vs. 500 ms) in the new context, and for L+H* than H* in the contrastive context (around 300 ms vs. 450 ms), reflecting some pitch effect in both new and contrastive context conditions.

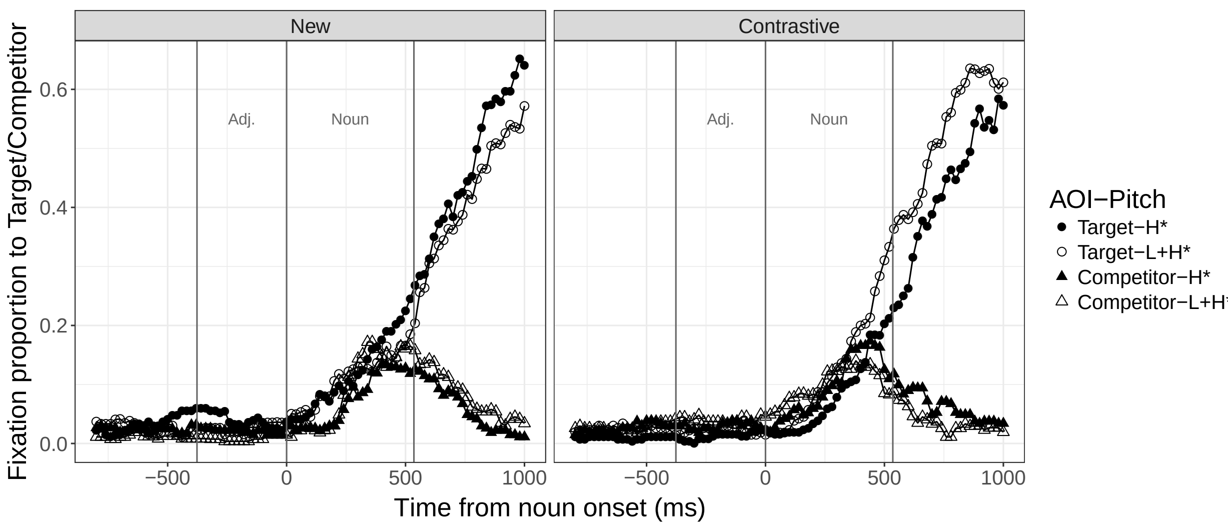


Figure 6.5: Fixation proportions to the target/competitor: L1 group

For statistical analysis on fixations to the target, we first collapsed data from the four conditions and plotted a fixation logit graph (see Figure J.1 in Appendix J), in order to determine the onset and the offset of a steady increase in fixation to the target object regardless of the pitch and the context conditions (Barr, 2008; K. Ito et al., 2012). For L1 and L2 listeners, the rise in fixation proportions began at approximately -100 ms and ended around 900 ms. Based on this observation, gaze data from -100 ms to 900 ms were used for further analyses.

Following Barr (2008), we conducted empirical logit analyses for participants and for items separately. For the participants analysis, we first aggregated all of the trials for each condition for each participant, and calculated empirical logit and weights for each combination of condition and participant. We then ran a model containing pitch, context, and the interaction of pitch and context as fixed factors and a random intercept by participants. Similarly, for the item analysis, we aggregated all the trials within a given condition for each item, calculated empirical logit and weights for each combination of condition and item, then ran a model with pitch, context, and the pitch-by-context interaction as fixed factors and a random intercept by items.[21] Figure 6.6 represents mean fixation proportion to the target object within the critical time window extending from -100 ms to 900 ms, and Table 6.4 summarizes the empirical logit model for the L1 group.

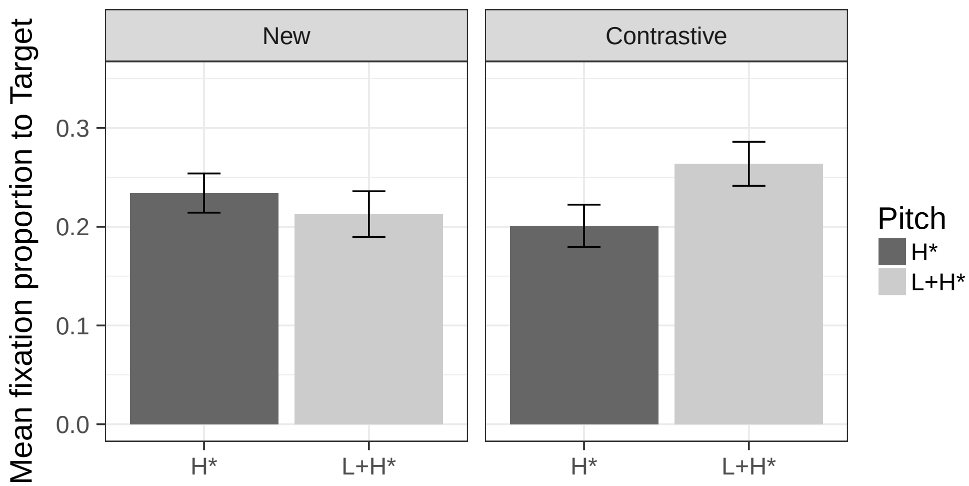


Figure 6.6: Mean fixation proportion to the target: L1 group

[21]For empirical logit and weights, the formulae were adopted from Barr (2008): elog = log((fixations to target + 0.5) / (all possible fixations - fixations to target + 0.5)), weights = 1 / (fixations to target + 0.5) + 1 / (all possible fixations - fixations to target + 0.5). Then using the *lmer* function in the *lme4* package, we submitted elog values to the following formula: lmer(elog ∼ pitch * context + (1 | participant), weights = 1/wts). Pitch and context were contrast coded: H* = -0.5, L+H* = 0.5; New = -0.5, Contrastive = 0.5.

Table 6.4: Empirical logit model for the target: L1 group

L1 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.22	0.08	-15.40	$< .001$***	-1.19	0.06	-20.08	$< .001$***
Pitch	0.08	0.07	1.15	0.25	0.11	0.08	1.39	0.17
Context	0.05	0.07	0.76	0.45	0.05	0.09	0.68	0.50
Pitch : Context	0.43	0.14	3.13	$< .01$**	0.43	0.15	2.89	$< .01$**

The analysis revealed a significant interaction effect of pitch and context with a positive coefficient [b_1 = 0.43, t_1 = 3.13, $p_1 < .01$; b_2 = 0.43, t_2 = 2.89, $p_2 < .01$]. Tukey's post-hoc analyses revealed a significant pairwise difference between H* and L+H* in the contrastive context [t = -3.05, $p < .05$], indicating that L+H* produced more looks to the target than H* did when the target was a contrastive referent.

Next, in order to examine whether L1 participants wrongly fixated on the competitor in the L+H*–new context and the H*–contrastive context, we plotted mean fixation proportion for the competitor with an analysis window extending from 100 to 800 ms (Figure 6.7) as determined based on the grand mean function (Figure J.2 in Appendix J). The mean fixation proportion is slightly higher for L+H* than for H* in the new context, while the two pitch conditions are very similar in the contrastive context.

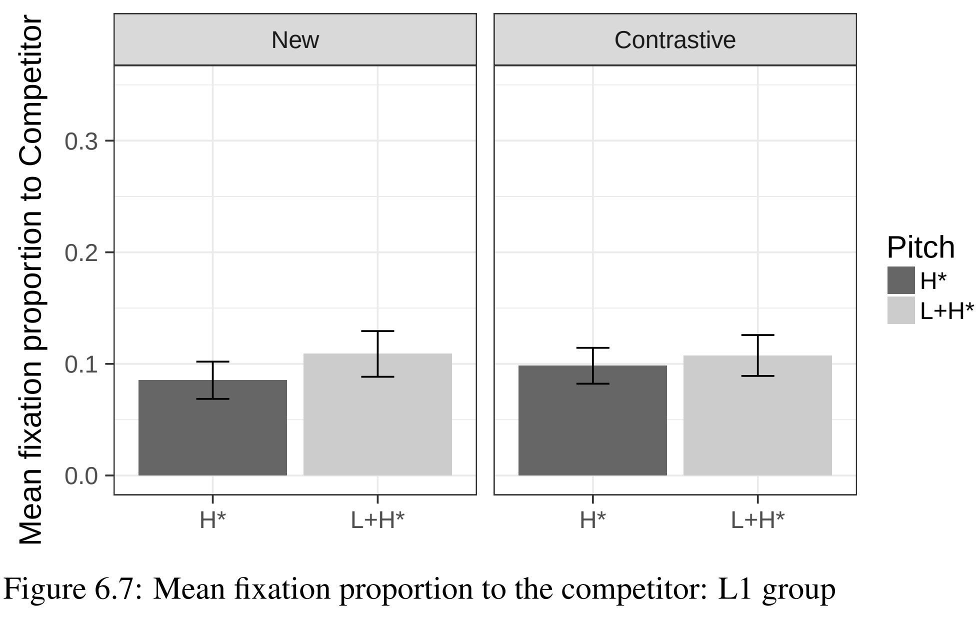


Figure 6.7: Mean fixation proportion to the competitor: L1 group

We then submitted the fixation data for the competitor within the analysis time window to empirical logit analyses. The model outcome shown in Table 6.5 revealed no main or interaction effect of pitch and context. This suggests that neither the discourse status of target referents nor pitch type affected native listeners' fixation pattern to the competitor.

Table 6.5: Empirical logit model for the competitor: L1 group

L1 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-2.02	0.07	-29.36	<.001***	-2.18	0.09	-23.44	<.001***
Pitch	0.17	0.10	1.73	0.08	0.19	0.12	1.63	0.10
Context	-0.02	0.10	-0.25	0.81	0.15	0.12	1.25	0.21
Pitch : Context	-0.23	-1.15	0.25	0.51	-0.19	0.24	-0.78	0.43

To sum up the L1 results, a strong facilitative effect of L+H* was observed in fixations to contrastive target referents, while H* did not increase looks to new referents. In addition, we found no evidence that either L+H* or H* affects fixation patterns on the incorrect target.

Turning now to the L2 results, the fixation proportion graph for the L2 group is shown in Figure 6.8 (see Appendix I, Figure I.2, for fixation proportions to all the areas of interest). In the new context, there was an initial increase in fixations to the target for L+H* starting at approximately 250 ms, and then the fixation line for H* rapidly rose around 400ms, making it difficult to determine the effect of either one of the two pitch conditions over the analysis time window. In the contrastive context, L+H* seems to show a slight advantage over H* in looks to the target starting from 100 ms to 600 ms. The proportion of the competitor is clearly higher in the contrastive context than in the new context over -200 ms to 500 ms. However, the difference between H* and L+H* within each context condition is very small.

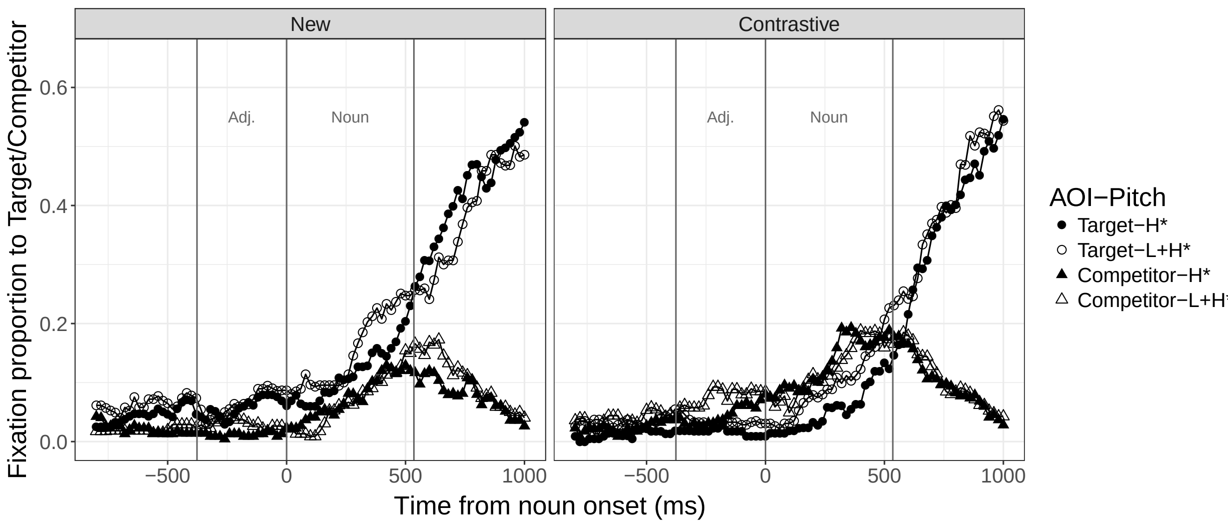


Figure 6.8: Fixation proportions to the target/competitor: L2 group

The mean fixation proportion to the target and statistical results over -100-900 ms are shown in Figure 6.9 and Table 6.6 respectively. There was a main effect of context [b_1 = -0.27, t_1 = -3.06, $p_1 < .01$; b_2 = -0.28, t_2 = -3.01, $p_2 < .01$] by participant and item analyses, which is reflected in the increased looks to the target for the new condition compared to the contrastive condition, regardless of pitch type. This means that L2 learners tend to look at a new referent no matter whether they hear H* or L+H* on the prenominal color adjective. No interaction between pitch and context was found.

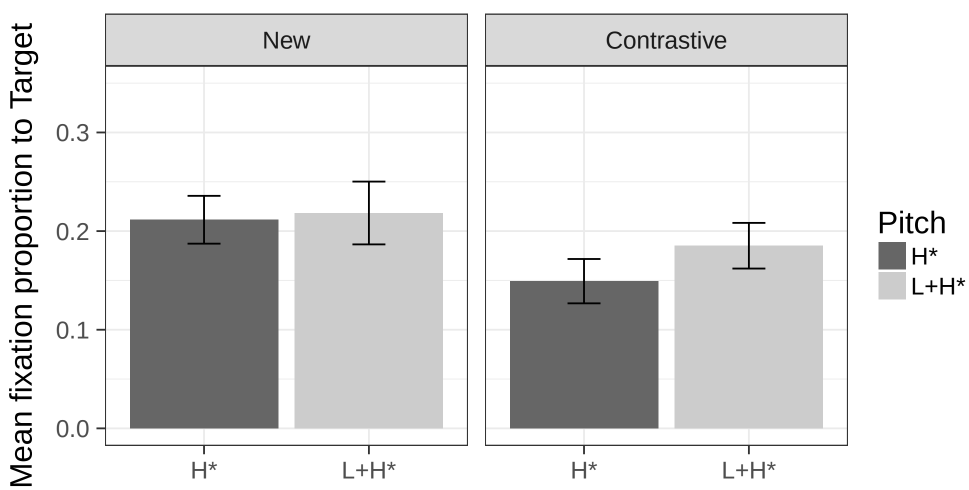


Figure 6.9: Mean fixation proportion to the target: L2 group

Table 6.6: Empirical logit model for the target: L2 group

L2 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.34	0.08	-17.57	<.001***	-1.40	0.07	-20.64	<.001***
Pitch	0.12	0.09	1.35	0.18	0.12	0.09	1.28	0.20
Context	-0.27	0.09	-3.06	<.01**	-0.28	0.09	-3.01	<.01**
Pitch : Context	0.11	0.18	0.63	0.53	0.21	0.19	1.11	0.27

Figure 6.10 and Figure 6.11 represent fixation proportions to the target analyzed by L2 proficiency. While the time-course fixation proportion graph in Figure 6.10 does not show any obvious patterns except that the lower group is slower in fixating to the target than the upper group, inspection of the mean fixation proportion graph in Figure 6.11 suggests that in both lower and upper groups, proportion of fixations to the target is higher in the new context than in the contrastive context. Furthermore, in the lower group L+H* seems to have generated more looks to the target than H* in both new and contrastive contexts, while in the upper group there appears to be no clear difference between H* and L+H*.

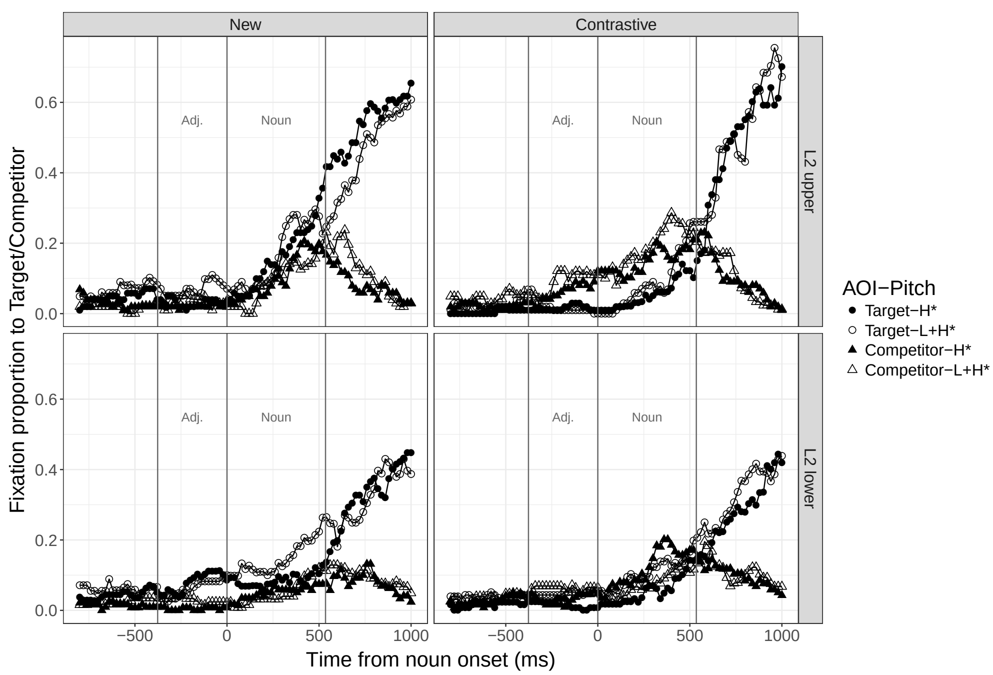


Figure 6.10: Fixation proportions to the target/competitor: L2 lower and upper groups

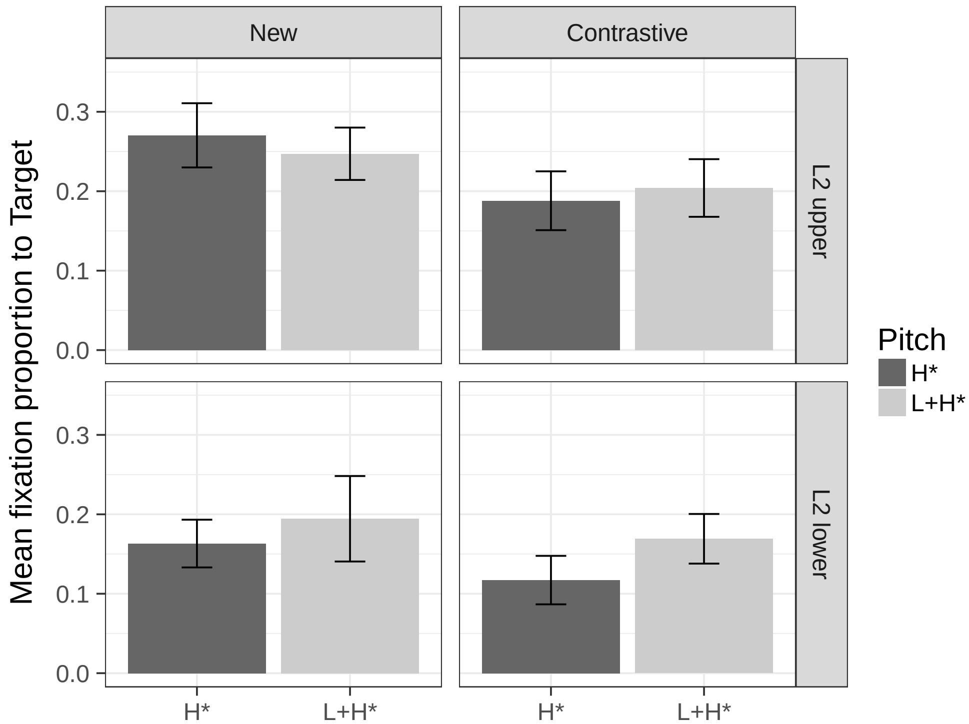


Figure 6.11: Mean fixation proportion to the target: L2 lower and upper groups

These observations were confirmed by statistical analyses summarized in Table 6.7 for the lower proficiency group and Table 6.8 for the upper proficiency group. In the lower group, there was a main effect of pitch with a positive coefficient by participant and item analyses [b_1 = 0.34, t_1 = 2.48, $p_1 < .05$; b_2 = 0.33, t_2 = 2.63, $p_2 < .01$], meaning that L+H* triggered looks to the target more than H* did regardless of context. There was also a main effect of context by item analysis with a negative coefficient, which suggests more fixations to the target in the new context than in the contrastive context regardless of pitch type. No pitch-by-context interaction was found. The lower-level learners thus preferred to fixate to discourse new referents over contrastive referents independent of pitch, and L+H* on the prenominal adjective accelerated the identification of the target object compared to H* independent of the information status of the target referent.

Table 6.7: Empirical logit model for the target: L2 lower group

L2 lower	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.49	0.10	-14.77	<.001***	-1.52	0.07	-22.00	<.001***
Pitch	0.34	0.14	2.48	<.05*	0.33	0.13	2.63	<.01**
Context	-0.25	0.14	-1.81	0.07	-0.25	0.13	-2.00	<.05*
Pitch : Context	0.06	0.28	0.22	0.82	0.22	0.25	0.89	0.37

Table 6.8: Empirical logit model for the target: L2 upper group

L2 upper	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.17	0.10	-11.48	<.001***	-1.11	0.08	-13.94	<.001***
Pitch	-0.08	0.10	-0.77	0.44	-0.05	0.11	-0.47	0.64
Context	-0.29	0.10	-2.75	<.01**	-0.22	0.11	-2.02	<.05*
Pitch : Context	0.16	0.21	0.78	0.44	0.18	0.22	0.82	0.41

Analysis of the data for the upper group revealed a significant main effect of context with a negative coefficient by both participant and item analyses [b_1 = -0.29, t_1 = -2.75, $p_1 < .01$; b_2 = -0.22, t_2 = -2.02, $p_2 < .05$]. No main effect of pitch or interaction between context and pitch reached significance. This indicates that the L2 upper-level learners are biased toward new referents, and their fixation pattern is not affected by pitch accent during visual search.

Fixation patterns to the competitor also reflect L2 bias due to new vs. contrastive status of discourse referents. In Figure 6.12, the mean fixation proportion to the competitor is much higher for the contrastive context than for the new context, which was confirmed by the main effect of context by participant and item analyses as shown in Table 6.9 [b_1 = 0.38, t_1 = 3.27, $p_1 <$ .01; b_2 = 0.41, t_2 = 0.11, $p_2 < .001$]. This means that when the target was a contrastive referent, L2 learners were more likely to look at a new referent (e.g., *green paintbrush* $\rightarrow$ *blue crayon* instead of *blue paintbrush*), and when the target was a new referent, they were less likely to fixate wrongly on a contrastive referent (e.g., *green crayon* $\rightarrow$ *blue paintbrush*, not *blue crayon*).

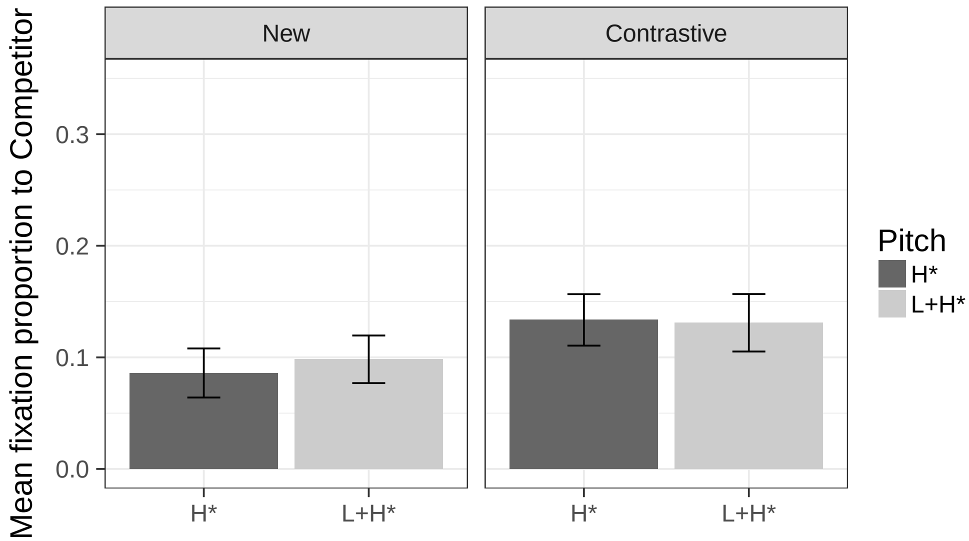


Figure 6.12: Mean fixation proportion to the competitor: L2 group

Table 6.9: Empirical logit model for the competitor: L2 group

L2 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.87	0.08	-22.54	<.001***	-2.07	0.10	-20.70	<.001***
Pitch	-0.04	0.11	-0.38	0.71	0.05	0.11	0.50	0.62
Context	0.38	0.11	3.27	<.01**	0.41	0.11	3.79	<.001***
Pitch : Context	-0.06	0.23	-0.26	0.80	-0.09	0.21	-0.41	0.68

When the data were analyzed by proficiency, we found the same context effect in both lower and upper groups. Figure 6.13 shows higher mean proportion fixation to the competitor for the contrastive context than for the new context in both proficiency groups. This was reflected in the significant main effect of context for the lower group [b_1 = 0.41, t_1 = 2.52, $p_1 < .05$; b_2 = 0.44, t_2 = 3.15, $p_2 < .01$ in Table 6.10] and for the upper group [b_1 = 0.33, t_1 = 2.07, $p_1 < .05$; b_2 = 0.30, t_2 = 2.23, $p_2 < .05$ in Table 6.10]. No main effect of pitch or interaction between pitch and context was found in either group.

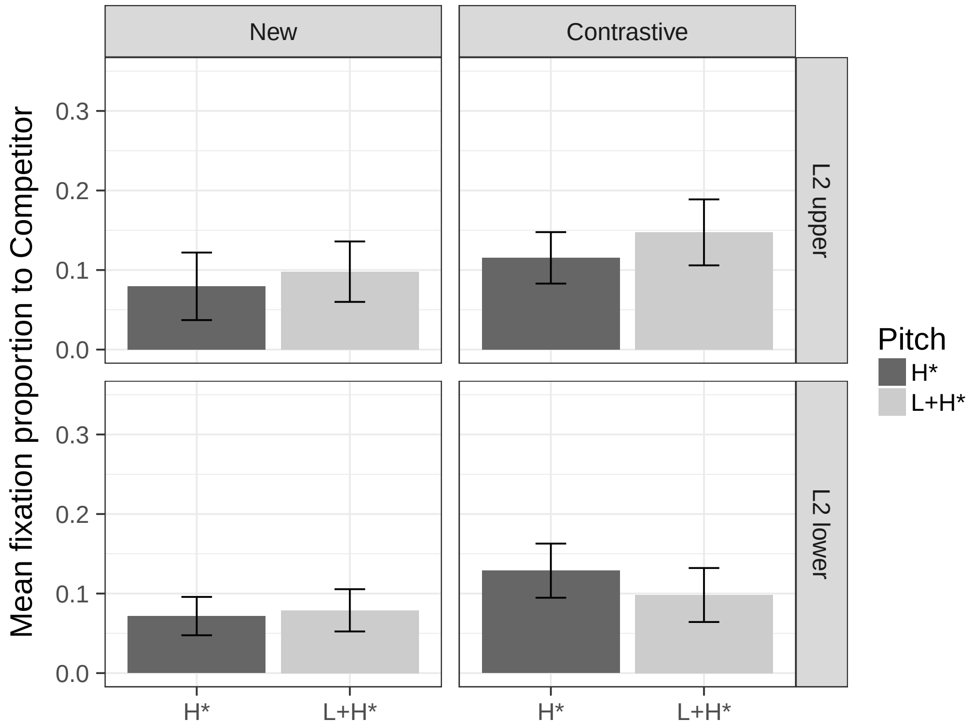


Figure 6.13: Mean fixation proportion to the competitor: L2 lower and upper groups

Table 6.10: Empirical logit model for the competitor: L2 lower group

L2 lower	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-2.00	0.12	-16.35	<.001***	-2.12	0.12	-17.80	<.001***
Pitch	-0.17	0.16	-1.05	0.29	-0.08	0.14	-0.59	0.56
Context	0.41	0.16	2.52	<.05*	0.44	0.14	3.15	<.01**
Pitch : Context	-0.36	0.33	-1.11	0.27	-0.40	0.28	-1.44	0.15

Table 6.11: Empirical logit model for the competitor: L2 upper group

L2 upper	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.75	0.11	-15.59	<.001***	-1.72	0.09	-19.71	<.001***
Pitch	0.08	0.16	0.54	0.59	0.13	0.13	0.95	0.34
Context	0.33	0.16	2.07	<.05*	0.30	0.14	2.23	<.05*
Pitch : Context	0.21	0.32	0.68	0.50	0.14	0.27	0.53	0.60

All in all, the L2 results in the new–contrastive experiment exhibited a strong bias to discourse new entities for all proficiency levels. They fixated more on the target picture when the target was a new referent and its competitor was a contrastive referent than when the target was a contrastive referent and its competitor was a new referent. Similarly, the L2 learners fixated more on the competitor when the target was a contrastive referent and its competitor was a new referent than when the target was a new referent and its competitor was a contrastive referent. In addition to their overall preference for new entities, the lower-level learners detected the target object faster with L+H* than H* independent of new vs. contrastive status of the target object.

6.2.4 Discussion

The new–contrastive eye-tracking experiment aimed to test whether L2 learners are able to associate H* and L+H* to new and contrastive information, respectively. The L1 English group demonstrated the facilitative effect of L+H* on the identification of contrastive referents: The presence of L+H* on the prenominal adjective (e.g., *Use the green paintbrush to color the cow. Now, use the BLUE paintbrush to color the dolphin.*) accelerated fixations to the contrastive target object compared to when H* was on the same adjective. In contrast to the clear effect of L+H* observed in the fixation patterns to the contrastive target, the analysis of fixations to the competitor provided no robust evidence of such pitch effects. In the L+H*–new condition, in which L+H* was predicted to wrongly guide listeners to the competitor (contrastive object), the mean fixation proportion to the competitor was numerically higher than that of the H*–new condition, reflecting a weak garden-path effect. However, this difference did not reach a statistical significance.

Our finding that L+H* facilitates looks to the contrastive target is consistent with Ito and Speer (2008). Experiment 2 in Ito and Speer shows that L+H* in comparison to neutral H* induced significantly more fixations to the contrastive target. What differed in our study is that L+H* in the new target trials did not result in garden-path looks. Our results are more similar to those in Watson, Tanenhaus, and Gunlogson (2008) with respect to both target and competitor fixation patterns Watson et al. found that L+H* on a noun whose information status is temporarily ambiguous between new and contrastive (e.g., *Click on the camel and the dog. Move the dog to the right of the square. Now, move the* <u>*camel/candle*</u> *below the triangle.*) generated significantly more fixations to the contrastive referent than the new referent, while H* did not result in a significant difference in looks to the new referent vs. the contrastive referent. The same kind of fixation pattern was found in our study. When L+H* was on the prenominal adjective, the fixation proportion for the contrastive target was significantly higher than that for the new target. When H* was on the adjective, there was no difference in fixation proportion to the contrastive target vs. to the new target. Furthermore, although Watson et al. do not report statistical analysis on fixations to the competitor, their mean fixation proportion graph shows no garden-path effect in the L+H*–new referent condition. In fact, the mean fixation proportions to the new target and the competitor remain equally high (approximately 0.3), until looks to the new target begin to rise at around the offset of the critical noun (The mean duration of the noun with L+H* was 475 ms). In our study, too, no garden-path effect was observed when the new referent was infelicitously marked with L+H*. Ito and Speer (2008), on the other hand, revealed both the facilitative effect of L+H* on contrastive targets and the infelicitous (garden-path) contrastive interpretation of L+H* for new targets.

There are several reasons why we think Watson et al. and our study showed only the facilitative effect of L+H* on contrastive targets, and not infelicitous garden-path effects on new targets. First, our visual display was much simpler than Ito and Speer's, and so was Watson et al.'s. In our study, there was only one competitor that carried the same color as the target object, and in Watson et al., too, only one competitor was presented along with the target. Ito and Speer's visual stimuli, on the other hand, contained multiple competitors of the same color. When multiple candidates are presented to listeners in an online task with time restrictions, listeners have to be more selective as to which object to look at, and therefore they may be more likely to rely on available cues such as prosody in making a decision in their eye gaze. When there are only two possible candidates (i.e., target and competitor) as in the present study and in Watson et al., listeners can more easily shift attention between the two objects and continue this until disambiguating information becomes available. As mentioned above, in infelicitous (L+H*) new referent trials in Watson et al., the mean fixation proportions to the new target and the (contrastive) competitor remain equally high until the offset of the critical noun by which point listeners have processed segmental information that disambiguates the target referent and launched eye movement accordingly. In our study, too, looks to the new target and to the (contrastive) competitor continue to rise till about 500 ms into the disambiguating noun, which is much later than the divergence point in Ito and Speer (approximately 150 ms prior to the disambiguating noun onset). Thus, the simple visual display that allows listeners to pay attention to both of the potential targets may have reduced facilitative and infacilitative eye gaze triggered by L+H* that may have been observable with a more complex visual display.

The second possibility is that more complex audio stimuli used in our study and in Watson et al. resulted in outcomes somewhat different from those in Ito and Speer, which used much simpler stimuli. In both our study and Watson et al., the target stimuli contained additional discourse entities besides the critical object. Our stimuli, such as *Use the green paintbrush to color the dog. Now, use the blue paintbrush to color the dolphin*, include animals in addition to drawing instruments, and the stimuli in Watson et al. contain geometric shapes as well as animals, as in *Click on the camel and the dog. Move the dog to the right of the square. Now, move the camel below the triangle.* In Ito and Speer, on the other hand, the target sentence contained only the critical object and no other discourse entity, as in *Hang the green drum. Now, hang the blue drum*. It is possible that in our study and in Watson et al.'s, additional discourse entities made the discourse structure more complex, making the contrast between the target entity and the preceding discourse entity less salient for the listener.

As for the L2 learners, neither proficiency group demonstrated sensitivity to prosodic cues in the interpretation of contrastive discourse referents. This was unexpected because in Japanese, contrast is marked prosodically with an expanded pitch range similar to English, and also because in the prosody-in-context naturalness rating task, the same group of L2 participants were able to successfully associate L+H* with contrastive information regardless of proficiency. Why did the L+H*–contrastive mapping not surface in the L2 online comprehension? One possible reason is that task difficulty affects L2 learners' processing of prosody and its mapping to information status, especially for lower proficiency learners. One major difference between the prosody-in-context naturalness rating task and the eye-tracking task is that in the former, prosody-discourse mapping was the explicit focus of attention and participants were able to take as much time to make conscious decisions, while in the latter, contrastive focus was implicit and participants had to process linguistic information while performing visual search under time constraints. Although the mouse click accuracy in the eye-tracking experiment is more than 90% for both proficiency groups, overall fixation speed to the target object is clearly slower for the lower proficiency group than for the upper proficiency group, as we can see in Figure 6.10.

Despite the simple vocabulary and simple syntactic structure used in the experimental stimuli, those less proficient learners may still have experienced a processing burden due to their slower acoustic perception, lexical access, and syntactic parsing, not being able to make full use of prosodic cues available to them.

As for the upper proficiency learners, their fixation speed to the target differed very little from that of the native group in the new referent condition, which implies that those learners had no difficulty in comprehending the target sentences and performing the online visual search task. And yet, they still could not process prosodic information and its association with discourse referents in the same way as the native speakers did. This could be due to less efficient processing of the prosody-discourse interface in non-native listening than in native listening, as suggested in Akker and Cutler (2003). Akker and Cutler observed that L1-Dutch L2-English listeners were not able to exploit prosodic cues to focus structure in TL English even though when tested in Dutch, they in principle showed the same prosody-focus interaction effects as native English listeners. Our results are similar in that L1 Japanese listeners' rapid use of pitch in contrastive referent resolution has been previously reported in Ito, Jincho, Minai, Yamane, and Mazuka (2010), but the Japanese listeners in the current study nevertheless did not display the parallel processing patterns when a similar task was conducted in L2 English. Since our upper-level learners did not show the main effect of pitch type nor the interaction between pitch type and discourse status, we are uncertain whether those learners did not process the prosodic information itself, or whether they did process the prosodic information but had difficulty in the mapping between prosody and information status. The interpretation of contrastive prosody indeed involves complex processes. First, listeners need to detect acoustic cues to prominence in the TL (e.g., high pitch, longer duration, increased pitch excursion, etc.) and develop the semantic saliency of the relevant discourse entity in their discourse model. Then, by virtue of contrastive prosody, the listeners evoke a set of alternatives to the salient referent, and anticipate an upcoming contrast. If prosody for focus interpretation is less efficiently processed by non-native listeners than by native listeners, it is possible that the time constraints in the eye-tracking task prevented our L2 learners from successfully performing the processing of L+H*-contrastive information in the same way as they did in the prosody-in-context naturalness rating task.

In contrast to the L+H*–contrastive association, the processing of null accent–given (repeated) information may be less complex as a fewer steps are involved in the latter. When listeners detect lack of acoustic prominence (e.g., low pitch, short duration, reduced vowel quality, etc.), they should map the expression on to a set of discourse entities that are currently activated, and anticipate the given referent to appear. Given the relative simplicity of the deaccentuation–given referent association, our upper-level L2 learners, who demonstrated their knowledge of such mapping in the prosody-in-context rating task, may succeed in using null accent prosody for discourse given interpretations during an online task. In the next section, we present the new–given experiment and examine this possibility.

6.3 Experiment 2b: New–given experiment

6.3.1 Methods

Task

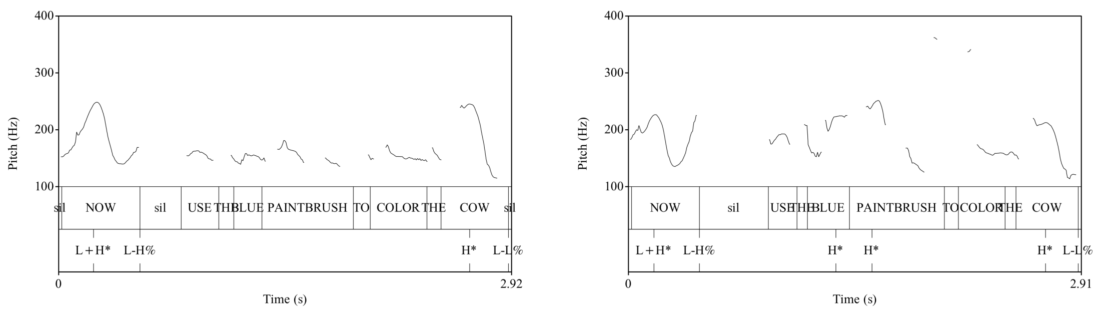
The same animal-coloring task as in the new–contrastive experiment was used in the new–given experiment.

Auditory and visual stimuli

Like the new–contrastive experiment, each trial consisted of two auditory instructions (e.g., *Use the green crayon to color the cow. Now, use the blue paintbrush to color the dolphin.*). In the second instruction, the target object was either new (new adjective and new noun) or given (given adjective and given noun). Within each of these two types of context, the pitch accent on the adjective was produced as either H* or null accent (Table 6.12). When H* was on the adjective, H* was also assigned to the following noun; when the adjective had no accent, the noun was also deaccented. Example sentences and their tonal contours are shown in Figure 6.14.

Table 6.12: Example of new vs. given experimental conditions

	Context	Pitch accent	Instruction 1	Instruction 2
(a)	New	Null accent	green crayon	blue paintbrush
				∅ ∅
(b)	New	H*	green crayon	blue paintbrush
				H* H*
(c)	Given	Null accent	blue paintbrush	blue paintbrush
				∅ ∅
(d)	Given	H*	blue paintbrush	blue paintbrush
				H* H*



(a) Null accent condition

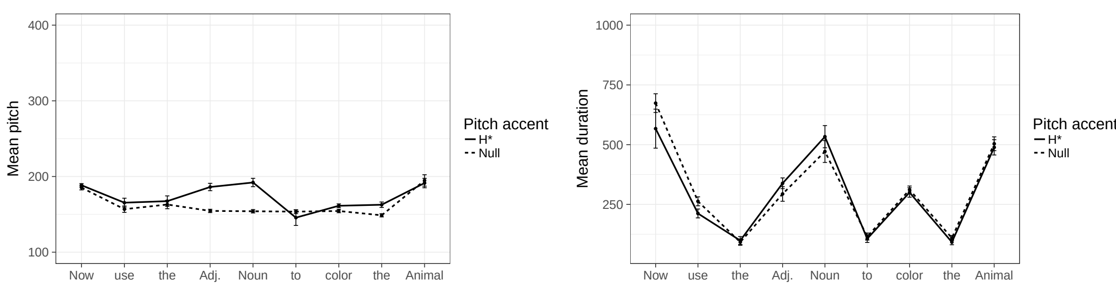
(b) H* condition

Figure 6.14: Example pitch contours of Instruction 2

Table 6.13 summarizes mean pitch, minimum and maximum pitch, pitch excursion, and mean duration of the target audio stimuli. Figure 6.15a and Figure 6.15b visualize mean pitch and duration in each word region.

Table 6.13: Acoustic analysis of Instruction 2

	Pitch accent	*Now* (& silence)	*use*	the_1	adjective	noun	*to color*	the_2	animal
F0 mean									
(Hz)	Null accent	185.33	156.83	163.00	154.56	154.09	154.03	148.76	194.82
	H*	188.46	165.43	167.46	186.16	192.10	153.50	162.70	191.37
F0 max									
(Hz)	Null accent	239.11	166.92	167.58	163.94	170.06	167.07	156.31	235.76
	H*	230.29	182.16	177.80	208.90	239.77	161.19	175.79	223.92
F0 min									
(Hz)	Null accent	132.16	145.50	156.72	144.54	144.93	148.16	140.77	131.25
	H*	147.02	141.95	158.00	146.27	129.51	145.03	149.10	142.20
F0 excursion									
(Hz)	Null accent	106.95	21.42	10.86	19.40	25.13	18.91	15.54	104.51
	H*	83.28	40.22	19.80	62.63	110.26	16.16	26.69	81.73
Duration									
(ms)	Null accent	673.75	262.92	92.08	293.75	471.25	213.33	109.58	503.75
	H*	567.08	212.08	98.75	338.33	533.33	203.13	92.08	488.75



(a) Mean pitch (b) Duration

Figure 6.15: Mean pitch and duration of Instruction 2

Mean pitch and durational values for each word of the experimental stimuli were submitted to a linear mixed effects regression analysis to further investigate acoustic differences between the H* and the null accent conditions.[22] Mean pitch differed significantly between H* and null accent in *use* [b = 8.60, $t(23)$ = 2.59, $p < .05$], adjective [b = 31.60, $t(23)$ = 12.00, $p < .001$], noun [b = 38.01, $t(23)$ = 13.07, $p < .001$], *color* [b = 6.90, $t(23)$ = 6.00, $p < .001$], and the_2

[22]We used the *lme* function in the R *nlme* package, and performed separate analysis for each word region using the following formula: lme(mean pitch $\sim$ pitch accent, random = $\sim$ 1 | item / pitch accent). Reference condition = null accent.

[b = 13.95, $t(23)$ = 6.62, $p < .001$]. Mean duration differed significantly in *now* [b = -106.67, $t(23)$ = -2.61, $p < .05$], *use* [b = -50.83, $t(23)$ = -6.58, $p < .001$], adjective [b = 44.58, $t(23)$ = 3.93, $p < .001$], noun [b = 62.08, $t(23)$ = 5.43, $p < .001$], and the_2 [b = -17.50, $t(23)$ = -5.12, $p < .001$] (see Appendix G, Table G.2 for the full regression table). This confirms that the H* condition was clearly more prominent than the null accent condition in both mean pitch and duration at the critical adjective-noun region.

These auditory stimuli were played while a visual display like Figure 6.16 was presented on a computer screen, as in the new–contrastive experiment. Besides the target object, there was only one other drawing implement that had the same color as the target implement (competitor). A sample item list can be found in Appendix H, Table H.2.

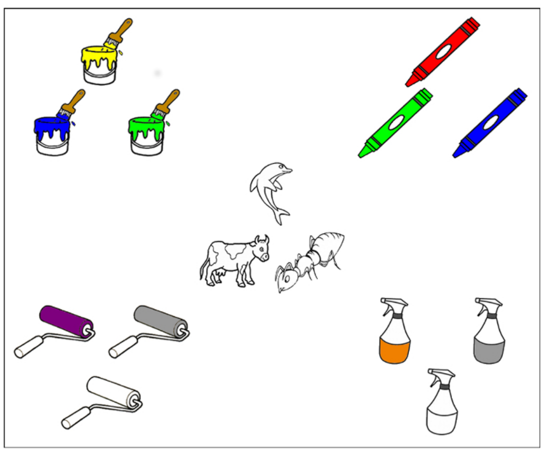


Figure 6.16: Example visual display

Procedure

The procedure was the same as in the new–contrastive experiment.

6.3.2 Predictions

If native speakers of English and L2 learners associate accentuation with new information and null accent with given information during the task, H* in the new context and null accent in the given context, i.e., conditions (b) and (c) in Table 6.14, should produce more looks to the target object.

Table 6.14: Predictions for the trial *blue paintbrush*

	Context	Pitch accent	Instruction 1	Instruction 2	Preductions
(a)	New	Null accent	green crayon	blue paintbrush ∅ ∅	infelicitous
(b)	New	H*	green crayon	blue paintbrush H* H*	felicitous
(c)	Given	Null accent	blue paintbrush	blue paintbrush ∅ ∅	felicitous
(d)	Given	H*	blue paintbrush	blue paintbrush H* H*	infelicitous

In Chapter 3, we described Terken and Noteboom's (1987) proposal that the processing mechanisms for accented and unaccented expressions are essentially different: The presence of an accent drives the listener's attention to acoustic details while the lack of an accent leads the listener to map the expression on to a set of discourse entities that are currently activated. If this claim holds true, as soon as our listeners hear H* on the new color in the new context condition (condition (b)), they should direct their attention to the acoustic details of the segmental information and shift their eye gaze quickly to the new target object (e.g., *green crayon* → *blue paintbrush*). When listeners hear null accent on the new color (condition (a)), on the other hand, they should initially search for the object of that color within a set of discourse entities that are already activated in their mind. It is only after the search fails that listeners redirect their attention to segmental cues, add a new referent to the current discourse set, and execute their eye gaze to the target object of that new color. This whole process should cause a delay in the identification of the target object, reflected as fewer looks to the target compared to what is expected in the

H*–new condition.[23] As for the given context, when listeners detect the lack of an accent on the given adjective (condition (c)), they consult the currently available discourse entities in their cognitive space, identify the matching referent, and fixate on the corresponding object in the visual display fairly quickly. When H* is on the given adjective (condition (d)), listeners first process the segmental information and try to add a new referent to the discourse model, only realizing that the referent has already been activated. This process may cause confusion and a delay in listeners' eye fixations to the given target object compared to when null accent is used.[24]

As for L1-Japanese L2 learners of English, the prosody-in-context rating task in Experiment 1 demonstrated that both upper and lower proficiency learners are able to associate null accent with given information. If they can apply their knowledge to online discourse processing, they should be able to identify the given target object more quickly when it is deaccented than when it is accented, and conversely detect the new target object more slowly when it is deaccented than when it is accented. If, on the other hand, the learners cannot make use of their knowledge during online discourse reference resolution, no prosody-by-discourse interaction is expected: Their eye-movements should pattern the same within each context condition, i.e., same fixation patterns between (a) and (b) and between (c) and (d), regardless of the two different types of pitch accent.

6.3.3 Results

Mouse click mean accuracy for the target trials was 99.66% (SD: 2.79, range: 97.92-100%) for the native group and 99.18% (SD: 6.40, range: 95.34-100%) for the L2 group. Mean accuracy for the filler trials was 98.56% (SD: 8.20, range: 93.48-100%) for the native speakers and 97.47% (SD: 13.34, range: 89.36-100%) for the L2 learners. All of the 51 participants had a mean

[23]While Terken and Noteboom's model presupposes a serial language-processing mechanism and a direct relationship between prosody and Information Structure, more recent research supports a probabilistic view (Calhoun, 2007, 2010a; Im et al., 2018). Calhoun (2007), for instance, proposes that information status is not directly signaled by prosody because the latter is also influenced by lexical, syntactic, and rhythmical constraints and other factors such as illocutionary force and emotion. Rather, likelihood of information status interpretation is determined by all these factors.

[24]It is possible that the lack of an accent on the color adjective triggers a noun contrastive interpretation (given adjective and new noun) rather than the given adjective-noun interpretation (given adjective and given noun). For instance, upon hearing *green crayon* $\rightarrow$ *green*$_{\emptyset}$...., with no accentuation on *green* in Instruction 2, listeners may look at the green paintbrush (noun contrastive) rather than the green crayon (noun given). Contrary to this expectation, Speer and Ito's (2011) spontaneous production study revealed that adjectives in noun contrastive utterances were marked with H* much more often than with no pitch accent. We therefore predict that listeners will be biased toward the given adjective-noun interpretation over the noun contrastive interpretation when they hear an unaccented prenominal adjective in Instruction 2.

accuracy of 90% or above, and therefore they were all included in the subsequent gaze data analyses.

As in the new–contrastive eye-tracking experiment, we first removed all the trials in which participants selected incorrect target objects, reframed the remaining gaze data from 4 ms to 50 ms time bins, coded fixations to the target and to the competitor as 1 or 0, and realigned the data from the onset of the critical noun where the segmental information starts to disambiguate the target object. We collapsed data from all the conditions and plotted a fixation logit graph (Figure J.3 in Appendix J) to identify the onset and the offset of a steady increase in fixation proportion to the target object. We then determined -100 to 900 ms to be the analysis window for empirical logit analyses. Models for this experiment were parallel to those used in the new–contrastive experiment: We specified pitch, context, and the interaction of pitch and context as fixed factors and random intercepts by participants and by items.[25]

Figure 6.17 presents L1 fixation proportions to the target and the competitor in the new and given conditions over a time window extending from -800 to 1000 ms (see Appendix I, Figure I.3, for fixation proportions to all the areas of interest). In the new context, the target was a new referent (e.g., *green crayon* → *blue paintbrush*) and its competitor shared the same color as the target (e.g., *blue crayon*). In the given context, the target was the same object as in the first instruction (e.g., *blue paintbrush* → *blue paintbrush*) and its competitor had the same color as the target (e.g., *blue crayon*).

[25]We used the following formula: lmer(elog ∼ pitch * context + (1 | participant), weights = 1/wts). Pitch and context were contrast coded independently: H* = -0.5, Null accent = 0.5; New = -0.5, Given = 0.5.

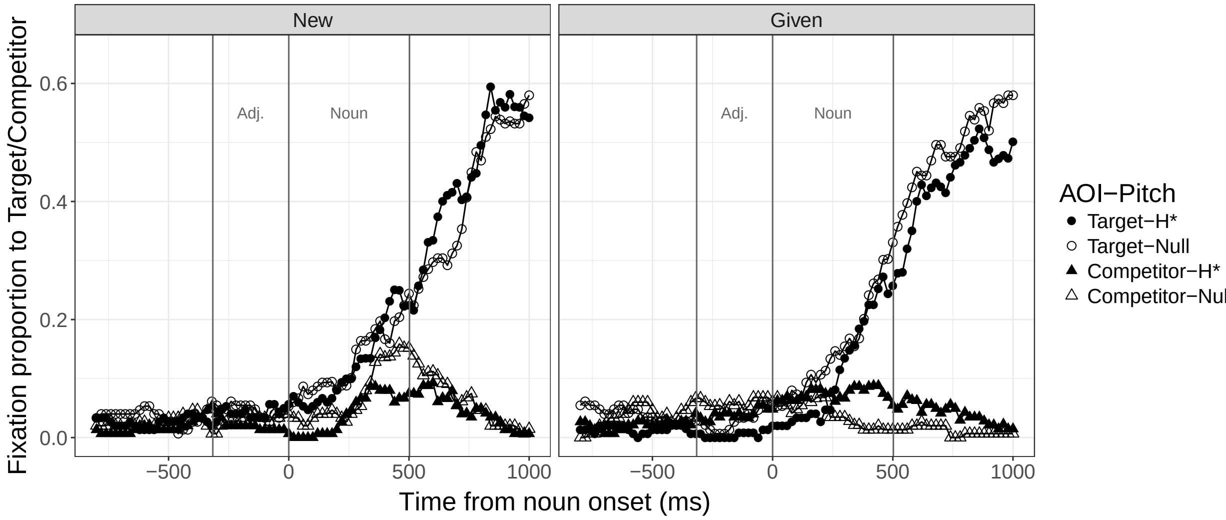


Figure 6.17: Fixation proportions to the target/competitor: L1 group

Visual inspection of the L1 graph revealed a slight null accent advantage over H* in proportion of looks to the target in the given context around -200 to 250 ms, while no difference in looks to the target appeared between the two pitch accent conditions in the new context. The effect of pitch accent on fixations to the competitor starts to emerge around 300 ms into the noun in both the new and given context conditions. In the new context condition, the proportion of fixations to the competitor increases more sharply at 300 ms with null accent than H*. In the given context, the fixation proportion to the competitor remains higher with H* than null accent, starting from 250 ms till 800 ms. The timing of divergence between the target and the competitor fixation lines is earlier with H* than null accent in the new context (approximately 300 ms vs. 400 ms), and only slightly earlier with null accent than H* in the given context (approximately 200 ms vs. 250 ms). The bar graph in Figure 6.18 shows the mean fixation proportion to the target within the analysis time window, and Table 6.15 summarizes the empirical logit model.

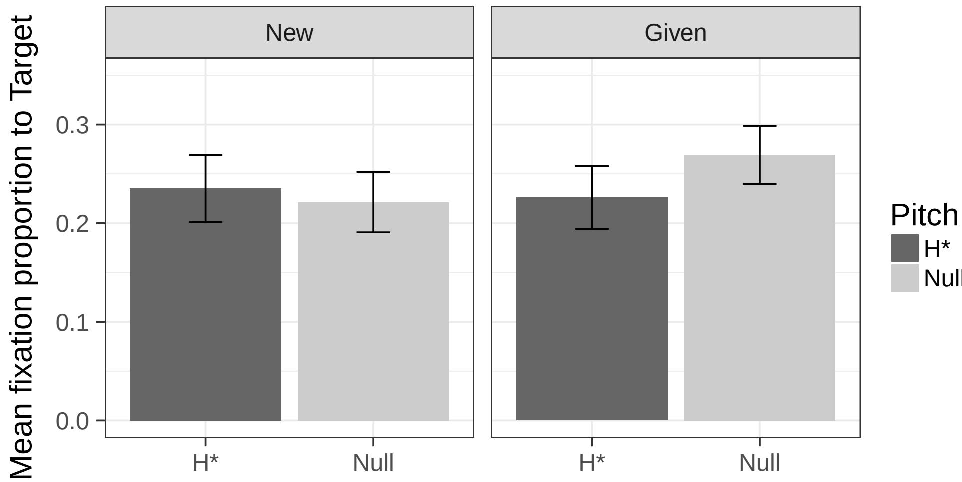


Figure 6.18: Mean fixation proportions to the target: L1 group

Table 6.15: Empirical logit model for the target: L1 group

L1 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.12	0.12	-9.50	$<$.001***	-1.12	0.07	-15.58	$<$.001***
Pitch	0.08	0.10	0.80	0.42	0.08	0.10	0.81	0.42
Context	0.09	0.10	0.85	0.39	0.07	0.10	0.74	0.46
Pitch : Context	0.41	0.20	2.02	$<$.05*	0.30	0.20	1.53	0.13

There was no main effect of pitch or context. The interaction between pitch and context was significant by participant analysis, but not by item analysis [b_1 = 0.41, t_1 = 2.02, $p_1 < .05$; b_2 = 0.30, t_2 = 0.20, p_2 = 0.13]. Inspection of the bar graph in Figure 6.18 suggests that the identification of the target was facilitated more with null accent than H* in the given context, while no clear difference between null accent and H* appeared in the new context. However, this observation was not confirmed in the post-hoc pairwise comparisons, which showed no statistical difference between any pairs of the four conditions. The significant pitch-by-context interaction effect in the main participant analysis thus provides only weak evidence for the effect of null accent on the interpretation of given referents for the L1 group.

Next, we examined fixation patterns for the competitor. In the same way as we analyzed the target looks, fixation data for the competitor were submitted to empirical logit models performed separately for participants and for items. The analysis window starting from 100 to 800 ms was determined based on the grand mean function (Figure J.4 in Appendix J). The L1 mean fixation proportion to the competitor within this time window is shown in Figure 6.19, and the statistical results are summarized in Table 6.16.

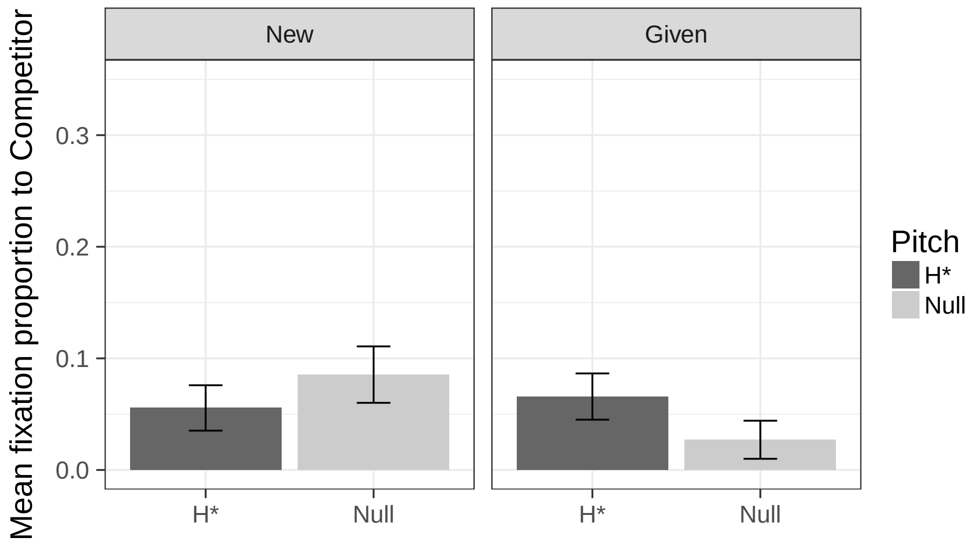


Figure 6.19: Mean fixation proportion to the competitor: L1 group

Table 6.16: Empirical logit model for the competitor: L1 group

L1 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-2.49	0.11	-22.26	<.001***	-2.64	0.11	-24.41	<.001***
Pitch	-0.21	0.18	-1.16	0.63	-0.07	0.15	-0.44	0.66
Context	-0.31	0.18	-1.74	0.35	-0.33	0.15	-2.11	<.05*
Pitch : Context	-1.08	0.36	-3.01	<.01**	-0.87	0.31	-2.76	<.01**

The model revealed a significant main effect of context with a negative coefficient by item analysis [b_1 = -0.31, t_1 = -1.74, p_1 = 0.35; b_2 = -0.33, t_2 = -2.11, $p_2 < .05$], which means fewer fixations on the competitor in the given context than in the new context. There was also a significant interaction effect between pitch and context by both participant analysis and item analysis [b_1 = -1.08, t_1 = -3.01, $p_1 < .01$; b_2 = -0.87, t_2 = -2.76, $p_2 < .01$]. Tukey's post-hoc comparisons indicated a statistical difference between H* and null accent in the given context by participant analysis [b_1 = 0.75, t_1 = 2.66, $p_1 < .05$; b_2 = 0.53, t_2 = 2.18, p_2 = 0.13], meaning that in the given context, the mean fixation proportion to the competitor was lower when the color adjective was deaccented (null accent) than when it was accented with H*. In the new context, the mean fixation proportion to the competitor was not statistically different between H* and null accent.

Summarizing the L1 results thus far, while null accent in the given context condition speeds up fixations to the given target only slightly, it does reduce incorrect looks to the new competitor. Null accent and H* did not show differing effects on fixations to the new target or the given competitor in the new context condition.

Turning now to the L2 results, fixation proportions for the target and the competitor are shown in Figure 6.20 (see Appendix I, Figure I.4, for fixation proportions to all the areas of interest). Visual inspection of the fixation line for the target in the new context reveals a slight null accent advantage over -250-200 ms, followed by a sharp increase with H* from 200 ms till 700 ms. In the given context, a small null accent advantage appears over -250-500 ms. The proportion of fixations to the competitor in the new context increases more rapidly with null accent than with H* around 400 ms, while in the given context, looks to the competitor increase faster with H* than with null accent starting around -200 ms. In addition, fixation proportion lines for the target and the competitor diverge earlier for H* than null accent in the new context (200 ms vs. 500 ms), and for null accent than H* in the given context (200 ms vs. 500 ms).

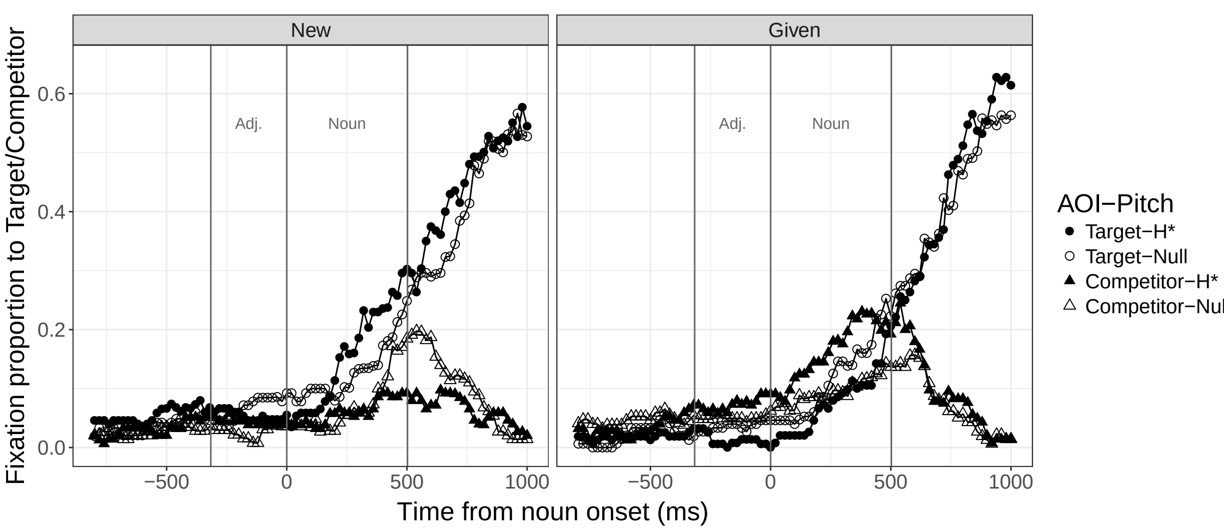


Figure 6.20: Fixation proportions to the target/competitor: L2 group

L2 fixation data for the target were submitted to empirical logit analyses, using the same critical time window (-100-900 ms) as in the L1 data analysis. The mean fixation proportion to the target within the critical time range, shown in Figure 6.21, indicates little difference between null accent and H* in both new context and given context, which was confirmed by the statistical analyses yielding no main effect or interaction of pitch and context by either participant analysis or item analysis (Table 6.17).

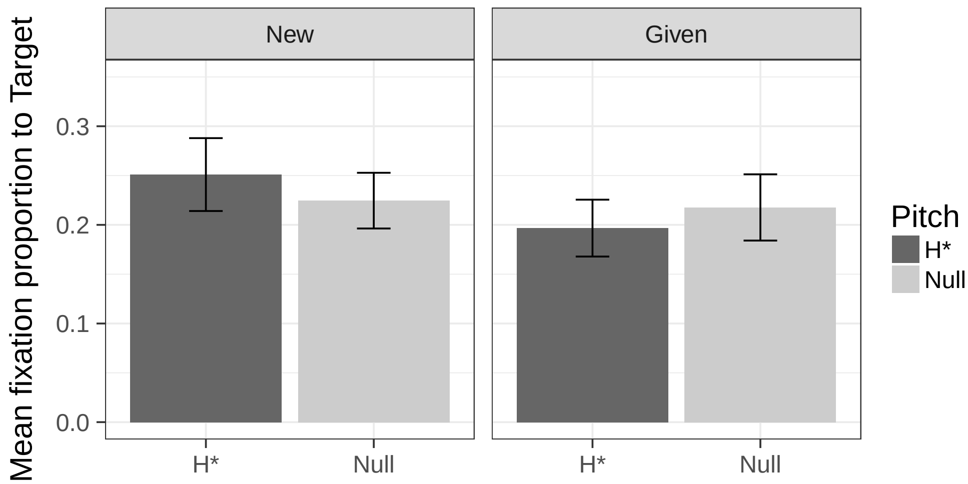


Figure 6.21: Mean fixation proportion to the target: L2 group

Table 6.17: Empirical logit model for the target: L2 group

L2 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.14	0.07	-15.57	<.001***	-1.20	0.08	-15.21	<.001***
Pitch	-0.03	0.11	0.80	-0.33	0.01	0.10	0.12	0.90
Context	-0.10	0.11	-0.91	0.39	-0.09	0.10	-0.88	0.38
Pitch : Context	0.23	0.21	1.06	0.29	0.22	0.20	1.12	0.26

Even when the L2 data were analyzed by proficiency, the null results remained the same. The fixation proportions to the target and the competitor for each proficiency group are shown in Figure 6.22. The mean proportion of fixations to the target over -100-900 ms in Figure 6.23 does not show any difference between the two pitch conditions in either context for either proficiency group, which was confirmed by statistical analyses (Table 6.18 and Table 6.19) showing no main effect or interaction between pitch and context for each proficiency group analyzed independently.

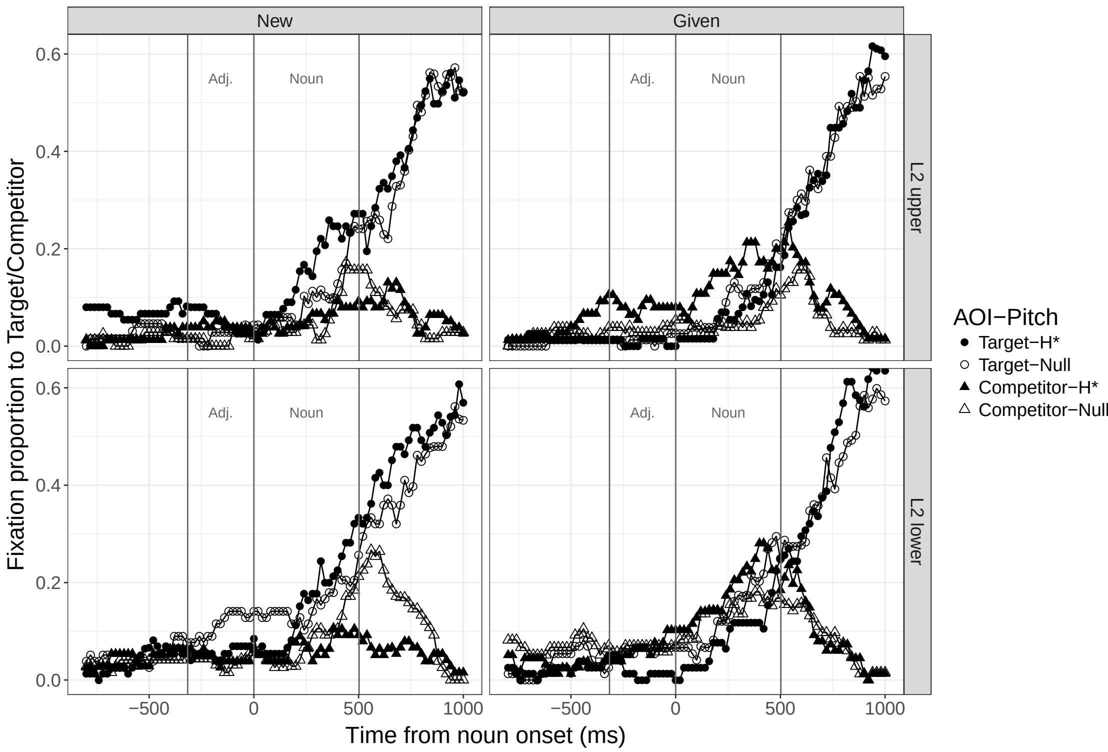


Figure 6.22: Fixation proportions to the target/competitor: L2 lower and upper groups

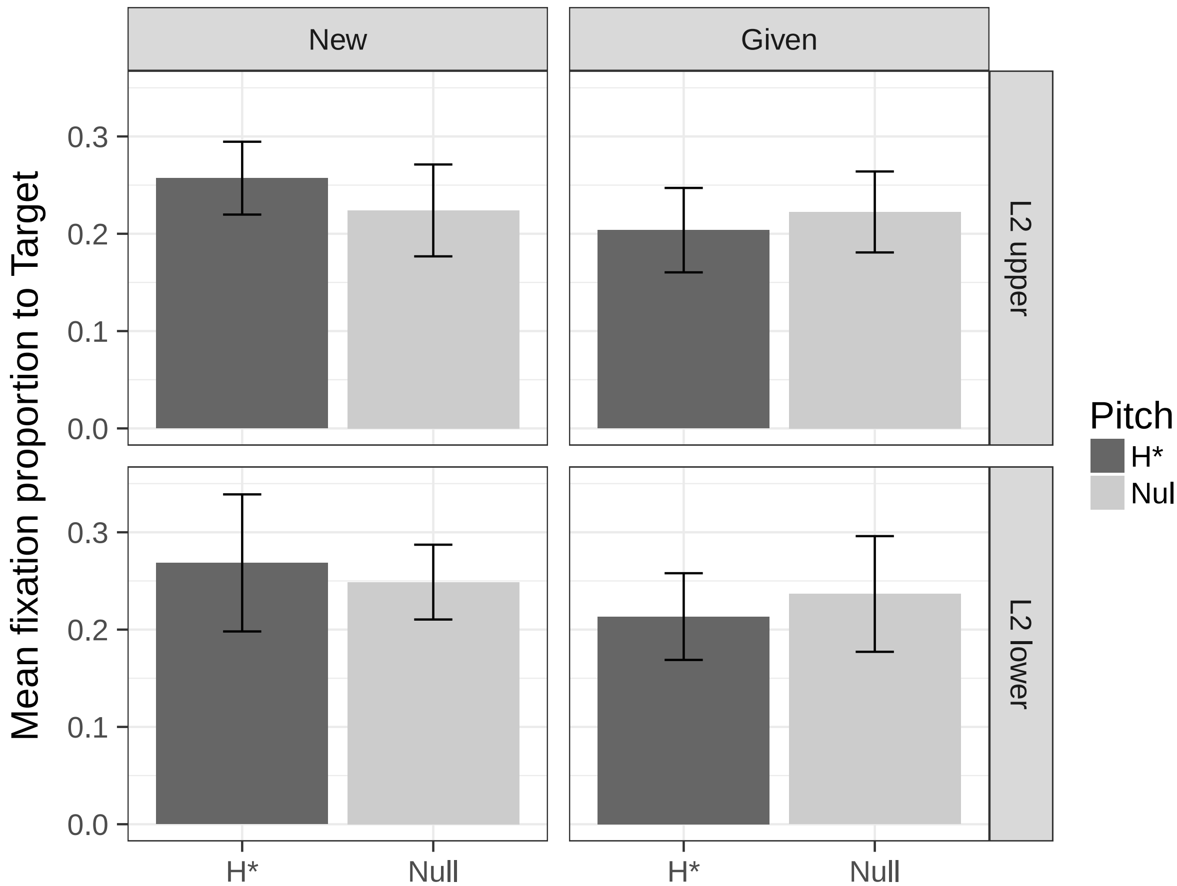


Figure 6.23: Mean fixation proportion to the target: L2 lower and upper groups

Table 6.18: Empirical logit model for the target: L2 lower group

L2 lower	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-0.98	0.09	-11.1	$< .001$***	-1.06	0.09	-11.23	$< .001$***
Pitch	-0.11	0.17	-0.63	0.53	0.04	0.13	0.36	0.72
Context	-0.13	0.17	-0.76	0.45	-0.02	0.13	-0.13	0.90
Pitch : Context	0.18	0.34	0.53	0.59	0.15	0.25	0.59	0.56

Table 6.19: Empirical logit model for the target: L2 upper group

L2 upper	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.26	0.09	-13.62	$< .001$***	-1.26	0.09	-13.88	$< .001$***
Pitch	0.01	0.14	0.09	0.93	0.00	0.12	0.01	1.00
Context	-0.06	0.14	-0.43	0.67	-0.13	0.12	-1.09	0.28
Pitch : Context	0.29	0.29	0.99	0.32	0.28	0.23	1.19	0.24

In contrast to the results for fixations to the target, the results for the competitor provide a clearer picture of fixation patterns in the L2 group. Figure 6.22 presents the L2 mean fixation proportion to the competitor over a time span of 100-800 ms. Empirical logit analyses summarized in Table 6.20 revealed a main effect of context by item analysis with a positive coefficient [b_1 = 0.25, t_1 = 1.51, p_1 = 0.13; b_2 = 0.32, t_2 = 2.54, $p_2 < .05$], suggesting that the given context condition induced more looks to the competitor than the new condition. There was also a significant interaction effect between pitch and context by item analysis [b_1 = -0.50, t_1 = -1.55, p_1 = 0.12; b_2 = -0.72, t_2 = -2.86, $p_2 < .01$]. Tukey's post-hoc tests verified a significant pairwise difference between null accent and H* in the given context, but not in the new context [b_1 = 0.35, t_1 = 1.65, p_1 = 0.36; b_2 = 0.47, t_2 = 2.84, $p_2 < .05$]. What these results imply is that when L2 listeners hear the same color again in the second instruction, they are more likely to fixate on the object of that same color that was not mentioned in the first instruction (competitor), and this pattern was more notable when the color adjective in the second instruction was marked with H* than with null accent. In the new context where the target object in the second instruction had a different color from the object mentioned in the first instruction, the difference in pitch type in the second instruction did not influence fixations to the competitor.

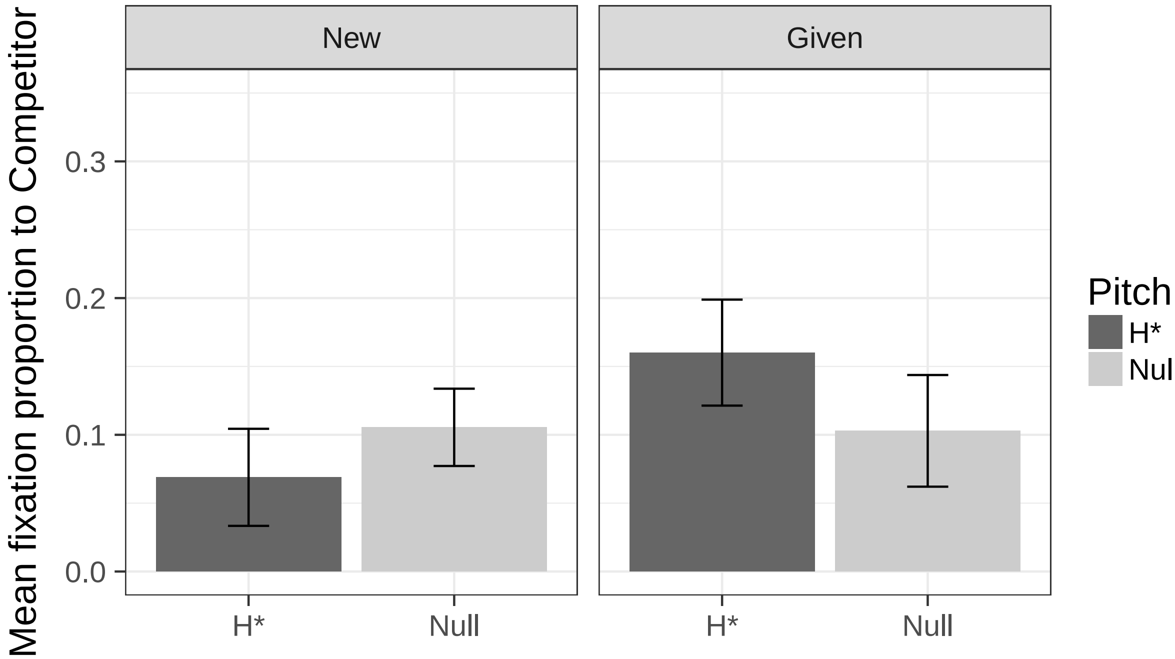


Figure 6.24: Mean fixation proportion to the competitor: L2 group

Table 6.20: Empirical logit model for the competitor: L2 group

L2 group	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.94	0.11	18.37	< .001***	-2.15	0.10	-22.58	< .001***
Pitch	-0.10	0.16	-0.61	0.54	-0.11	0.13	-0.91	0.36
Context	0.25	0.17	1.51	0.13	0.32	0.12	2.54	< .05*
Pitch : Context	-0.50	0.32	-1.55	0.12	-0.72	0.25	-2.86	< .01**

When we analyzed the L2 data for the competitor by proficiency (Figure 6.25), a context effect appeared in the L2 lower group. Context was statistically significant with a positive coefficient by item analysis (Table 6.21), confirming the lower-level L2 learners' bias towards fixating on the alternative (non-mentioned) object of the target color when the target object is repeated from the first to the second instruction.

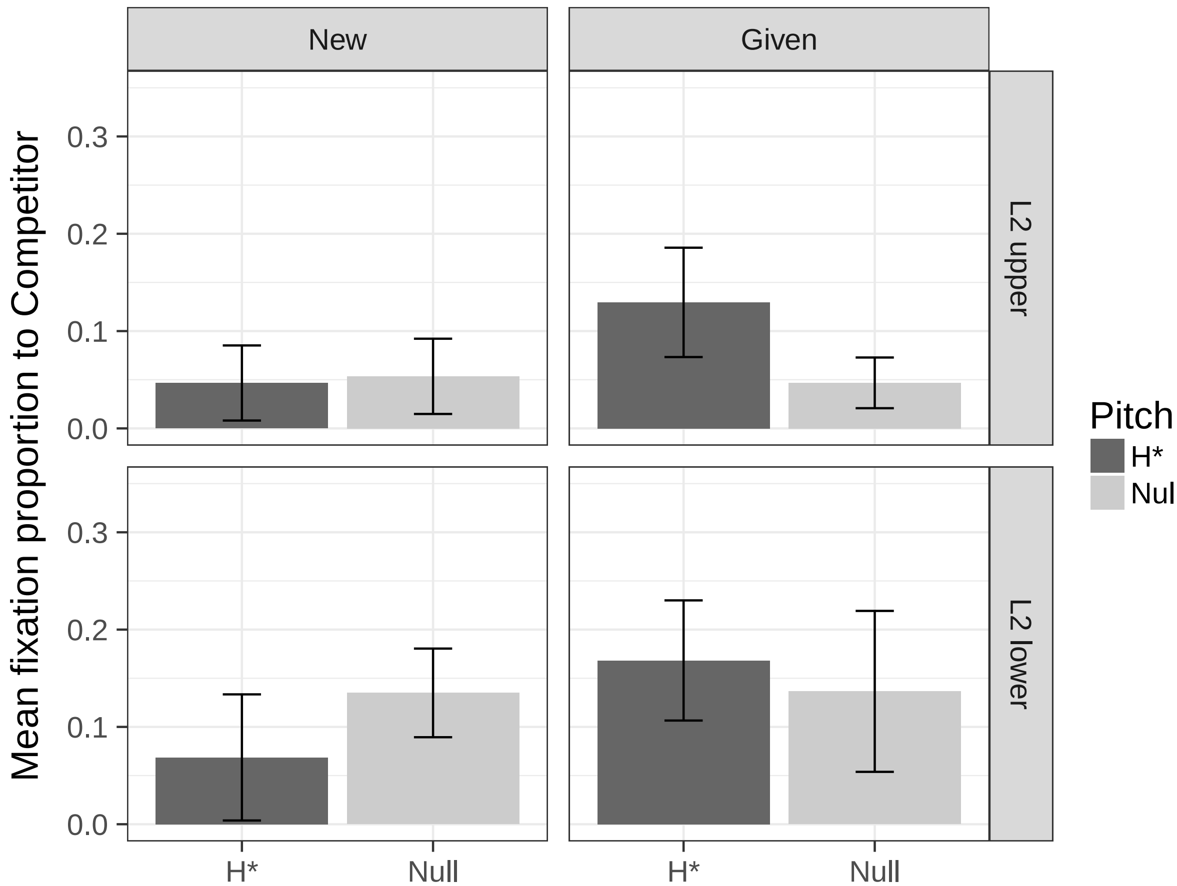


Figure 6.25: Mean fixation proportion to the competitor: L2 lower and upper groups

Table 6.21: Empirical logit model for the competitor: L2 lower group

L2 lower	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-1.77	0.17	-10.47	$< .001$***	-1.87	0.11	-16.29	$< .001$***
Pitch	0.13	0.24	0.53	0.60	0.08	0.16	0.48	0.63
Context	0.33	0.25	1.32	0.19	0.39	0.16	2.45	$< .05$*
Pitch : Context	-0.05	0.48	-0.09	0.92	-0.48	0.31	-1.54	0.12

In contrast to the results for the L2 lower group, results for the L2 upper group provide some evidence for the effect of pitch accent. Analysis (Table 6.22) indicated a main effect of context with a positive coefficient by item analysis [b_1 = 0.16, t_1 = 0.76, p_1 = 0.44; b_2 = 0.36, t_2 = 2.22, $p_2 < .05$], replicating the results for the lower group that showed a bias toward the unmentioned object of the repeated color in the given context condition. What differed from the lower group is that this bias was reduced when the color adjective in the second instruction was deaccented than when it was accented with H*. This is reflected in a significant interaction between pitch and context by participant analysis [b_1 = -1.02, t_1 = -2.39, $p_1 < .05$; b_2 = -0.54, t_2 = -1.64, p_2 = 0.10].

Table 6.22: Empirical logit model for the competitor: L2 upper group

L2 upper	Participant analysis				Item analysis			
	Est.	SE	*t*	*p*	Est.	SE	*t*	*p*
(Intercept)	-2.11	0.11	-19.75	<.001***	-2.08	0.10	-21.69	<.001***
Pitch	-0.41	0.21	-1.91	0.05	-0.32	0.17	-1.95	0.05
Context	0.16	0.21	0.76	0.44	0.36	0.16	2.22	< .05*
Pitch : Context	-1.02	0.43	-2.39	< .05*	-0.54	0.33	-1.64	0.10

To sum up the L2 results, our L2 lower-level learners showed no evidence for online use of prosodic cues when they process discourse referents. Instead, they demonstrated a bias due to the new vs. given status of referents. When the referent was repeated in a series of instructions, the lower-level learners displayed a tendency to look at the same-colored object that was not mentioned in the first instruction. The same bias was observed in the L2 upper-level learners; however, preferential looks to the non-mentioned object decreased when the modifying adjective had no pitch accent compared to when it carried H*. These patterns derive from the fixation data for the competitor as in the L1 group, and no clear effect of discourse status, pitch, or their interaction was found in the fixation data for the target.

6.3.4 Discussion

The new–given experiment compared online use of H* and null accent in anticipating and recognizing discourse new vs. given referents. The L1 group showed some facilitative effect of null accent on the detection of given target referents. Eye-movements to the target showed a significant pitch-by-context interaction in the by-item analysis which, in combination with visual observation of the mean fixation proportion graph, indicated more fixations on given referents when the prenominal adjective was deaccented than when it was marked with H*; however, this was not confirmed in post-hoc statistical comparisons. Eye-movements to the competitor, on the other hand, revealed a statistically robust effect of null accent. In the given context condition where the target was a given referent (*blue paintbrush* → *blue paintbrush*) and the competitor was a new entity bearing the same color as the target (*blue crayon*), null accent on the prenominal adjective induced fewer fixations to the competitor than H* did.

One possible reason for the reduced effect of null accent in given target trials is that our relatively simple display led to a ceiling effect in fixations to the target. As mentioned in the discussion of the new–contrastive experiment, Ito and Speer (2008), whose study revealed facilitative effects of L+H* on L1 English speakers' processing of contrast, used a visual display containing 11 cells with three to five objects in each so that there were multiple competitors of the same color as the target. This visual complexity likely prevented participants from establishing a display-oriented referential bias and guessing the next target object based solely on visual information. Our visual display contained four sets of objects with three drawing instruments in each set placed in each corner, plus animals in the center, and there was only one competitor object that had the same color as the target. In the given referent condition, as participants heard the color in the first instruction (*blue* in *blue paintbrush*), both the target (*blue paintbrush*) and the competitor (*blue crayon*) must have been fixated on in the visual display and activated in listeners' discourse model. When the repeated color in the second instruction (*blue*) was deaccented, listeners interpreted the pitch cue as a signal for given information and directed their eye gaze immediately to the previously mentioned object (*blue paintbrush*) without looking at the competitor of the same color (*blue crayon*). When, on the other hand, the repeated color in the second instruction was accented, listeners momentarily shifted their eye gaze to the competitor, i.e., the non-mentioned object of the same color (*blue crayon*), yet quickly redirected their gaze back to the given target object (*blue paintbrush*). The simple display design may have made this attention shift easy and quick, causing no delay in fixations to the given target object while reflecting brief looks to the competitor when the pronominal adjective was produced with the H* accent.

The facilitative effect of pitch accent was also confirmed in the L2 upper proficiency group. Similar to what happened in the native group, fixations to the target were not affected by pitch accent, while fixations to the competitor indicated that incorrect looks to a new referent were suppressed when the prenominal adjective had no accent compared to when it had H*. It is likely that null accent on the adjective kept listeners' eye gaze on the same discourse entity, while H* led their gaze to the competitor, i.e., the object that had the same color as the target but was not mentioned in the first instruction. Even when eye gaze was diverted (wrongly) to the new competitor due to infelicitous H*, the simple visual display likely made it easy for the L2 upper listeners to quickly shift their gaze back to the given target without a significant delay.

In contrast to the L2 advanced group, the lower proficiency group showed no evidence of pitch effects during the online processing of discourse referents. Their mouse click accuracy was 99.16%, which means they had no problem understanding segmental information and completing the task. Despite the high comprehension accuracy, neither information status nor pitch type affected their fixations to the target. Fixations to the competitor demonstrated a bias for a non-repeated object (i.e., object that was not referred to in the first instruction) regardless of pitch accent. The bias due to the given vs. new status of the visual stimuli has been reported previously in Dahan et al. (2002). In their eye-tracking study, native English listeners hearing sentences such as *Put the candy below the triangle. Now put the candle/candy above the square* fixated on the target object more when it was new information (*candle*) than when it was given information (*candy*) before segmental and suprasegmental information of the critical noun (*can-*) comes into play. Similarly, the listeners fixated more on the competitor when it was new information than when it was given information. This means that listeners are initially biased toward fixating on a picture associated with a new entity, trying to guess the next target object based on the visual contrast within the display. The same bias seems to appear in our L2 upper proficiency group as well, since in the analysis of fixations to the competitor, they showed a main effect of context reflecting preferential looks to non-repeated discourse entities. It could be that L2 learners tend to rely more on visual than verbal cues in anticipating and detecting upcoming referents. While the bias toward non-repeated referents was the only effect found in the L2 lower group, the L2 advanced group demonstrated a significant interaction effect between context and pitch accent as well, suggesting that this bias towards non-repeated discourse entities was reduced when the critical region was deaccented compared to when it was accented with H*.

In summary, the new–given experiment indicated that L2 advanced learners are able to acquire the null accent–given mapping, which is not instantiated in their native language, and use such knowledge to facilitate real-time processing of discourse referents. By contrast, L2 lower-level learners are not sensitive to English pitch cues conveying information status of discourse entities. Rather, they tend to rely on non-linguistic visual contrasts and favor new discourse referents that have not been mentioned previously. These findings provide an additional piece of evidence that the degree to which L2 learners make use of discourse prosody in the TL largely depends on their language proficiency level. Furthermore, processing difficulty also seems to affect L2 learners' sensitivity to prosodic marking of discourse status. In the discussion section of the new–contrastive eye-tracking experiment, we made a proposal that the processing of null accent–given (repeated) information may be less complex and thus easier for L2 learners to execute than the processing of L+H*–contrastive information. Even though the lower-level learners were highly accurate in clicking correct target objects in both new–contrastive and new–given experiments, the overall fixation speed in the new–contrastive experiment was slower than that in the new–given experiment, which suggests that the mapping of null accent–given information imposes less processing demand for non-native listeners. As we mentioned already, this may also be a reason why our upper-level learners could not exploit contrastive prosody in the new–contrastive experiment, while they successfully made use of deaccentuation to help identify given referents in the same way as native speakers did in the new–given experiment.

In the next chapter, we examine whether the same group of L2 learners (and native speakers) mark information status using prosody during a semi-spontaneous dyadic speaking task. If language proficiency and processing difficulty play essential roles in production as well as in comprehension of discourse prosody, we expect to see outcomes similar to those in the eye-tracking experiment: The upper-level learners will have more difficulty indicating contrastive referents than given referents, while the lower-level learners will mark neither given nor contrastive referents prosodically. If, on the other hand, the L2 learners can apply their knowledge of the mappings between null accent and given information and between L+H* and contrastive information as they did in the rating task, they should be able to utilize prosody to indicate contrastive referents and given referents in production.

CHAPTER 7
EXPERIMENT 3: PRODUCTION TASK

7.1 Introduction

As described in Chapter 3, a number of studies investigating L2 prosody identified influence from L1 prosody in the production of L2 discourse marking. In the course of the L2 development, overuse of accentuation is typical of the early stages of L2 acquisition. Rasier and Hiligsmann (2007) suggest that L2 learners tend to emphasize almost every single word due to the difficulty in distinguishing the given vs. new information status of discourse referents.[26] It is not until later stages that L2 learners restructure their interlanguage to approximate to the target language norm. Investigation into utterance-level accentuation (by natives and) by L2 learners at beginning and advanced proficiency should help to reveal the processes by which L2 learners both develop sensitivity to discourse structure and make use of prosody for conveying such information.

Experiment 3 examines whether speakers use prosody to reliably indicate information status during a semi-spontaneous speaking task which is the production counterpart of the eye-tracking animal-coloring comprehension task: Native and non-native speakers of English provided oral instructions to a confederate, specifying which instrument to use to color which animal. In the prosody-in-context naturalness rating task, both upper- and lower-level learners demonstrated their knowledge of the null accent–given mapping and the L+H*–contrastive mapping. In the eye-tracking listening comprehension task, while the upper-level learners made use of the link between null accent and given information, neither they nor the lower-level learners were able to draw on the link between L+H* and contrastive information. We designed Experiment 3 to be parallel to Experiment 2 so as to compare the findings from the two experiments with the same group of learners and to gain a better understanding of the relationship between production and comprehension of discourse prosody within the same learner population.

[26]We assume in our study that Information Structure is a universal phenomenon in natural language, while the ways in which information status is encoded vary from one language to another. See Section 7.5.

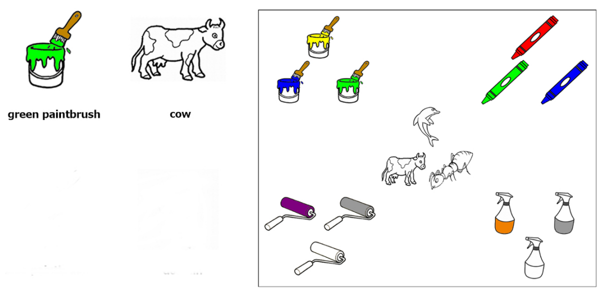
7.2 Methods

7.2.1 Participants

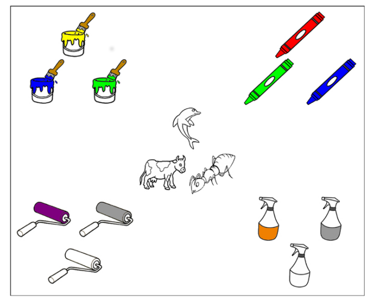
The participants were the same 70 native speakers of English and 64 Japanese-speaking L2 learners of English (34 lower-level and 30 upper-level learners) who completed the prosody-in-context naturalness rating task and the eye-tracking listening comprehension task in Experiments 1 and 2. The production task was in actuality conducted prior to the other tasks in order to avoid any carry-over effect from the audio stimuli used in the listening tasks to the participants' performance in the speaking task.

7.2.2 Visual stimuli and procedure

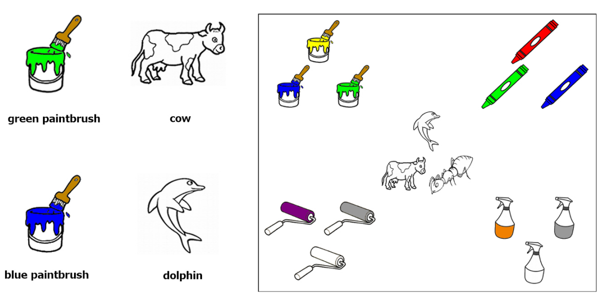
In this dyadic speaking task, participants were seated in front of a computer approximately 10 feet away from a confederate, sitting at a different computer so they could not see each other's computer display. An experimental trial consisted of participants seeing a series of two slides, as in Figure 7.1. First, participants saw a display with an image of a drawing tool plus an animal with a label under each image on the left side of the screen. The right-hand side was the same visual display as presented to the confederate, which contained three uncolored animals located in the center and four sets of drawing tools (paintbrush, crayon, roller, spray bottle) in eight different colors (blue, green, orange, purple, red, silver, white, yellow) located in the four corners (Figure 7.1a). On the second slide, another pair of drawing instrument and animal appeared as shown in Figure 7.1c. The display for the confederate remained unchanged within each trial (Figure 7.1b and Figure 7.1d).



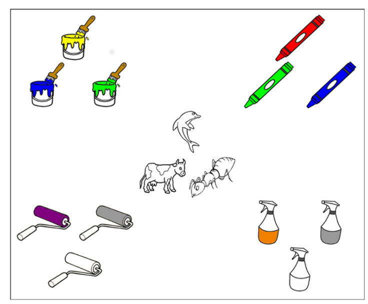
(a) Instruction 1: Participant



(b) Instruction 1: Confederate



(c) Instruction 2: Participant



(d) Instruction 2: Confederate

Figure 7.1: Example visual display in the production task

Using the visual stimuli with particular drawing instruments and animals, participants gave instructions to the confederate, directing her to color the designated animals using the designated instruments. The confederate was a native speaker of Japanese at an advanced level of English proficiency, and she kept interactions with all participants to a minimum in order to avoid any influence on their task performance. During the practice trials, model sentences (e.g., *Use the green paintbrush to color the cow. Now, use the blue paintbrush to color the dolphin.*) appeared at the bottom of the screen to help participants, L2 learners in particular, construct sentences. Because the prosody of the adjective-noun sequence was the object of enquiry, participants were instructed to articulate clearly, making sure to name drawing tools and their color as well as the animals; this was done to minimize the use of pronouns instead of full NPs [Det + Adj + N] (see Appendix K for the actual directions given to participants for this task). In order to elicit the participants' most natural discourse prosody, no model sentences were displayed during the experimental trials.

We systematically varied the sequential order of drawing tools and their colors presented on the two presentation slides in order to elicit targeted descriptions from the participants in three contexts: new (different color and different tool), adjective contrastive (different color but same tool), and given (same color and same tool), as laid out in Table 7.1. Since the NP containing a color adjective and a noun in the second instruction (e.g., *blue paintbrush*) was the critical region to be submitted to acoustic analyses, we kept this adjective and noun sequence identical across the three conditions and manipulated the color adjective and the noun in the first instruction (e.g., *green crayon*, *green paintbrush*, or *blue paintbrush*). Eight sets of experimental stimuli were created and distributed in a Latin square design across three lists. Each list was combined with six practice trials, yielding 30 trials in total (24 experimental items, 6 practice items, and no fillers). A sample item list can be found in Appendix L, Table L.1.

Table 7.1: Examples of three conditions for the trial *blue paintbrush*

	Conditions	Instruction 1	Instruction 2
(a)	New	green crayon	blue paintbrush
(b)	Adjective contrastive	green paintbrush	blue paintbrush
(c)	Given	blue paintbrush	blue paintbrush

7.3 Predictions

Based on the findings from the past L1 English production/corpus studies (Hedberg & Sosa, 2008; Sityaev, 2000; Terken & Hirschberg, 1994), we predict that for our native controls, new information will be produced with H* or accentuation with elevated pitch and elongated duration, and contrastive information with L+H* (higher mean F0 and larger F0 excursion for contrastive information than for new information). As for given information, since the given elements in our test stimuli are repeated nouns that appear in the same syntactic position bearing the same grammatical role as the previous mention, we expect native speakers to produce given information with deaccentuation (low mean F0 and reduced duration).

As for the L2 learners, if they (especially lower-level learners) are more susceptible to L1 influence in production than in comprehension/perception, they should mark contrastive referents with L+H* or something similar, but neither given nor new referents will be marked prosodically. If, on the other hand, both upper- and lower-level learners can apply their knowledge of prosodic marking of given information and contrastive information (as demonstrated in the rating task) to TL production, they should indicate given referents and contrastive referents with deaccentuation and L+H*, respectively. Furthermore, if language proficiency and processing difficulty matter in semi-spontaneous production as well as in the eye-tracking listening comprehension task, we expect to see some correlation between the production task and the comprehension task: The upper-level learners will have more difficulty indicating contrastive and new referents than given referents, while the lower-level learners will not mark given, new or contrastive referents prosodically.

7.4 Results

7.4.1 Acoustic analysis

We recorded a total of 1680 target descriptions from the L1 group and 1536 from the L2 group during the experiment. Of those, utterances that contained disfluency or self-correction at the critical NP were removed, which left 1676 L1 utterances and 1520 L2 utterances (712 utterances for the upper proficiency group and 808 utterances for the lower proficiency group) for data analysis. They were first segmented word-by-word using the Prosodylab-Aligner (Gorman, Howell, & Wagner, 2011) and readjusted by a native speaker of English who had training in English phonetics and phonology. Then for acoustic analyses, mean pitch, pitch excursion (maximum pitch minus minimum pitch), and duration of the prenominal adjective in the second instruction were measured using Praat and submitted to mixed effects linear regression models with context as a fixed factor, participants and items as random factors on the intercept, and context as a random subject and item factor on the slope.[27] As we expect the effect of discourse context on prosody to appear most clearly on the adjective, our discussion in this section focuses on the adjective region only. Statistical results of the noun region are provided in Appendix M.[28]

[27]We used the following formula with the *lmer* function in the *lme4* package: lmer(mean pitch ∼ context + (1 + context | subject) + (1 + context | item)). Context was Treatment coded with the given condition as the base level.

[28]For acoustic analyses of the noun, we used the same formula as the one for the adjective region, except that the dependent variable was mean pitch, pitch excursion or duration of the noun.

In the following pitch graphs, individual variability in pitch was adjusted by subtracting mean pitch values of the initial verb *use* from the raw mean pitch of the prenominal adjective region within the same utterance. The mean pitch for the L1 group in Figure 7.2 shows pitch increasing from the given condition to the new condition, and then again to the adjective contrastive condition at the adjective region. The mean pitch values in the new condition and the adjective contrastive condition were indeed significantly higher than the mean pitch in the given condition [$b = 7.87$, $t = 3.44$, $p < .01$ for given vs. new; $b = 13.77$, $t = 6.33$, $p < .001$ for given vs. adjective contrastive]. Furthermore, Tukey's post-hoc pairwise comparisons revealed significantly higher mean pitch for the adjective contrastive condition than for the new condition. [$b = 5.90$, $z = 3.07$, $p < .01$]. These results confirmed the three-way distinction among the given, new, and adjective contrastive conditions with increasing mean pitch. In the L2 graph, the mean pitch values for the new condition and the adjective contrastive condition appear higher than the mean pitch for the given condition at the adjective region. Statistical analyses confirmed that the mean pitch for the new condition was significantly higher than that for the given condition [$b = 3.68$, $t = 2.07$, $p < .05$], while the difference between the adjective contrastive condition and the given condition was marginally significant [$b = 4.01$, $t = 1.88$, $p = .06$]. The post-hoc test revealed no significant difference between the new and contrastive conditions.[29]

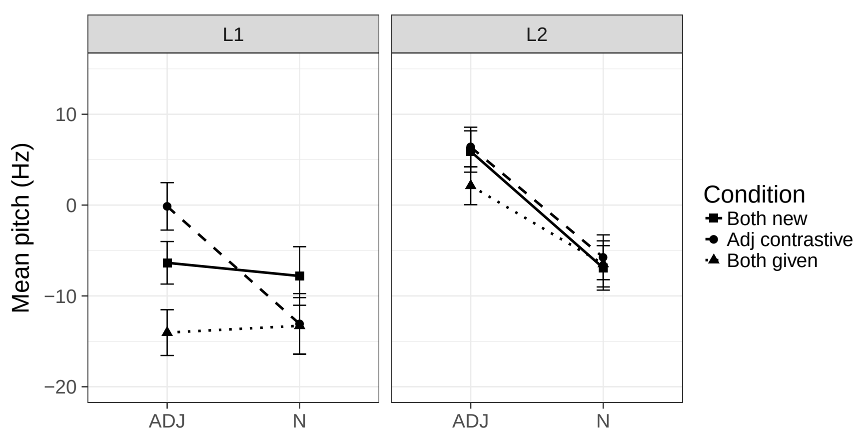


Figure 7.2: Mean pitch for the adjective + noun region

[29]A statistical analysis including language group as a fixed factor was also conducted using the following formula: lmer(mean pitch ∼ context * language group + (1 + context | subject) + (1 + context | item)). Context and language group were Treatment coded with the given condition and the L1 group as the base levels. The results showed a main effect of language group [$b = 31.95$, $t = 3.80$, $p < .001$] as well as an interaction between context and language group [$b = -9.82$, $t = -3.43$, $p < .001$]. We then conducted further statistical analysis for each language group separately.

When the L2 data are analyzed by proficiency (Figure 7.3), a different pattern emerged. For the upper L2 group, the difference between the adjective contrastive condition and the given condition was significant [$b = 7.18$, $z = 2.33$, $p < .05$], while no difference was found between the adjective contrastive and new conditions or between the new and given conditions. This suggests that the upper level learners were able to distinguish contrastive information from given information in their production. For the L2 lower proficiency group, no pitch difference was found between any pairs of the three context conditions.

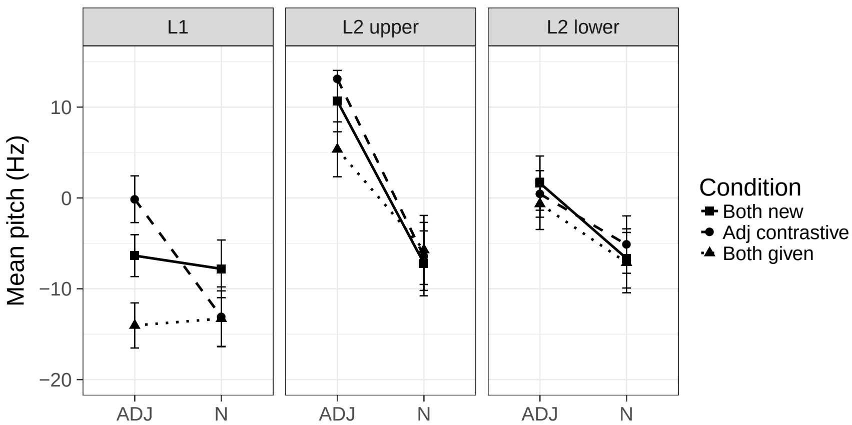


Figure 7.3: Mean pitch for the adjective + noun region by L2 proficiency

As for pitch excursion, which was calculated by subtracting the minimum pitch from the maximum pitch within the critical region for each participant and then averaged within the group, the L1 graph in Figure 7.4 appears to show larger pitch excursion for the adjective contrastive context than for the new or given contexts at the adjective region. However, the main analysis revealed no significant difference among the three context conditions. The L2 group seems to show larger excursion for the new context than for the adjective contrastive or given contexts, but as with the L1 group, no significant difference was found in the statistical analysis.[30]

[30]No statistical difference between the L1 and L2 groups was found when language group was included in the analysis as a fixed factor in the formula: lmer(pitch excursion ∼ context * language group + (1 + context | subject) + (1 + context | item)).

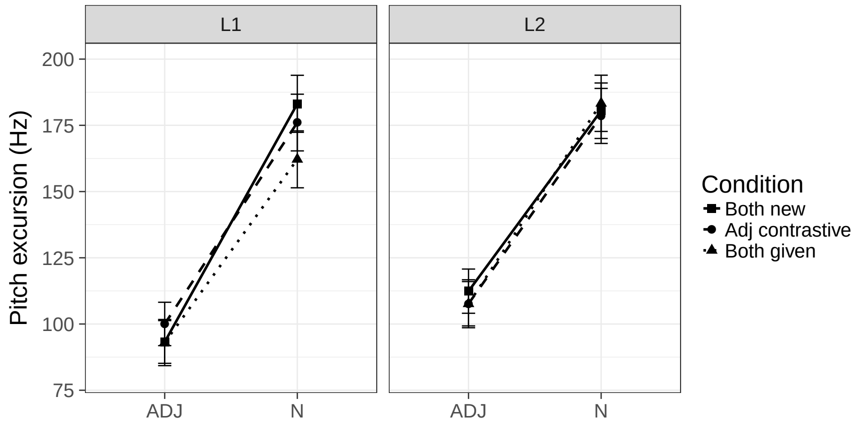


Figure 7.4: Pitch excursion for the adjective + noun region

Even when the L2 data were analyzed by proficiency (Figure 7.5), the results remained the same. Pitch excursion was not statistically different among the new, adjective contrastive, and given conditions for either the upper or lower proficiency groups.

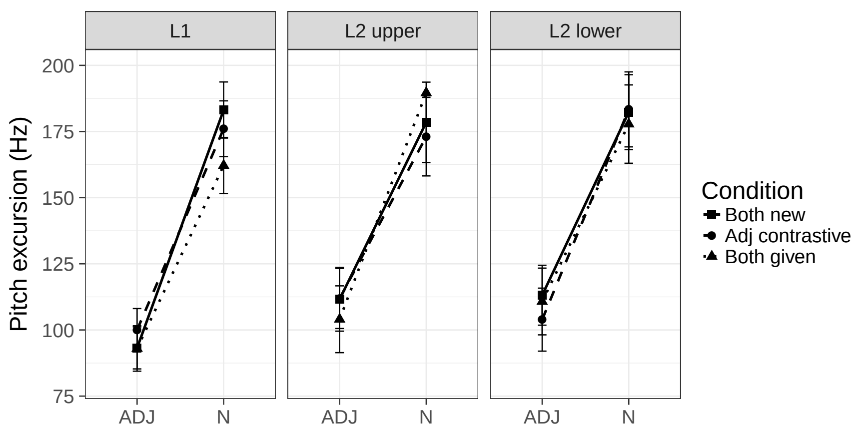


Figure 7.5: Pitch excursion for the adjective + noun region by L2 proficiency

Analysis of duration (Figure 7.6) revealed that for the L1 group, lengthening of the adjective region was more prominent in the new and adjective contrastive conditions than in the given condition [b = 26.79, t = 5.58, $p < .001$ for new vs. given; b = 16.96, t = 3.57, $p < .001$ for adjective contrastive vs. given]. The post-hoc analysis showed only marginally longer duration for the new condition than for the adjective contrastive condition [b = -9.83, t = -2.10, $p =$ .09]. These results provide robust evidence that L1 speakers reliably mark new and contrastive information with increased duration compared to given information. There was also a tendency that L1 speakers lengthen new information more than contrastive information. A similar pattern was observed in the L2 data. Both new and adjective contrastive conditions induced significantly longer duration of the adjective than in the given condition [b = 20.90, t = 3.30, $p < .01$ for new vs. given; b = 13.66, t = 2.03, $p < .05$ for adjective contrastive vs. given] in the main analysis, and the post-hoc comparisons revealed no significant difference between the new condition and the adjective contrastive condition.[31]

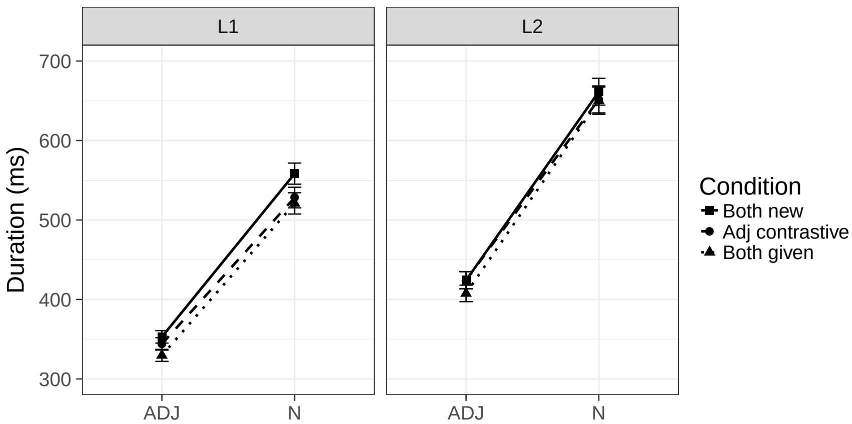


Figure 7.6: Mean duration for the adjective + noun region

Analysis by proficiency (Figure 7.7) indicated that the L2 upper group produced new information with significantly longer duration than given information [b = 22.86, t = 2.43, $p < .05$], and adjective contrastive information was marginally longer than given information [b = 17.26, t = 1.85, $p = .07$]. New information and adjective contrastive information were not significantly different from each other. The upper-level L2 learners are similar to the L1 English speakers in that new and contrastive information had longer duration than given information, and especially in that new information tends to be produced with the longest duration. For the lower level learners, no significant difference among the three conditions was found in either the main analysis or the post-hoc comparisons.

[31]There was a main effect of language group [b = 79.54, t = 7.70, $p < .001$] in the statistical analysis including language group as a fixed factor: lmer(duration $\sim$ context * language group + (1 + context | subject) + (1 + context | item)).

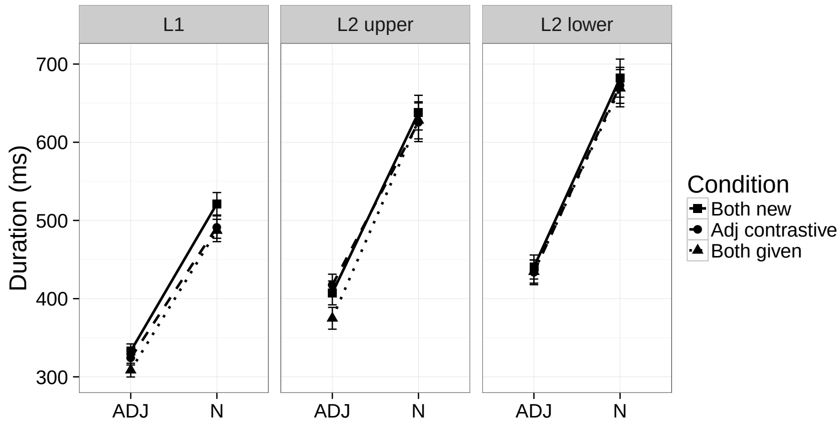


Figure 7.7: Mean duration for the adjective + noun region by L2 proficiency

To sum up the acoustic analyses, the L1 English controls marked given, new, and contrastive information with increasing pitch. They also used longer duration to mark new and contrastive information and shorter duration for given information. Furthermore, the L1 group showed a tendency to lengthen new information more than contrastive information. The upper proficiency L2 group was able to differentiate contrastive information from given information using pitch, while no difference between new and given information or between contrastive and new information was observed. They also differentiated new/contrastive information from given information using duration, and the difference between new information and given information was especially notable. The lower proficiency L2 group did not make use of pitch or durational cues to indicate new, contrastive, and given information as reflected in non-significant results among the three context conditions. Unlike mean pitch and duration, pitch excursion was not affected by information status of discourse entities for any language group in the present experiment.

7.4.2 Phonological analysis

For phonological analysis, six participants from each of the three language/proficiency group were randomly selected, and the critical NP (color adjective + noun) in the 428 utterances from those 18 participants were ToBI annotated by a trained native speaker of English who was blind to the context condition of each utterance.[32] The mean proficiency scores for those sub-groups of participants were 9.29 (range: 8.75-9.63) for the L1 group, 6.38 (range: 5-8.38) for the upper-level L2 group, and 4.12 (range: 3.38-4.5) for the lower-level L2 group. Table 7.2 summarizes pitch accent count with percentage breakdown for each group based on the ToBI annotations.

Table 7.2: Pitch accent percentage (raw count)

		Adjective				Noun				
Group	Condition	L+H*	H*	L*	Deaccented	L+H*	H*	L*	!H*	Deaccented
L1										
(*k* = 143)	(a) New	55% (26)	32% (15)	11% (5)	2% (1)	6% (3)	21% (10)	17% (8)	43% (20)	13% (6)
	(b) Adj. contrastive	69% (33)	29% (14)	2% (1)	-	-	8% (4)	23% (11)	31% (15)	38% (18)
	(c) Given	29% (14)	25% (12)	44% (21)	2% (1)	4% (2)	23% (11)	38% (18)	15% (7)	21% (10)
L2										
(*k* = 285)	(a) New	52% (49)	31% (29)	18% (17)	-	4% (4)	19% (18)	47% (45)	21% (20)	8% (8)
	(b) Adj. contrastive	40% (38)	40% (38)	16% (15)	3% (3)	7% (7)	26% (24)	46% (43)	19% (18)	2% (2)
	(c) Given	36% (35)	33% (32)	29% (28)	1% (1)	10% (10)	24% (23)	39% (37)	21% (20)	6% (6)
L2 upper										
(*k* = 144)	(a) New	52% (25)	38% (18)	10% (5)	-	-	19% (9)	46% (22)	19% (9)	17% (8)
	(b) Adj. contrastive	44% (21)	42% (20)	8% (4)	6% (3)	6% (3)	19% (9)	56% (27)	15% (7)	4% (2)
	(c) Given	33% (16)	40% (19)	25% (12)	2% (1)	10% (5)	25% (12)	29% (14)	23% (11)	13% (6)
L2 lower										
(*k* = 141)	(a) New	51% (24)	23% (11)	26% (12)	-	9% (4)	19% (9)	49% (23)	23% (11)	-
	(b) Adj. contrastive	37% (17)	39% (18)	24% (11)	-	9% (4)	33% (15)	35% (16)	24% (11)	-
	(c) Given	40% (19)	27% (13)	33% (16)	-	10% (5)	23% (11)	48% (23)	19% (9)	-

The L1 results revealed clear differences among the three context conditions in the types of pitch accent assigned to the critical prenominal adjective.[33] The adjective in the contrastive condition was much more frequently produced with L+H* than H* (69% vs. 29%), and the following noun was most frequently deaccented (38%), reflecting enhanced prosodic prominence of contrastive information realized with the L+H* focus accent and post-focus deaccentuation. The adjective in the new condition was most frequently marked with L+H* as well, rather than H* (55% vs. 32%). However, the difference between the two pitch accents in the new context is smaller than that in the contrastive condition (23% vs. 40% difference), and the following noun

[32]Of 432 utterances produced by the 18 participants, four utterances were excluded from the analysis due to disfluency or self-correction at the critical NP.

[33]We did not run statistical analysis on the ToBI annotations because the number of participants in each group was very small, and because in the phonological analysis, we aimed to identify characteristic prosodic patterns for each group rather than test for statistical differences among the experimental conditions.

in the new context condition bears the H* or !H* accent more frequently than less prominent L* or deaccentuation. This indicates increased use of H* rather than L+H* on the adjective and its reduced prominence due to the accentuation on the following noun in the new context. In the given context condition, the most frequent annotation for the adjective and the noun was L* (44% and 38%, respectively). The frequent use of L* rather than deaccentuation on the adjective was somewhat unexpected. This could be due to L1 speakers' effort to avoid an unnaturally long string of words with no accentuation, such as *NOW, use the blue paintbrush to color the DOLPHIN*, while maintaining the prosodic indication of information status for the adjective region. In addition, speaking to an unfamiliar listener in an experimental situation after receiving the task instructions to articulate clearly may have encouraged some tendency for hyperarticulated prosody.

Contrary to the L1 group, the upper L2 group produced contrastive adjectives equally often with L+H* and with H* (44% vs. 42%), and deemphasized the post-focus noun with either L* (56%) or deaccentuation (2%). Furthermore, for the new context condition, they used L+H* more often than H* on the adjective (52% vs. 38%), and the prosodic prominence of the following new noun was reduced with L* (46%) or deaccentuation (17%). These results imply that the upper-level learners do not seem to have differentiated the L+H*–contrastive association and the H*–new association in their production. In the given context condition, adjectives are still marked with L+H* and H* most frequently, yet there is an increase in the percentage of L* compared to the contrastive or the new context conditions (25% for the given condition vs. 8% for the contrastive condition and 10% for the new condition) along with a slight decrease in the percentage of L* on the following noun. Thus, the upper-level learners are more likely to deemphasize adjectives when they are repeated than when they are new, which may be an indication of those learners coming to acquire the deaccentuation (or deamphasis)–given mapping, but unlike the native group, they did not deemphasize the following given noun.

The L2 lower group is similar to the L2 upper group in that they use H* (39%) and L+H* (39%) equally often in marking contrastive adjectives, while L+H* (51%) was more frequently used than H* (23%) for new adjectives. For the given condition, the percentage of L* is greater than that for the new or the contrastive condition (33% for the given condition vs. 24% for the contrastive condition and 26% for the new condition) although the increase is small in comparison to the upper proficiency group (16% for the upper group vs. 8% for the lower group).

Another important observation is that the lower-level learners tend to underuse deaccentuation regardless of the information status of the critical discourse entities. Of 141 spoken utterances recorded from six lower-level learners, none had a deaccentuated adjective or noun, which is clearly different from the L1 group and the L2 upper group who showed more frequent use of deaccentuation, especially for the noun.

All in all, the ToBI analysis revealed English native speakers' preference for L+H* to signal contrastive information and for L* to mark given information. L+H* was also preferred for new information, yet the percentage of L+H* decreases while that of H* increases in the new context condition compared to the contrastive context condition. Unlike the L1 English group, both upper and lower L2 groups used H* and L+H* equally often for contrastive information, and more L+H* than H* for new information. For given information, use of L* on the critical adjective increased only slightly in comparison to new and contrastive information for both upper and lower L2 groups. Lastly, the L2 lower group was different from the L1 and L2 upper groups in that they used deaccentuation less frequently across the three context conditions.

7.5 Discussion

The purpose of Experiment 3 was to examine whether native speakers and L2 learners of English use prosodic cues to signal information status of discourse referents manipulated by the prenominal adjective during a semi-spontaneous interactive speaking task. The results from our L1 acoustic analyses provided further evidence for native English speakers' reliable use of mean pitch to indicate new, contrastive, and given status of discourse referents. Contrastive referents were signaled with higher mean pitch than new referents, which in turn marked higher mean pitch than given referents. This suggests that in native production, mean pitch is the key acoustic cue to prosodic prominence that signals different types of information status in discourse. This three-way distinction among new, contrastive, and given information was also reflected in the ToBI annotations. Contrastive information was most frequently produced with L+H* on the adjective, new information with reduced prominence on the adjective (fewer L+H* instances) as well as reduced post-focus pitch reduction on the noun, and given information with L*. In contrast to mean pitch, the analysis of pitch excursion did not reveal any significant difference among the three conditions.

In terms of duration, native speakers of English produced new and contrastive referents with longer duration on the adjective than for given referents. Furthermore, they tended to elongate new referents more than contrastive referents, which was unexpected as L+H* has been reported to bear longer duration than H* (Breen et al., 2010; K. Ito & Speer, 2008; Katz & Selkirk, 2011). This could be explained by Watson's (2010) Multiple Source view, briefly described in Chapter 5 on the prosody-in-context naturalness rating task, which holds that longer duration is attributable to the speaker's difficulty in the production process. The speaker has most difficulty producing new information since the novel referent has to be activated and added to the discourse model, and contrastive information is less so because it has been partially activated as part of an alternative set. Given (repeated) information is the easiest to produce by virtue of having been just uttered. In this way, speaker-centered processing demands affect word duration in production. Our production results are consistent with this view in that new, contrastive, and given information were marked with decreasing duration. If duration indeed reflects the speaker's processing difficulty rather than his/her signaling of Information Structure for the listener, the listener must discount durational effects when interpreting acoustic prominence for information status of discourse referents. As discussed in Chapter 5, this may be the reason why our native speakers did not rely on durational cues as much as pitch cues when they were asked to judge appropriateness of prosody within a context. The combined results from the rating task and the production task thus provide additional support for the claim that the variance in duration is the result of speaker-oriented production process.

Turning now to the L2 results, we saw that less proficient learners produced the three types of information status similarly, without using distinct pitch or durational cues. In addition, the ToBI analysis revealed no use of deaccentuation by the lower-level learners for all types of information status. These results are compatible with Rasier and Hiligsmann's observation (2007) that L2 learners, especially beginners, tend to emphasize every single word and have difficulty in distinguishing given vs. new information status of discourse referents. We, however, argue that less proficient learners' overuse of accents is due to the difficulty in marking rather than due to distinguishing given vs. new information during production since our lower-level L2 learners did demonstrate their knowledge of given vs. new information in the naturalness rating task.

The upper-level learners, on the other hand, were able to mark given information with lower pitch and contrastive information with higher pitch, while there was no pitch difference between new vs. contrastive information. Furthermore, the ToBI annotations showed their tendency to deemphasize (with L*) repeated words. These results imply that the upper-level learners were able to signal given or repeated information with less salient pitch cues. Use of duration was also observed in the upper-level L2 group. Like the native group, they used longer duration for new/contrastive information and shorter duration for given information, and especially new information was marked with the longest duration. The fact that new information, contrastive information, and given information were signaled with decreasing duration by the upper-level learners suggests that the speaker-oriented production account can be applied to not only native speakers but also non-native speakers. Since L2 learners may be more constrained than native speakers in lexical access, morphological and syntactic processes, and the formation of discourse structure, it makes sense that the production of repeated words considerably reduces processing burden, resulting in faster speech and shorter duration of the words.

To summarize, the finding from the acoustic analysis that the upper-level L2 learners successfully marked given information and contrastive information with distinct prosodic cues parallel the results from the prosody-in-context naturalness rating task. The phonological analysis, on the other hand, has shown that the upper-level learners prosodically marked given information with reduced prominence while they made no phonological distinction between new and contrastive information just like in the eye-tracking listening experiment; but as the ToBI analysis dealt with only a subset of the data, we are uncertain whether or not these results from the phonological analysis represent the behavior of the entire upper-level L2 participant population. The lower-level L2 learners did not use distinct phonetic or phonological cues when indicating given, new, and contrastive information. It is unexpected that these lower-level learners did not indicate contrastive information with elevated pitch as they can do in their L1 Japanese. It seems that L1 prosodic transfer did not emerge in our production experiment unlike the previous studies reporting influence of L1 prosody in L2 production. This finding, in combination with the fact that the upper-level learners produced given information with reduced acoustic prominence, which is not instantiated in the L1, suggests that L2 learners do not always transfer their L1 prosodic marking of discourse and that they are capable of acquiring a new prosody-discourse mapping in TL production as proficiency increases.

In the next chapter, we will summarize the results from all of the three experiments and discuss how the results support or do not support the LILt, and whether there is a relationship between L2 learners' comprehension and production. The chapter further provides some implications of the current study in terms of L2 acquisition research as well as the teaching of L2 prosody.

CHAPTER 8
GENERAL DISCUSSION AND CONCLUSION

8.1 Introduction

The purpose of this dissertation research was to investigate whether L1-Japanese L2 learners of English can perceive and produce prosodic cues that signal Information Structure in English. The first experiment, utilizing a prosody-in-context naturalness rating task, tested whether L2 learners possess the knowledge of the mapping between null accent and given information, H* and new information, and L+H* and contrastive information. The second experiment employed the eye-tracking technique to investigate whether learners can apply their knowledge of prosody-discourse mapping to online comprehension. The third experiment examined whether L2 learners can use discourse prosody in production during a semi-spontaneous speaking task. The results of these three experiments are summarized in Table 8.1. Below we discuss the results of native English speakers first, then compare the native results with those of the L2 learners.

Table 8.1: Summary table of the results

Group	Mapping	Naturalness rating	Eye-tracking listening	Production
L1				
	H*–new	◯	◯	✓
	L+H*–contrastive	✓	✓	✓
	null accent–given	✓	✓	✓
L2 upper				
	H*–new	◯	◯	×
	L+H*–contrastive	✓	×	✓
	null accent–given	✓	✓	✓
L2 lower				
	H*–new	×	×	×
	L+H*–contrastive	✓	×	×
	null accent–given	✓	×	×

Note. ✓: successful mapping, ◯: somewhat successful, ×: unsuccessful

8.2 Summary and discussion of the L1 results

8.2.1 Null accent–given information mapping

In the prosody-in-context naturalness rating task, native speakers demonstrated a strong association between null accent (especially low pitch cue) and given information. To native ears, null accent sounded much more natural than H* and L+H* did in the given context. In the new–given eye-tracking experiment, the facilitative effect of null accent on the processing of given referents was confirmed partially by increased looks to discourse-given target objects and more robustly by reduced looks to discourse-new competitors. These results add to previously found evidence indicating that deaccentuation signals native English listeners about the given status of discourse referents (Arnold, 2008; Terken & Nooteboom, 1987). Native English listeners prefer an unaccented expression to refer to a given referent, and they can indeed comprehend a given referent more quickly when it is unaccented than when it is not.

In the production experiment, the same native participants produced repeated, given referents with lower pitch and shorter duration than for new or contrastive referents. This is in line with Terken and Hirschberg's finding that a referring expression tends to be deaccented when its previous mention appears in the same syntactic position as the current mention (Terken & Hirschberg, 1994). In addition, the ToBI annotations in our study showed that given referents were most likely to be produced with reduced acoustic prominence such as L*. Although this finding is somewhat different from the claim by Terken and Hirschberg, L* marking of given referents is indeed consistent with Pierrehumbert and Hirschberg's description that native speakers employ L* when they believe that the referent is already in the hearer's discourse model (Pierrehumbert & Hirschberg, 1990). The current study thus found strong evidence supporting native English speakers' association between null accent or reduced acoustic prominence and given information in the rating task, the online comprehension task as well as the production task.

8.2.2 L+H*–contrastive information mapping

In the prosody-in-context naturalness rating task, the L+H*–contrastive referent pair was rated significantly higher than the H*–contrastive referent pair. In addition, the response breakdown showed more "very natural" responses for the L+H*–contrastive pair than for the H*–contrastive pair. These results support native speakers' preference for L+H* over H* for marking contrastive information.

A processing advantage of L+H* over H* for contrastive referents was also observed in the online eye-tracking experiment. Native speakers of English fixated more on the contrastive target object when it is produced with L+H* than H*. Unlike the previous findings by Ito and Speer (2008), however, L+H* in the new context did not "garden-path" listeners to the incorrect target. The reduced effect of L+H* triggering a contrastive interpretation in our study is likely an artifact of the experimental design: a relatively simple visual display with more complex auditory stimuli. Our results parallel those in Watson, Tanenhaus, and Gunlogson (2008), which also utilized a less complex visual display and more complex target sentence structures involving more discourse entities than Ito and Speer (2008) did. Watson et al. report that native listeners fixated on the contrastive target more quickly with L+H* than with H*, yet the listeners were not garden-pathed to the contrastive competitor when the discourse-new target object was infelicitously marked with L+H*. While their study provides additional evidence for the facilitative effect of L+H* on the processing of contrastive referents, the authors did not conduct statistical analyses on looks to the competitor, leaving unclear the reason for the lack of a garden-path effect caused by infelicitous pitch accent. We suspect that the simple visual display with only one competitor, as in our study, allowed Watson et al.'s listeners to shift their attention between the two target candidates (actual target and competitor) more easily and quickly, not fully committed to a decision until the disambiguating segmental cue becomes available. If the visual display had been more complex with multiple competitors, then listeners would have had to limit their attention to the most relevant entities and thus might have relied more on available non-visual cues such as prosody. In addition, the complex discourse structure in our and Watson et al.'s audio stimuli may have reduced the facilitative effect of L+H* on contrastive interpretation. While Ito and Speer used a target sentence containing only the target object and nothing else, our stimuli contained the target object in addition to an animate entity (i.e., animal). This additional discourse entity intervening between the two contrasting referents may have made the discourse contrast less salient to the native listeners.

In our production experiment, native speakers of English marked contrastive information with elevated pitch and elongated duration. Mean pitch was significantly higher for the contrastive context than for the new or given contexts. Duration was significantly longer for the contrastive and new contexts than for the given context. The former finding is consistent with the claim that contrastive accents are more emphatic and acoustically more prominent with boosted pitch (Ladd, 1983). Duration also seems to distinguish new and contrastive information from given information; but somewhat unexpected was the fact that native speakers tended to elongate new information more than contrastive information. This could be due to the speaker having a more difficult time generating a new discourse entity than a contrastive entity during the production processes (speaker-centered production process). ToBI annotations further revealed that native speakers produced contrastive information with L+H* more often than with H*. These acoustic and phonological analyses suggest that native speakers of English reliably utilize more prominent prosodic cues to express contrast in production.

In sum, all three experiments in the current study provide empirical support for native speakers' mapping between L+H* and contrastive information. They consider L+H* to be most appropriate for marking contrast, and during online discourse comprehension, they immediately integrate the L+H* pitch accent to identify the contrastive target referent. In production, native speakers reliably mark contrast with more salient acoustic cues such as increased mean pitch and duration. These results lend further support to Pierrehumbert and Hirschberg's proposal (1990) that the L+H* pitch accent conveys a contrastive meaning, highlighting the accented discourse entity which stands in contrast with other plausible alternatives.

8.2.3 H*–new information mapping

Among the three mappings investigated in the present study, the one between H* and new information was least clear in native speakers' comprehension and production. In the prosody-in-context naturalness rating task, both H* and L+H* were rated higher than null accent in the new context, with little difference between H* and L+H*, suggesting that for new information, H* does not sound particularly more natural than L+H*. As explained in Chapter 5 on the rating task, we suspect this is due to our experimental material in which pitch manipulation was conducted on the sentence-final content word (i.e., the compound noun *banana bread*), which is the default nuclear pitch accent position. In the new context such as *What did Mariana do? – Mariana made the banana bread*, the nuclear pitch accent falls on *banana* so that the last content word *banana bread* sounds most prominent in the response sentence. It could be that the acoustic prominence caused by L+H* in the new context condition did not sound too unnatural due to the fact that it corresponds to the nuclear pitch accent position, minimizing the difference in naturalness ratings between L+H* and H* in the new context condition. A similar finding is reported in Welby (2003), which used a question-answer pair such as *How do you keep up with the news? – I read Dispatch*, with H* or L+H* on the final word of the response sentence *Dispatch*, and found that native listeners' ratings were not affected by the difference between H* and L+H*. Their conclusion was that listeners' interpretation of focus structure depends more on the presence or absence of pitch accent rather than the type of pitch accent on the focused material. The results in Welby's study could support our claim that the difference between H* and L+H* is difficult to perceive and interpret or at least both H* and L+H* sound natural when they appear in the default nuclear pitch accent position in the sentence responding to a broad focus question.

In the new–contrastive eye-tracking experiment, H* in the new context did not induce faster fixation to the discourse-new target than L+H* did, the latter being a garden-path condition. This could be partially due to our visual display being much simpler and our auditory stimuli being more complex than Ito and Speer's (2008), as explained in the section on the L+H* and contrastive information mapping above. Besides the lack of a garden-path effect in the L+H*–new condition, target fixation in the H*–new condition was not any faster than the fixation in the H*–contrastive condition, replicating the results in Watson, Tanenhaus, and Gunlogson (2008). Watson et al. claim that H* is compatible with either new referents or contrastive referents; but our study further indicates that compared to the mapping between L+H* and contrastive referents, the associations between H* and new referents and between H* and contrastive referents are rather weak. In the new–given eye-tracking experiment, H* and null accent in the new target condition did not show different fixation patterns for the target or the competitor. However, in the given context, native speakers looked more at the discourse-new competitor when the target was produced with H* vs. with null accent, indicating that they were garden-pathed. These results suggest that though H* may not be strongly associated with new information, listeners still prefer to look at a new referent when they hear H* compared to null accent.

Combining the results from the naturalness rating and the eye-tracking listening experiments, native English speakers clearly prefer accentuation over null accent as a prosodic indicator of new information. In these tasks, both H* and L+H* were associated with new information, possibly due to our experimental design, although H*–new and L+H*–new associations seem weaker than those between null accent–given or L+H*–contrastive information.

In production, native speakers marked new information with a mid-height pitch that is lower than the pitch for contrastive information but higher than the pitch for given information. This suggests that they are able to make use of pitch cues to signal new information. New information and contrastive information were produced with longer duration compared to that for given information, resulting in greater prominence. Furthermore, new information was marginally longer than contrastive information, which may be due to the result of speaker-oriented processing difficulty if new information is indeed harder to process than given or contrastive information.

All in all, our naturalness rating and eye-tracking listening experiments did not find evidence for a strong association between H* and new information. Instead, both H* and L+H* (accentuation) were compatible with new information. In the production experiment, on the other hand, native speakers did distinguish new information from contrastive or given information using pitch cues.

8.3 Summary and discussion of the L2 results

8.3.1 Null accent–given information mapping

In the prosody-in-context naturalness rating task, both upper- and lower-level L2 learners were able to associate null accent to given information in the same way as the native speakers. They rated null accent much higher than H* or L+H* in the given context. In the new and the contrastive contexts, the upper-level learners rated null accent low, while the lower-level learners rated null accent relatively high, which suggests that for lower-level learners, the association between null accent and given information is not as robust compared to that of the upper-level L2 learners and the native speakers of English.

In the new–given eye-tracking task, the upper-level learners looked less at the incorrect new referent instead of the correct given referent when they heard null accent compared to H*. This fixation pattern is parallel to that of the native speaker group. The lower-level learners, on the other hand, simply showed a bias toward a non-repeated referent regardless of pitch accent type.

Combining the results from both experiments, the upper-level learners are similar to native speakers in that they are able to match null accent to given information in perception, and use such knowledge to help them process Information Structure in a timely manner during online comprehension. The lower-level learners' association between null accent and given information is less robust in naturalness rating, and when it comes to online comprehension, they are not able to make use of pitch cues quickly enough to process Information Structure.

In the production experiment, for the upper L2 group, the difference in mean pitch between contrastive information and given information was significant, while no difference was found between contrastive and new information or between new information and given information. This means that the upper-level learners, like native speakers, are able to make use of pitch cues to signal given and contrastive information status: They use higher pitch for contrastive information and lower pitch for given information. In contrast to the upper-level learners, the lower-level learners did not distinguish among new, contrastive, and given information using pitch cues. As for duration, the upper-level learners lengthened new and contrastive information more than given information, like native speakers did. Contrary to the L1 and upper L2 groups, no such durational difference among the three types of information stats was found for the lower-level group.

Summing up, as for deaccentuation of given information, the upper-level L2 learners performed similarly to native speakers in naturalness rating, online listening comprehension, and production. They consider deaccented given referents to be more natural, perceive deaccentuation cues immediately in order to identify given referents, and produce given referents with less acoustic prominence. The lower-level learners seem to possess some knowledge of the mapping between deaccentuation and given information as shown in the naturalness rating experiment. However, they are not able to utilize such knowledge during real-time comprehension or semi-spontaneous production.

8.3.2 L+H*–contrastive information mapping

In the naturalness rating task, both lower- and upper-level L2 groups rated L+H* significantly higher than H* or null accent in the contrastive context, which confirms the L2 learners' knowledge of the L+H*–contrastive mapping in English regardless of proficiency level.

Contrary to the rating results, the new–contrastive eye-tracking experiment revealed no L2 sensitivity to the L+H* marking of contrastive referents. Both upper-level and lower-level learners instead showed a strong bias toward discourse new entities without being affected by the type of pitch accent. In addition, although the lower-level learners were in general slower than the upper-level learners in identifying the correct target, their fixation speed was accelerated with L+H* compared with H* regardless of the discourse status of the target object. These results imply that language proficiency affects fixation patterns: Less proficient learners perform the task more slowly and they benefit more from increased acoustic saliency during online listening. Furthermore, when it comes to real-time comprehension of contrast, even advanced learners, who demonstrated the mapping between L+H* and contrastive information in the offline rating task, were not able to draw on such knowledge. This is surprising given, first, that the native Japanese listeners in Ito et al. (2012) were sensitive to their L1 contrastive prosodic cue in an experimental paradigm similar to ours, and, second, that our upper-level learners did show online sensitivity to the null accent–given information, which is not instantiated in the L1. We suspect it is because the real-time processing of discourse contrast involves more a complex operation than the processing of discourse given (repeated) referents. When repeated information is deaccented, listeners simply search for the most accessible referent in their discourse model. For contrastive information signaled with L+H*, they need to activate a set of alternative referents and pick the most likely one from the contrast set. This extra processing step combined with time constraints imposed by the visual search task may have delayed L2 learners' comprehension, resulting in no sensitivity to the L+H* marking of contrastive information for either proficiency group.

In the production experiment, however, the upper-level learners did make a distinction between contrastive and given information with high vs. low mean pitch. Thus, they seem to be able to signal contrastive information with a more prominent acoustic cue (i.e., higher pitch) like native speakers. The upper-level learners also used elongated duration for new and contrastive information. New information was especially long, which may be the result of speaker-oriented processing difficulty as has been argued for native speakers. The lower-level learners tended to emphasize every single word in their production. They did not produce contrastive information with elevated pitch, elongated duration or distinct phonological cues, unlike the upper-level group.

In sum, the upper-level learners demonstrated their knowledge of the L+H*–contrastive mapping in the offline rating task and were able to utilize such knowledge in production. In online comprehension, however, prosodic cues did not affect their interpretation of contrastive information, which could be due to the difficulty involved in contrastive processing under a time restriction. The lower-level learners did show knowledge of the L+H*–contrastive mapping in the rating task, but such knowledge was not put to use either in online listening comprehension or in production.

8.3.3 H*–new information mapping

Though our study did not find strong evidence for the mapping between H* and new information for either upper- or lower-level L2 learners, the upper-level learners seemed to prefer accentuation for signaling new information in the prosody-in-context naturalness rating experiment and in the new–given eye-tracking experiment. For the new context in the rating task, the upper-level group rated both H* and L+H* much higher than null accent, while the lower-level group did not rate the three types of pitch accent differently. This suggests that only the upper-level learners are sensitive to accentuation and the association between accentuation and discourse new referents.

In the new–contrastive eye-tracking experiment, the L2 learners, regardless of proficiency, did not show sensitivity to H*-marking of new information. When the target object was a new referent, H* did not result in faster fixation to the target compared to L+H*. These results are in line with the L2 results from the naturalness rating task showing no significant differences between H* and L+H* in the new context. In the new–given eye-tracking experiment, both lower-level and upper-level learners showed an overall bias toward new referents over given referents. The upper-level group, however, looked more at the discourse-new competitor when the discourse-given target was marked with H* vs. with null accent, the same pattern found in the native group. This implies that although H* may not be strongly associated with new referents, advanced L2 listeners still prefer to look at a new referent when they hear accentuation.

In the production experiment, neither the upper- nor lower-level groups used prosodic cues to indicate new information, as reflected in no pitch difference between the new and adjective contrastive context conditions or between the new and given context conditions. However, the upper-level learners marked new information with longer duration, which again may be a reflection of speaker-oriented processing difficulty. According to the ToBI annotations, the upper-level group used both H* and L+H* to indicate new information, supporting the view that advanced L2 learners may be able to associate new information to accentuation, either H* or L+H*.

In sum, the upper-level L2 learners showed some evidence that they link accentuation (both H* and L+H*) to new information like native speakers do. The lower-level learners did not associate H* with any specific type of information status. These findings are not surprising given that even native speakers of English did not demonstrate a strong association between H* and new information in the rating and the eye-tracking tasks.

8.3.4 Research questions and L2 Intonation Learning Theory

Research questions

In this dissertation, we addressed the following research questions and tested whether L1-Japanese L2-English learners are able to acquire and utilize three types of prosody-discourse mapping (H*–new, L+H*–contrastive, null accent–given) in a prosody-in-context naturalness rating task, an online eye-tracking listening task, and a semi-spontaneous production task.

1. Can L2 learners associate a particular TL pitch accent pattern with an appropriate discourse meaning when focusing on intonation?

2. Can L2 learners make immediate use of TL pitch accent cues to identify discourse referents during real-time comprehension?

3. Can L2 learners employ TL pitch accent cues to express information status of discourse referents in naturalistic production?

In an attempt to answer these research questions, we framed our study within the L2 Intonation Learning theory (LILt) and made predictions as follows:

1. The L+H*–contrastive association will be easiest to master for L1-Japanese learners of English due to L1 transfer.
2. The associations between H* and new information and between deaccentuation and given information will be challenging for Japanese speakers.
3. L2 learners are expected to acquire a new link between a particular pitch pattern and its discourse meaning in the course of learning, showing more target-like performance as their proficiency increases.

As for the first two predictions, we had anticipated that the L+H*–contrastive association would be easier to acquire than the H*–new association or the null accent–given association because in Japanese contrast can be signaled with expanded pitch range, similar to English. However, this prediction was not supported in our study. In the naturalness rating task, both the upper-level and lower-level groups demonstrated the mapping between L+H* and contrastive information as well as between null accent and given information. In the production task, only the upper-level group demonstrated those two types of associations. Crucially, in online eye-tracking comprehension, the upper-level learners in the new–given experiment showed the facilitative effect of deaccentuation of given referents but those who are equally proficient in the new–contrastive experiment were not sensitive to L+H* marking of contrastive referents.

These findings point to several conclusions. First, L2 listeners do not always apply their L1 knowledge and online processing strategies when it comes to the mapping of prosodic information to discourse meaning. Online tasks such as eye-tracking are thought to reflect L2 learners' underlying language processing mechanism better and are less susceptible to metalinguistic knowledge and general problem-solving strategies compared to offline tasks as they reveal learners' unconscious behavior (Marinis, Blom, & Unsworth, 2010). As we described in Chapter 2, the L1 Japanese visual world eye-tracking study by Ito et al. (2012) has shown that native speakers of Japanese rapidly integrate expanded pitch cues for the processing of contrastive referents. Given that the mapping between prosodic information and discourse meaning is similar in English and Japanese in this respect, Japanese speakers should employ their L1 knowledge and processing strategy if L2 intonation is largely influenced by L1 discourse prosody as predicted by the LILt. Nevertheless, the L2 listeners in our study, even the advanced ones, did not show sensitivity to L+H* marking of contrast in the eye-tracking task, but instead demonstrated a native-like processing pattern for the mapping that is not instantiated in their L1, namely, the deaccentuation–given information mapping.

In fact, lack of L1 transfer between two languages that share similar phonological structures has been reported previously in Akker and Cutler (2003) as described in Chapter 3. In their phoneme-detection task, L1-Dutch L2-English listeners were able to detect the phoneme-bearing target word more quickly when it was focused by means of a question preceding the response sentence containing the target word vs. when it was not focused, and when the target word was accented vs. when it was not accented. However, only native English speakers showed an interaction effect of focus and accent: When the target word was unfocused, accentuation accelerated phoneme detection speed, but when the target word was focused, the presence or absence of accent did not affect detection speed. In other words, the processing of prosody was modulated by semantic information for native speakers but not for L2 learners. Akker and Cutler conclude that non-native listeners' mapping between prosodic information and semantic structure is not as efficient as that of native processing even when the prosodic structures and the prosodic marking of focus are similar between learners' L1 and the TL. Our study, too, involves two levels of processing (i.e., prosody and discourse), and we found some evidence in the eye-tracking experiments that the lower-level L2 learners process prosody and discourse independently: They preferred to look at discourse-new referents over contrastive or given referents, while they identified the target referent faster with L+H* than H*. Crucially, these two types of effects did not interact. The upper-level L2 learners, on the other hand, showed a native-like interaction effect of prosody and discourse when processing null accent for given vs. new referents. Thus, discrepancies between native and non-native processing can emerge even when the TL and learners' L1 are similar with respect to prosody-discourse associations.

Based on these findings, we propose that although L2 learners do not necessarily apply their L1 knowledge to the processing of L2 discourse prosody, it is possible for learners, especially advanced ones, to approximate native efficiency in the mapping of prosody to Information Structure when the prosody-discourse interface involves less complex processing as in the case of deaccentuation and given (repeated) information. As described in the eye-tracking chapter, the processing of deaccentuation for repeated referents may be cognitively less demanding than L+H* for contrastive referents since the former involves fewer steps. In the case of our eye-tracking task, a discourse entity that had been previously mentioned was already activated in the listener's mental space, and its semantic and phonological representations were also easily accessible. When the L2 listener perceived segmental information with no accentuation, s/he could quickly map it onto the set of discourse entities that were active at the moment (Terken & Nooteboom, 1987). The processing of contrastive referents may have been more complex. When the listener heard the contrastive intonation contour, s/he had to identify the most salient entity in her/his discourse model, activate an alternative that stood in contrast to the salient entity and establish a contrast set, and search for the alternative member in the visual display. Lee and Fraundorf (2017), which was introduced in Chapter 3, have shown that the interpretation of contrastive focus is cognitively demanding and that non-native processing of contrast is fundamentally different from native processing due to reduced cognitive resources. In their probe recognition task testing whether the contrastive pitch accent (L+H*) benefits listeners' memory for discourse entities compared to the non-contrastive pitch accent (H*), native speakers of English successfully distinguished the contrastively accented referential expression from its alternative in the recognition test. Advanced L2 learners, on the other hand, better recognized the two members of the contrast set with L+H* than H*, but were not able to remember which was the correct entity and which was the incorrect alternative. This led to the shallow representation account, which holds that the mechanism by which L2 learners use contrastive pitch accents to encode discourse referents differs from that of native speakers. Even though L2 learners can utilize L+H* to establish a contrast set, they are not capable of fully integrating information conveyed by the pitch accent into their discourse representation because of limited cognitive resources for L2 processing.

If prosody for discourse interpretation is indeed processed differently by native and non-native listeners due to cognitive constraints, it makes sense that our L2 speakers were more successful in the prosody-discourse mapping during the offline naturalness rating task. When processing demand is reduced by eliminating time constraints and explicitly directing listeners' attention to prosody, L2 listeners can allocate more of their attention and limited cognitive resources to intonation rather than to other kinds of linguistic (e.g., lexis, syntax, semantics) and non-linguistic (e.g., visual display) information. This may in turn lead to effective processing of more difficult mappings such as L+H* to contrastive information and accentuation to new information.

Without time restrictions, L2 learners are able to consider all aspects of the given stimulus including prosody, spend enough time to execute the necessary processes (e.g., perceive L+H* on a prenominal adjective, establish a contrast set, identify the most salient member of the set, and select another member of the set as the target), and think about the meaning before making a decision. Explicitly manipulating the focus of learners' attention may also help them concentrate on the key information and use such information more effectively. Previous research investigating the effect of intonation on the offline interpretation of ambiguous relative clause sentences such as *We adore the secretary of the psychologist who takes a walk* (Dekydtspotter et al., 2008) has actually shown that only a subset of L2 participants, regardless of proficiency, demonstrated sensitivity to intonation contour, implying that there is some individual variation in the degree to which prosodic information is taken into account during sentence processing. One reason could be that some learners are more attentive to prosodic features than others.

The role of attention on L2 prosodic processing has been investigated in a memory study by Pennington and Ellis (2000), which found that L2 learners are likely to remember contrastive vs. non-contrastive prosodic contours better (e.g., *Is HE driving the bus?* vs. *Is he driving the bus?*) when their attention is explicitly drawn to prosody. In the experiment, L1-Cantonese learners of English heard a set of English sentences (study phase), and after a short break, they heard another set of sentences (recognition phase) that were identical to or different from the first set in terms of prosody and lexis. In the study phase, learners heard sentences either without any instructions about prosody or with instructions to focus on intonation and consider two possible interpretations of each sentence provided by the researchers (e.g., "special attention on *HE* for emphasis or contrast" or "no special emphasis or contrast," after hearing *Is HE driving the bus?*). During the recognition phase, the learners indicated whether each sentence was exactly the same as or different from those they had heard before. The results showed that the learners' memory for prosodic cues improved when their attention was directed to intonation during the study phase, compared to when no specific directions were given. Based on the results, the authors proclaim the importance of explicit instructions and raising listeners' awareness about intonation. In our study, too, L2 performance in the discourse interpretation of prosody was better in the naturalness rating task, which required explicit focus of attention on intonation, than in the online eye-tracking task, in which intonational manipulation as the focus of the task was less obvious to the participants.

In the semi-spontaneous speaking task, the learners were under the pressure of speech production but no time limit was imposed. While the upper-level L2 group was able to mark contrastive referents with elevated pitch and given referents with low pitch, the lower-level group did not distinguish among new, contrastive, and given referents with different acoustic cues. In addition, the upper-level learners performed similarly in the rating task and the production task. In both tasks, they were successful in mapping contrastive information to L+H* (high pitch) and given information to null accent (low pitch). This hints that there may be some implicational relationship between offline prosody-discourse mapping and production at least for advanced learners. The lower-level learners, on the other hand, performed better in the rating task than in the production task. During production, their speech rate, regardless of context conditions, was overall slower than that of (the native controls and) the upper-level L2 learners as reflected in longer word duration, and they showed a tendency to emphasize every word with pitch accent, which has been reported to be a common characteristic of L2 speech (Rasier & Hiligsmann, 2007). It appears that advanced learners are able to apply what they know about the prosody-discourse relationships to production using appropriate prosodic cues for particular discourse status, while less advanced learners are not able to do so perhaps because of the difficulty in lexical retrieval and slow syntactic/semantic processing obstructing target-like realization of discourse prosody.

Regarding the third prediction about proficiency, the outcomes of our study clearly indicate that language proficiency plays an important role in the L2 acquisition of discourse prosody, as Mennen proposes (2015). In all three experiments, the upper-level learners outperformed the lower-level learners, and their performance was more native-like in many respects. For instance, in the prosody-in-context naturalness rating task, the response patterns were quite similar between the upper-level L2 group and the native group. Both groups rated accentuation (i.e., H* and L+H*) much higher than null accent for new information, while the lower-level group did not show such a preference. In the given–new eye-tracking experiment, the upper-level L2 group and the native group fixated less to the incorrect, discourse-new competitor when they heard null accent vs. H*, implying that they can quickly perceive and integrate the link between null accent and given referents and between accentuation and new referents. The lower-level group, on the other hand, did not demonstrate use of prosodic cues for discourse interpretations during real-time processing. In the production task, too, the native controls and the upper-level learners performed similarly in the marking of contrastive information and given information, while no indication for prosodic marking of information status was found among the lower-level learners. It is thus essential to take into consideration L2 learners' proficiency level when studying comprehension and production of L2 prosody.

L2 Intonation Learning Theory

As discussed in Chapter 3, L2 Intonation Learning theory (LILt) follows several underlying theoretical assumptions. First, it assumes that TL intonation categories are automatically perceived in relation to learners' existing L1 categories, and that both phonetic and semantic/discourse dimensions need to be considered when identifying similarities and differences between L1 and TL phonological categories. Second, the LILt takes on the assumption that L1 influence occurs at both the phonological and phonetic levels. Third, the LILt considers age of onset to be one of the determining factors for successful L2 acquisition of intonation, and fourth, the theory maintains that as learners receive more exposure to the TL, L2 intonation becomes fine-tuned to approach the native norm. Lastly, the LILt holds that L1 and L2 intonational categories co-exist in a shared perceptual space, allowing bi-directional influence between the two languages.

Relevant to our study are the first, third and fourth assumptions. As for the first assumption, our study has shown that even when the TL tonal category is similar to an L1 category in both form and meaning, L2 learners do not necessarily interpret the TL category in the same way as the corresponding L1 category, especially when the relevant form-meaning association involves more complex processing steps and the task itself imposes cognitive demands as in the case of the L+H*–contrastive information mapping during an online eye-tracking task. Thus, other factors besides similarities/differences in L1-TL intonation categories seem to have considerable impact on to what degree L2 learners assimilate TL categories into their L1 intonation system.

Our study does not directly speak to the third assumption as we tested only late L2 learners who grew up in a non-English speaking country and had no substantial amount of exposure to English before age 11. However, our findings clearly suggest that even late learners can successfully master target-like discourse intonation as they receive more TL input and improve proficiency, supporting the LILt's fourth assumption.

Whether the second and fifth assumptions apply to L2 acquisition of discourse prosody remains to be seen in future research. As for the second assumption, since the main purpose of our study was to examine L2 mappings of prosody and discourse meaning, manipulation of the experimental stimuli in the rating and eye-tracking tasks was done at the phonological level (i.e., null accent, H*, L+H*) only, which leaves us with no information on how phonetic details (e.g., exact timing of pitch rise onset, dip depth before pitch rise, or alignment of pitch peak) would have affected learners' perception of the target tonal categories in relation to discourse meaning. In the production experiment, we conducted both acoustic and ToBI analyses on learners' utterances. While the acoustic analysis identified a similar pattern between the upper-level L2 learners and the native English speakers for the null accent–given information mapping and the L+H*–information mapping, the ToBI annotations indicated no clear evidence for such mappings. As the ToBI analysis was performed on only a subset of all participant data and by only one annotator, further analysis is called for in order to better understand how L2 intonation reflects L1 influence phonetically and phonologically.

Regarding the fifth assumption, one way to test it is to investigate whether L1-Japanese advanced learners of English who demonstrate native-like sensitivity to prosodic marking of Information Structure behave differently from L1-Japanese monolingual speakers when tested in the Japanese-equivalent version of the experiments. The results should reveal how L2 tonal categories affect L1 perception and production of intonation.

8.4 Implications for the teaching of L2 prosody

Our L2 listeners, especially the less proficient ones, in general had less difficulty in associating prosody with discourse when they had no time pressure and focused on intonational cues when making decisions about prosodic naturalness. They were even successful in mapping null accent to given information, which does not exist in their L1. This means that if L2 learners of English are given explicit instructions to guide their attention to intonation in relation to discourse context and enough time to complete the task, they are more likely to recognize prosodic distinctions and make better use of prosodic cues in understanding discourse information. Even without explicit instruction, advanced learners seem to come to master a new mapping that is relatively easy to process (e.g., null accent–given information mapping) and apply such knowledge quickly during real-time processing. For L2 learners with limited proficiency, processing linguistic information in real time is such a challenging task that they may not be able to direct their attention to suprasegmentals and exploit prosodic information. However, they may approach native-like perception and comprehension of discourse prosody as proficiency increases. In production, our advanced L2 learners used prosodic cues to indicate contrastive and given information successfully, while the less proficiency learners did not, perhaps due to lack of fluency caused by slow lexical access and morphological/syntactic processes in speech production.

These findings can have several implications for language teachers and practitioners. First, elimination of time pressure and explicit instruction improve learners' recognition of discourse-prosody associations at least in offline tasks, which in turn could lead to a better understanding of messages being conveyed. And setting no time pressure and explicit instruction could benefit not only advanced learners but also lower proficiency learners. In previous research on L2 prosodic comprehension reviewed in Chapter 3, some offline studies (Klassen, 2013; Yoshimura, Fujimori, & Shirahata, 2015) identified native-like performance of discourse prosody interpretations in L2 learners, while none of the online studies (Akker & Cutler, 2003; Braun & Tagliapietra, 2011; Chen & Lai, 2011; E.-K. Lee & Fraundorf, 2017) have revealed such results. This implies that L2 learners can successfully map prosodic information to appropriate discourse meaning during offline judgment tasks while they may not make full use of available information during online tasks. Whether or how explicit instruction affects learners' online processing of discourse prosody is still an unresearched issue as our study did not investigate L2 performance in eye-tracking with vs. without explicit focus on prosody. However, we predict that when it comes to a complex mapping (e.g., L+H*–contrastive information mapping), explicit task directions to pay attention to intonation would still not lead to target-like task performance since time constraints are likely to keep learners from fully integrating prosodically marked elements into their discourse model in a time fashion.

Secondly, teaching prosodic marking of discourse for speaking may not be effective for beginners, as their speech production tends to be slow (with a number of errors, self-corrections, and hesitations), making it difficult to produce natural-sounding intonational contours. As learners become more fluent, they may benefit more from explicit instruction focusing on appropriate use of intonation and other prosodic cues to better communicate their messages. In L2 pronunciation research, it has been long debated whether teaching pronunciation in formal instruction is useful (Algethami, 2017; J. Lee, Jang, & Plonsky, 2014; Thomson & Derwing, 2014). While most studies have shown that pronunciation instruction bears some positive effect in general (Saito & Lyster, 2012), other studies have found little or no improvement in learners' pronunciation after formal instruction (Derwing, Munro, Foote, Waugh, & Fleming, 2014; Derwing & Rossiter, 2003; Saito, 2007). While many of the L2 pronunciation studies have focused on the effect of teaching individual sounds or segments on pronunciation accuracy, our study has shown that explicitly drawing learners' attention to intonation can improve their perception of prosody in relation to information status of discourse referents. Our next step is to further test this possibility empirically and validate the current findings as more studies on the learning and teaching of suprasegmental features are called for in order to reach any firm conclusions about instructional effects on both offline and online L2 prosodic listening as well as speaking.

8.5 Concluding remarks

The present research sought to investigate prosodic marking of Information Structure by native English speakers and L1-Japanese L2 learners of English. In general, it was found that native speakers associate given information to deaccentuation, contrastive information to L+H*, and new information to accentuation (both H* and L+H*) in both comprehension and production. Advanced L2 learners could map given information to deaccentuation and contrastive information to L+H* in the naturalness rating task as well as in the production task, but only the mapping of given information to deaccentuation was exhibited in the online eye-tracking task. Lower-level learners were able to associate given information to deaccentuation and contrastive information to L+H* in the rating task, but no evidence for prosody-discourse mapping was found in production or online eye-tracking comprehension. The outcomes of the present study suggest that it is not the case for all tasks that L2 learners necessarily transfer L1 prosody-discourse mapping, and that they can master L2 prosodic marking of discourse structure that is not instantiated in their L1. However, how successfully they map prosody to discourse in the TL depends on several factors: (a) learners' language proficiency, (b) the processing difficulty of a particular mapping (deaccentuation–given information being the easiest, L+H*–contrastive information being rather challenging, H*–new information being most difficult), (c) difficulty of tasks (offline vs. online), and (d) whether learners' attention is explicitly drawn to prosody. Further studies with improved experimental designs with a larger number of L2 participants are necessary in order to investigate more thoroughly L2 learners' sensitivity to pitch accent cues signaling discourse structure. Especially for the eye-tracking experiments, a more complex visual scene with multiple competitor objects and simpler auditory stimuli containing fewer discourse entities may lead to stronger effects of pitch accents on the interpretation of information status when tested with more L2 participants at a higher level of language proficiency.

APPENDIX A
C-TEST

Word Completion Exercise (To be completed within 15 minutes)

Directions: The two texts below contain gaps where parts of some words have been left out (no whole words are missing, though). In the blanks provided, please complete the words so that the sentences and texts make sense. Note that in each blank, you should only write one word; do not add extra words.

- **Text 1:**
  We all live with other people's expectations of us. These are a refle_____ of th_____ trying to under_____ us; th_____ are predic_____ of wh_____ they th_____ we will think, d_____ and feel. Gene_____ we acc_____ the sta_____ quo, but these expec_____ can be ha_____ to han_____ when they co_____ from our fami_____ and can be diff_____ to ign_____, especially wh_____ they come from our par_____.

- **Text 2:**
  The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction to helping children make better choices when it comes to what they eat and drink. Childhood obe_____ has bec_____ a ser_____ problem in th_____ country a_____ children cons_____ more sugar-based fo_____ and sp_____ less ti_____ getting the nece_____ exercise. Many par_____ have quest_____ schools' deci_____ to al_____ vending machines which disp_____ candy and so_____ drinks. Many schools, tho_____ , have co_____ to re_____ on the mo_____ these machines generate through agreements with the companies which make soft drinks and junk food.

C-test scoring criteria and answers

Scoring criteria:

- 1 point per correct blank
- Mis-spelling did not count (so participants received the point even if they mis-spelled the word).
- Everything else did matter, so if they made any mistake other than spelling they got 0 point: If they provided a continuation other than the one(s) provided, they did not get a point (example: writing: *display* instead of *dispense* was scored as wrong).
- Only the last blank in Text 1 allows two options: Either *partner* or *parents* is acceptable.
- Morphology matters, so if they forgot the *-ly* ending, plural marking, used *this* instead of *the*, etc., they did not get the point.

Answer key:

- Text 1:
  We all live with other people's expectations of us. These are a reflection of them trying to understand us; they are predictions of what they think we will think, do and feel. Generally we accept the status quo, but these expectations can be hard to handle when they come from our family and can be difficult to ignore, especially when they come from our parents/partner.
- Text 2:
  The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction to helping children make better choices when it comes to what they eat and drink. Childhood obesity has become a serious problem in this country as children consume more sugar-based food and spend less time getting the necessary exercise. Many parents have questioned schools' decisions to allow vending machines which dispense candy and soft drinks. Many schools, though, have come to rely on the money these machines generate through agreements with the companies which make soft drinks and junk food.

APPENDIX B
READ-ALOUD TASK

Elicitation paragraph from Weinberger (2015)

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

APPENDIX C
BACKGROUND QUESTIONNAIRE

The purpose of this survey is to collect information about your second language learning experiences. Your information will be used for the research purpose only and will not be reported nor provided to anybody besides the researcher. Please fill out your answers in the given spaces, or mark the appropriate answer if choices are given.

1. Participant ID

2. Sex

3. Age

4. Home city

5. Home country

6. University level
   (Undergraduate / Graduate / Other)

7. Major

8. Is your hearing normal?
   (Yes / No)

9. Is your vision normal or corrected to normal?
   (Yes / No)

10. Do you need vision correction to perform tasks on a computer?
    (Yes / No)

11. Please indicate which form(s) of vision correction you use to perform tasks on a computer monitor.
    (Eye glasses / Soft contact lenses / Hard contact lenses / Other / Not applicable)

12. What was the first language you learned as a child?

13. What language(s) was/were spoken in your home when you were growing up?

14. At what age did you start learning English?

15. How long in total have you lived in an English-speaking country (in years)?

16. How long have you lived in Hawaii (in years)?

17. Have you taken or are you currently taking ESL classes?

18. Have you taken a TOEFL test?
    (Yes / No)

19. If "Yes", which version of the test did you take?
    (PBT / CBT / iBT)

20. What was your total score?

21. When did you take the TOEFL test? (Please indicate year)

22. Have you taken a STEP Eiken test?
    (Yes / No)

23. If "Yes", what is the highest Eiken level you have passed?
    (Grade 1 / Grade pre-1 / Grade 2 / Grade pre-2 / Grade 3 / Grade pre-3)

24. When did you take the STEP Eiken test? (Please indicate year)

25. Please list all the languages you know, other than English, below. For each language, please indicate your proficiency on a scale from 'novice' to 'superior'

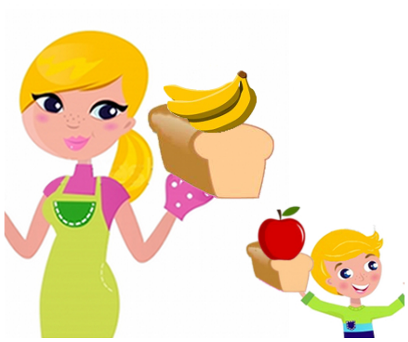
26. If there is any additional information you wish to share with us, please write it in the box below.

APPENDIX D
PROSODY-IN-CONTEXT NATURALNESS RATING TASK: INSTRUCTIONS

Directions

In each trial, you will hear a dialog: one of the 3 questions followed by the answer "Mariana made the banana bread." The answer is always the same sentence but is spoken in a different intonational pattern each time.

Your task is to listen to each dialog carefully and rate the intonational naturalness of the answer, "Mariana made the banana bread" in that dialog. Each dialog will be played only once.



Did Mariana make the apple bread?

Mariana made the banana bread.

What did Mariana do?

Mariana made the banana bread.

Who made the banana bread?

Mariana made the banana bread.

APPENDIX E
PROSODY-IN-CONTEXT NATURALNESS RATING TASK: ANALYSIS INCLUDING CONTEXT, PITCH, AND DURATION

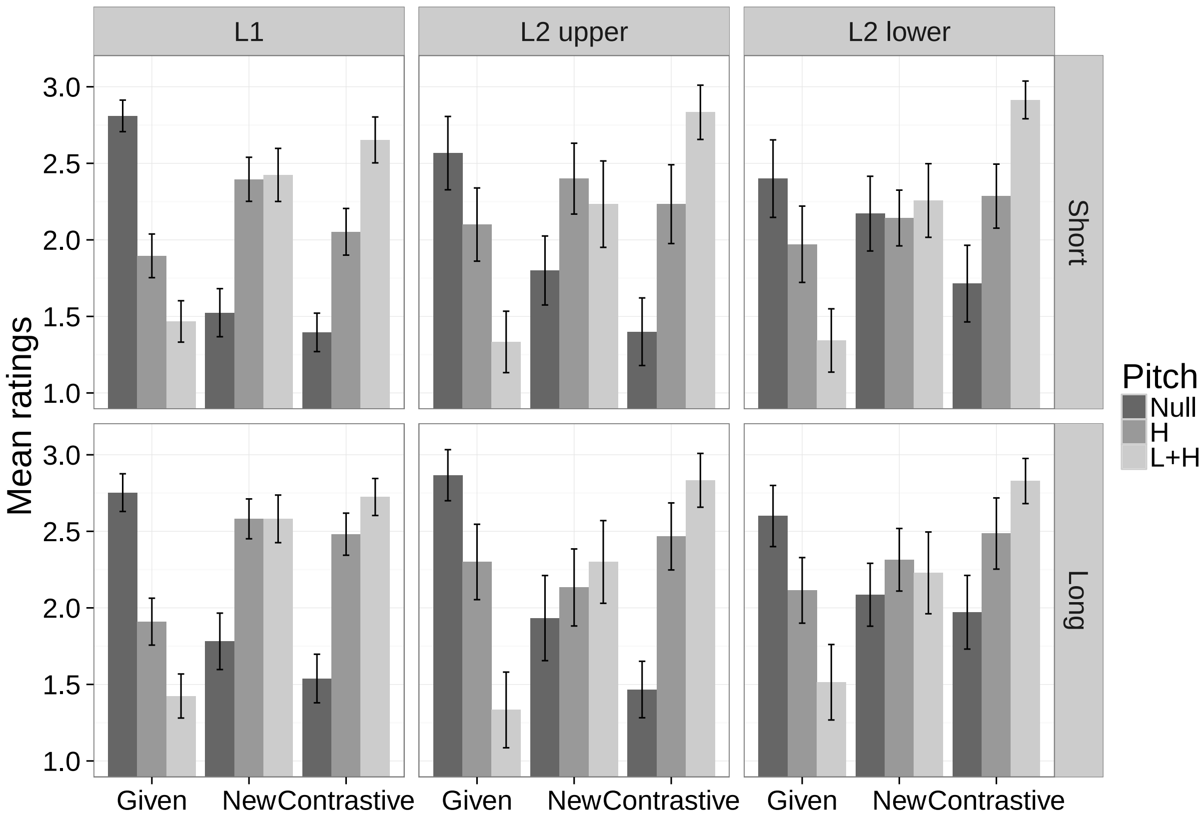


Figure E.1: L1 and L2 naturalness ratings by context, pitch, and duration

Table E.1: Cumulative link mixed model including context, pitch, and duration: L1 group

Fixed factor	Est.	SE	*z*	*p*
Null vs. H/L+H	0.65	0.13	4.94	$< .001$***
H vs. L+H	0.11	0.14	0.80	0.42
Given vs. New/Contrastive	0.34	0.13	2.60	$< .01$**
New vs. Contrastive	-0.14	0.15	-0.97	0.33
Short vs. Long	0.35	0.12	2.92	$< .01$**
Null vs. H/L+H : Given vs. New/Contrastive	6.25	0.33	19.00	$< .001$***
H vs. L+H : Given vs. New/Contrastive	2.21	0.30	7.35	$< .001$***
Null vs. H/L+H : New vs. Contrastive	0.57	0.30	1.88	0.05
H vs. L+H : New vs. Contrastive	1.38	0.36	3.81	$< .001$***
Null vs. H/L+H : Short vs. Long	0.05	0.26	0.19	0.85
H vs. L+H : Short vs. Long	-0.46	0.29	-1.62	0.10
Given vs. New/Contrastive : Short vs. Long	0.79	0.26	3.04	$< .01$**
New vs. Contrastive : Short vs. Long	-0.03	0.29	-0.09	0.93
Null vs. H/L+H : Given vs. New/Contrastive : Short vs. Long	-0.24	0.58	-0.41	0.68
H vs. L+H : Given vs. New/Contrastive : Short vs. Long	-0.47	0.59	-0.80	0.42
Null vs. H/L+H : New vs. Contrastive : Short vs. Long	0.47	0.60	0.78	0.44
H vs. L+H : New vs. Contrastive : Short vs. Long	-0.85	0.72	-1.18	0.24

Table E.2: Cumulative link mixed model including context, pitch, and duration: L2 upper group

Fixed factor	Est.	SE	*z*	*p*
Null vs. H/L+H	0.51	0.20	2.50	$<$ .01*
H vs. L+H	-0.16	0.24	-0.68	0.49
Given vs. New/Contrastive	0.29	0.21	1.41	0.16
New vs. Contrastive	0.43	0.23	1.85	0.06
Short vs. Long	0.29	0.19	1.51	0.13
Null vs. H/L+H : Given vs. New/Contrastive	5.47	0.50	11.04	$<$ .001***
H vs. L+H : Given vs. New/Contrastive	3.62	0.50	7.20	$<$ .001***
Null vs. H/L+H : New vs. Contrastive	2.50	0.47	5.28	$<$ .001***
H vs. L+H : New vs. Contrastive	1.98	0.59	3.34	$<$ .001***
Null vs. H/L+H : Short vs. Long	-0.61	0.41	-1.50	0.13
H vs. L+H : Short vs. Long	-0.12	0.47	-0.26	0.79
Given vs. New/Contrastive : Short vs. Long	-0.39	0.41	-0.94	0.35
New vs. Contrastive : Short vs. Long	0.42	0.46	0.92	0.36
Null vs. H/L+H : Given vs. New/Contrastive : Short vs. Long	0.89	0.91	0.98	0.33
H vs. L+H : Given vs. New/Contrastive : Short vs. Long	0.89	0.98	0.91	0.36
Null vs. H/L+H : New vs. Contrastive : Short vs. Long	0.62	0.93	0.67	0.51
H vs. L+H : New vs. Contrastive : Short vs. Long	-1.60	1.18	-1.36	0.17

Table E.3: Cumulative link mixed model including context, pitch, and duration: L2 lower group

Fixed factor	Est.	SE	*z*	*p*
Null vs. H/L+H	0.16	0.18	0.91	0.36
H vs. L+H	0.05	0.22	0.23	0.81
Given vs. New/Contrastive	0.93	0.18	5.14	$<$ .001***
New vs. Contrastive	0.78	0.21	3.61	0.10
Short vs. Long	0.28	0.17	1.63	0.72
Null vs. H/L+H : Given vs. New/Contrastive	3.70	0.39	9.41	$<$ .001***
H vs. L+H : Given vs. New/Contrastive	2.90	0.45	6.39	$<$ .001***
Null vs. H/L+H : New vs. Contrastive	2.22	0.43	5.16	$<$ .001***
H vs. L+H : New vs. Contrastive	2.02	0.56	3.62	$<$ .001***
Null vs. H/L+H : Short vs. Long	-0.21	0.35	-0.59	0.56
H vs. L+H : Short vs. Long	-0.65	0.44	-1.47	0.14
Given vs. New/Contrastive : Short vs. Long	-0.33	0.36	-0.92	0.36
New vs. Contrastive : Short vs. Long	0.16	0.43	0.37	0.71
Null vs. H/L+H : Given vs. New/Contrastive : Short vs. Long	0.33	0.75	0.44	0.66
H vs. L+H : Given vs. New/Contrastive : Short vs. Long	-0.96	0.90	-1.07	0.28
Null vs. H/L+H : New vs. Contrastive : Short vs. Long	-0.96	0.85	-1.12	0.26
H vs. L+H : New vs. Contrastive : Short vs. Long	-0.77	1.11	-0.69	0.49

APPENDIX F
PROSODY-IN-CONTEXT NATURALNESS RATING TASK: STATISTICAL ANALYSIS TABLE INCLUDING CONTEXT, PITCH, AND L2 PROFICIENCY GROUP

Table F.1: Cumulative link mixed model including context, pitch, and L2 proficiency group

Fixed factor	Est.	SE	*z*	*p*
Null vs. H/L+H	0.45	0.23	1.94	0.05
H vs. L+H	0.12	0.30	0.41	0.68
Given vs. New/Contrastive	0.87	0.26	3.36	$<$ .001***
New vs. Contrastive	0.98	0.35	2.83	$<$ .01**
Proficiency	-0.08	0.22	-0.36	0.72
Null vs. H/L+H : Given vs. New/Contrastive	5.88	0.53	11.12	$<$ .001***
H vs. L+H : Given vs. New/Contrastive	4.54	0.56	8.16	$<$ .001***
Null vs. H/L+H : New vs. Contrastive	3.15	0.55	5.66	$<$ .001***
H vs. L+H : New vs. Contrastive	3.11	0.85	3.66	$<$ .001***
Null vs. H/L+H : Proficiency	0.50	0.39	1.27	0.20
H vs. L+H : Proficiency	-0.21	0.42	-1.43	0.61
Given vs. New/Contrastive : Proficiency	-0.64	0.45	-1.82	0.15
New vs. Contrastive : Proficiency	-0.30	0.56	-0.54	0.60
Null vs. H/L+H : Given vs. New/Contrastive : Proficiency	2.25	0.84	2.68	$<$ .01**
H vs. L+H : Given vs. New/Contrastive : Proficiency	1.09	0.83	1.31	0.19
Null vs. H/L+H : New vs. Contrastive : Proficiency	0.48	0.92	0.53	0.60
H vs. L+H : New vs. Contrastive : Proficiency	0.24	1.18	0.22	0.84

APPENDIX G
EYE-TRACKING LISTENING COMPREHENSION TASK: ACOUSTIC ANALYSIS OF TARGET STIMULI

Table G.1: Acoustic analysis of the target stimuli for the new–contrastive experiment

		Mean pitch					Pitch excursion					Duration				
		Est.	SE	DF	*t*	*p*	Est.	SE	DF	*t*	*p*	Est.	SE	DF	*t*	*p*
Now	(Intercept)	185.69	0.79	23	233.78	<.001***	85.26	2.15	23	39.72	<.001***	678.96	25.88	23	26.24	<.001***
	H* vs. L+H*	-5.53	1.59	23	-3.48	<.01**	3.96	3.31	23	1.20	0.24	223.75	42.10	23	5.31	<.001**
use	(Intercept)	168.04	1.62	23	103.60	<.001***	36.00	5.17	23	6.97	<.001***	214.17	4.40	23	48.68	<.001***
	H* vs. L+H*	5.23	3.24	23	1.61	0.12	-8.44	8.40	23	-1.00	0.33	4.17	6.48	23	0.64	0.53
the	(Intercept)	170.36	2.70	23	63.14	<.001***	23.61	3.77	23	6.25	<.001***	99.17	5.49	23	18.07	<.001***
	H* vs. L+H*	5.80	5.40	20	1.08	0.30	7.61	7.55	20	1.01	0.33	0.83	5.35	23	0.16	0.88
Adj.	(Intercept)	193.98	1.91	23	101.42	<.001***	93.89	4.53	23	20.71	<.001***	376.67	10.64	23	35.42	<.001***
	H* vs. L+H*	15.65	3.83	23	4.09	<.001***	62.52	9.07	23	6.90	<.001***	76.67	16.38	23	4.68	<.001***
Noun	(Intercept)	165.29	2.27	23	72.94	<.001***	87.54	6.49	23	13.49	<.001***	535.21	23.66	23	22.62	<.001***
	H* vs. L+H*	-53.63	3.86	23	-13.90	<.001***	-45.43	10.63	23	-4.27	<.001***	3.75	15.25	23	0.25	0.81
to	(Intercept)	145.24	3.02	21	48.16	<.001***	12.49	4.85	21	2.57	<.05*	108.54	3.75	23	28.92	<.001***
	H* vs. L+H*	-0.82	6.03	9	-0.14	0.90	-0.91	1.49	9	-0.61	0.56	4.58	4.81	23	0.95	0.35
color	(Intercept)	163.45	0.97	23	167.98	<.001***	27.36	2.21	23	12.36	<.001***	300.83	4.53	23	66.39	<.001***
	H* vs. L+H*	4.21	1.84	23	2.28	<.05*	6.87	4.27	23	1.61	0.12	1.67	7.58	23	0.22	0.83
the	(Intercept)	166.52	1.94	23	85.72	<.001***	28.49	4.28	23	6.65	<.001***	93.96	3.60	23	26.11	<.001***
	H* vs. L+H*	7.64	3.19	23	2.39	<.05*	3.59	5.04	23	0.71	0.48	3.75	3.45	23	1.09	0.29
Animal	(Intercept)	194.70	2.35	23	82.86	<.001***	85.88	5.82	23	14.76	<.001***	487.71	15.28	23	31.91	<.001***
	H* vs. L+H*	6.65	3.50	23	1.90	0.07	8.31	7.72	23	1.08	0.29	-2.08	13.66	23	-0.15	0.88

Table G.2: Acoustic analysis of the target stimuli for the new–given experiment

		Mean pitch					Duration				
		Est.	SE	DF	*t*	*p*	Est.	SE	DF	*t*	*p*
Now	(Intercept)	186.89	0.90	23	208.00	<.001***	620.42	26.96	23	23.01	<.001***
	H* vs. Null accent	3.12	1.78	23	1.75	0.09	-106.67	40.79	23	-2.61	<.05*
use	(Intercept)	161.13	1.66	23	96.96	<.001***	237.50	4.58	23	51.81	<.001***
	H* vs. Null accent	8.60	3.32	23	2.59	<.05*	-50.83	7.73	23	-6.58	<.001***
the	(Intercept)	165.23	2.11	23	78.18	<.001***	95.42	4.75	23	20.09	<.001***
	H* vs. Null accent	4.45	4.23	20	1.05	0.30	6.67	6.13	23	1.09	0.29
Adj.	(Intercept)	170.36	1.32	23	129.42	<.001***	316.04	10.06	23	31.42	<.001***
	H* vs. Null accent	31.60	2.63	23	12.00	<.001***	44.58	11.36	23	3.93	<.001***
Noun	(Intercept)	173.10	1.45	23	119.05	<.001***	502.29	24.53	23	20.48	<.001***
	H* vs. Null accent	38.01	2.91	23	13.07	<.001***	62.08	11.44	23	5.43	<.001***
to	(Intercept)	149.63	2.13	23	70.34	<.001***	110.83	3.21	23	34.52	<.001***
	H* vs. Null accent	-7.95	4.25	14	-1.87	0.08	-9.17	6.42	23	-1.43	0.17
color	(Intercept)	157.90	1.00	23	158.07	<.001***	305.63	4.87	23	62.70	<.001***
	H* vs. Null accent	6.90	1.15	23	6.00	<.001***	-11.25	7.28	23	-1.55	0.14
the	(Intercept)	155.73	1.05	23	208.00	<.001***	100.83	4.07	23	24.79	<.001***
	H* vs. Null accent	13.95	2.11	23	6.62	<.001***	-17.50	3.42	23	-5.12	<.001***
Animal	(Intercept)	193.09	3.06	23	63.15	<.001***	496.25	15.04	23	32.99	<.001***
	H* vs. Null accent	-3.45	3.51	23	-0.98	0.34	-15.00	12.13	23	-1.24	0.23

APPENDIX H
EYE-TRACKING LISTENING COMPREHENSION TASK: ITEM LIST

Table H.1: Example item list for the new–contrastive experiment

	Condition		Instruction 1		Instruction 2	
			Tool	Animal	Tool	Animal
1	Filler		white paintbrush	iguana	red spray bottle	parrot
2	Target	Contrastive–L+H*	orange crayon	bird	green crayon	bee
3	Filler		white spray bottle	goldfish	white roller	flamingo
4	Target	Contrastive–H*	yellow crayon	fish	white crayon	penguin
5	Filler		blue crayon	cockroach	red paintbrush	chicken
6	Target	New–L+H*	orange paintbrush	frog	green roller	lion
7	Filler		orange crayon	wolf	while paintbrush	peacock
8	Target	New–L+H*	white spray bottle	cat	silver paintbrush	dolphin
9	Target	Contrastive–H*	purple spray bottle	horse	orange spray bottle	gorilla
10	Filler		purple roller	shrimp	red crayon	lizard
11	Filler		white crayon	alligator	purple paintbrush	fox
12	Target	Contrastive–L+H*	blue spray bottle	gorilla	yellow spray bottle	sheep
13	Filler		purple spray bottle	bat	purple roller	owl
14	Filler		orange paintbrush	squirrel	yellow spray bottle	zebra
15	Target	New–H*	green crayon	dolphin	blue paintbrush	cow
16	Target	New–H*	purple paintbrush	lion	orange crayon	snail
17	Filler		yellow roller	oyster	blue crayon	goldfish
18	Target	Contrastive–H*	purple paintbrush	octopus	orange paintbrush	snake
19	Filler		purple paintbrush	ostrich	blue spray bottle	pelican
20	Target	New–H*	yellow spray bottle	koala	white roller	rabbit
21	Filler		green roller	flamingo	orange crayon	jellyfish
22	Target	New–L+H*	white roller	bee	silver spray bottle	cat
23	Filler		yellow spray bottle	pelican	yellow roller	donkey
24	Target	Contrastive–L+H*	white roller	snake	silver roller	shark
25	Filler		red crayon	zebra	green paintbrush	deer
26	Filler		red spray bottle	donkey	red roller	squirrel
27	Target	Contrastive–L+H*	red paintbrush	cow	purple paintbrush	elephant
28	Target	Contrastive–L+H*	blue crayon	shark	yellow crayon	pig
29	Filler		silver paintbrush	hippo	purple spray bottle	wolf
30	Target	Contrastive–H*	yellow spray bottle	rabbit	white spray bottle	frog

	Condition		Instruction 1		Instruction 2	
			Tool	Animal	Tool	Animal
31	Filler		green paintbrush	parrot	white spray bottle	cockroach
32	Target	Contrastive–L+H*	red spray bottle	goat	purple spray bottle	mouse
33	Target	New–H*	silver crayon	duck	red spray bottle	bird
34	Filler		green crayon	owl	silver paintbrush	hippo
35	Target	New–L+H*	orange roller	mouse	green paintbrush	giraffe
36	Filler		blue roller	mosquito	green crayon	raccoon
37	Target	Contrastive–H*	green crayon	pig	blue crayon	horse
38	Target	New–H*	silver roller	tiger	red paintbrush	dog
39	Filler		silver roller	jellyfish	white crayon	iguana
40	Filler		orange spray bottle	jellyfish	blue roller	beetle
41	Target	New–L+H*	red roller	snail	purple crayon	bear
42	Filler		yellow paintbrush	peacock	silver spray bottle	shrimp
43	Target	Contrastive–H*	silver roller	penguin	red roller	butterfly
44	Filler		silver crayon	ant	orange paintbrush	turkey
45	Target	New–H*	green spray bottle	giraffe	blue roller	panda
46	Filler		blue spray bottle	hamster	blue roller	alligator
47	Target	New–L+H*	blue paintbrush	elephant	yellow roller	goat
48	Filler		red roller	caterpillar	silver crayon	bat

Table H.2: Example item list for the new–given experiment

	Condition		Instruction 1		Instruction 2	
			Tool	Animal	Tool	Animal
1	Filler		white paintbrush	iguana	red spray bottle	parrot
2	Target	Given–Null	green crayon	bird	green crayon	bee
3	Filler		white spray bottle	goldfish	white roller	flamingo
4	Target	Given–H*	white crayon	fish	white crayon	penguin
5	Filler		blue crayon	cockroach	red paintbrush	chicken
6	Target	New–Null accent	orange paintbrush	frog	green roller	lion
7	Filler		orange crayon	wolf	while paintbrush	peacock
8	Target	New–Null accent	white spray bottle	cat	silver paintbrush	dolphin
9	Target	Given–H*	orange spray bottle	horse	orange spray bottle	gorilla
10	Filler		purple roller	shrimp	red crayon	lizard
11	Filler		white crayon	alligator	purple paintbrush	fox
12	Target	Given–Null accent	yellow spray bottle	gorilla	yellow spray bottle	sheep
13	Filler		purple spray bottle	bat	purple roller	owl
14	Filler		orange paintbrush	squirrel	yellow spray bottle	zebra
15	Target	New–H*	green crayon	dolphin	blue paintbrush	cow
16	Target	New–H*	purple paintbrush	lion	orange crayon	snail
17	Filler		yellow roller	oyster	blue crayon	goldfish
18	Target	Given–H*	orange paintbrush	octopus	orange paintbrush	snake
19	Filler		purple paintbrush	ostrich	blue spray bottle	pelican
20	Target	New–H*	yellow spray bottle	koala	white roller	rabbit
21	Filler		green roller	flamingo	orange crayon	jellyfish
22	Target	New–Null accent	white roller	bee	silver spray bottle	cat
23	Filler		yellow spray bottle	pelican	yellow roller	donkey
24	Target	Given–Null accent	silver roller	snake	silver roller	shark
25	Filler		red crayon	zebra	green paintbrush	deer
26	Filler		red spray bottle	donkey	red roller	squirrel
27	Target	Given–Null accent	purple paintbrush	cow	purple paintbrush	elephant
28	Target	Given–Null accent	yellow crayon	shark	yellow crayon	pig
29	Filler		silver paintbrush	hippo	purple spray bottle	wolf
30	Target	Given–H*	white spray bottle	rabbit	white spray bottle	frog

	Condition		Instruction 1		Instruction 2	
			Tool	Animal	Tool	Animal
31	Filler		green paintbrush	parrot	white spray bottle	cockroach
32	Target	Given–Null accent	purple spray bottle	goat	purple spray bottle	mouse
33	Target	New–H*	silver crayon	duck	red spray bottle	bird
34	Filler		green crayon	owl	silver paintbrush	hippo
35	Target	New–Null accent	orange roller	mouse	green paintbrush	giraffe
36	Filler		blue roller	mosquito	green crayon	raccoon
37	Target	Given–H*	blue crayon	pig	blue crayon	horse
38	Target	New–H*	silver roller	tiger	red paintbrush	dog
39	Filler		silver roller	jellyfish	white crayon	iguana
40	Filler		orange spray bottle	jellyfish	blue roller	beetle
41	Target	New–Null accent	red roller	snail	purple crayon	bear
42	Filler		yellow paintbrush	peacock	silver spray bottle	shrimp
43	Target	Given–H*	red roller	penguin	red roller	butterfly
44	Filler		silver crayon	ant	orange paintbrush	turkey
45	Target	New–H*	green spray bottle	giraffe	blue roller	panda
46	Filler		blue spray bottle	hamster	blue roller	alligator
47	Target	New–Null accent	blue paintbrush	elephant	yellow roller	goat
48	Filler		red roller	caterpillar	silver crayon	bat

APPENDIX I
EYE-TRACKING LISTENING COMPREHENSION TASK: PROPORTIONS OF FIXATIONS TO TARGET, COMPETITOR, OTHER DRAWING TOOLS, AND ANIMALS (CONDITIONS COLLAPSED)

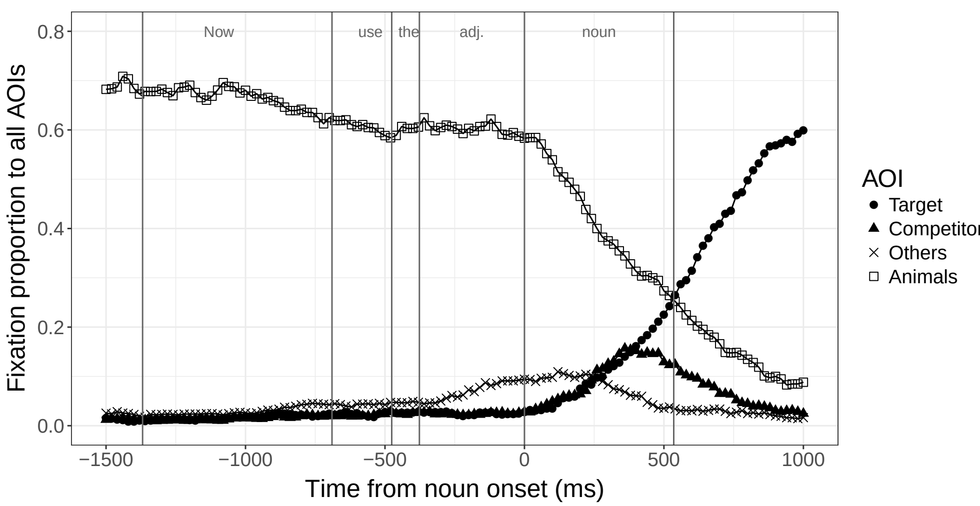


Figure I.1: New–contrastive experiment: Fixation proportions to AOIs for the L1 group

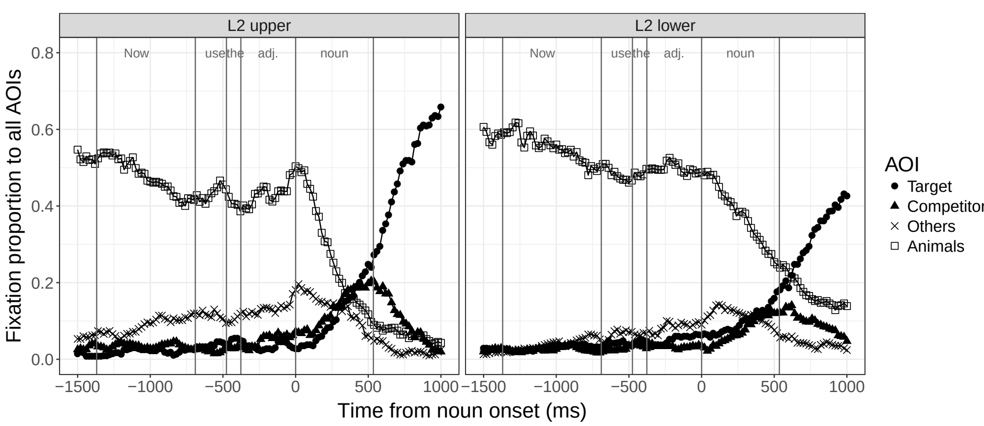


Figure I.2: New–contrastive experiment: Fixation proportions to AOIs for the L2 group

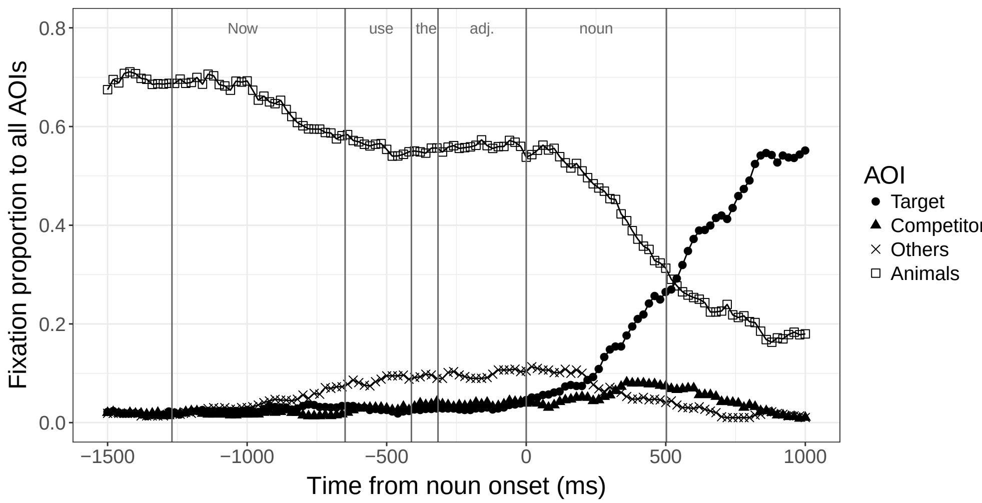


Figure I.3: New–given experiment: Fixation proportions to AOIs for the L1 group

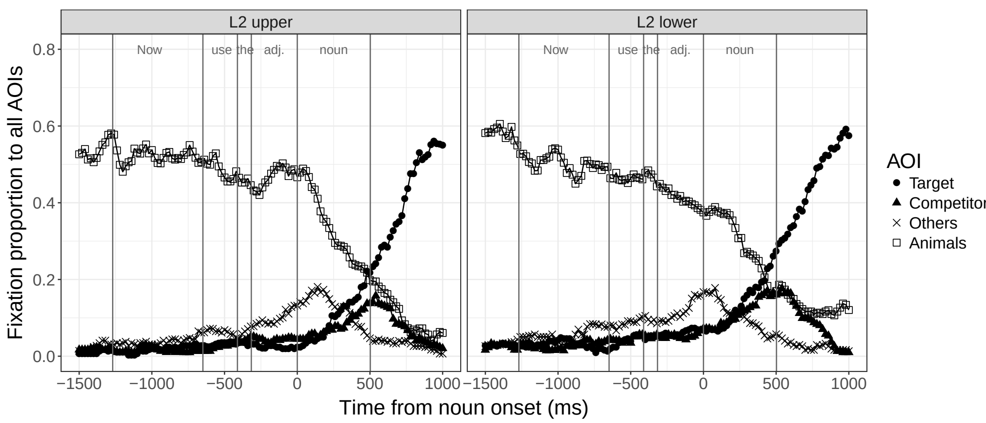


Figure I.4: New–given experiment: Fixation proportions to AOIs for the L2 group

APPENDIX J
EYE-TRACKING LISTENING COMPREHENSION TASK: FIXATION LOGIT FOR THE TARGET/COMPETITOR (CONDITIONS COLLAPSED)

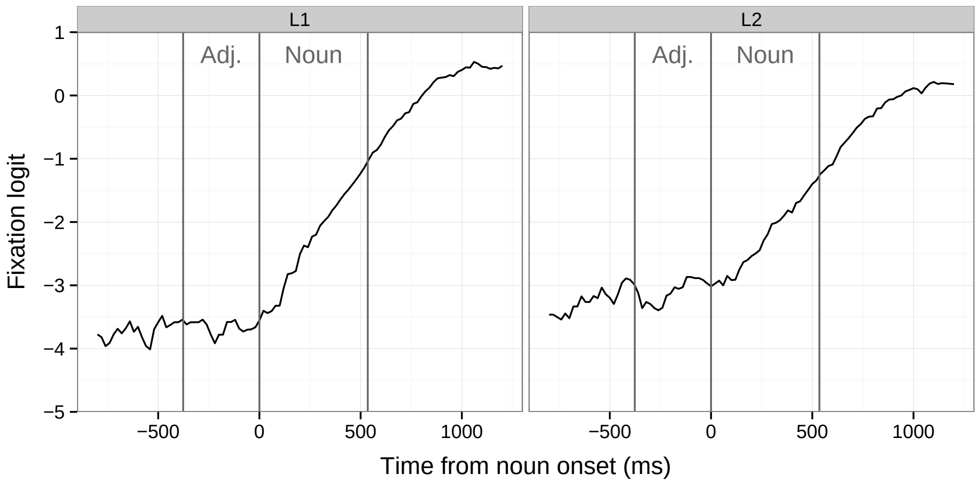


Figure J.1: New–contrastive experiment: Fixation logit for the target across experimental conditions

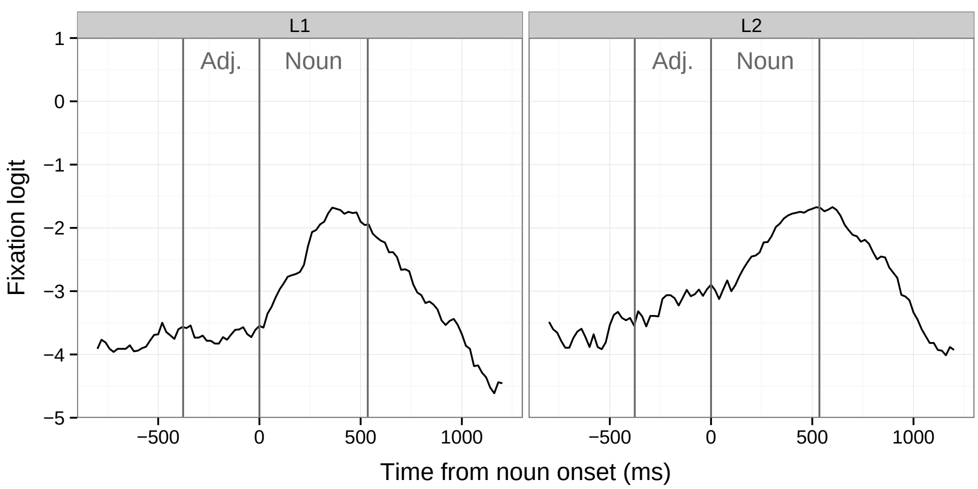


Figure J.2: New–contrastive experiment: Fixation logit for the competitor across experimental conditions

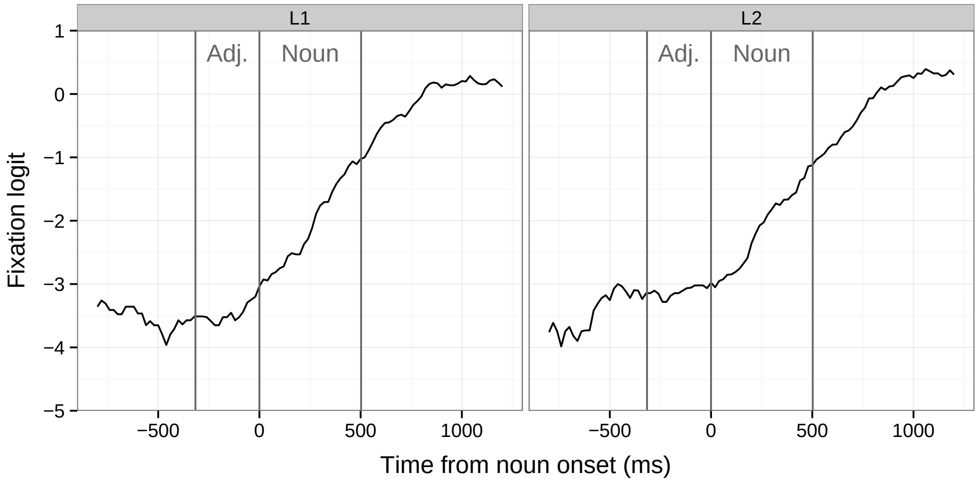


Figure J.3: New–given experiment: Fixation logit for the target across experimental conditions

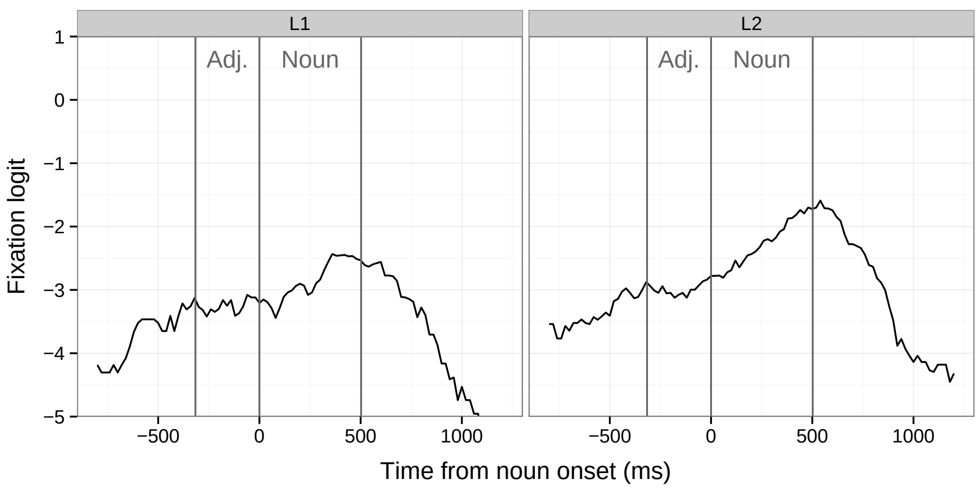
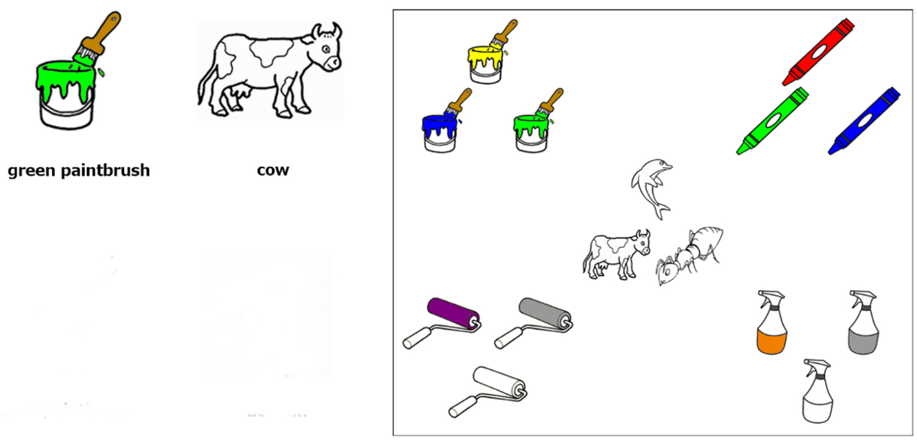


Figure J.4: New–given experiment: Fixation logit for the competitor across experimental conditions

APPENDIX K
PRODUCTION TASK: INSTRUCTIONS

Directions

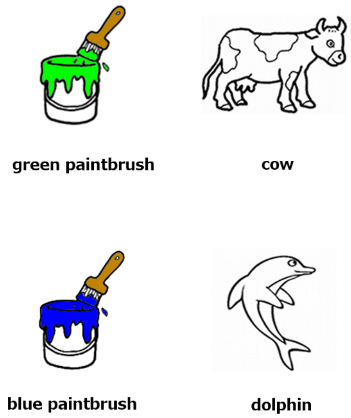
1. At first, you will see an animal and a drawing instrument on the left side of your screen. The right side of the screen is the display that your partner sees on her computer.

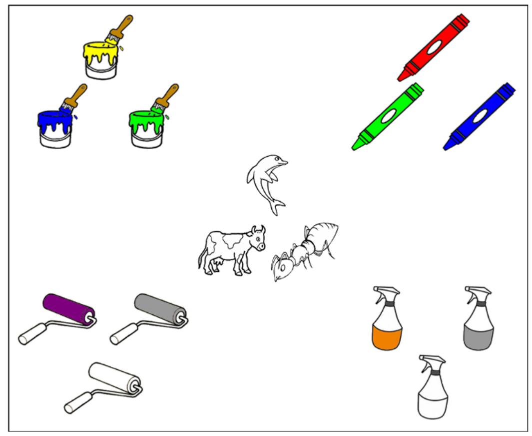


2. Your task is to direct your partner to color the specified animals using the specified drawing instruments. For example, if you see a green paintbrush and a cow on the left side of the screen, say:

   Use the green paintbrush to color the cow.

3. After your partner has clicked on the instrument and the animal, press the space bar. You will see another pair of a drawing instrument and an animal.





4. Using a similar sentence pattern, give directions to your partner. For example, say:

   Now, use the blue paintbrush to color the dolphin.

APPENDIX L
PRODUCTION TASK: ITEM LIST

Table L.1: Example item list

	Condition	Instruction 1		Instruction 2	
		Tool	Animal	Tool	Animal
1	Given	yellow paintbrush	koala	yellow paintbrush	turtle
2	Adjective contrastive	white crayon	sheep	silver crayon	tiger
3	New	green paintbrush	octopus	blue spray bottle	kangaroo
4	New	yellow spray bottle	butterfly	white roller	penguin
5	Adjective contrastive	white paintbrush	dolphin	silver paintbrush	cat
6	New	yellow roller	fish	white crayon	monkey
7	New	green spray bottle	panda	blue roller	giraffe
8	Adjective contrastive	white spray bottle	duck	silver spray bottle	bee
9	Given	orange spray bottle	mouse	orange spray bottle	goat
10	New	yellow crayon	dog	white paintbrush	tiger
11	Adjective contrastive	orange spray bottle	turtle	green spray bottle	bear
12	Given	yellow spray bottle	sheep	yellow spray bottle	gorilla
13	New	yellow paintbrush	cat	white spray bottle	bird
14	Adjective contrastive	orange paintbrush	giraffe	green paintbrush	mouse
15	Adjective contrastive	white roller	shark	silver roller	snake
16	New	green crayon	cow	blue paintbrush	dolphin
17	Adjective contrastive	orange roller	lion	green roller	frog
18	Given	yellow crayon	pig	yellow crayon	shark
19	New	purple paintbrush	ostrich	blue spray bottle	pelican
20	Given	orange paintbrush	elephant	orange paintbrush	cow
21	Given	yellow roller	goat	yellow roller	elephant
22	Adjective contrastive	orange crayon	bee	green crayon	bird
23	Given	orange roller	monkey	orange roller	dog
24	Given	orange crayon	bear	orange crayon	snail

APPENDIX M
PRODUCTION TASK: ACOUSTIC ANALYSIS ON THE NOUN REGION

Table M.1: Mean pitch: L1 group

Fixed factor	Est.	SE	*t*	*p*
Intercept	186.94	6.26	29.88	<.001***
Given vs. New	5.45	2.15	2.53	<.05*
Given vs. Contrastive	0.08	1.97	0.04	0.97

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-5.37	2.30	-2.33	0.05

Table M.2: Mean pitch: L2 upper group

Fixed factor	Est.	SE	*t*	*p*
Intercept	218.18	6.89	31.68	<.001***
Given vs. New	-2.13	2.22	-0.96	0.34
Given vs. Contrastive	1.63	2.23	0.73	0.75

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-5.37	2.30	-2.33	0.05

Table M.3: Mean pitch: L2 lower group

Fixed factor	Est.	SE	*t*	*p*
Intercept	202.01	8.59	23.51	<.001***
Given vs. New	0.14	2.65	0.05	0.95
Given vs. Contrastive	1.70	2.27	0.75	0.45

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	1.56	2.33	0.67	0.78

Table M.4: Pitch excursion: L1 group

Fixed factor	Est.	SE	*t*	*p*
Intercept	162.31	15.31	10.60	<.001***
Given vs. New	20.46	6.40	3.20	<.01**
Given vs. Contrastive	13.27	6.77	1.96	<.05*

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-7.18	6.29	-2.33	0.49

Table M.5: Pitch excursion: L2 upper group

Fixed factor	Est.	SE	*t*	*p*
Intercept	188.87	16.32	11.58	<.001***
Given vs. New	-9.63	11.23	-0.86	0.39
Given vs. Contrastive	-15.63	10.64	-1.47	0.14

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-6.00	9.97	-0.60	0.82

Table M.6: Pitch excursion: L2 lower group

Fixed factor	Est.	SE	*t*	*p*
Intercept	179.38	16.23	11.05	<.001***
Given vs. New	3.11	8.93	0.35	0.73
Given vs. Contrastive	3.62	9.63	0.38	0.71

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	0.50	8.86	0.06	1.00

Table M.7: Duration: L1 group

Fixed factor	Est.	SE	*t*	*p*
Intercept	520.96	29.24	17.81	<.001***
Given vs. New	37.15	4.97	7.48	<.001***
Given vs. Contrastive	6.68	5.50	1.21	2.25

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-5.37	2.30	-2.33	0.05

Table M.8: Duration: L2 upper group

Fixed factor	Est.	SE	*t*	*p*
Intercept	626.09	35.03	17.87	<.001***
Given vs. New	11.45	8.36	1.37	0.17
Given vs. Contrastive	4.82	9.64	0.50	0.62

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-6.63	8.89	-0.75	0.74

Table M.9: Duration: L2 lower group

Fixed factor	Est.	SE	*t*	*p*
Intercept	672.52	37.21	18.07	<.001***
Given vs. New	10.55	9.95	1.06	0.29
Given vs. Contrastive	-2.29	10.47	-0.22	0.83

Tukey's post-hoc pairwise comparison

Fixed factor	Est.	SE	*z*	*p*
New vs. Contrastive	-12.84	9.86	-1.30	0.39

REFERENCES

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81–96.

Algethami, G. (2017). The effects of explicit pronunciation instruction on the degree of perceived foreign accent in the speech of EFL learners. *Research in Language*, *15*(3), 253–263.

Archibald, J. (1997). The acquisition of second language phrasal stress: A pilot study. *Language Acquisition and Language Disorders*, *16*, 263–290.

Arnold, J. E. (2008). *THE BACON* not *the bacon*: How children and adults understand accented and unaccented noun phrases. *Cognition*, *108*(1), 69–99.

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*(3), 505–512.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, *4*(2), 239–265.

Baker, R. E. (2010). *The acquisition of English focus marking by non-native speakers* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. *The Journal of the Acoustical Society of America*, *95*(5), 2973–2973.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, The Netherlands: Foris Publications.

Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions*. Columbus, OH: Ohio State University.

Beckman, M. E., Hirschberg, J. B., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). New York: Oxford University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(1), 255–309.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365–391.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*(1), 64–76.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer. Version 6.0.14.* Retrieved from `http://www.praat.org/`

Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, *16*(1), 71–77.

Braun, B., & Chen, A. (2010). Intonation of ‘now’ in resolving scope ambiguity in English and Dutch. *Journal of Phonetics*, *38*(3), 431–444.

Braun, B., & Tagliapietra, L. (2011). On-line interpretation of intonational meaning in L2. *Language and Cognitive Processes*, *26*(2), 224–235.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098.

Breen, M., Kurumada, C., Wagner, M., Watson, D. G., & Yu, K. M. (2018). Introducing prosodic variability. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 1–2.

Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–77). New York: Springer.

Büring, D. (2006). Focus projection and default prominence. *The Architecture of Focus*, *82*(321), 365–391.

Büring, D. (2016). *Intonation and meaning*. Oxford, United Kingdom: Oxford University Press.

Calhoun, S. (2007). *Information structure and the prosodic structure of English: A probabilistic relationship* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh, United Kingdom.

Calhoun, S. (2010a). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*(1), 1–42.

Calhoun, S. (2010b). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7-9), 1099–1140.

Carroll, S. E. (2006). Salience, awareness and SLA. In L. Dekydtspotter, B. D. Schwartz, R. A. Sprouse, M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 17–24). Somerville, MA: Cascadilla Press.

Carroll, S. E., & Shea, C. (2007). Salience, focus and second language acquisition. *Nouveaux Cahiers de Linguistique Française*, *28*, 99–106.

Chafe, W. L. (1974). Language and consciousness. *Language*, *50*(1), 111–133.

Chafe, W. L. (1987). Cognitive constraints on information flow. *Coherence and Grounding in Discourse*, *11*, 21–51.

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning*, *59*(2), 367–409.

Chen, A., Den Os, E., & De Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review*, *24*(2-3), 317–344.

Chen, A., & Lai, V. T. (2011). *Comb* or *coat*: The role of intonation in online reference resolution in a second language. In W. Zonneveld & H. Quené (Eds.), *Sound and sounds. Studies presented to M.E.H. (Bert) Schouten on the occasion of his 65th birthday* (pp. 57–68). Utrecht: UiL OTS.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *The Journal of the Acoustical Society of America*, *117*(6), 3867–3878.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam, The Netherlands: John Benjamins Publishing.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*, 287–299.

Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience*, *30*(1-2), 1–31.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142–2156.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language*, *12*(3), 643–661.

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1–23.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*(1), 1–10.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature*, *304*, 159–160.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*(4), 385–400.

Cutler, A., Mehler, J., Otake, T., & Hatano, G. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*(2), 258–278.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, *33*(6), 824–844.

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 441–452.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314.

Dehé, N. (2009). Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics*, *45*(3), 569–615.

De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, *97*(1), 491–504.

Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition*, *30*(4), 453.

de Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism*, *16*(1), 101–116.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*(3), 526–548.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, *13*(1), 1–17.

Dilley, L. C. (2010). Pitch range variation in English tonal contrasts: Continuous or categorical? *Phonetica*, *67*(1-2), 63–81.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682–706.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, *80*(2), 402–415.

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., & Lotts, D. W. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, *29*(3), 233–251.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of psycholinguistic research*, *24*(6), 409–436.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315–330.

Eckman, F. R. (2008). Typological markedness and second language phonology. *Phonology and Second Language Acquisition*, *36*, 95–115.

Eliasson, S. (1997). Tone in second language acquisition. *Trends in Linguistics Studies and Monographs*, *101*, 1273–1290.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore, MD: York Press.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386.

Gili Fivela, B. (2012). Testing the perception of L2 intonation. In M. G. Busá & A. Stella (Eds.), *Methodological perspectives on second language prosody. Papers from ML2P 2012* (p. 17-30). Padova: Cleup.

Goad, H., & White, L. (2004). Ultimate attainment of L2 inflection: Effects of L1 prosodic structure. *EUROSLA Yearbook*, *4*, 119–145.

Goad, H., & White, L. (2008). Prosodic structure and the representation of L2 functional morphology: A nativist approach. *Lingua*, *118*(4), 577–594.

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, *39*(3), 192–193.

Gussenhoven, C. (2005). Transcription of Dutch intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 118–145). Oxford: Oxford University Press Oxford.

Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition*, *36*(2), 283–302.

Gut, U., Pillai, S., & Don, Z. M. (2013). The prosodic marking of information status in Malaysian English. *World Englishes*, *32*(2), 185–197.

Harley, B., Howard, J., & Hart, D. (1995). Second language processing at different ages: Do younger learners pay more attention to prosodic cues to sentence structure? *Language Learning*, *45*(1), 43–71.

He, X., Hanssen, J., van Heuven, V. J., & Gussenhoven, C. (2011). Phonetic implementation must be learnt: Native versus Chinese realization of focus accent in Dutch. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 843–846). Hong Kong: City University of Hong Kong.

Hedberg, N., & Sosa, J. M. (2008). The prosody of topic and focus in spontaneous English dialogue. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 101–120). New York: Springer.

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, *34*(3), 103–123.

Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. In A. W. Biermann (Ed.), *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics* (pp. 136–144). Stroudsburg, PA: Association for Computational Linguistics.

Huang, B. H., & Jun, S.-A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech*, 387–414.

Hwang, H. (2007). *Prosodic phrasing in sentence comprehension: Evidence from native English speakers and native Korean-speaking second language learners of English* (Unpublished doctoral dissertation). University of Hawaii, Honolulu, HI.

Im, S., Cole, J., & Baumann, S. (2018). The probabilistic relationship between pitch accents and information status in public speech. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 508–511). Poznań, Poland: Adam Mickiewicz University.

Isaacs, A. M., & Watson, D. G. (2010). Accent detection is a slippery slope: Direction and rate of F0 change drives listeners' comprehension. *Language and Cognitive Processes*, *25*(7-9), 1178–1200.

Ito, J., & Mester, A. (2013). Prosodic subcategories in Japanese. *Lingua*, *124*, 20–40.

Ito, K. (2002). *The interaction of focus and lexical pitch accent in speech production and dialogue comprehension: Evidence from Japanese and Basque* (Unpublished doctoral dissertation). University of Illinois Urbana–Champaign, Urbana, IL.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, *66*(1), 265–284.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573.

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT Press.

Jenner, B. R. (1976). Interlanguage and foreign accent. *Interlanguage Studies Bulletin*, *1*(2-3), 166–195.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL-International Review of Applied Linguistics in Language Teaching*, *28*(2), 99–118.

Jun, S.-A. (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*(3), 402–420.

Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience*, *30*(1-2), 103–119.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, *87*(4), 771–816.

Kelm, O. R. (1987). An acoustic study on the differences of contrastive emphasis between native and non-native Spanish speakers. *Hispania*, *70*(3), 627–633.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

Klassen, J. (2013). Second language acquisition of English focus prosody: Evidence from Spanish native speakers. In J. Cabrelli Amaro, T. Judy, & D. Pascual y Cabo (Eds.), *Proceedings of the 12th Generative Approaches to Second Language Acquisition Conference (GASLA 2013)* (pp. 76–84). Somerville, MA: Cascadilla Press.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, *34*(4), 391–405.

Kubozono, H. (1987). *The organisation of Japanese prosody* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh, United Kingdom.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore, MD: York Press.

Kuno, S. (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry*, *3*(3), 269–320.

Kuno, S. (1973). *The structure of the Japanese language* (Vol. 3). Cambridge, MA: MIT Press.

Ladd, D. R. (1996). *Intonational phonology* (1st ed.). New York: Cambridge University Press.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). New York: Cambridge University Press.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, *25*(3), 313–342.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*(1), 81–112.

Lee, E.-K., & Fraundorf, S. (2017). Effects of contrastive accents in memory for L2 discourse. *Bilingualism: Language and Cognition*, *20*(5), 1063–1079.

Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*(1), 380–380.

Ma, J. H. (2007). The role of prosody in English sentence processing: Evidence from L1-Korean EFL learners. *Language Research*, *43*(1), 175–201.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, *19*(2), 144–161.

Marinis, T., Blom, E., & Unsworth, S. (2010). Using on-line processing methods in language acquisition research. *Experimental methods in language acquisition research*, 139–162.

McCawley, J. D. (1968). *The phonological component of a grammar of Japanese*. The Hague, The Netherlands: Mouton.

Mennen, I. (1999). The realisation of nucleus placement in second language intonation. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 555–558). Oakland, CA: University of California.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543–563.

Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 171–188). Berlin, Heidelberg: Springer.

Mennen, I., & de Leeuw, E. (2014). Beyond segments. *Studies in Second Language Acquisition*, *36*(2), 183–194.

Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, *36*(2), 303.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, *131*(3), 2249–2260.

Most, R. B., & Saltz, E. (1979). Information structure in sentences: New information. *Language and Speech*, *22*(1), 89–95.

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. mora rhythm. *Language and Speech*, *50*(1), 77–99.

Nagahara, H. (1994). *Phonological phrasing in Japanese* (Unpublished doctoral dissertation). University of California Los Angeles, Los Angeles, CA.

Nakanishi, K. (2001). Prosody and information structure in Japanese: A case study of topic marker *wa*. *Japanese/Korean Linguistics*, *10*, 434–447.

Nakatani, C. H. (1997). Discourse structural constraints on accent in narrative. In J. P. Van Santen, R. Sproat, J. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis* (pp. 139–156). New York: Springer Science & Business Media.

Nava, E. (2008). Prosody in L2 acquisition. In R. Slabakova (Ed.), *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)* (pp. 155–164). Somerville, MA: Cascadilla Press.

Nibert, H. J. (2005). The acquisition of the phrase accent by intermediate and advanced adult learners of Spanish as a second language. In D. Eddington (Ed.), *Selected proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 108–122). Somerville, MA: Cascadilla Press.

Nibert, H. J. (2006). The acquisition of the phrase accent by beginning adult learners of Spanish as a second language. In M. Díaz-Campos (Ed.), *Selected proceedings of the 2nd Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 131–148). Somerville, MA: Cascadilla Press.

Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *The Journal of the Acoustical Society of America*, *82*(5), 1512–1524.

O'Brien, M., & Gut, U. (2010). Phonological and phonetic realisation of different types of focus in L2 speech. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech: New Sounds 2010* (pp. 331–336). Poznań, Poland: Adama Mickiewicz University.

O'Brien, M., Jackson, C. N., & Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Applied Psycholinguistics*, *35*(1), 27–70.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, *84*(3), 372–389.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic Inquiry Monographs*(15), 1–282.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (p. 271-311). Cambridge, MA: MIT Press.

Piske, T. (2012). Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 41–59). Berlin, Germany: Springer.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, *90*(6), 2956–2970.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, *28*, 41–66.

Rasier, L., & Hiligsmann, P. (2009). Exploring the L1-L2 relationship in the L2 acquisition of prosody. In *Proceedings of first and second languages: Exploring the relationship in pedagogy related context.* Oxford, United Kingdom. Retrieved from http://www.education.ox.ac.uk/research/applied-linguistics/events/events-archive/first-and-second-languages/

Ratner, N. B., Rooney, B., & MacWhinney, B. (1996). Analysis of stuttering using CHILDES and CLAN. *Clinical Linguistics & Phonetics*, *10*(3), 169–187.

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. *Studies in Visual Information Processing*, *6*, 3–22.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116.

Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation in EFL settings: The case of English vowels and Japanese learners of English. *Linguistics Journal*, *2*(3).

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*(2), 595–633.

Schafer, A. J., Takeda, A., Rohde, H., & Grüter, T. (2015, November). *Mapping prosody to reference in L2.* Poster presented at the 40th Boston University Conference on Language Development. Boston, MA.

Schulz, B. (2006). *Wh-scope marking in English interlanguage grammars: Transfer and processing effects on the second language acquisition of complex wh-questions* (Unpublished doctoral dissertation). University of Hawaii, Honolulu, HI.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *The handbook of phonological theory* (p. 550-569). Cambridge, MA: Blackwell.

Selkirk, E. (2009). On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu*, *136*, 35–73.

Selkirk, E., et al. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (Vol. 2, pp. 435–483). Oxford: Blackwell Publishing.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247.

Shibatani, M. (1990). *The languages of Japan*. Cambridge, UK: Cambridge University Press.

Sityaev, D. (2000). The relationship between accentuation and information status of discourse referents: A corpus-based study. *UCL WorkingPapers in Linguistics*, *12*, 285–304.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293.

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*(2), 195–210.

Speer, S. R., & Blodgett, A. (2006). Prosody. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 505–537). New York: Elsevier.

Speer, S. R., & Ito, K. (2011). Prosodic properties of contrastive utterances in spontaneous speech. In W.-S. Lee & E. See (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1890–1893). Hong Kong: City University of Hong Kong.

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology*, *2*(1), 35–98.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*(4), 629–654.

Swerts, M., & Zerbian, S. (2010). Intonational differences between L1 and L2 English in South Africa. *Phonetica*, *67*(3), 127–146.

Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition*, *7*(3), 159–165.

Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, *27*(3), 269–289.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing ‘given’ information: Effects of persistence of grammatical function and surface position. *Language and Speech*, *37*(2), 125–145.

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, *2*(3-4), 145–163.

’t Hart, J., & Collier, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics*, *3*(4), 235–255.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344.

Tomioka, S. (2016). Information structure in Japanese. In C. Féry & S. Ishihara (Eds.), *The Oxford handbook of information structure* (pp. 753–773). Oxford: Oxford University Press.

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics*, *29*(4), 553–584.

Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the native language on the learning of fundamental frequency in second-language speech segmentation. *Frontiers in Psychology*, *7*. Retrieved from `https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00985/full`

Trofimovich, P., & Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, *28*(02), 251–276.

Trouvain, J., & Gut, U. (2007). *Non-native prosody: Phonetic description and teaching practice*. Berlin, Germany: Walter de Gruyter.

Turco, G., Dimroth, C., & Braun, B. (2015). Prosodic and lexical marking of contrast in L2 Italian. *Second Language Research*, *31*(4), 465–491.

Vallduví, E. (1992). *The information component* (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia, PA.

van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning*, *66*(1), 124–158.

van Maastricht, L., Zee, T., Krahmer, E., & Swerts, M. (2017). L1 perceptions of L2 prosody: The interplay between intonation, rhythm, and speech rate and their contribution to accentedness and comprehensibility. In W. Hess & M. Cooke (Eds.), *Proceedings of*

Interspeech 2017 (pp. 364–368). Stockholm, Sweden: International Speech Communication Association.

Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). *Transcribing prosodic structure of spoken utterances with ToBI.* Retrieved from http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172–200). Oxford: Oxford University Press.

Venditti, J. J., & Hirschberg, J. (2003). Intonation and discourse processing. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 107–114). Barcelona, Spain: Universitat Autonoma de Barcelona. Retrieved from https://www.researchgate.net/profile/Julia_Hirschberg/publication/240057737_Intonation_and_discourse_processing/links/0c960529dc9db3b8ae000000.pdf

Venditti, J. J., Maekawa, K., & Beckman, M. E. (2008). Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese Linguistics* (pp. 456–512). New York: Oxford University Press.

Venditti, J. J., & Swerts, M. (1996). Intonational cues to discourse structure in Japanese. In *Proceedings of the 4th International Conference on Spoken Language Processing* (Vol. 2, pp. 725–728). New York: IEEE. Retrieved from http://www1.cs.columbia.edu/~jjv/pubs/icslp.pdf

Verdugo, D. R. (2003). Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English. *ICAME Journal*, *26*, 115–132.

Verdugo, D. R. (2006). Prosodic realization of focus in the discourse of Spanish learners and English native speakers. *Estudios Ingleses de la Universidad Complutense*(14), 9–32.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905–945.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Warren, P. (1996). Prosody and parsing: An introduction. *Language and Cognitive Processes*, *11*(1-2), 1–16.

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *Psychology of Learning and Motivation*, *52*, 163–183.

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science*, *32*(7), 1232–1244.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392.

Weinberger, S. (2015). *Speech accent archive.* Retrieved from `http://accent.gmu.edu`

Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech*, *46*(1), 53–81.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, *15*(4), 399–420.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition*, *20*(1), 1–25.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.

Ying, H. (1996). Multiple constraints on processing ambiguous sentences: Evidence from adult L2 learners. *Language Learning*, *46*(4), 681–711.

Yoshimura, N., Fujimori, A., & Shirahata, T. (2015). Focus and prosody in second language acquisition. *Journal of International Relations and Comparative Culture*, *13*(2), 21–36.